離散型確率変数

樋口さぶろお https://hig3.net

龍谷大学 先端理工学部 数理・情報科学課程

確率統計 I L01(2025-04-14 Mon)

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

今日の目標

- 科目の目標/合格条件を説明できる, Moodle 使える
- 久保川 統計学入門 §4.1 離散型確率変数とは何か説明できる
- 久保川 統計学入門 §4.2 離散型確率変数の確率，母平均値, 母分散, 母期待値が計算できる

ここまで来たよ

- はじめに
  - この授業どんなのり?

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

学習目標

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

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

きょうは 1 変数離散型

確率統計 I を履修してはいけない理由

次のどれも響かない人は履修しないことを奨めます.

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

こんなことに答えます

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

(再掲) 到達目標 → シラバス

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

ピーナツカウント (成績計算) ののり

成績計算
ムズくないけどだるい科目です
科目の成績 100 ピーナツは

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

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

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

週サイクルののり

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

- 2025-04-14 月 対面授業 来週の trial を予告, チーム課題で練習
- 2025-04-15 火 9:20 ごろ –2025-04-18 金 20:00 LearnMoodle 練習問題
- 2025-04-17 木昼 オフィスアワー 1-539
- 2025-04-21 月 09:15(10 分間くらい) trial 非参照 非相談

数学系の 2 年次以上科目は週 1 コマしかない!

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

授業ののり (教科書やその他の準備)

https://hig3.net

- → 確率統計 I (配布資料).
- → LearnMoodle https://learn.hig3.net (Google でログイン) → 確率統計 I
  - → Teams

https://hig3.net

§1,2: 記述統計. データ分析 相当.

§3.1, §3.2: 大事だけど今日圧縮してやります

§3.3, §3.4: 必要な時に戻ってきます

§4–9,12.1,12.2: とばすところもあるけどこの科目でカバー

教科書 必須です. 久保川 統計学入門 (3.2) は式 (3.2) ということ.

久保川達也 公式と例題で学ぶ 統計学入門, 共立出版 (2024)

演習解答・正誤表

座席指定

チーム別エリア座席指定．チームはプロジェクト演習と同じはず．

ノート

教科書や紙配布資料に書き込む + 自分で問題を解いた過程をノートやルーズリーフに残す，ことをお奨めします．

ノート PC

Google Colab や動画を使います．ノート PC とイヤフォンを毎回持参していただきます．

担当者ののり

- なまえ: 樋口さぶろお
- へや: 1-507
- Web ページ. `https://hig3.net`
- オフィスアワー 前木昼, 1-539 or Teams chat a00010

相談できるところ

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

ここまで来たよ

- はじめに
  - この授業どんなのり?

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

事象と確率 久保川 統計学入門 §3.1

確率は集合位相の言葉で語られる. 集合位相 (2 年後期)

定義

全事象 $\Omega =$(トランプのカード全体の集合)

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

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

試行 基本事象のひとつを選ぶこと (トランプからカード 1 枚引くこと)

全事象 $\Omega \subset \Omega$

空事象 $\emptyset \subset \Omega$

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

積事象 $A \cap B$ ‘かつ’.

和事象 $A \cup B$ ‘または’.

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

定義 (確率の公理 久保川 統計学入門 定義 3.1)

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

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

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

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

事象と確率

公理から導ける公式

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

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

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

積事象 $A \cap B$ ‘かつ’.

和事象 $A \cup B$ ‘または’.

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

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

Ω が有限集合で, 基本事象が同様に確からしい, なら, 確率は

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

確率変数

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

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

ここまで来たよ

- はじめに
  - この授業どんなのり?

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

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	$\frac{1}{52}$
$\vdots$	$\vdots$
♠13	$\frac{1}{52}$
計	1

偶数のカードのでる確率は? $\frac{24}{52}$.

離散型確率変数 久保川 統計学入門 §4.1

高校数学でよく見る確率の問題 高校 数学 A

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

X は離散型 の確率変数 離散型 $\approx$ 実現値が可算個（例えば整数全体）
全事象が整数全体と思える.

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

言葉 確率分布 (確率関数, 確率質量関数)

久保川 統計学入門 §4.1

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

$$p_k = p(x_k).$$

高校と大学の, 確率の問題の違い

高校 数学 A ではこの表を作るまでを考える

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

確率 (質量) 関数と累積確率分布関数

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

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

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

累積分布関数の性質は来週．

久保川 統計学入門 図 4.1

ここまで来たよ

- はじめに
  - この授業どんなのり?

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

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

定義 (関数 $g(x)$ の母期待値 久保川 統計学入門 公式 4.4)

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

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

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

命題 (母期待値の性質)

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

定義 (母平均値, 母分散, 母標準偏差)

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

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

L01-Q1

Quiz(離散的な確率変数の母平均・母分散・母標準偏差)

整数に値をとる離散型確率変数 X は次の確率関数を持つ.

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

1. 母期待値 $\mathrm{E}[3 + 12\mathrm{e}^X]$ を求めよう.
2. X の母平均値を求めよう.
3. X の母分散を求めよう.

久保川 統計学入門 問 1–4(p.100)

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

幾何分布 久保川 統計学入門 §4.3.4

L01-Q2

Quiz(離散的な確率変数の累積分布関数)

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

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

1. 累積分布関数 $F(x)$ を，整数 x に対して求めよう．

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(2025-04-21 Mon)

最終更新: Time-stamp: "2025-04-21 Mon 11:47 JST hig"

今日の目標

- 久保川 統計学入門 §5.1 連続型確率変数とは何か説明できる
- 久保川 統計学入門 §5.2 連続型確率変数の母平均値, 母分散, 母期待値が計算できる

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 $\mathrm{E}[3+12\mathrm{e}^X] = \frac{4}{12}\cdot(3+12\mathrm{e}^{-1}) + \frac{5}{12}\cdot(3+12\mathrm{e}^0) + \frac{3}{12}\cdot(3+12\mathrm{e}^2) = 8+4\mathrm{e}^{-1}+3\mathrm{e}^2.$
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 $\mathrm{Var}(X) = \mathrm{E}[(X-\mu)^2] = \frac{4}{12}\cdot(-1-\frac{1}{6})^2 + \frac{5}{12}\cdot(0-\frac{1}{6})^2 + \frac{3}{12}(2-\frac{1}{6})^2 = \frac{47}{36}.$

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 久保川 統計学入門 公式 4.4(p.85)**)**

確率変数 X，$a, b \in \mathbb{R}$: 定数，関数 g, h に対して，

$$\begin{aligned}
\mathrm{E}[ag(X) + bh(X)] &= \sum_x (ag(x) + bh(x)) \times p(x) = a\mathrm{E}[g(X)] + b\mathrm{E}[h(X)] \\
\mathrm{E}[a] &= a, \\
\mathrm{E}[1] &= 1, \\
\mathrm{E}[aX + b] &= a\mathrm{E}[X] + b
\end{aligned}$$

定義 (k 次のモーメント 久保川 統計学入門 p.121**)**

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

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.

$\mathrm{E}[X^2] \neq \mathrm{E}[X]^2$

母分散の性質 久保川 統計学入門 公式 4.5(p.86)

分散の計算方法（分散公式） 高校 数学 B 久保川 統計学入門 公式 4.5(p.86),p.106

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

$$\mathrm{Var}(X) = \mathrm{E}[X^2] - (\mathrm{E}[X])^2$$
$$\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X).$$

証明 $\mu = \mathrm{E}[X]$ とする.
$\mathrm{Var}(X) = \mathrm{E}[(X-\mu)^2] = \mathrm{E}[X^2 - 2\mu X + \mu^2] =$
$\mathrm{E}[X^2] - 2\mu\mathrm{E}[X] + \mu^2\mathrm{E}[1] = \mathrm{E}[X^2] - 2\mu^2 + \mu^2.$
証明 $\mu = \mathrm{E}[X]$ と, $\mathrm{E}[Y] = \mu_{\mathrm{Y}} = a\mu + b$ とする.

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

ここまで来たよ

1. 離散型確率変数
   - 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
   - 連続型確率変数

確率変数の標本抽出

久保川 統計学入門 §7 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 回行って得られる実現値 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

```
import numpy as np # ライブラリの読込
from scipy import stats

xk = (0,  1 ,  2  , 3  ,4   , 5  ,6)    # x_1, x_2,..,x_k
pk = (0.1, 0.2, 0.3, 0.1, 0.1, 0.0, 0.2) # p_1, p_2,..,p_k
rvx = stats.rv_discrete(values=(xk, pk)) # 離散型確率変数 X を定義
sample=rvx.rvs(size=30) # サイズ30の標本抽出(毎回違った結果になる)

# 母ナントカ. 答はいつも同じ
rvx.mean()# 母平均値
rvx.var() # 母分散
rvx.std() # 母標準偏差

# データ分析 でやってた標本平均値
## 標本抽出のたびに結果が変わる
sample.mean() # 標本平均値
sample.var()  # 不偏標本平均値
sample.std()  # 標本標準偏差
```

ここまで来たよ

1. 離散型確率変数
   - 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
   - 連続型確率変数

場合分けとそうでない確率変数

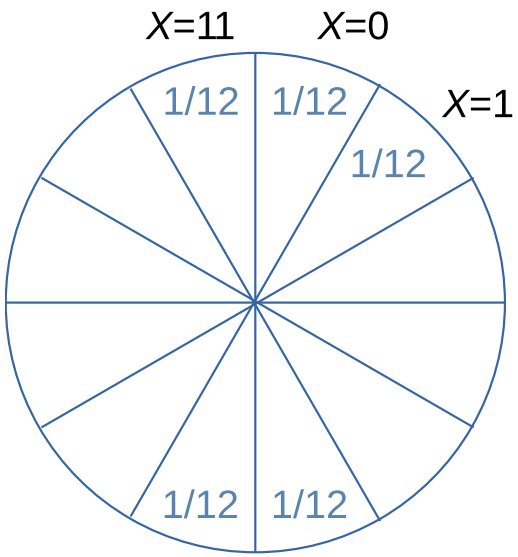
整数に値をとる離散型確率変数 X が次の確率関数 $p(x)$ を持つ.
(同じこと)

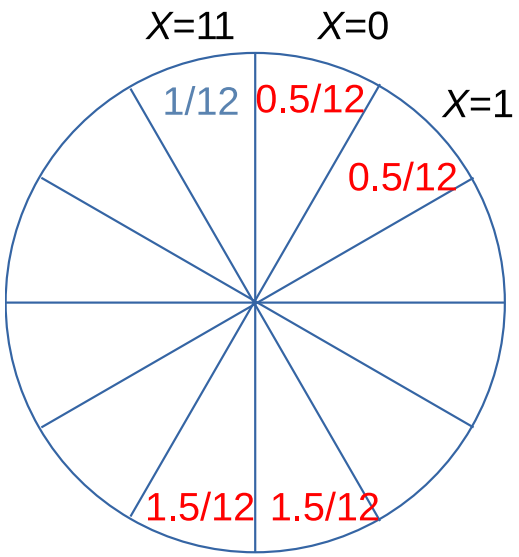
$$p(x) = \begin{cases} 0.1 & (x = 6) \\ 0.2 & (x = 7) \\ 0.3 & (x = 8) \\ 0.4 & (x = 9) \\ 0 & (\text{他}) \end{cases}$$

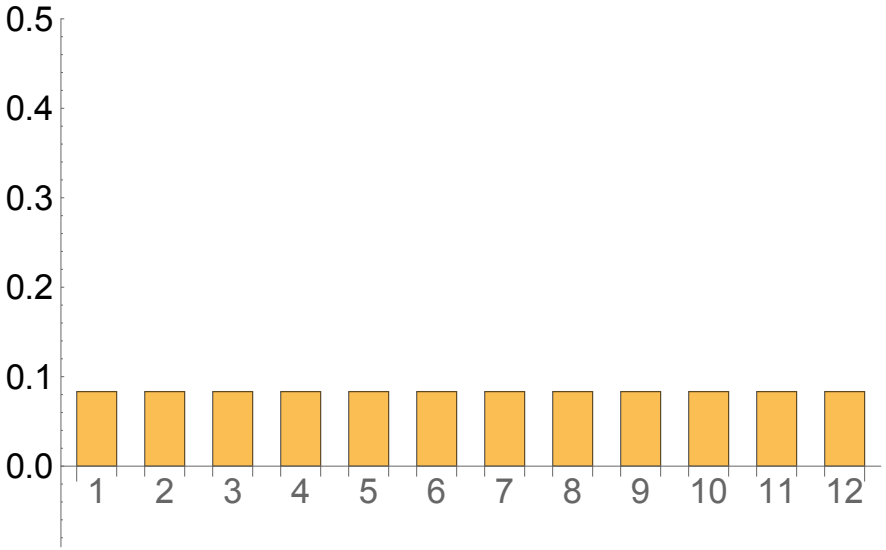
$$p(x) = \begin{cases} (x-5)/10 & (6 \leq x \leq 9) \\ 0 & (\text{他}) \end{cases}$$

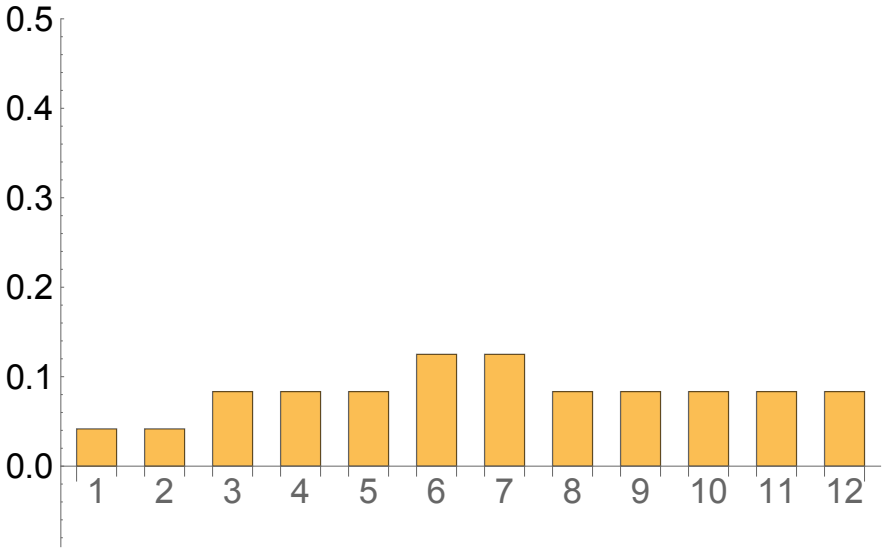
離散型確率変数の例：離散ルーレット

$P(X=1)=1/12, P(0 \le X \le a)=1, \ldots, a$ の棒の高さの和



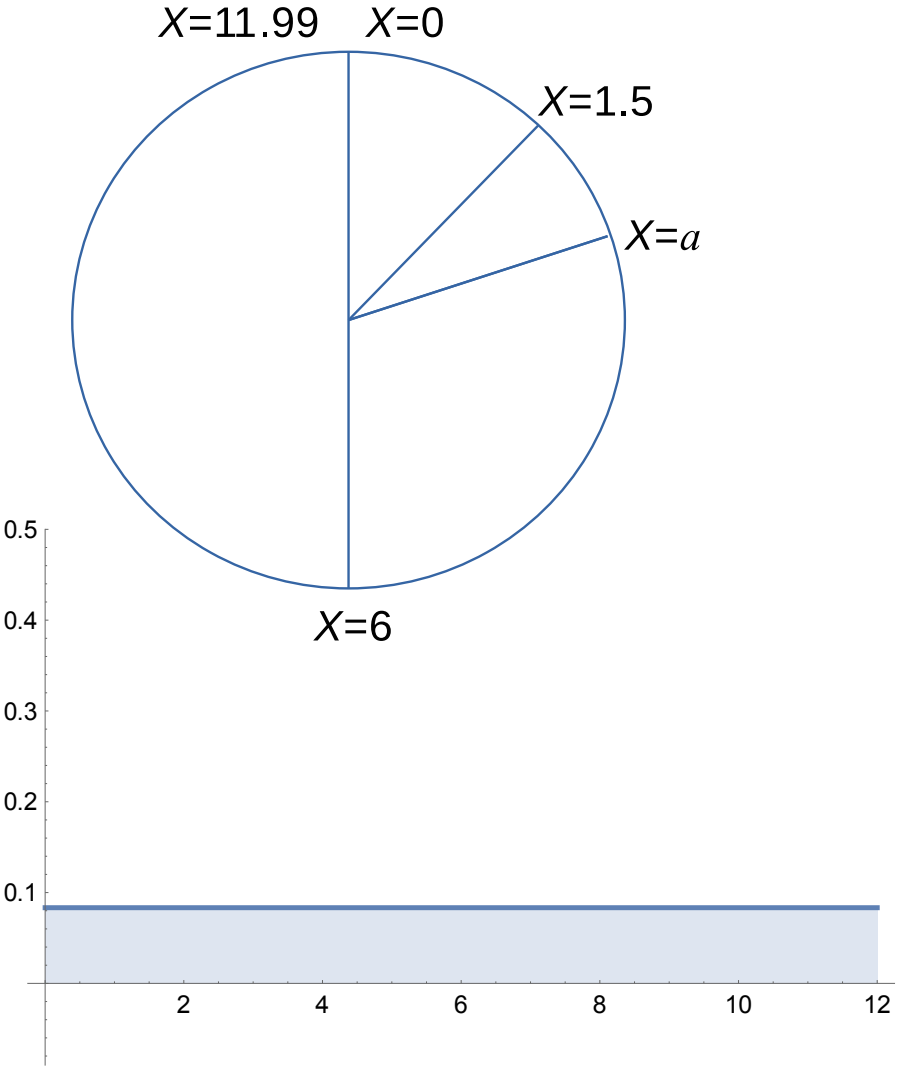


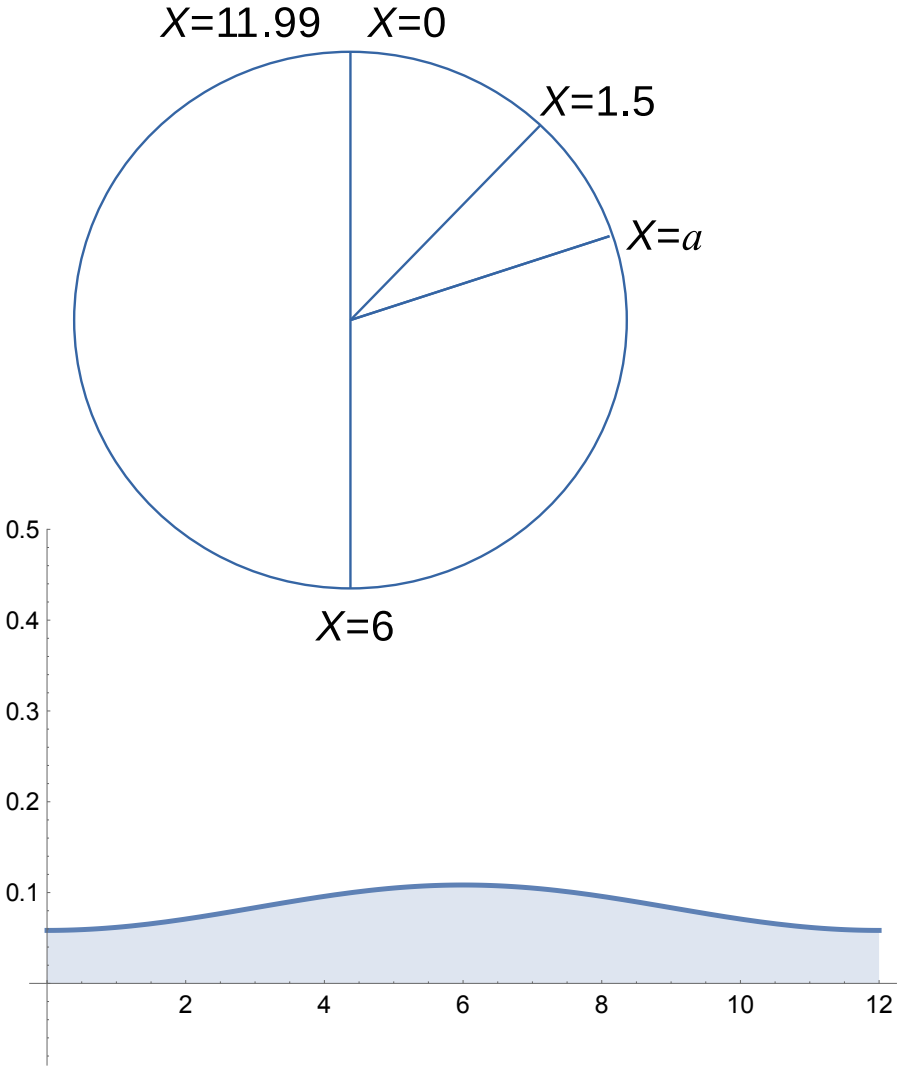




連続型確率変数の例：連続型ルーレット

$$P(X=1)=0, P(0 \leq X \leq a) = a/12 = (0 \text{ から } a \text{ までのグラフの面積})$$





連続型確率変数 久保川 統計学入門 §5.1

定義 (連続型確率変数 久保川 統計学入門 p.104)

連続型確率変数 X とは, 実数のように連続値をとる確率変数のこと.

$\rightsquigarrow$ X の実現値の出やすさは x のある関数 $f(x)$ 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)

定義 (確率密度関数 久保川 統計学入門 p.104)

次の性質を持つ関数 $f(x)$ を，連続型確率変数の確率密度関数という．X の確率分布を指定できる．

$$f(x) \geq 0$$

$$\int_{-\infty}^{+\infty} f(x)\,\mathrm{d}x = 1$$

$f(x) \leq 1$ とはいえない！

離散的

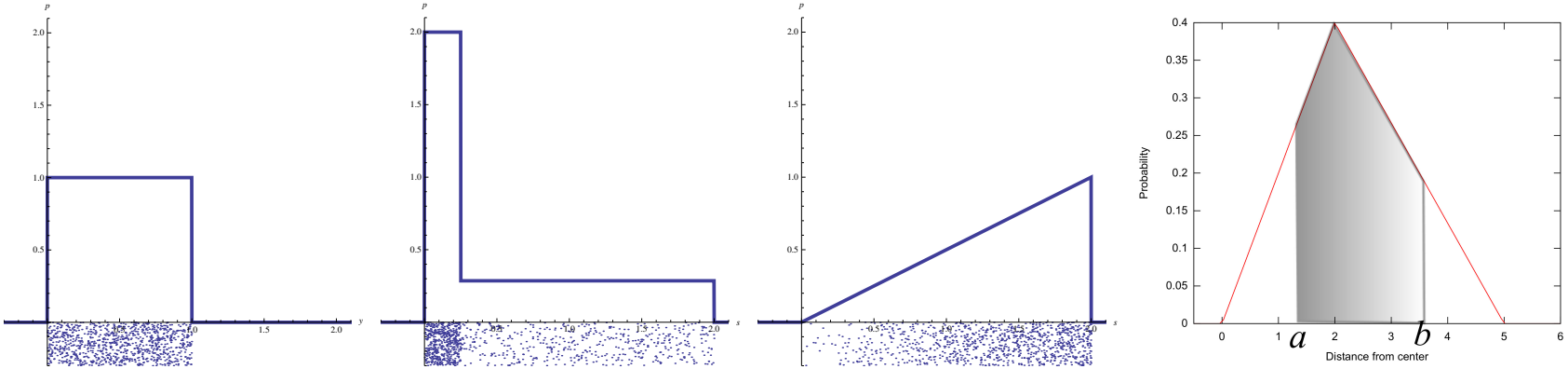
得点 x	確率 $p(x)$
0	0.1
1	0.3
$\vdots$	
x	$p(x)$

連続的

- $f(x)$ が大きいほど，その値 x がでやすい

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

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)

命題 (確率密度関数と確率 久保川 統計学入門 p.104 **)**

$$P(c < X \leq d) = \int_c^d f(x)\,\mathrm{d}x \quad \text{(下側面積)}$$

連続型確率変数の母期待値

定義 (母期待値 久保川 統計学入門 定義 5.3, 5.4 **)**

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

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

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

定義 (k 次のモーメント ($k = 0, 1, 2, 3, \cdots$) 久保川 統計学入門 p.121 **)**

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

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 $\mathrm{Var}(X)$
4. 母期待値 $\mathrm{E}[(2X+3)^2]$
5. 母分散 $\mathrm{Var}(2X+3)$
6. 確率 $P(-\frac{1}{4} < X \leq \frac{1}{4})$

久保川 統計学入門 例題 5.6,5.7(p.107),§5 基本問題問 2,3(p,124)

累積分布関数・分位数・指数分布・幾何分布

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 L03(2025-04-28 Mon)

最終更新: Time-stamp: "2025-04-28 Mon 11:55 JST hig"

今日の目標

- 連続型確率変数の累積分布関数と分位点関数が計算できる 久保川 統計学入門 §5.1
- 指数分布 久保川 統計学入門 §5.3.3 の意味が説明でき計算できる
- 幾何分布 久保川 統計学入門 §4.3.4 の意味が説明でき計算できる

L02-Q1

Quiz 解答: 連続型確率変数

1. k 次のモーメントは,

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

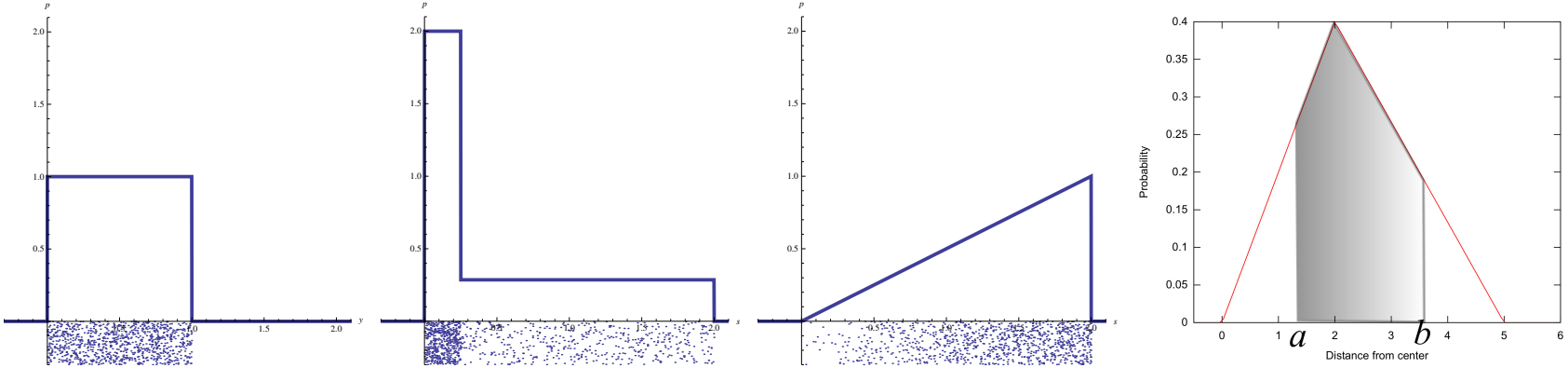
   $\mathrm{E}[X^0] = 1$ が確認できる.
2. $\mathrm{E}[X^1] = \frac{1}{3}$.
3. $\mathrm{Var}(X) = \mathrm{E}[X^2] - \mathrm{E}[X^1]^2 = \frac{1}{8} - (\frac{1}{3})^2 = \frac{1}{72}$.
4. $\mathrm{E}[(2X+3)^2] = 4\mathrm{E}[X^2] + 12\mathrm{E}[X^1] + 9\mathrm{E}[X^0] = 4 \times \frac{1}{8} + 12 \times \frac{1}{3}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

ここまで来たよ

2 連続型確率変数

3 累積分布関数・分位数・指数分布・幾何分布
- 累積分布関数
- 指数分布
- 分位数

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)

命題 (確率密度関数と確率 久保川 統計学入門 p.104 **)**

$$P(c < X \leq d) = \int_c^d f(x)\,\mathrm{d}x \quad \text{(下側面積)}$$

定義 (累積分布関数 久保川 統計学入門 §4.1(p.83)§5.1(p.104))

$F(x) = P(X \leq x)$ を累積分布関数（分布関数, 確率分布関数）という．

命題 (累積分布関数の性質 久保川 統計学入門 p.104)

(C1) $F(x)$ は広義単調増加関数
(C2) $F(-\infty) = 0, F(+\infty) = 1.$
(C3) $P(c < X \leq d) = F(d) - F(c).$

定義 (連続型確率変数の累積分布関数 久保川 統計学入門 p.104)

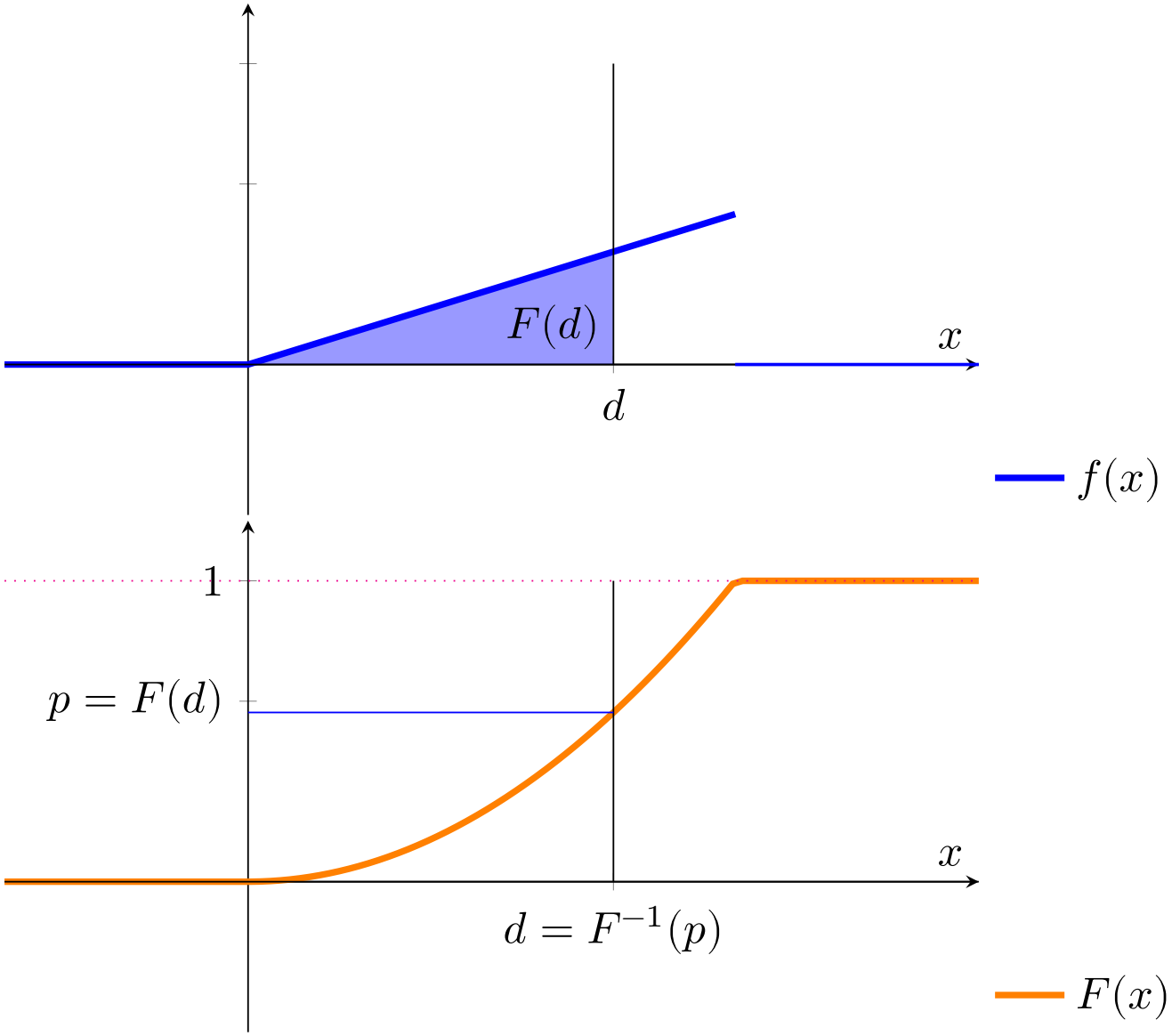
$$F(x) = \int_{-\infty}^{x} f(x')\,\mathrm{d}x'$$

命題 (久保川 統計学入門 公式 5.1(p.105))

F が微分可能なとき，$f(x)$ は $F(x)$ の導関数

確率密度関数 f $\underset{\text{微分}}{\overset{\text{積分}}{\rightleftarrows}}$ 累積分布関数 F

久保川 統計学入門 例題 5.2



久保川 統計学入門 図 5.4

ここまで来たよ

2 連続型確率変数

3 累積分布関数・分位数・指数分布・幾何分布
- 累積分布関数
- 指数分布
- 分位数

指数分布 久保川 統計学入門 §5.3(p.116) 久保川 統計学入門 §5.3.3

連続型確率変数の分布の, 名前がつく有名な例．パラメタ（母数）$\lambda > 0$ を持つ family (パラメタは確率変数でない．$8x^1$ の 8 や 1 のこと).

定義 (指数分布)

連続型確率変数 X が確率密度関数

$$f(x;\lambda) = \begin{cases} \lambda \mathrm{e}^{-\lambda x} & (x > 0) \\ 0 & (x \leq 0) \end{cases}$$

を持つとき, X はパラメタ $\lambda > 0$ の指数分布に従うといい, 記号 ∼(従う) で $X \sim Ex(\lambda)$ と書く ($Ex()$ は 久保川 統計学入門 ローカル記号)

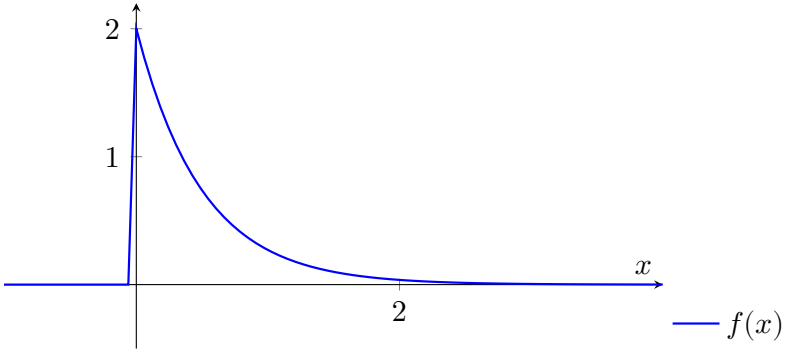
性質 実は $\mathrm{E}[X^k] = \lambda^{-k}$ (多数回部分積分) 久保川 統計学入門 公式 5.19

意味時間的にランダムに起きる事象，例えば，「ある機械が x 秒後に初めて故障する」「サッカーで x 秒後に次のゴールが起きる」のような事象の，時間間隔 x のしたがう分布．母平均値 $\mathrm{E}[X] = 1/\lambda$.

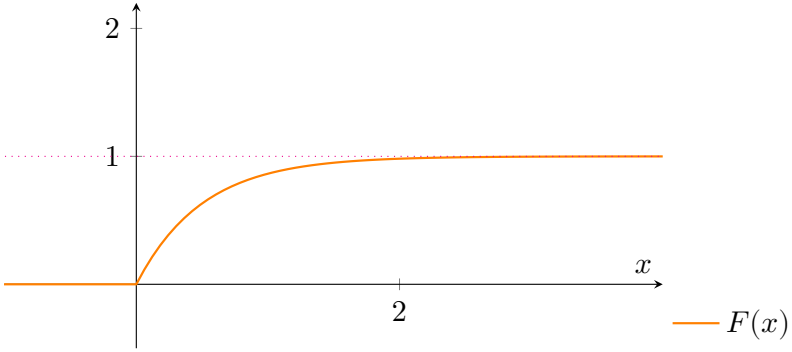
$X \sim Ex(\lambda)$, 確率密度関数 $f(x) = \lambda \mathrm{e}^{-\lambda x}$ のとき，累積分布関数

$$F(x) = \int_{-\infty}^{x} f(x')\mathrm{d}x' = \begin{cases} \int_{-\infty}^{x} 0\ \mathrm{d}x' = 0, & (x \leq 0) \\ \int_{-\infty}^{0} 0\ \mathrm{d}x' + \int_{0}^{x} \lambda \mathrm{e}^{-\lambda x'} \mathrm{d}x' = 1 - \mathrm{e}^{-\lambda x}. & (x > 0) \end{cases}$$

確率密度関数



累積分布関数



L03-Q1

Quiz(指数分布の累積分布関数・確率・分位数)

連続型確率変数 X はパラメタ $\lambda = 3$ の指数分布 $Ex(3)$ にしたがう．確率密度関数 $f(x)$ は次で与えられる．

$$f(x) = \begin{cases} 3\mathrm{e}^{-3x} & (x \geq 0) \\ 0 & (x < 0) \end{cases}.$$

1. X の累積分布関数 $F(x)$ を求めよう．
2. 確率 $P(X \leq 4)$ を求めよう．
3. 確率 $P(-5 < X \leq 2)$ を求めよう．
4. 確率 $P(X \leq d) = \frac{1}{7}$ となる $d \in \mathbb{R}$ を求めよう．

指数・対数関数の値は Windows の電卓．iPhone を横向きにして関数電卓で．または, Google Colab で import numpy as np のあと，np.exp(数値), np.log(数値).

L03-Q2

Quiz(指数分布)

あるサッカーチームが，開始後 X 分で最初の得点をするとしたとき，この X はパラメタ $\lambda = 1/30$ の指数分布にしたがう（開始直後の特別なイベントは無視してゲームは時間的に一様とする．毎週の試合をつなげて考える, 本当は相手のチームによるだろうけど無視）.

1. 開始後 5 分から 10 分に最初の得点する確率を求めよう．有効数字 3 桁で答えよう．
2. 開始後 80 分たっても得点していない確率を求めよう．有効数字 3 桁で答えよう．
3. 得点と次の得点の間の時間の母平均値を求めよう．有効数字 3 桁で答えよう．
4. 1 点目がはいる確率が $2/3$ を超える時刻を求めよう．

指数・対数関数の値は Windows の電卓．iPhone を横向きにして関数電卓で．または, Google Colab で import numpy as np のあと，np.exp(数値), np.log(数値).

ここまで来たよ

2 連続型確率変数

3 累積分布関数・分位数・指数分布・幾何分布
- 累積分布関数
- 指数分布
- 分位数

連続型確率変数の分位数関数

定義 (分位数関数 久保川 統計学入門 p.109 **)**

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-(下側) 分位点 $F^{-1}(1-q)$ を, q-上側分位点 という.

意味

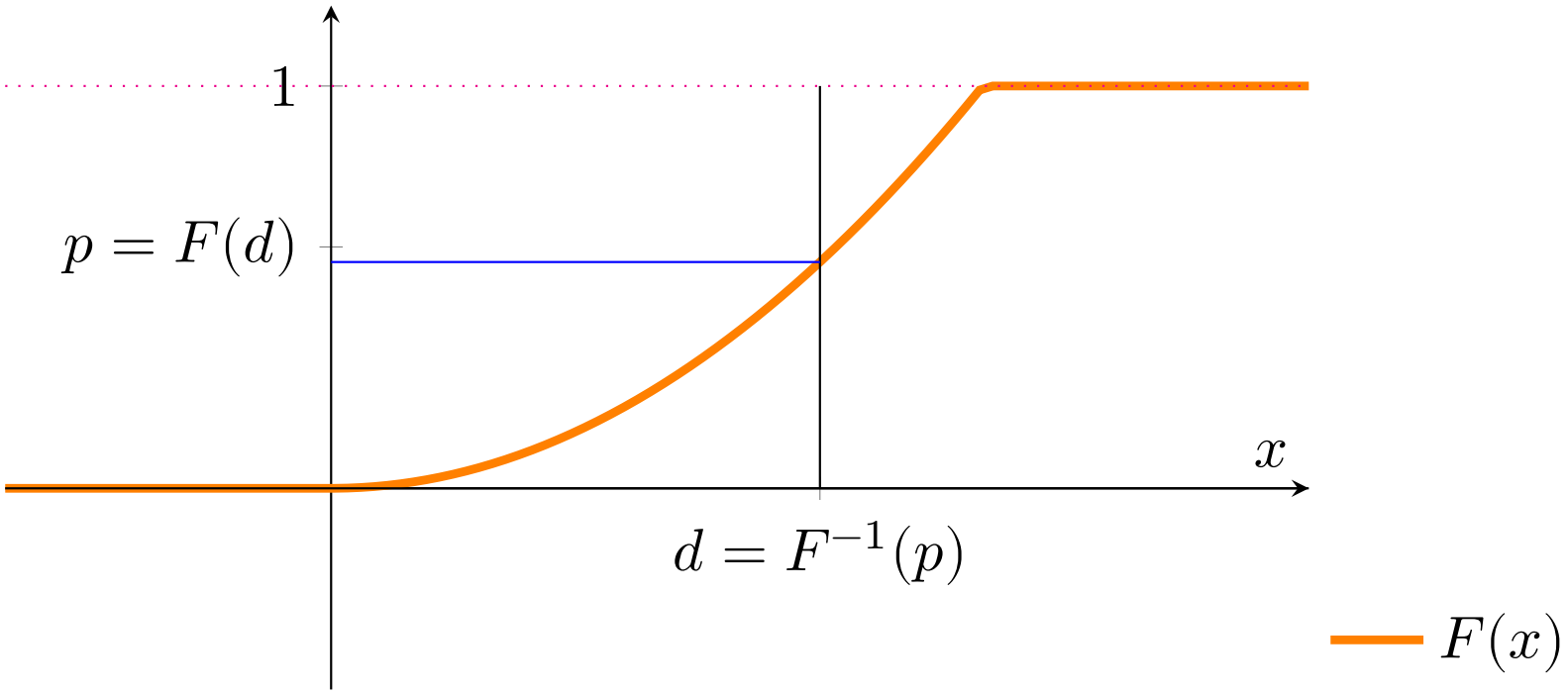
'q-下側分位点とは，確率 $P(X \leq d)$ が q に等しくなる**最小ぎりぎりの**境い目 $x = d$'

$F^{-1}(y)$ 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.

'q-上側分位点とは，確率 $P(X \geq d)$ が q に等しくなる**最大ぎりぎりの**境い目 $x = d$'

言い訳

定義域 $[0,1]$ の端は含められない場合もある．$F(x)$ が狭義単調増加でない限り逆関数 F^{-1} は定義できない．これらのややこしいケースには深入りしない．離散型にも深入りできない．



例 (中央値・四分位点)

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

データ分析

(標本) 中央値，四分位点 久保川 統計学入門 §1.2.3

scipy による連続型確率変数

https://colab.research.google.com

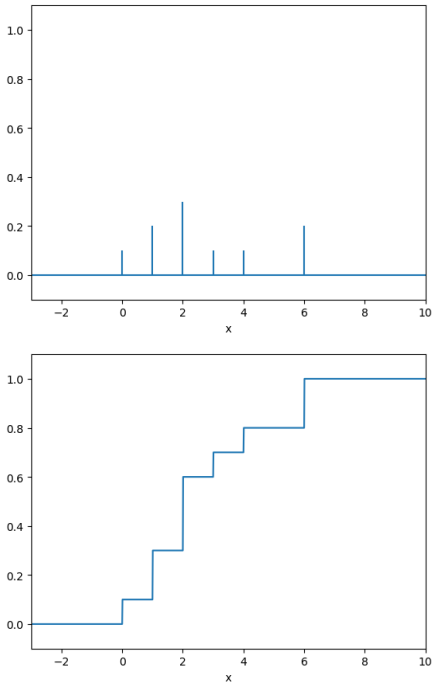
```
import numpy as np # ライブラリの読込
from scipy import stats

rvx = stats.expon(loc=0,scale=0.2) # 指数分布，専用の名前がある．
                                   # scale=1/λ
# 母ナントカ．答はいつも同じ
rvx.mean()# 母平均値
rvx.var() # 母分散
rvx.std() # 母標準偏差
rvx.moment(k) k 次のモーメント
rvx.pdf(x) 確率密度関数 probability density function
rvx.cdf(x) 累積分布関数 cumurative density function
rvx.ppf(p) 分位数関数 percent point function

# データ分析 でやってた標本ナントカ 毎回違った結果
sample=rvx.rvs(size=30) # サイズ30の標本抽出
sample.mean() # 標本平均値
sample.var()  # 不偏標本平均値
sample.std()  # 標本標準偏差
```

離散型確率変数の累積分布関数

連続型と同様の定義 $F(x) = P(X \leq x)$

命題 (離散型確率変数の累積分布関数 久保川 統計学入門 p.83 **)**

累積分布関数の累積分布関数は，確率関数 $p(x)$ で，$F(x) = \sum_{x' \leq x} p(x')$ と書ける．



久保川 統計学入門 図 4.1

幾何分布 久保川 統計学入門 §4.3.4(p.97)

定義 (幾何分布 久保川 統計学入門 §4.3.4(p.97))

整数値をとる離散型確率変数 X が，パラメタ $0 < p < 1$ として，確率関数

$$p(x) = \begin{cases} 0 & (x < 0) \\ (1-p)^x p & (0 \leq x) \end{cases}$$

を持つとき，X はパラメタ p の幾何分布 $Geo(p)$ にしたがうという．

表の出る確率が p のコインを繰り返し投げ続ける中で，初めて表が出るまでに要した裏の回数 X のしたがう分布．

L03-Q3

Quiz(離散的な確率変数の累積分布関数)

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

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

1. 累積分布関数 $F(x)$ を，整数 x に対して求めよう．

連続型一様分布・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(2025-05-12 Mon)

最終更新: Time-stamp: "2025-05-17 Sat 08:29 JST hig"

今日の目標

- 母平均値母標準偏差の意味，scipy.stats の位置母数 location 尺度母数 scale との対応を説明できる
- 一様分布 久保川 統計学入門 §5.3.1 の意味が説明できる
- 確率変数の標準化 久保川 統計学入門 p.120 の意味が説明できる

L03-Q1

Quiz 解答: 指数分布の累積分布関数・確率・分位数

1.

$$
\begin{aligned}
F(x) =& P(X \le x) = \int_{-\infty}^{x} f(x')\,\mathrm{d}x' \\
=& \begin{cases} \int_{-\infty}^{x} 0\mathrm{d}x' = 0, & (x < 0) \\ \int_{-\infty}^{0} 0\mathrm{d}x' + \int_{0}^{x} 3\mathrm{e}^{-3x'}\mathrm{d}x' = 0 + (1 - \mathrm{e}^{-3x}) & (x \ge 0). \end{cases}
\end{aligned}
$$

emark 広義単調増加, $F(-\infty) = 0, F(+\infty) = +1, F'(x) = f(x)$ 成立？

2. $P(X \le 4) = F(4) - F(-\infty) = 1 - \mathrm{e}^{-3\cdot 4} - 0 = 1 - \mathrm{e}^{-12}$.

emark $\int_{-\infty}^{4} f(x)\,\mathrm{d}x = \int_{0}^{4} 3\mathrm{e}^{-3x}\mathrm{d}x$ を直接計算しても求められが，$F(x)$ を求めておくと，積分は一度で済む．

3. $P(-5 < X \le 2) = F(2) - F(-5) = 1 - \mathrm{e}^{-3\cdot 2} - 0 = 1 - \mathrm{e}^{-6}$.

emark $\int_{-5}^{2} f(x)\,\mathrm{d}x = \int_{0}^{2} 3\mathrm{e}^{-3x}\mathrm{d}x$ を直接計算しても求められる．

4. $F(d) = \frac{1}{7}$ を解いて，$d = -\frac{1}{3}\log\frac{6}{7} = \frac{1}{3}\log\frac{7}{6} (> 0)$.

emark $\int_{-5}^{d} f(x)\,\mathrm{d}x = \int_{0}^{d} 3\mathrm{e}^{-3x}\mathrm{d}x = \frac{1}{7}$ を直接解いても求められる.

L03-Q2

Quiz 解答: 指数分布

間隔 X 分 は, パラメタ $\lambda = 1/30$ の指数分布にしたがう.

1. $P(5 < X \leq 10) = F(10) - F(5) = \int_{5}^{10} \frac{1}{30}\mathrm{e}^{-\frac{1}{30}x}\,\mathrm{d}x = \mathrm{e}^{-5/30} - \mathrm{e}^{-10/30} =$ 数値.
2. $P(X > 80) = 1 - F(80) = \int_{80}^{+\infty} \frac{1}{30}\mathrm{e}^{-\frac{1}{30}x}\,\mathrm{d}x = \mathrm{e}^{-80/30} =$ 数値.
3. $\mathrm{E}[X^1] = \int_{-\infty}^{+\infty} x \cdot f(x)\mathrm{d}x = \int_{0}^{+\infty} x \cdot \lambda\mathrm{e}^{-\lambda x}\,\mathrm{d}x = \lambda = 30.$
4. 分位数関数は $q = F(x) = \mathrm{e}^{-x/30}$ を解いて，
   $x = F^{-1}(q) = -30\log q$. よって，$30\log(3/2)$ 分.

ここまで来たよ

3 累積分布関数・分位数・指数分布・幾何分布

4 連続型一様分布・確率変数の標準化
- グラフの平行移動拡大縮小・確率分布の位置と広がり
- 連続型一様分布
- 確率変数の変数変換
- 確率変数の標準化

復習：関数のグラフの平行移動拡大縮小

次の 2 次関数のグラフを考えよう．$x = b$ が対称軸．$a = 1/\sqrt{a'}$ は「広がり」

$$y = -a'(x-b)^2 + c = -(\tfrac{x-b}{a})^2 + c.$$

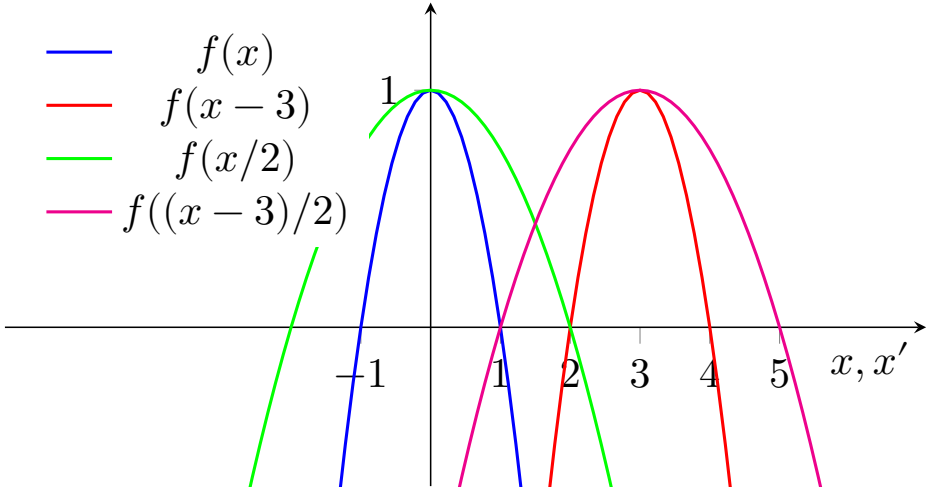
基本のグラフ $a = 1, b = 0 \rightsquigarrow y = f(x) = -x^2 + c$

x 方向に b だけ平行移動 $x = x' - b \quad y = f(x' - b)$

$x = 0$ を中心に x 方向に a 倍だけ拡大 $x = \frac{x'}{a} \quad y = f(\frac{x'}{a})$

まず $x = 0$ を中心に x 方向に a 倍だけ拡大，次に

x 方向に b だけ平行移動 $x = \frac{x'-b}{a} \quad y = f(\frac{x'-b}{a})$



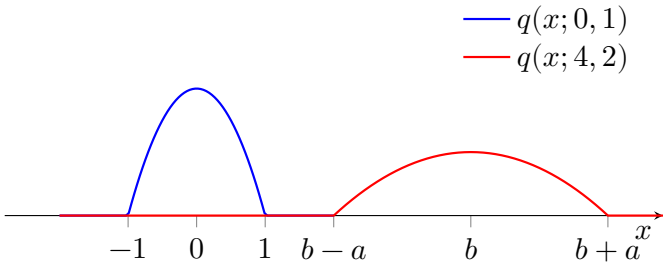
L04-Q1

Quiz(位置とスケールを持つ連続型確率変数)

連続型確率変数 X が，次の確率密度関数 $f(x) = q(x; b, a)$ を持つとき，分布 $Q(b, a)$ にしたがうと言うことにする．ここで，$b \in \mathbb{R}, a > 0$ の定数（パラメタ）である．

$$q(x; b, a) = \begin{cases} \frac{3}{4a}(1 - (\frac{x-b}{a})^2) & (|x - b| \leq a) \\ 0 & (\text{他}) \end{cases}$$

1. $b = 0$ のとき モーメント $\mathrm{E}[X^k]$ を求めよう．
2. $b = 0$ のとき母標準偏差 $(\mathrm{Var}(X))^{1/2}$ を求めよう．
3. 一般の b で母平均値 $\mathrm{E}[X]$ を求めよう．


$q(x;0,1)$
$q(x;4,2)$
-1
0
1
$b-a$
b
$b+a$
x

確率分布の位置と広がり

$X' \sim q(x;0,a)$ は，$q(x;0,1)$ を $x=0$ を中心に x 方向に a 倍に拡大したもの．
母標準偏差 $\mathrm{Var}(X')^{1/2}$ は，$q(x';0,a)$ の「広がり」を表している．尺度母数

$q(x;b,1)$ は，$q(x;0,1)$ を x 軸の方向に b だけ平行移動したもの．
$X' \sim q(x;b,1)$ のとき，変数変換 $x+b=x'$ により，

$$\mathrm{E}[X'] = \int_{-\infty}^{\infty} x' q(x';b,1)\mathrm{d}x' \int_{-\infty}^{\infty} (x+b)q(x;0,1)\mathrm{d}x = 0+b.$$

母平均値 $\mathrm{E}[X']$ は，$q(x';b,1)$ の「位置」を表している．位置母数
母平均値は分布の位置，母標準偏差は分布の広がりを表す．

ここまで来たよ

3 累積分布関数・分位数・指数分布・幾何分布

4 連続型一様分布・確率変数の標準化
- グラフの平行移動拡大縮小・確率分布の位置と広がり
- 連続型一様分布
- 確率変数の変数変換
- 確率変数の標準化

連続型一様分布 久保川 統計学入門 §4.3.1 I

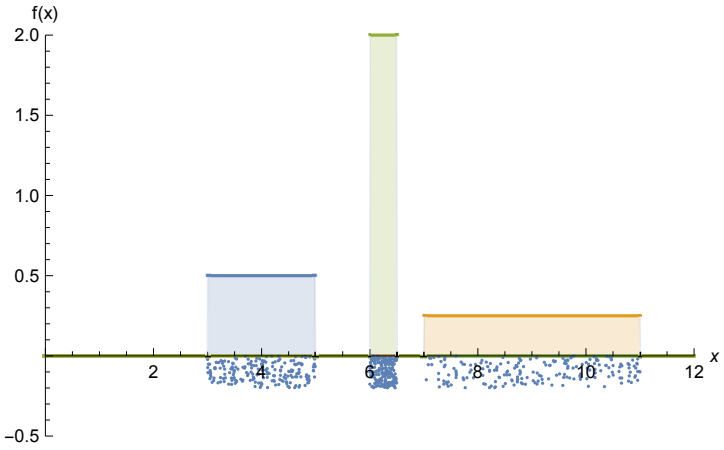
(連続型) 一様分布 $U(c,d)$

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 $U(c,d)$ にしたがうという.

$$f(x) = u(x;b,a) = \begin{cases} \frac{1}{a} & (b - \frac{1}{2}a \leq x \leq b + \frac{1}{2}a) \\ 0 & (\text{他}) \end{cases},$$

$$u(x;\tfrac{c+d}{2}, d-c) = \begin{cases} \frac{1}{d-c} & (c \leq x \leq d) \\ 0 & (\text{他}) \end{cases}$$

連続型一様分布 久保川 統計学入門 §4.3.1 II



$$U(4,6), U(6,6.5), U(7,11).$$

連続型一様分布

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

Colab LearnMoodle 連続型確率変数-連続型一様分布.ipynb loc=位置，scale=尺度

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 $U(c,d)$ 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 $\mathrm{Var}(X)^{1/2}$ を求めよう.

さっき想像した，母平均値・母標準偏差の意味とマッチしてる?

		母ナントカ	scipy.stats
位置	位置母数	$\mathrm{E}[X^1] = \frac{c+d}{2}$	loc=c
広がり	尺度母数	$\mathrm{Var}(X)^{1/2} = \frac{d-c}{\sqrt{12}}$	scale=$d-c$

$U(c,d)$ に対するこの結果は, 公式のように記憶して使おう.

久保川 統計学入門 発展問題問 6(p.125)

ここまで来たよ

3 累積分布関数・分位数・指数分布・幾何分布

4 連続型一様分布・確率変数の標準化
- グラフの平行移動拡大縮小・確率分布の位置と広がり
- 連続型一様分布
- 確率変数の変数変換
- 確率変数の標準化

変数変換 $X' = 2X + 4$

変換前 X, 変換後 X'

$X \sim U(0,3)$

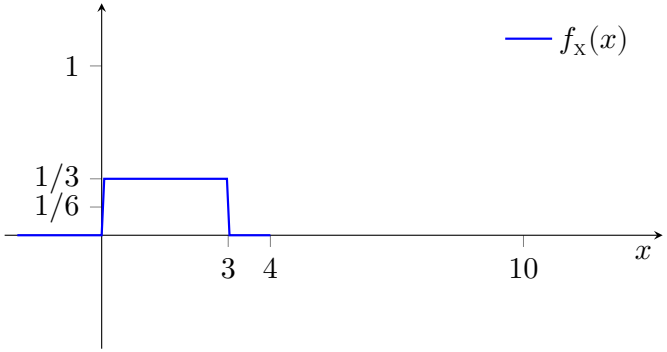
$$f_{\mathrm{X}}(x) = \begin{cases} \frac{1}{3} & (0 \le x \le 3) \\ 0 & (\text{他}) \end{cases}$$

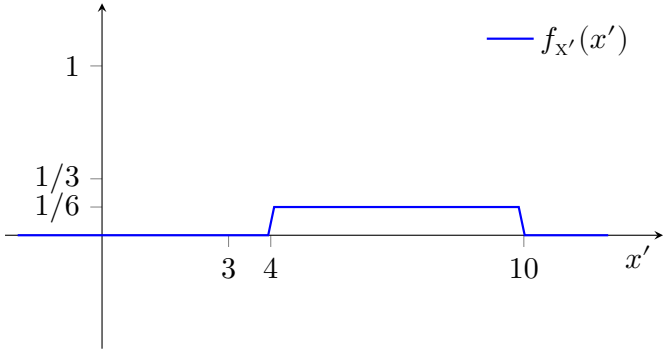
$X = \frac{X'-4}{2}$,　$X' = 2X + 4$, 2 倍して 4 平行移動

$X' \sim U(4,10)$ 一様分布のまま!

$$f_{\mathrm{X}'}(x') = \begin{cases} \frac{1}{2} \cdot \frac{1}{3} & (0 \le \frac{x'-4}{2} \le 3) \\ 0 & (\text{他}) \end{cases}$$

$$= \begin{cases} \frac{1}{6} & (4 \le x' \le 10) \\ 0 & (\text{他}) \end{cases}$$





x	0	3
x'	4	10

解説

復習

変換前 X, 変換後 $X' = Y = aX + b$.

命題 ($Y = aX + b$ の母平均値と母分散)

$$\begin{aligned}
\mathrm{E}[Y] =& \mathrm{E}[aX + b] = a\mathrm{E}[X] + b \\
\mathrm{Var}(Y) =& \mathrm{Var}(aX + b) = a^2\mathrm{Var}(X) \\
\mathrm{Var}(Y)^{1/2} =& \mathrm{Var}(aX + b)^{1/2} = |a|\mathrm{Var}(X)^{1/2}
\end{aligned}$$

これを使って，Y の母平均値母分散求められる？

母期待値の 2 つの計算方法

変換前 $X \sim U(0,3)$

計算方法 1 一様分布だから…

$\mathrm{E}[X] = \frac{0+3}{2}$

$\mathrm{Var}(X) = \frac{(3-0)^2}{12}$.

変換後 $Y = 2X + 4$

計算方法 1 変換後も一様分布になるから…$Y = 2X + 4$, 2 倍して 4 平行移動.

$Y \sim U(4,10)$

$\mathrm{E}[Y] = \frac{4+10}{2} = 7$

$\mathrm{Var}(Y) = \frac{(10-4)^2}{12} = 3$.

計算方法 2 ぜんぶ X の母期待値で書き直す

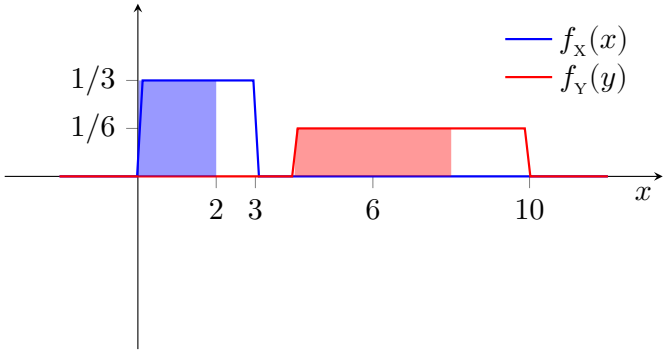
$\mathrm{E}[Y] = \mathrm{E}[2X+4] = 2\mathrm{E}[X] + 4 = 7$.

$\mathrm{Var}(Y) = \mathrm{Var}(2X+4) = 2^2\mathrm{Var}(X) = 3$.

確率の計算方法

$P(X \leq 2) = ?$
変換前 $X \sim U(0,3)$
計算方法 1 一様分布だから…
$P(X \leq 2) = \int_{-\infty}^{2} f_{\mathrm{X}}(x)\mathrm{d}x = \int_{0}^{2} \frac{1}{3}\mathrm{d}x = \frac{2}{3}$.

x	0	2	3
y	4	8	10



$P(Y \leq 8) = ?$
変換後 $Y = 2X + 4$
計算方法 1 変換後も一様分布になるから…$Y = 2X + 4$, $Y \sim U(4,10)$
$P(Y \leq 8) = \int_{-\infty}^{8} f_{\mathrm{Y}}(y)\mathrm{d}y = \int_{4}^{8} \frac{1}{6}\mathrm{d}y = \frac{2}{3}$.
計算方法 2 X の確率で書き直す
$P(Y \leq 8) = P(X \leq 2) = \int_{-\infty}^{2} f_{\mathrm{X}}(x)\mathrm{d}y = \int_{0}^{2} \frac{1}{3}\mathrm{d}x = \frac{2}{3}$.

ここまで来たよ

3 累積分布関数・分位数・指数分布・幾何分布

4 連続型一様分布・確率変数の標準化
- グラフの平行移動拡大縮小・確率分布の位置と広がり
- 連続型一様分布
- 確率変数の変数変換
- 確率変数の標準化

確率変数の標準化 久保川 統計学入門 p.120

定義 (標準化された確率変数)

確率変数 Z が, $\mathrm{E}[Z]=0, \mathrm{Var}(Z)=1^2$ を満たすとき, 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 次式で，標準化された確率変数に変換できる．
このとき，確率密度関数のグラフは，拡大，移動しただけで，形は同じ．

定義 (確率変数の標準化 久保川 統計学入門 p.120)

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

$\mathrm{E}[Z]=\mathrm{E}[\frac{X-\mu}{\sigma}]=\boxed{0}$,
$\mathrm{Var}(Z)=\mathrm{Var}(\frac{X-\mu}{\sigma})d=\boxed{1}$.

L04-Q3

連続型一様分布の標準化

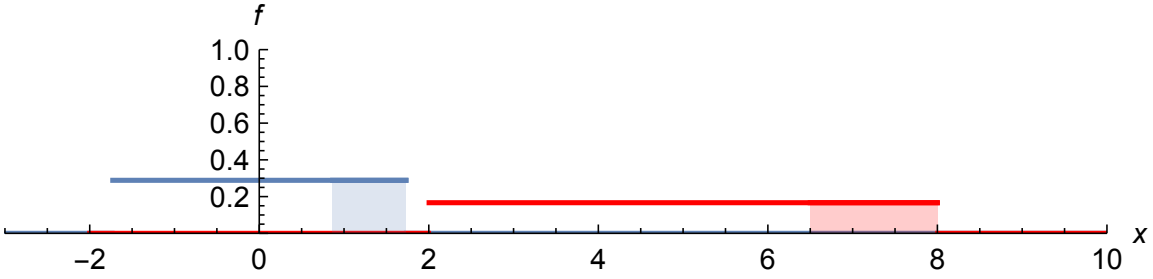
$X \sim U(2,8)$ を標準化しよう.

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

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

$$f_{\mathrm{Z}}(z) = \begin{cases} \frac{1}{2\sqrt{3}} & (-\sqrt{3} \leq z \leq \sqrt{3}) \\ 0 & (\text{他}) \end{cases}$$

標準連続型一様分布 $Z \sim U(-\sqrt{3}, +\sqrt{3})$. 確率密度関数 $u(x; 0, \sqrt{3})$.



x	2	8
z	$-\sqrt{3}$	$+\sqrt{3}$

確率変数の標準化とは

確率密度関数 $f_{\mathrm{Y}}(x)$ 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

逆に，標準化された 変換前 Z から 変換後 $X = aZ + b$ を考えると，
$\mathrm{E}[X] = b$, $\mathrm{V}[X] = a^2$.

命題

標準化された連続型一様分布 $Z \sim U(-\sqrt{3}, +\sqrt{3})$ から，$X = aZ + b$ で，すべての連続型一様分布 $X \sim U(c, d)$ を作れる．

L04-Q4

Quiz(連続型一様分布)

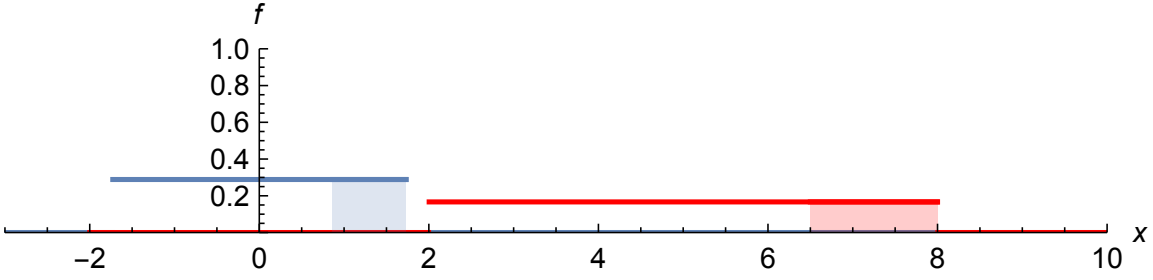
連続型確率変数 $Z \sim U(-\sqrt{3}, \sqrt{3})$ に対して, $X = \sqrt{3}Z + 5$ を考える.

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. $\mathrm{Var}(X)$ を求めよう.

それぞれ 2 つの方法で.

確率変数の変数変換 $X = aZ + b$ の意味

変換前 $X \sim U(0,3)$, 変換後 $Y = 2X + 4 \sim U(4,10)$ の話を,
変換前 $Z \sim U(-\sqrt{d}, +\sqrt{3})$, 変換後 $X = aZ + b$ でリプレイ.
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)
左 $Z \sim U(-\sqrt{3}, \sqrt{3})$ 右 $X = aZ + b = \sqrt{3}Z + 5 \sim U(2,8)$ の確率密度関数



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

対応する確率=塗った箇所の面積は同じ

$$P\left(\frac{13}{2} < X \le 8\right) = P\left(\frac{\frac{13}{2} - 5}{\sqrt{3}} < Z \le \frac{8-5}{\sqrt{3}}\right).$$

z	$-\sqrt{3}$	$\frac{\sqrt{3}}{2}$	$\sqrt{3}$
x	2	$\frac{13}{2}$	8

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(2025-05-19 Mon)

最終更新: Time-stamp: "2025-05-27 Tue 07:34 JST hig"

今日の目標

- 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． 久保川 統計学入門 §5.3.2

L04-Q1

Quiz 解答: 位置とスケールを持つ連続型確率変数

1. $q(x;0,s)$ は偶関数であることに注意すると，k 次のモーメントの被積分関数は，k が奇のとき奇関数で $\mathrm{E}[X^k]=0$.
   $k=2\ell$ が偶関数のとき，

$$\begin{aligned}\mathrm{E}[X^k] &= \int_{-a}^{+a} x^k \tfrac{3}{4a}(1-(\tfrac{x}{a})^2)\mathrm{d}x = \frac{3}{4a}\cdot 2\int_0^a x^k(1-(\tfrac{x}{a})^2)\mathrm{d}x \\ &= \frac{3}{2a}[\tfrac{1}{k+1}x^{k+1} - a^{-2}\tfrac{1}{k+3}x^{k+3}]_0^a \\ &= \frac{3a^k}{(k+1)(k+3)}\end{aligned}$$

2. $\mathrm{E}[X^1]=0$. $\mathrm{Var}(X)=\mathrm{E}[X^2]-\mathrm{E}[X^1]^2=\frac{3}{15}a^2-0^2$. よって標準偏差は $\mathrm{Var}(X)=5^{-1/2}a\simeq 0.45a$.
3. $b\neq 0$ のとき，$\mathrm{E}[X^1]=b$, $\mathrm{E}[X^2]=\frac{1}{5}a^2+b^2$.

$\mathrm{Var}(X) = \mathrm{E}[(X-\mu)^2] = 5^{1/2}a^2, (\mathrm{Var}(X))^{1/2} = 5^{-1/2}a.$

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. $\mathrm{Var}(X) = \mathrm{E}[X^2] - \mathrm{E}[X^1]^2 = \frac{(d-c)^2}{12}.$ $\mathrm{Var}(X)^{1/2} = \frac{d-c}{\sqrt{12}} \simeq \frac{d-c}{3.5}.$

L04-Q3

Quiz 解答: 連続型一様分布

1. $X \sim U(\sqrt{3}(-\sqrt{3})+5, \sqrt{3}(+\sqrt{3})+5)$, すなわち，$X \sim U(2,8)$. 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. $\mathrm{Var}(X) = \frac{(8-2)^2}{12} = 3$, または, $\mathrm{Var}(X) = (\sqrt{3})^2\mathrm{Var}(Z) = 3\cdot 1.$

ここまで来たよ

4 連続型一様分布・確率変数の標準化

5 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 $N(0,1^2)$ の性質

定義 (標準正規分布 $N(0,1^2)$ 久保川 統計学入門 式 (5.2))

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

$$f_{\mathrm{Z}}(z) = \phi(z) = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{z^2}{2}}.$$

累積分布関数

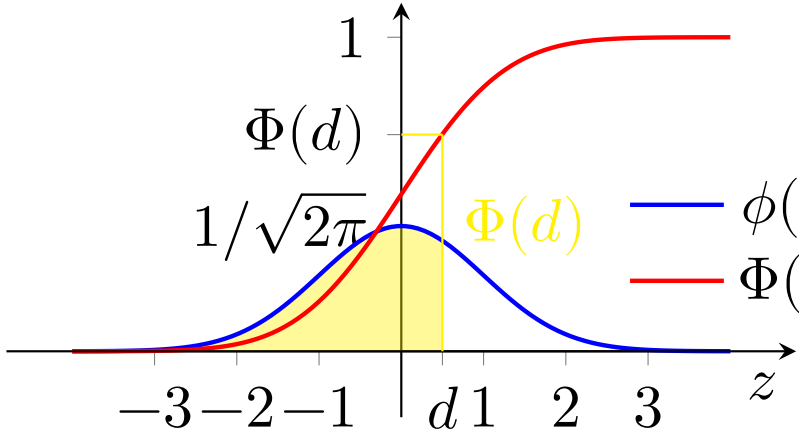
$$F_{\mathrm{Z}}(z) = \Phi(z) = \int_{-\infty}^{z} \phi(z')\mathrm{d}z'.$$

ϕ, Φ は f, F のギリシャ文字．あまりに有名分布なので専用文字を使う．
$\Phi(z)$ の積分は具体的に書けない．
Python で `scipy.stats.norm(loc=0,scale=1)`
Excel で`=norm.s.dist(z,FALSE)`

標準正規分布 $Z \sim N(0, 1^2)$ の確率密度関数と母期待値

$\phi(z)$ は偶関数（y 軸に関して対称），$|z| \to \pm\infty$ で z 軸に漸近．

$\phi(0)\times$	$\phi(\pm1)/\phi(0)$	$\phi(\pm2)/\phi(0)$	$\phi(\pm3)/\phi(0)$	$\phi(\pm\infty)/\phi(0)$
$(2\pi)^{-1/2}\times$	$\mathrm{e}^{-1/2}$	$\mathrm{e}^{-4/2}$	$\mathrm{e}^{-9/2}$	$\mathrm{e}^{-\infty}$
$0.4\times$	0.6	0.14	0.01	$\to 0$

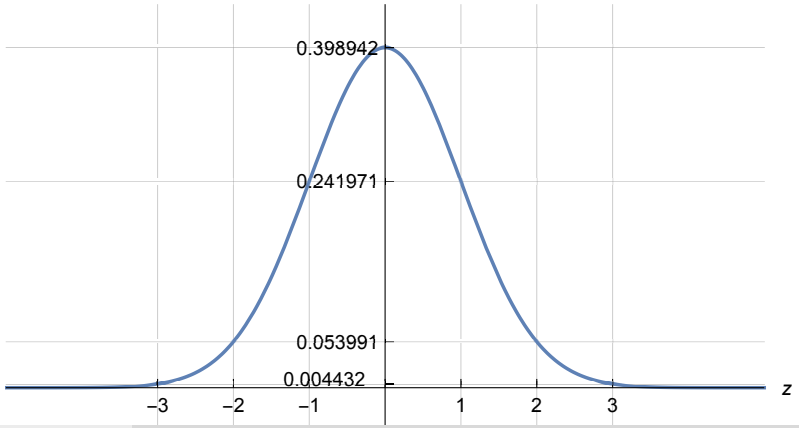


k 次のモーメント (k: 自然数)

$$
\begin{aligned}
\mathrm{E}[Z^{2k-1}] &= 0, \quad \text{奇関数} \\
\mathrm{E}[Z^{2k}] &= (2k-1)!! \text{部分積分} \\
&= 1 \cdot 3 \cdot 5 \cdots \cdot (2k-1), \\
\mathrm{E}[Z^0] &= 1, \\
\mathrm{E}[Z] &= 0, \\
\mathrm{Var}(Z) &= 1.
\end{aligned}
$$

久保川 統計学入門 p.111 微積分 II

久保川 統計学入門 公式 5.17



たしかに，Z は標準化された確率変数．

標準正規分布の確率と $\Phi(z)$ の数値

$Z \sim N(0, 1^2)$ のとき,

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

$\Phi(z)$, 分位数関数 確率統計 I(2025)L03 $\Phi^{-1}(q)$ は LearnMoodle Python による正規分布や 久保川 統計学入門 公式 5.9, 数値表 (p.290) から．数値表は $z \geq 0$ だけなので $z < 0$ は，$\Phi(-z) + \Phi(z) = 1$ を利用して求める． 高校 数学 C

```
from scipy import stats
rvz=stats.norm(loc=0,scale=1) # Z ~ N(0,1^2)
rvz.cdf(d) # Φ(d)
rvz.ppf(q) # Φ^{-1}(q)
```

z	$\Phi^{-1}(q)$	$-\infty$	$-z'$	0	$+\infty$
$\Phi(z)$	q	0	$1 - \Phi(z')$	$1/2$	1

上側分位点 $z_p = \Phi^{-1}(1-p)$ は $P(Z > z_p) = p$ となる点 久保川 統計学入門 公式 5.9

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 $P(Z \leq 1.23)$ を累積分布関数 $\Phi(z)$ 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 $\Phi(z)$ で表そう.
3. 確率 $P(-1.23 < Z \leq 0)$ を，累積分布関数 $\Phi(z)$ ただし, $z > 0$ で表そう.
4. 確率 $P(Z > d) = 0.025$ となる d を累積分布関数 $\Phi(z)$ の逆関数で表そう．Python または表で求めよう.

久保川 統計学入門 例題 5.12

ここまで来たよ

4. 連続型一様分布・確率変数の標準化

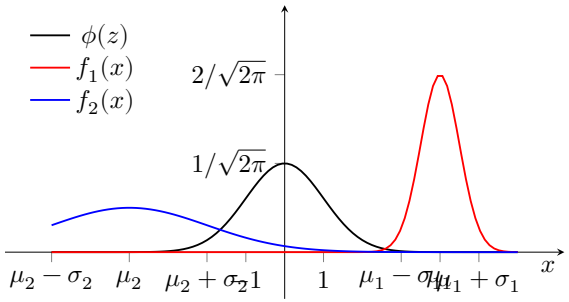
5. 正規分布
   - 標準正規分布
   - 一般の正規分布

一般の正規分布 $N(b, a^2)$

$Z \sim N(0, 1^2)$ に対して，$X = aZ + b$ を考える． 久保川 統計学入門 公式 5.17

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

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



定義 (一般の正規分布 $N(\mu, \sigma^2)$ 久保川 統計学入門 式 (5.1))

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 $N(\mu, \sigma^2)$ (normal distribution) にしたがうという．

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

Python で `scipy.stats.norm(loc=`b`,scale=`a`)`.
$X \sim N(b, a^2)$ を標準化すると $Z \sim N(0, 1^2)$.

L05-Q2

Quiz(正規分布の確率)

連続型確率変数 X が, 次の確率密度関数を持つ.

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

1. $\mathrm{E}[X]$ を求めよう.
2. $\mathrm{Var}(X)$ 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.

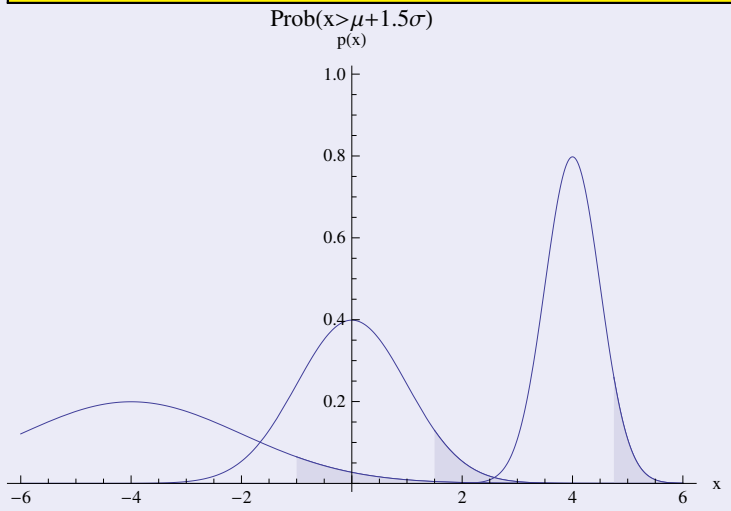
一般の正規分布 $N(\mu, \sigma^2)$ の確率

命題 (変数変換前後の確率は同じ 久保川 統計学入門 公式 5.10)

確率は変数変換，特に標準化 $Z = \frac{X-\mu}{\sigma}$ の前後で変わらない．

斜線部の面積はどれも同じ

$$
\begin{aligned}
P(c < X \leq d) &= P\left(\frac{c-\mu}{\sigma} < \frac{X-\mu}{\sigma} \leq \frac{d-\mu}{\sigma}\right) \\
&= P\left(\frac{c-\mu}{\sigma} < Z \leq \frac{d-\mu}{\sigma}\right) \\
&= \Phi\left(\frac{d-\mu}{\sigma}\right) - \Phi\left(\frac{c-\mu}{\sigma}\right)
\end{aligned}
$$



変数				下限	上限	母平均値	母分散
z	-2	0	2	$\frac{c-\mu}{\sigma}$	$\frac{d-\mu}{\sigma}$	0	1^2
x	$\mu - 2\sigma$	μ	$\mu + 2\sigma$	c	d	μ	σ^2

変数変換しても確率が同じことの別説明 (置換積分)

$X \sim N(\mu, \sigma^2)$ 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 $X \sim N(3, 2^2)$, $Z \sim N(0, 1^2)$ 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.

久保川 統計学入門 例題 5.13

多変数の確率変数

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(2025-05-26 Mon)

最終更新: Time-stamp: "2025-05-26 Mon 08:22 JST hig"

今日の目標

- 同時分布, 周辺分布を説明できる 久保川 統計学入門 §6.1,§6.3
- 同時分布から周辺分布, 母期待値, 母共分散を求められる 久保川 統計学入門 §6.2
- 独立性を使って計算できる 久保川 統計学入門 §6.4

L05-Q1

Quiz 解答: 標準正規分布の確率

`rvz=stats.norm(loc=0,scale=1)` とする.

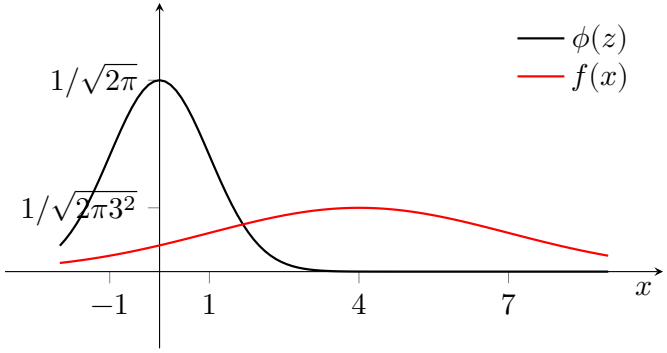
1. $P(Z \leq 1.23) = \int_{-\infty}^{1.23} \phi(z)\,\mathrm{d}z = \Phi(1.23) - \Phi(-\infty) = \Phi(1.23) - 0 =$ `rvz.cdf(1.23)` $= 0.5708$.
2. $P(-0.56 < Z \leq +1.23) = \int_{-0.56}^{1.23} \phi(z)\,\mathrm{d}z = \Phi(1.23) - \Phi(-0.56) =$ rvz.cdf(1.23) − rvz.cdf(-0.56) $= 0.6030$.
3. $\Phi(0) - \Phi(-1.23) = \frac{1}{2} - (1 - \Phi(1.23)) = \Phi(1.23) - \frac{1}{2}$.
4. $0.025 = P(Z > d) = P(d < Z < +\infty) = \Phi(+\infty) - \Phi(d) = 1 - \Phi(d)$.
   よって, $\Phi(d) = 1 - 0.025$.
   $d = F^{-1}(1 - 0.025) =$ `rvz.ppf(0.975)` $= 1.960$.

L05-Q2

Quiz 解答: 正規分布の確率

1. $\mathrm{E}[X] = 4$.

2. $\mathrm{Var}(X) = 3^2$.
3. $x = 4$ を対称軸とする，幅 3 くらいの正規分布の確率密度関数のグラフ．標準正規分布のグラフを, 横に 3 倍, 縦に 1/3 倍, 右に 4 平行移動.

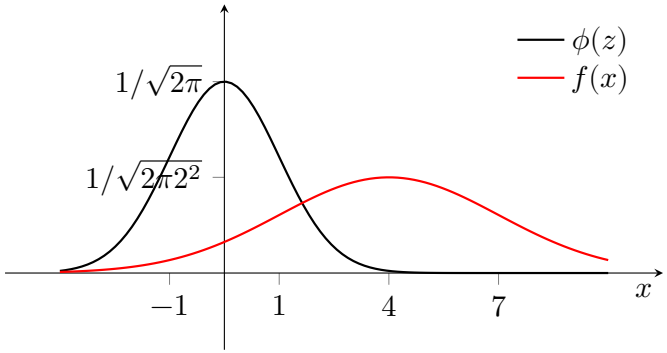


L05-Q3

Quiz 解答: 正規分布の確率

X を標準化すると, $Z = \frac{X-3}{2}$.

1. $f_{\mathrm{X}}(x) = \frac{1}{\sqrt{2\pi\cdot 2^2}}\mathrm{e}^{-\frac{(x-3)^2}{2\cdot 2^2}}$, $\phi(z) = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{z^2}{2}}$. 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), $\Phi(z)$ =scipy.stats.norm(loc=0,scale=1).cdf(z) とかく.



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

ここまで来たよ

5 正規分布

6 多変数の確率変数
- 2 変数の確率分布
- 確率変数の独立性

2 つの離散型確率変数の同時確率分布 高校 数学 B 久保川 統計学入門 §6.1

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.

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. P(条件) 型の表記で $P(X = x, Y = y)$ とも. 表で書いた方がまし. ここでは,「他」は省略.

表の縦軸横軸に意味なし. ラベル X, Y をよく見て.

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

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

今まで時々見た，「$X = 2Y + 7$」のような式は，同時分布が次の形であることを意味する．

$y \backslash x$	7	8	9
0	p	0	0
1	0	0	$1-p$

定義 (同時分布の確率 久保川 統計学入門 §6.1)

条件 $a < X \leq b$ かつ $c < Y \leq d$ の成り立つ確率は，

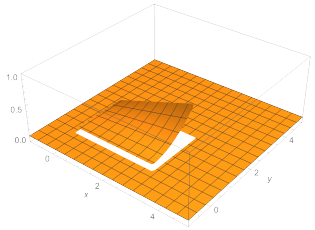
$$P(a < X \leq b \text{ かつ } c < Y \leq d) = \sum_{a<x\leq b} \sum_{c<y\leq d} p(x, y)$$

定義 (同時分布の母期待値 久保川 統計学入門 §6.2)

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

多変数の連続型確率分布

2 変数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 (連続型同時分布の確率)

$$P(a < X \leq b \text{ かつ } c < Y \leq d) = \int_b^a \int_c^d f(x, y)\, \mathrm{d}x\mathrm{d}y$$

平面の領域上にある部分の体積

定義 (連続型同時分布の母期待値)

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

L06-Q1

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 (多変数の確率分布の母期待値の性質高校 数学 B 久保川 統計学入門 §6.2**)**

$$\mathrm{E}[g_1(X,Y)+g_2(X,Y)] = \mathrm{E}[g_1(X,Y)] + \mathrm{E}[g_2(X,Y)] \tag{6.1}$$

$$\text{特に}\mathrm{E}[X+Y] = \mathrm{E}[X] + \mathrm{E}[Y] \tag{6.2}$$

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 久保川 統計学入門 §6.1

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,

$$\text{離散型} p_{\mathrm{X}}(x) = p(x,\bullet) = \sum_y p(x,y), \quad p_{\mathrm{Y}}(y) = p(\bullet,y) = \sum_x p(x,y)$$

$$\text{連続型} f_{\mathrm{X}}(x) = \int_{-\infty}^{+\infty} p(x,y)\,\mathrm{d}y, \quad f_{\mathrm{Y}}(y) = \int_{-\infty}^{+\infty} p(x,y)\,\mathrm{d}x$$

要するに 一方を無視した分布. 小計.

$y\backslash x$	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-Q2

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 $\mathrm{Var}(X)$ を求めよう.

母共分散 高校 数学 C 久保川 統計学入門 §6.2

定義 (母共分散 covariance 久保川 統計学入門 §6.2)

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

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

$\mathrm{Var}(X) = \mathrm{E}[(X-\mu_{\mathrm{X}})(X-\mu_{\mathrm{X}})] = \mathrm{E}[X^2] - \mathrm{E}[X]\mathrm{E}[X].$
$\mathrm{Var}(X) = \mathrm{Cov}(X,X).$
L06-Q3

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 変数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 変数の確率分布の母分散のよくある間違い

$$\mathrm{Var}(X+Y) \neq \mathrm{Var}(X) + \mathrm{Var}(Y)$$

母分散を, 母期待値と区別しよう.

ここまで来たよ

5 正規分布

6 多変数の確率変数
- 2 変数の確率分布
- 確率変数の独立性

確率変数の独立性 高校 数学 C 久保川 統計学入門 §6.3

定義 (事象の独立性 久保川 統計学入門 定義 3.11)

事象 A, B が独立 (independent) とは次が成立すること.

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

例:A=赤のサイコロの目が 1, B=白のサイコロの目が 1

定義 (確率変数の独立性 久保川 統計学入門 定義 6.9)

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，直観的に

- 同時分布が, 周辺分布 だけから (積で) 決まっちゃうこと
- 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
- X, Y が互いに「無関係」であること

独立である例

例

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

$=$

$y\backslash x$	$p_{\mathrm{X}}(0)$	$p_{\mathrm{X}}(1)$
$p_{\mathrm{Y}}(10)$	$p_{\mathrm{X}}(0)p_{\mathrm{Y}}(10)$	$p_{\mathrm{X}}(1)p_{\mathrm{Y}}(10)$
$p_{\mathrm{Y}}(11)$	$p_{\mathrm{X}}(0)p_{\mathrm{Y}}(11)$	$p_{\mathrm{X}}(0)p_{\mathrm{Y}}(11)$

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

L06-Q4

Quiz(離散型確率変数の独立性)

独立な離散型確率変数 X, Y がある. 同時分布 $p(x, y)$ は次の表で与えられる. 空欄の確率を求めよう.

$y\backslash x$	3	4	計
13			$\frac{2}{7}$
14			
計		$\frac{7}{10}$	

命題 (復習:X, Yが独立でなくても成立する性質 久保川 統計学入門 §6.2)

$$\mathrm{E}[g_1(X,Y)+g_2(X,Y)] = \mathrm{E}[g_1(X,Y)] + \mathrm{E}[g_2(X,Y)] \tag{6.3}$$

$$\text{特に}\mathrm{E}[X+Y] = \mathrm{E}[X] + \mathrm{E}[Y] \tag{6.4}$$

$$\mathrm{E}[g_1(X)\times g_2(Y)] \neq \mathrm{E}[g_1(X)]\times \mathrm{E}[g_2(Y)] \tag{6.5}$$

$$\text{母共分散}\mathrm{Cov}(X,Y) \overset{\text{定義}}{=} \mathrm{E}[(X-\mu_{\mathrm{X}})(Y-\mu_{\mathrm{Y}})] \tag{6.6}$$

$$= \mathrm{E}[X\times Y] - \mathrm{E}[X]\times \mathrm{E}[Y] \tag{6.7}$$

L06-Q5

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が独立のときに追加で成立する性質 久保川 統計学入門 公式 6.13)

$$g(X,Y) = g_1(X) \times g_2(Y) \text{と因数分解されるとき}$$

$$\mathrm{E}[g(X,Y)] = \mathrm{E}[g_1(X) \times g_2(Y)] = \mathrm{E}[g_1(X)] \times \mathrm{E}[g_2(Y)] \tag{6.8}$$

$$\text{特に}\mathrm{E}[X \times Y] = \mathrm{E}[X] \times \mathrm{E}[Y] \tag{6.9}$$

$$0 = \mathrm{Cov}(X,Y) = \mathrm{E}[XY] - \mathrm{E}[X] \times \mathrm{E}[Y] \tag{6.10}$$

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) \tag{6.11}$$

$$\mathrm{Var}(aX+bY+c) = a^2\mathrm{Var}(X) + b^2\mathrm{Var}(Y) \tag{6.12}$$

$g_1(X)$ には Yが，$g_2(Y)$ には X が含まれない．こう書いたらそういう意味．

$\mathrm{Cov}(,)$ は「独立でない度」ただし，$\mathrm{Cov}(X,Y) = 0$ は独立の**必要条件**(だが**十分条件**ではない) 久保川 統計学入門 公式 6.14

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

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

$$
\begin{aligned}
\mathrm{Var}(X+Y) &=\mathrm{E}[(X+Y)^2]-\mathrm{E}[X+Y]^2 \\
&=\mathrm{E}[X^2]+2\mathrm{E}[XY]+\mathrm{E}[Y^2]-(\mathrm{E}[X]^2+2\mathrm{E}[X]\mathrm{E}[Y]+\mathrm{E}[Y]^2) \\
&=\mathrm{Var}(X)+2\mathrm{Cov}(X,Y)+\mathrm{Var}(Y) \\
&=\mathrm{Var}(X)+0+\mathrm{Var}(Y).
\end{aligned}
$$

L06-Q6

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.
$\mathrm{E}[X] = 2, \mathrm{E}[Y] = 3, \mathrm{Var}(X) = 5, \mathrm{Var}(Y) = 11$ 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. $\mathrm{Var}(-2X + 3Y)$ を求めよう.

二項分布・独立同一分布の和

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　L07(2025-06-02 Mon)

最終更新: Time-stamp: "2025-06-02 Mon 06:46 JST hig"

今日の目標

- 独立同一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 久保川 統計学入門 §6.5, p.152
- 二項分布の母期待値, 母分散, 確率を求められる

久保川 統計学入門 §4.3.1,§4.3.2

L06-Q1 L06-Q2 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. $\mathrm{E}[2X^2 + \mathrm{e}^Y] = 2\mathrm{E}[X^2] + \mathrm{E}[\mathrm{e}^Y] \overset{\text{周辺分布}}{=} 2(1^2 \cdot \frac{4}{12} + 2^2 \cdot \frac{2}{12} + 3^2 \cdot \frac{6}{12}) + (\mathrm{e}^0 \cdot \frac{3}{12} + \mathrm{e}^2 \cdot \frac{9}{12})$.
5. $\mathrm{Var}(X) = \mathrm{E}[X^2] - \mathrm{E}[X]^2 \overset{\text{周辺分布}}{=} \frac{66}{12} - (\frac{13}{12})^2 = \frac{29}{36}$.
6. $\mathrm{E}[XY] \overset{\text{同時分布}}{=} 1 \cdot 2 \cdot \frac{4}{12} + 3 \cdot 2 \cdot \frac{5}{12} = \frac{19}{6}$.
7. $\mathrm{Cov}(X, Y) = \mathrm{E}[XY] - \mathrm{E}[X]\mathrm{E}[Y] = \frac{19}{6} - \frac{26}{12} \cdot \frac{18}{12} = -\frac{1}{12}$.

L06-Q4

Quiz 解答: 離散型確率変数の独立性

$y\backslash x$	3	4	計
13	$\frac{3}{10}\cdot\frac{2}{7}$	$\frac{7}{10}\cdot\frac{2}{7}$	$\frac{2}{7}$
14	$\frac{3}{10}\cdot\frac{5}{7}$	$\frac{7}{10}\cdot\frac{5}{7}$	$\frac{5}{7}$
計	$\frac{3}{10}$	$\frac{7}{10}$	1

L06-Q5

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.

L06-Q6

Quiz 解答: 独立な確率変数の母期待値

1. X, Yは独立なので $\mathrm{E}[XY] \overset{\text{IE1}}{=} \mathrm{E}[X]\mathrm{E}[Y] = 2 \cdot 3$.
2. X, Yは独立なので $\mathrm{Cov}(X, Y) \overset{\text{IC1}}{=} 0$.
3. $\mathrm{E}[(-2X + 3Y)(X + 5Y)] \overset{\text{E5}}{=} \mathrm{E}[-2X^2] + \mathrm{E}[-7XY] + \mathrm{E}[15Y^2] \overset{\text{V1,IE1}}{=} -2(\mathrm{Var}(X) + \mathrm{E}[X]^2) - 7\mathrm{E}[X]\mathrm{E}[Y] + 15(\mathrm{Var}(Y) + \mathrm{E}[Y]^2) = 240$.
4. X, Yは独立なので $\mathrm{Var}(g_1(X) + g_2(Y)) \overset{\text{IC2}}{=} \mathrm{Var}(g_1(X)) + \mathrm{Var}(g_2(Y))$ であることに注意して, $\mathrm{Var}(-2X + 3Y) \overset{\text{IC2}}{=} \mathrm{Var}(-2X) + \mathrm{Var}(3Y) \overset{\text{V2}}{=} (-2)^2\mathrm{Var}(X) + 3^2\mathrm{Var}(Y) = 119$.

ここまで来たよ

6 多変数の確率変数

7 二項分布・独立同一分布の和
- 二項分布
- ベルヌーイ分布
- 独立同一分布
- 独立同分布にしたがう確率変数の和

復習+ちょっと (x で書かれた確率関数) I

L07-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 $\mathrm{Var}(X)$ を求めよう.

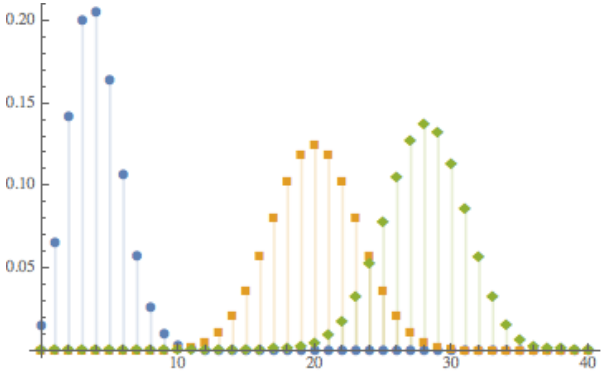
二項分布 高校 数学 B 久保川 統計学入門 §4.3.2

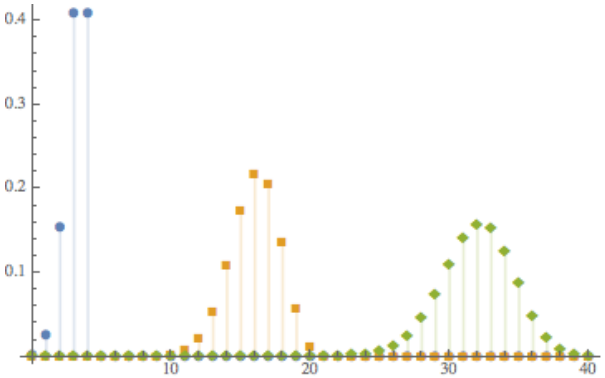
定義 (二項分布 久保川 統計学入門 §4.3.2)

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 $Bin(n,p)$ にしたがうという.

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

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.





二項係数 高校 数学 A

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

$Bin(40, 0.1)$, $Bin(40, 0.5)$, $Bin(40, 0.7)$, $Bin(4, 0.8)$, $Bin(20, 0.8)$, $Bin(40, 0.8)$

二項分布 $Bin(n,p)$ の Python での扱い

Google Colab on Moodle

$X \sim Bin(n,p) \leftrightarrow$ `rvx=scipy.stats.binom(n=`n`,p=`p`)`

```
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 久保川 統計学入門 公式 4.10)

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

(証明延期)

$\mathrm{E}[X^0] =$

定理 (二項定理 高校 数学 A 久保川 統計学入門 (4.4))

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

L07-Q2

Quiz(二項分布)

確率 $\frac{2}{3}$ で表がでるコインを 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，パラメタ○○の…分布のように，または記号で答えよう.
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, $\binom{n}{k}$ などの記号は使わないで答えること.
3. X の母平均値, 母分散を求めよう.

久保川 統計学入門 例題 4.11

L07-Q3

Quiz(二項分布または独立同一分布)

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 $\mathrm{Var}(Y)$ を求めよう.
4. 確率 $P(Y = 16)$ を求めよう.

ここまで来たよ

6 多変数の確率変数

7 二項分布・独立同一分布の和
- 二項分布
- ベルヌーイ分布
- 独立同一分布
- 独立同分布にしたがう確率変数の和

ベルヌーイ分布 久保川 統計学入門 §4.3.1

定義 (ベルヌーイ分布 久保川 統計学入門 §4.3.1)

$n=1$ の二項分布 $Bin(1,p)$ をパラメタ p のベルヌーイ分布 $Ber(p)$ 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と母分散 久保川 統計学入門 定理 4.8)

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

公式に頼らず言える.

命題 (ベルヌーイ分布と二項分布の関係 久保川 統計学入門 例題 6.20)

$X_1, \ldots, X_n$ が独立同一分布 $Ber(p)$ にしたがうとき,
$S_n = X_1 + \cdots + X_n$ は $S_n \sim Bin(n, p)$.

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.

ここまで来たよ

6 多変数の確率変数

7 二項分布・独立同一分布の和
- 二項分布
- ベルヌーイ分布
- 独立同一分布
- 独立同分布にしたがう確率変数の和

復習: 独立，非独立な確率変数の母期待値と母分散

X_1, X_2 が独立でも独立でなくても成立すること

$$
\begin{aligned}
\mathrm{E}[aX_1 + bX_2 + c] &= a\mathrm{E}[X_1] + b\mathrm{E}[X_2] + c, \\
\mathrm{Var}(aX_1 + c) &= a^2\mathrm{Var}(X_1), \\
\mathrm{Var}(aX_1 + bX_2 + c) &= a^2\mathrm{Var}(X_1) + 2ab\mathrm{Cov}(X_1, X_2) + b^2\mathrm{Var}(X_2), \\
\mathrm{Cov}(X_1, X_2) &= \mathrm{E}[X_1X_2] - \mathrm{E}[X_1]\mathrm{E}[X_2].
\end{aligned}
$$

X_1, X_2 が独立なとき成立すること

$$
\begin{aligned}
\mathrm{E}[g_1(X_1) \times g_2(X_2)] &= \mathrm{E}[g_1(X_1)] \times \mathrm{E}[g_2(X_2)], \\
\mathrm{Cov}(X_1, X_2) &= 0, \\
\mathrm{Var}(aX_1 + bX_2 + c) &= a^2\mathrm{Var}(X_1) + b^2\mathrm{Var}(X_2).
\end{aligned}
$$

独立同一分布

定義 (独立同一分布 久保川 統計学入門 p.143,p.150 **)**

確率変数 $X_1, \dots, X_n$ が互いに独立であり，同一の分布にしたがう（周辺分布の確率関数 p_{X_i} が同じである）とき，$X_1, \dots, X_N$ は独立同一分布 (i.i.d. independent and identically distributed) といい，次のように書く（たとえばその分布が $Ber(p)$ の場合）.

$$X_1, \dots, X_n, \text{i.i.d.} \sim Ber(p)$$

このとき，母期待値も等しい. $\mathrm{E}[X_i] = \mu, \mathrm{Var}(X_i) = \sigma^2$.

例: 形の同じ青色コイン X_1 と赤色コイン X_2.
表がでる $(X_i = 1)$ の確率 p
$X_1, X_2, \text{i.i.d.} \sim Ber(p)$.

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

ここまで来たよ

6 多変数の確率変数

7 二項分布・独立同一分布の和
- 二項分布
- ベルヌーイ分布
- 独立同一分布
- 独立同分布にしたがう確率変数の和

独立同分布にしたがう確率変数の和の性質 久保川 統計学入門 §6.5

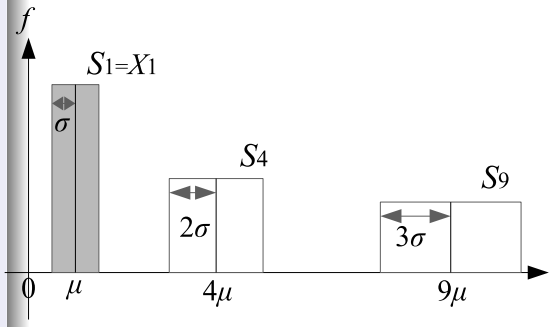
命題 (i.i.d にしたがう確率変数の和 久保川 統計学入門 公式 6.19)

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

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

$$\mathrm{Var}(S_n) \overset{\text{独立}}{=} \sum_{i=1}^{n} \mathrm{Var}(X_i) \overset{\text{同分布}}{=} n \times \sigma^2$$

S_n の確率 (密度) 関数はこんな感じ?



L07-Q4

Quiz(独立同一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, $\mathrm{E}[X_i] = 80, \mathrm{Var}(X_i) = 200$ 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. $\mathrm{E}[S_{10}]$, $\mathrm{Var}(S_{10})$ を求めよう.
2. $\mathrm{E}[\frac{1}{10}S_{10}]$, $\mathrm{Var}(\frac{1}{10}S_{10})$ を求めよう.
3. $\mathrm{E}[\frac{1}{100}S_{100}]$, $\mathrm{Var}(\frac{1}{100}S_{100})$ を求めよう.

L07-Q5

Quiz(独立同一分布にしたがう変数の和)

確率変数 $X_1, \ldots, X_n$ は $\mathrm{E}[X_i] = \mu, \mathrm{Var}(X_i) = \sigma^2$ の独立同分布に従う.

1. 確率変数 $S_n = X_1 + X_2 + X_3 + \cdots + X_n$ の母平均値と母分散を求めよう.
2. 確率変数 $U_n = \frac{1}{n}(X_1 + X_2 + X_3 + \cdots + X_n)$ の母平均値と母分散を求めよう.
3. 確率変数 U_n を標準化して, 確率変数 Y_n を, 母平均値 0, 母分散 1^2 になるように定義しよう.

二項分布の母平均値・母分散の導出

$X_1, \dots, X_n$, i.i.d. $\sim Ber(p)$ とすると, $\mathrm{E}[X_i] = p, \mathrm{Var}(X_i) = p(1-p)$.

このとき, $S_n = \sum_{i=1}^{n} X_i \sim Bin(n, p)$ に対して,

$$\mathrm{E}[S_n] \overset{\text{いつでも}}{=} \mathrm{E}[X_1] + \cdots + \mathrm{E}[X_n] \overset{\text{同一}}{=} n \times p$$

$$\mathrm{Var}(S_n) \overset{\text{独立}}{=} \mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_n) \overset{\text{同一分布}}{=} n \times p(1-p).$$

中心極限定理・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 L08(2025-06-09 Mon)

最終更新: Time-stamp: "2025-06-20 Fri 16:37 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) 久保川 統計学入門 §7.3, 大数の法則 久保川 統計学入門 §7.3, 中心極限定理 久保川 統計学入門 §7.4 が説明できる
- 独立同一分布の和の確率を正規近似できる

L07-Q1 略

L07-Q2

Quiz 解答: 二項分布

1. $X \sim Bin(100, \frac{2}{3}))$
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. $\mathrm{E}[X] = n \times p = \frac{200}{3}$. $\mathrm{Var}(X) = n \times p(1-p) = \frac{200}{9}$.

L07-Q3

Quiz 解答: 二項分布または独立同一分布

1. $X \sim Bin(10, 0.3)$ とすると, $Y = 2X$.
2. $\mathrm{E}[Y] = \mathrm{E}[2X] = 2 \times 10 \cdot 0.7 = 14$ 個.
3. $\mathrm{Var}(Y) = 2^2\mathrm{Var}(X) = 2^2 \times 10 \cdot 0.7 \cdot 0.3 = 8.4$ 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 は独立同一分布にしたがうので,
$\mathrm{E}[Y] = 10\mathrm{E}[Y_i], \mathrm{Var}(Y) = 10\mathrm{Var}(Y_i)$.
$Y_i = 2X_i$ 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, $X_i \sim Ber(0.7)$.
$\mathrm{E}[X_i] = 0.7$, $\mathrm{Var}(Y_i) = 0.7 \cdot (1 - 0.3)$ なので, $\mathrm{E}[Y_i] = 2 \cdot 0.7$,
$\mathrm{Var}(Y_i) = 2^2 \cdot 0.7 \cdot (1 - 0.3)$ である.
よって, $\mathrm{E}[Y] = 10\mathrm{E}[Y_i] = 14, \mathrm{Var}(Y) = 10\mathrm{Var}(Y_i) = 8.4$.

L07-Q4

Quiz 解答: 独立同一分布にしたがう確率変数の和

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

L08-Q5

Quiz 解答: 独立同一分布にしたがう変数の和

1. S_n は母平均値が $n\mu$, 母分散が $n\sigma^2$.
2. U_n は母平均値が μ, 母分散が $\frac{\sigma^2}{n}$.
3. $Y_n = \frac{U_n - \mu}{\sigma/\sqrt{n}} = \frac{1}{\sqrt{n}\sigma}(X_1 + X_2 + X_3 + \cdots + X_n - n\mu)$
   $Y_n = \frac{S_n - n\mu}{\sqrt{n}\sigma}$.

ここまで来たよ

7 二項分布・独立同一分布の和

8 中心極限定理・正規近似
- チェビシェフの不等式
- 統計モデル
- 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

久保川 統計学入門 §7.3

定理 (チェビシェフの不等式応用形は 久保川 統計学入門 公式 7.6(p.158))

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 $\sigma^2 = \mathrm{Var}(X)$, $a > 0$: 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)

「離れた x」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，

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

ここまで来たよ

7 二項分布・独立同一分布の和

8 中心極限定理・正規近似
- チェビシェフの不等式
- 統計モデル
- 大数の法則・中心極限定理
- 正規近似

統計モデル 久保川 統計学入門 §7.1

統計モデル 久保川 統計学入門 §7.1

$$\boxed{\text{構造}} + \boxed{\text{ゆらぎ=確率変数}} \quad \underset{\text{推定}}{\overset{\text{標本抽出}}{\rightleftarrows}} \quad \boxed{\text{観測データ}}$$

記述統計の母集団と標本，有限母集団 久保川 統計学入門 §7.1

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!
握手券を 5 枚買って質問する．標本 $X_1, \dots, X_5$
標本平均値 $\frac{1}{5}(X_1 + \cdots + X_5)$

データ分析

推測統計の母集団と標本，無限母集団 久保川 統計学入門 §7.1

某アイドルグループは，正規分布 $N(?,?)$ の身長をもって無限に生成される (→ 無限母集団)

```
rvx=stats.norm(loc=?,scale=?) #N(?,?^2) 実は身長は正規分布
rvx.rvs(size=5)
```

「確率変数 $X \sim N(?,?)$ にしたがう試行を 5 回して，5 個の数値を得る」
＝「5 個の確率変数 $X_1, \ldots, X_n$, i.i.d. $\sim N(?,?)$ の実現値を得る」
「平均値」$\overline{X} = \frac{1}{5}(X_1 + \cdots + X_5)$ も確率変数.

ここまで来たよ

7 二項分布・独立同一分布の和

8 中心極限定理・正規近似
- チェビシェフの不等式
- 統計モデル
- 大数の法則・中心極限定理
- 正規近似

独立同一分布にしたがう確率変数の和の性質

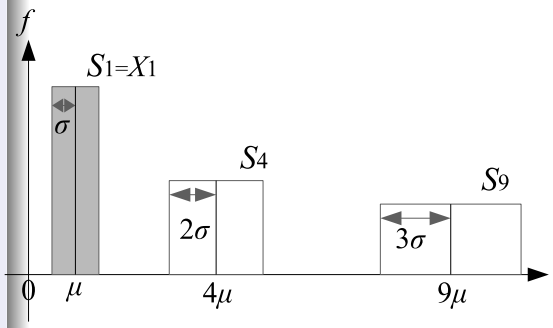
復習 確率統計 I(2025)L07 i.i.d にしたがう確率変数の和

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

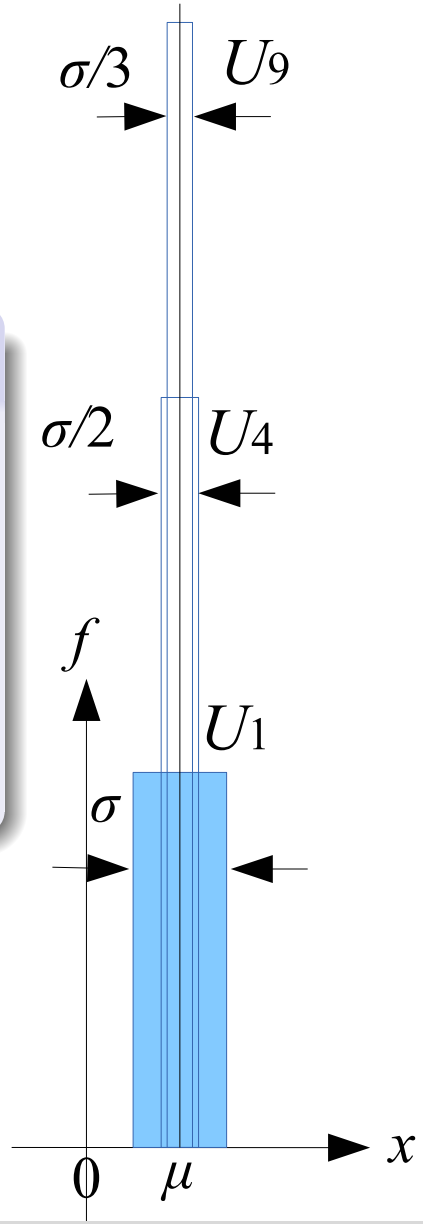
$$\mathrm{E}[S_n] = \sum_{i=1}^{n} \mathrm{E}[X_i] \overset{\text{同一分布}}{=} n \times \mu.$$

$$\mathrm{Var}(S_n) \overset{\text{独立}}{=} \sum_{i=1}^{n} \mathrm{Var}(X_i) \overset{\text{同一分布}}{=} n \times \sigma^2$$

S_n の確率密度関数はこんな感じ?



U_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 $1/n$

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

$$\mathrm{E}[U_n] = \mathrm{E}[\tfrac{1}{n} S_n] = \tfrac{1}{n} \times n \times \mu.$$

$$\mathrm{Var}(U_n) = \mathrm{Var}(\tfrac{1}{n} S_n) = \left(\tfrac{1}{n}\right)^2 \times n \times \sigma^2.$$

大数の (弱) 法則

$X_1, \ldots, X_n$ が独立同一分布にしたがい (n 回のくじの賞金), $\mathrm{E}[X_i] = \mu$, $\mathrm{Var}(X_i) = \sigma^2$ とする.

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

定理 (大数の (弱) 法則アバウト版 久保川 統計学入門 公式 7.8)

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 久保川 統計学入門 公式 7.8)

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

「U_n は μ に確率収束する」

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,

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

P の中の不等式の右辺を ϵ にしたい下心から $a = \frac{\epsilon}{\frac{\sigma}{\sqrt{n}}}$ とすると,

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

これが母平均値・母期待値の直観的意味. 要するに,

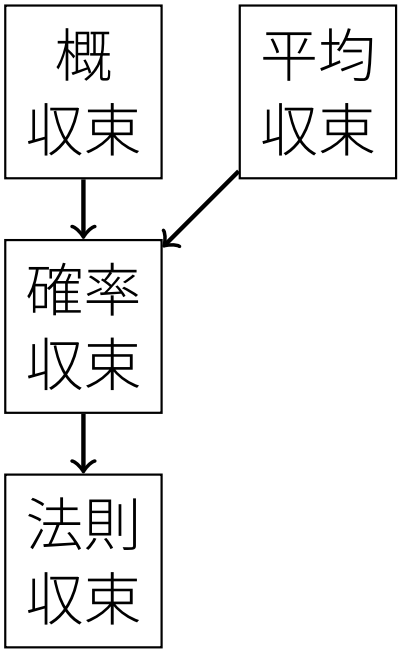
何回も宝くじを買って賞金の相加平均をとると, 必ず E[賞金] に近い

確率にまつわるいろいろな「収束」

定理 (大数の強法則)

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

「U_n は μ に概収束 する」(確率収束とは極限の位置が違う)
U_n の「近づく」(!) 先が μ でない確率はゼロである
確率にまつわる「収束」にはいろいろある…



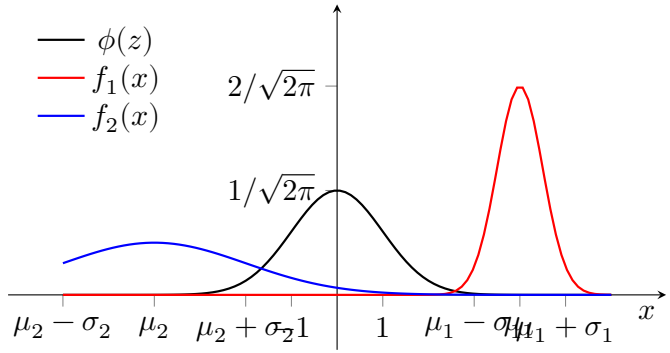
復習: 正規分布 久保川 統計学入門 §5.3.2

定義 ((一般の) 正規分布 $N(\mu, \sigma^2)$ 久保川 統計学入門 §5.3.2)

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

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

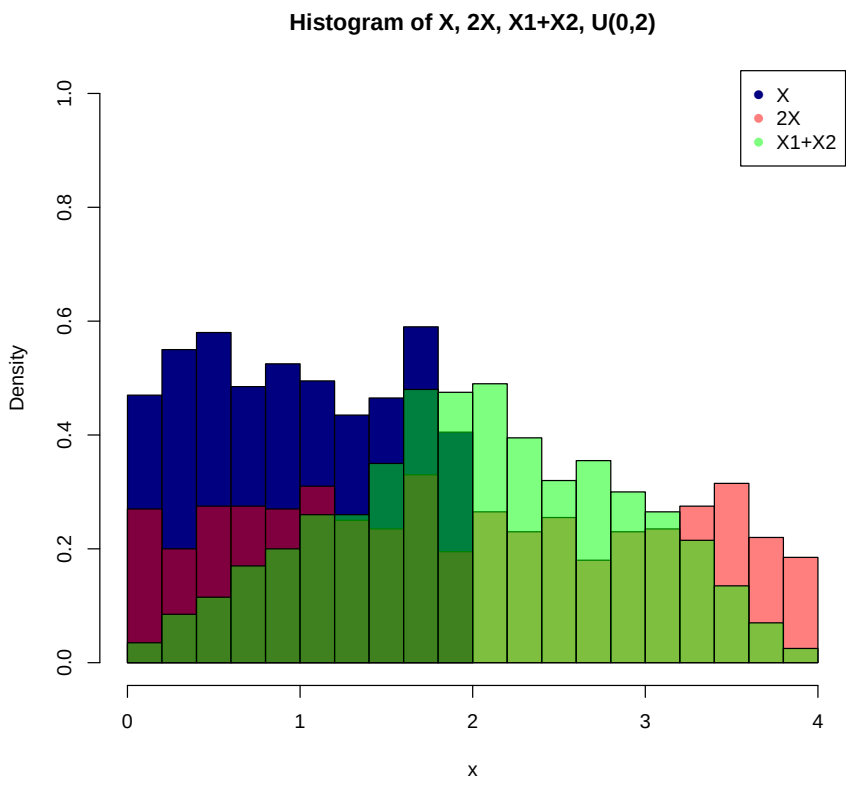
$f(x)$ =scipy.stats.norm(loc=μ,scale=σ).pdf(x)

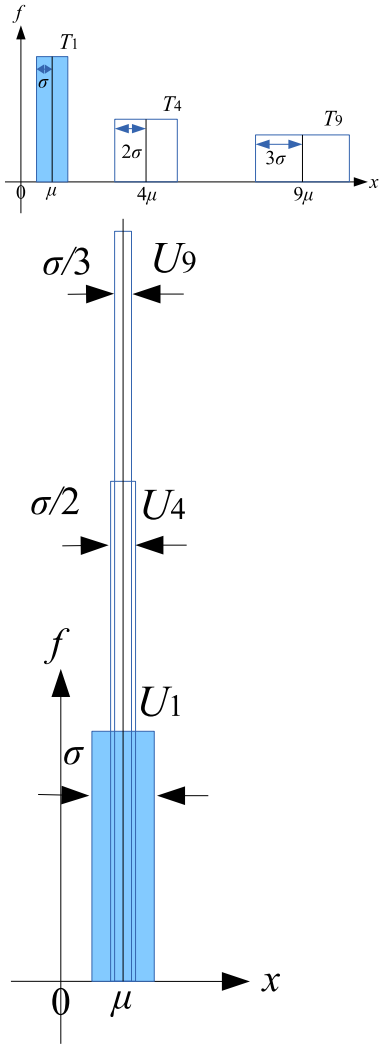


対称軸は $x = \mu$, $f(\mu) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma}$, σ 離れると 0.6 倍.

$Z = \frac{X-\mu}{\sigma} \sim N(0, 1^2)$ 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？

$X_1, X_2 \sim U(0,2)$, i.i.d.
$2 \times X_1 \sim$
$X_1 + X_2 \sim$

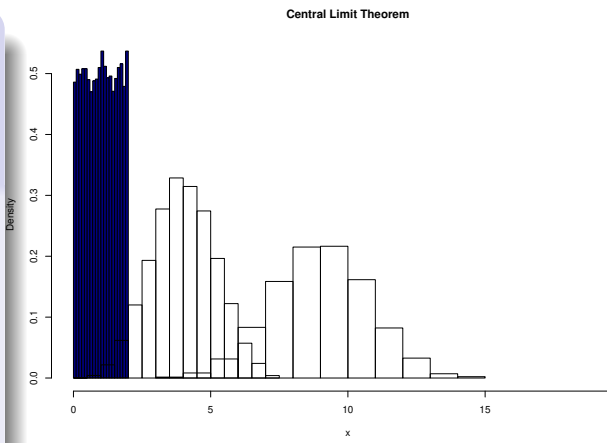
中心極限定理 久保川 統計学入門 §7.4

定理 (中心極限定理 (いいかげんバージョン) 久保川 統計学入門 公式 7.11)

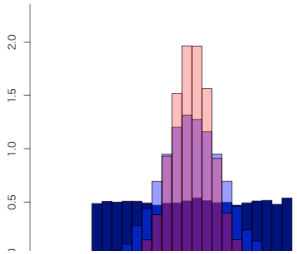
$X_1, \dots, X_n$: 独立同一分布にしたがう．$\mathrm{E}[X_i] = \mu$, $\mathrm{Var}(X_i) = \sigma^2$. $n \to +\infty$ で,

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

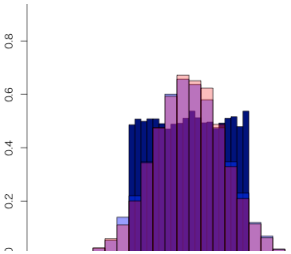
S



U



Z



ここまで来たよ

7 二項分布・独立同一分布の和

8 中心極限定理・正規近似
- チェビシェフの不等式
- 統計モデル
- 大数の法則・中心極限定理
- 正規近似

独立同一分布にしたがう確率変数の和の正規近似 I

正規近似 久保川 統計学入門 p.162, 例題 7.14

L08-Q1

Quiz(独立同一分布と中心極限定理)

確率変数 X_i $(i = 1, \ldots, 400)$ は独立同一分布にしたがい, $\mathrm{E}[X_i] = \frac{1}{10}$, $\mathrm{Var}(X_i) = \frac{9}{100}$ 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.

L08-Q2

Quiz(独立同一分布と中心極限定理)

確率変数 X_i $(i = 1, \ldots, 400)$ は独立同一分布にしたがい, $X_i \sim U(3, 5)$ 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

久保川 統計学入門 基本問題問 1(p.166)

L08-Q3

Quiz(ベルヌーイ分布の独立同一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.

二項分布の正規近似高校 数学 B 久保川 統計学入門 公式 7.13

憶え方：二項分布 $Bin(n,p)$ は, 正規分布 $N(np, np(1-p))$ で近似できる.

母集団・標本・標本抽出・推定・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 L09(2025-06-16 Mon)

最終更新: Time-stamp: "2025-06-16 Mon 11:34 JST hig"

今日の目標

- 母集団, 標本抽出, 推定を説明できる 久保川 統計学入門 §7.1
- 母平均値, 母比率, 母分散を点推定できる
  久保川 統計学入門 §7.2
- 不偏推定量，一致推定量の意味を説明できる

L08-Q1

Quiz 解答: 独立同一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 $N(n\mu, n\sigma^2)$ すなわち $N(40, 6^2)$ にしたがう. $Z = \frac{S-40}{6}$ は近似的に標準正規分布 $N(0, 1^2)$ にしたがう. よって, 求める確率は, $P(S > 31) = P(Z > -\frac{9}{6}) = P(-\frac{9}{6} < Z < +\infty) = \Phi(\infty) - \Phi(-\frac{9}{6}) = 1 - \Phi(-\frac{9}{6}) = 0.9332$.

L08-Q2

Quiz 解答: 独立同一分布と中心極限定理

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

$n = 400$ 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 $N(n\mu, n\sigma^2)$ すなわち $N(1600, \frac{400}{3})$ 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 $N(0, 1^2)$ にしたがう. よって, 求める確率は, $P(-\infty < Z \leq -5\sqrt{3}) = \Phi(-5\sqrt{3}) - \Phi(-\infty) = \Phi(-5\sqrt{3}) - 0 = 2.35 \times 10^{-18}$.

とりうる最小の値は $400 \cdot 3$ だから 1500 はありうる値だが，確率はこんなに小さい.

L08-Q3

Quiz 解答: ベルヌーイ分布の独立同一分布の和と中心極限定理

100 回中表の出る回数は，$Y = X_1 + \cdots + X_{100}$, $X_i \sim Bin(1, \frac{4}{5})$, 独立同一分布, $\mathrm{E}[X_i] = \frac{4}{5}, \mathrm{Var}(X_i) = \frac{4}{5}\frac{1}{5}$. よって, $\mathrm{E}[Y] = 80$, $\mathrm{Var}(Y) = 4^2$ である (これは, $Y \sim Bin(100, \frac{4}{5})$ 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 $N(80, 4^2)$ 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 $N(0, 1^2)$ にしたがう.

よって, 求める確率は,

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

ここまで来たよ

8 中心極限定理・正規近似

9 母集団・標本・標本抽出・推定・点推定
- 母集団と標本
- 母平均値・母比率の (点) 推定
- 母分散の (点) 推定

統計モデル 久保川 統計学入門 §7.1 確率統計 I(2025)L08

統計モデル 久保川 統計学入門 §7.1

$$\boxed{\text{構造}} + \boxed{\text{ゆらぎ}=\text{確率変数}} \quad \underset{\text{推定}}{\overset{\text{標本抽出}}{\rightleftarrows}} \quad \boxed{\text{観測データ}}$$

記述統計の母集団と標本，有限母集団 久保川 統計学入門 §7.1

某アイドルグループ全員 ($\to$ 有限母集団,$N = 46$)
身長 x_i の平均値 $\overline{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$ を求めたい!
某アイドルグループの滋賀県出身率 $p_1 = m/N$ を求めたい!
握手券を 5 枚買って質問する．標本 $X_1, \dots, X_5$, $Y_1, \dots, Y_5$
標本平均値 $\frac{1}{5}(X_1 + \cdots + X_5)$，標本比率 $\frac{1}{5}(Y_1 + \cdots + Y_5)$.　データ分析

- 母集団サイズ= $\boxed{46}$ の母集団から，
- 標本サイズ= $\boxed{5}$ の標本 (sample) を
- 標本の個数= $\boxed{1}$ 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.
- 握手券 46 枚買わず (実際は重複するかも), 5 枚で何とかすませたい.

推測統計の母集団と標本，無限母集団 久保川 統計学入門 §7.1

某アイドルグループの各メンバーの身長は，何かの分布にしたがう
滋賀県出身の正否は，ベルヌーイ分布 $Ber(p_1)$ にしたがう．
確率変数と思えば，何回でも引くことができる (46 を超えて)．(→ 無限母集団)
「確率分布 $Ber(p_1)$ にしたがう試行を 5 回する」
＝「5 個の確率変数 $Y_1, \dots, Y_n$, i.i.d. $\sim Ber(p_1)$ の実現値を得る」(p_1 は不明)
「平均値」$\overline{Y} = \frac{1}{5}(Y_1 + \cdots + Y_5)$ も確率変数．

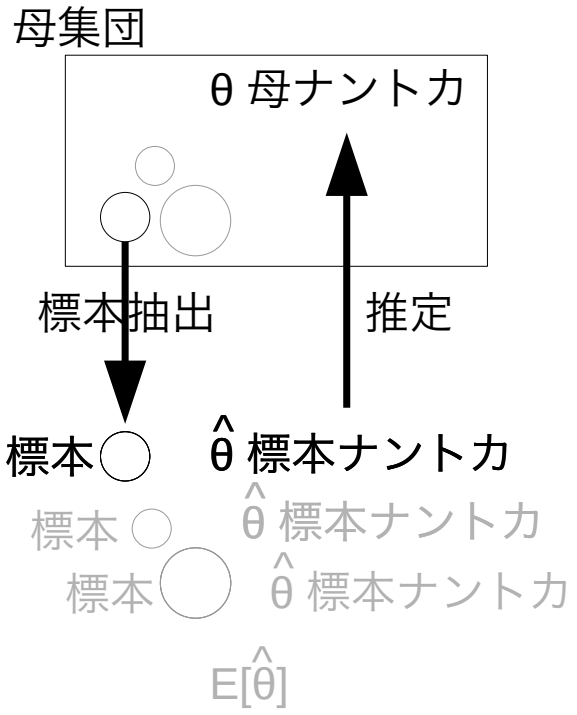
- 母集団サイズ＝$\boxed{+\infty}$ の母集団から，
- 標本サイズ＝$\boxed{5}$ の標本 (sample) を
- 標本の個数＝$\boxed{1}$ 個を標本抽出する

推定

- 分布の平均値を知りたい．
- 分布を知れば定義の式を使うだけ
- 分布を知ることはできない．
- 有限個の実現値から何とか「推測」

母集団・標本抽出・推定

- 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. パラメタ θ が備わる.
- 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 $\rightsquigarrow$ sampling (動名詞)
- 推定 する estimate(動詞) =標本を調べて母集団について，特に θ について，正しそうな事実を見つける $\rightsquigarrow$ estimation (名詞)
- 確率変数 X, $\overline{X}$ 分布をもつ変数
- 実現値, 観測値 x, $\overline{x}$ 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 を推定する式
- 点推定 1 個の値を提出する方式の推定



推定には誤差ある． 標本の選び方ごとに答は違う！

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 久保川 統計学入門 p.34
- W A,B,C,D スクショする？ 質的データ 順位尺度 久保川 統計学入門 p.34

→ 導かれるデータ「身長高い」

- $W = \begin{cases} 1 & (X \geq 170.7) \\ 0 & (\text{No}) \end{cases}$

統計モデル
$X_i \sim$ 何かの独立同一分布.
$Y_i \sim Ber(p_1))$.
スクショ $\sim$ 多項分布
$W_i \sim Ber(p_2)$. X_i の言い換え.
こういう文脈での，ベルヌーイ分布のパラメタ，母平均値を，母比率 (ratio) という.

- $p_1 = \mathrm{E}[Y] = P(Y = 1)$.
- $p_2 = \mathrm{E}[W] = P(X \geq 170.7)$.

母集団=回答者全体

1. 母集団サイズ 95 (クラス全体が回答すれば 118 だった)

サブチームに割り当られた標本

1. 1 サブチームに割り当てる標本の個数 1 個
2. 標本サイズ 様々

久保川 統計学入門 p.34

- よい標本抽出には実はいろいろなハイテクがある ⇝ 社会調査法, 社会統計学
- 本当は, 母集団が無限でないとき, 標本を作るのに重複選出を禁止するとき, 数学的にいろいろ複雑なことが起きるのだが, この授業では知らん顔する. ⇝ 確率・統計

ここまで来たよ

8 中心極限定理・正規近似

9 母集団・標本・標本抽出・推定・点推定
- 母集団と標本
- 母平均値・母比率の (点) 推定
- 母分散の (点) 推定

母平均値の (点) 推定 高校 数学 C 久保川 統計学入門 §7.2

組 $(X_1, X_2, \dots, X_n)$ はサイズ n の標本. 各 X_i は母平均値 $\mu = \mathrm{E}[X_i]$, 母分散 $\sigma^2 = \mathrm{Var}(X_i)$ の独立同一分布にしたがう確率変数.

定義 (標本平均値 久保川 統計学入門 公式 7.2)

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

母平均値 $\mu = \mathrm{E}[X_i]$ の 'よい' 推定量になっている. 'よい'≃ 不偏推定量
Pandas/Python では .mean(), Excel, Google スプレッドシートでは関数 average()

定義 (標本期待値)

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

$\mathrm{E}[g(X_i)]$ の 'よい' 推定量になっている.

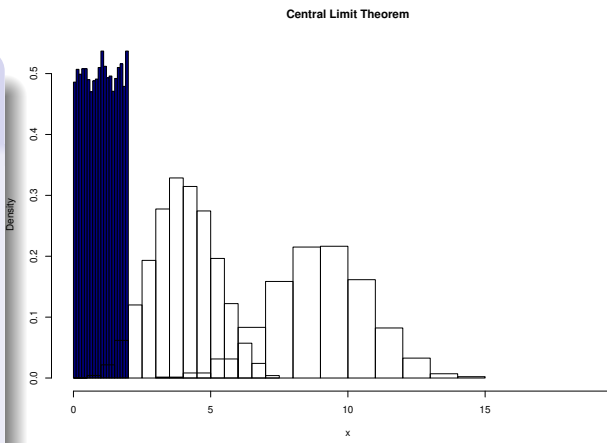
復習：中心極限定理 久保川 統計学入門 §7,4

定理 (中心極限定理 (いいかげんバージョン))

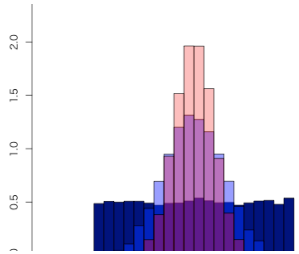
$X_1, \dots, X_n$ が母平均値 μ, 母分散 σ^2 の独立同分布に従うとき, $n \to +\infty$ で

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

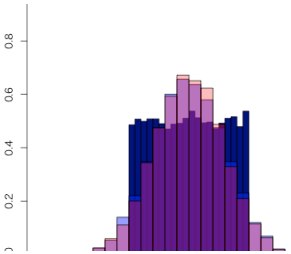
S



U



Z



‘よい’ 推定量の資格 1

パラメタ θ を持つ分布の
母平均値 $\mathrm{E}[X]$，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)
一般に推定量 $\hat{\theta}(X_1, \dots, X_n)$ は確率分布を持つ．

定義 (推定量の不偏性 久保川 統計学入門 定義 7.1 **)**

$\hat{\theta} = \hat{\theta}(X_1, \dots, X_n)$ が，パラメタ θ の不偏推定量であるとは，

$$\mathrm{E}[\hat{\theta}(X_1, \dots, X_n)] = \theta$$

が成り立つこと．

標本平均値 $\hat{\theta} = \overline{X}_n$ は，母平均値 $\theta = \mu$ の不偏推定量である: $\mathrm{E}[\overline{X}_n] = \mu$. 中心極限定理
また，$\mathrm{Var}(\overline{X}_n) = \sigma^2/n$. 中心極限定理

‘よい’ 推定量の資格 2

定義 (推定量の一致性 久保川 統計学入門 定義 7.5 **)**

$\hat{\theta} = \hat{\theta}(X_1, \dots, X_n)$ が，パラメタ θ の一致推定量であるとは，

$$\forall \epsilon > 0 \qquad \lim_{n \to +\infty} P(|\hat{\theta}(X_1, \dots, X_n) - \theta| > \epsilon) = 0.$$

推定量と母ナントカに一定の差がある確率が, 標本サイズ n を大きくすると zero になること.
大数の法則 確率統計 I(2025)L08 より，標本平均値 $\hat{\theta} = \overline{X}_n$ は，母平均値 $\theta = \mu$ の一致推定量である.

L09-Q1

Quiz(確率変数としての標本平均値の分布)

母平均値 10, 母分散 6^2 の分布にしたがう母集団から, 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.

L09-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.

118g,　109g,　109g,　119g,　101g,　113g.

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. 105g 以上のものの母比率を点推定しよう.

記述上の注意　厳しく弾圧します. × き

- 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 $= \sigma^2 = \mathrm{Var}(X) \neq$ 不偏標本分散 $V^2 = \frac{1}{n-1}(\cdots)$
- 母比率 $= p \neq$ 標本比率 $\hat{p} = \frac{k}{n}$.
- 「母ナントカを…と × 求めた○推定する」

比率=ratio 久保川 統計学入門 p.12

確率変数 $Y \sim Ber(p)$ 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)

$Ber(p)$ の p を母比率という. 母集団で X の条件「…」から $Y \sim Ber(p)$ を作ったとき, '母集団の「…である」ものの母比率 p', ともいう.

有限母集団なら,

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

母比率の (点) 推定

定義 (標本比率)

ベルヌーイ分布にしたがう $Y_1, \dots, Y_n \sim Ber(p)$ に対しては，（推定量である）標本平均値を，特に標本比率 $\hat{p}$ という．$Y_i = 1$ のものが K 個とすると，

$$\hat{p}(Y_1, \dots, Y_n) = \frac{1}{n}(Y_1 + \cdots + Y_n) = \frac{K}{n}$$

標本比率は，母比率 p の不偏推定量，一致推定量

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

ここまで来たよ

8 中心極限定理・正規近似

9 母集団・標本・標本抽出・推定・点推定
- 母集団と標本
- 母平均値・母比率の (点) 推定
- 母分散の (点) 推定

母分散の (点) 推定

母分散 $\mathrm{Var}(X) = \mathrm{E}[(X-\mu)^2]$ もしょせん母期待値じゃん．データ分析で使った

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

がよい推定値じゃないの？…ちょっと待った！
$\overline{X}$ の中には $X_1, \ldots, X_n$ がぜんぶ入ってるので，定数 μ であるかのような計算はできない．
分布の重心がちょっとずれるので修正が必要．

今後，その分散や分布を考えようと思っても，$(X_1 - \overline{X})^2, \ldots, (X_n - \overline{X})^2$ は互いに独立ではないから簡単ではない．

母分散の (点) 推定

定義 (不偏標本分散 久保川 統計学入門 公式 7.2)

$$
\begin{aligned}
\text{不偏標本分散 } \hat{\theta}(X_1, \dots, X_n) = V^2 &= \frac{1}{n-1}[(X_1 - \overline{X})^2 + \cdots + (X_n - \overline{X})^2] \\
&= \frac{n}{n-1}\left[\frac{1}{n}\sum_i X_i^2 - (\overline{X})^2\right]
\end{aligned}
$$

不偏標本分散は母分散 σ^2 の 'よい' 推定値 不偏性 $\mathrm{E}[V^2] = \sigma^2$.
不偏標本分散は 久保川 統計学入門 p.155 の不偏分散のこと. 久保川 統計学入門 p.154 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.
$n-1$ の理由 こうしないと不偏にならない
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

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

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

不偏標本分散は，Pandas/Python では .var(ddof=1), Excel では関数 var.s(), Google スプレッドシートでは, var(), var.s()s for sample(標本)
標本分散は，Pandas/Python では .var(), Excel では関数 var.p(), Google スプレッドシートでは, varp(), var.p()p for population(母集団)

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 L10(2025-06-23 Mon)

最終更新: Time-stamp: "2025-07-21 Mon 18:01 JST hig"

今日の目標

- 母平均値を近似的に区間推定できる 久保川 統計学入門 §8.3
- 母比率を区間推定できる 久保川 統計学入門 §8.3
- カイ二乗分布を説明できる 久保川 統計学入門 §8.1

L09-Q1

Quiz 解答: 確率変数としての標本平均値の分布

1. $\mathrm{E}[\overline{X}] = \frac{1}{4} \cdot 4\mathrm{E}[X_i] = 10.$
2. $\mathrm{Var}(\overline{X}) = \frac{1}{4^2} \cdot 4\mathrm{Var}(X_i) = 6^2/4.$
3. $Z = \frac{\overline{X}-10}{6/\sqrt{4}}$

L09-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 $\overline{X} = \frac{1}{6}(118 + \cdots + 113) = 111.5\mathrm{g}$ なので, 母平均値を $111.5\mathrm{g}$ と推定する.
2. 標本期待値 (の実現値) は $\overline{X^2} = \frac{1}{6}(118^2 \cdots + 113^2) = 12469.5\mathrm{g}^2$ なので, 母期待値を $12469.5\mathrm{g}^2$ と推定する.

3. 不偏標本分散 (の実現値) は,
$S^2 = \frac{1}{6-1}[(117-111.5)^2 + \cdots + (112-111.5)^2] = 37.25\mathrm{g}^2$ なので,
母分散を $37.25\mathrm{g}^2$ と推定する.
4. 標本比率 (の実現値) は, $\hat{p} = \frac{1}{6}[1+1+1+1+0+1] = 0.83$ なので,
母比率を 0.83 と推定する.

ここまで来たよ

9 母集団・標本・標本抽出・推定・点推定

10 区間推定
- 大標本の区間推定・近似的信頼区間
- 母比率の区間推定・信頼区間
- 母分散の区間推定・信頼区間とカイ二乗分布

点推定 対 区間推定

点推定 久保川 統計学入門 §7.2 「1 個の推定値を求めること」
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 久保川 統計学入門 §8.2 「信頼係数に応じた信頼区間を求めること」

「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

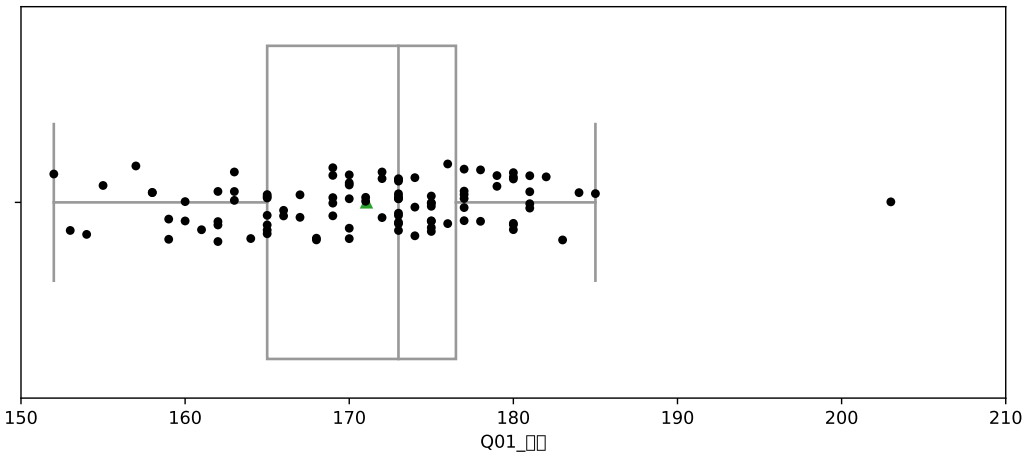
ここで '確率' というのは不誠実. 正しい言葉遣いは, 信頼係数=信頼度で

「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」

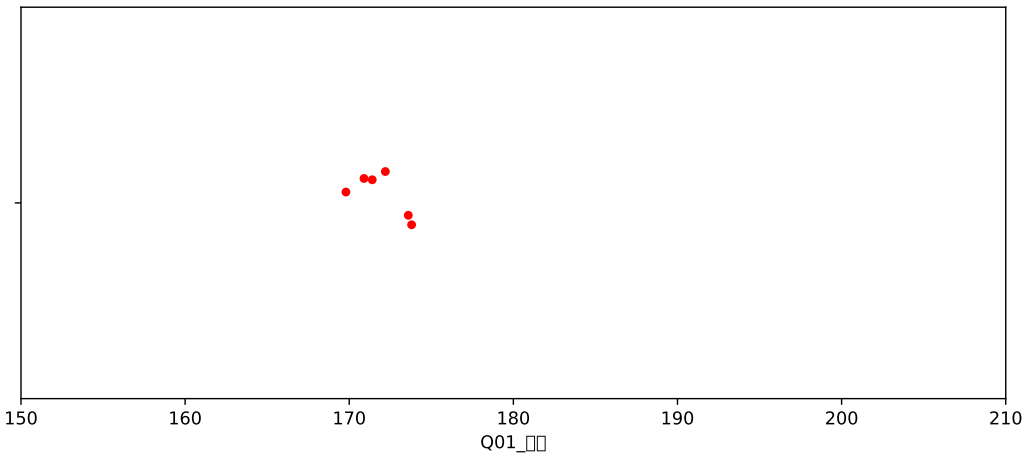
動く (確率変数である) のは母平均値 μ でなく, B, C のほう.

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



大標本からの母平均値特に母比率の近似的信頼区間

未知 母集団の母平均値 μ, 母分散 σ^2

$\downarrow$

既知 標本 $X_1, \ldots, X_n \rightsquigarrow$ 標本平均値 $\overline{X}$, 不偏標本分散 V^2.

$\boxed{\text{近似 1}}$ サンプルサイズ n が大きいとき，中心極限定理により，近似的に．

$$\overline{X} \sim N(\mu, \sigma^2/n)$$

$$\text{標準化} \quad \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1^2)$$

$\boxed{\text{近似 2}}$ σ のかわりに $\sqrt{V^2}$ を使っても近似的になりたつ．

$$\frac{\overline{X} - \mu}{\sqrt{V^2/n}} \sim N(0, 1^2)$$

近似=誤差がある．後から，一定の条件下で，これらの近似なしに，より正確に評価できるハッピーなケースを扱う

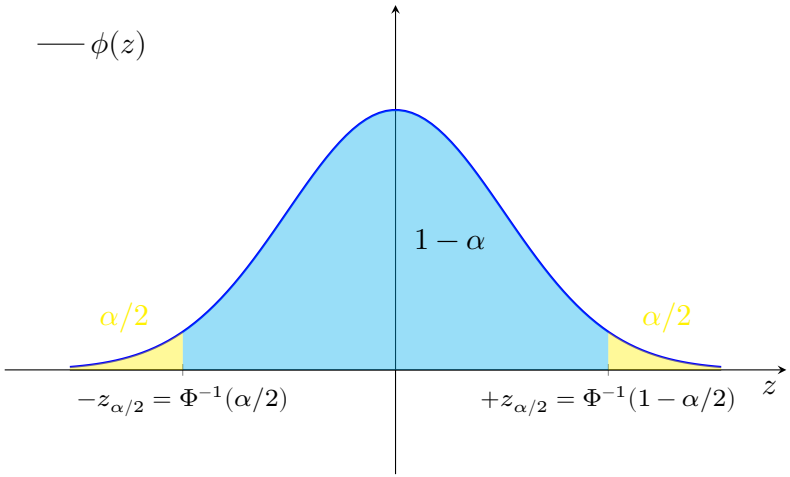
標準正規分布の計算の準備 I

L10-Q1

Quiz(正規分布の上側確率)

$Z \sim N(0, 1^2)$ 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_α と書く．
3. Z の確率密度関数 $\phi(z)$ 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. 「母平均値 0 から遠い確率」$P(|Z - 0| > u) = P(Z < -u) + P(u < Z) = \alpha$ となる u を α と $z_\bullet$ 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_α 久保川 統計学入門 公式 5.9,(8.5), 図 8.4)

$Z \sim N(0, 1^2)$ のとき，上側確率 $P(Z \geq z_{\alpha/2}) = \alpha/2$ となる境い目を $z_{\alpha/2}$ と定める．
Φ を累積分布関数，Φ^{-1} をその逆関数とするとき，
$z_{\alpha/2} = \Phi^{-1}(1-\alpha/2)$. 偶関数だから $z_{1-\alpha/2} = -z_{\alpha/2}$.
$P(-z_{\alpha/2} < Z < +z_{\alpha/2}) = \alpha$.

標本平均値と母平均値の差 $\overline{X}-\mu$ が 0 に近い確率は，正規分布が真ん中で盛り上がっているために大きい (ここでは $1-\alpha$) という式を書くと…

$$\boxed{\text{近似 1}}\, P\left(\Phi^{-1}(\tfrac{\alpha}{2}) < \frac{\overline{X}-\mu}{\sqrt{\sigma^2/n}} < \Phi^{-1}(1-\tfrac{\alpha}{2})\right) = 1-\alpha.$$

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

μ について不等式を解くと，

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

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

$$\boxed{\text{近似 2}}\, P\left(\overline{X} - z_{\alpha/2}\times\sqrt{V^2/n} < \mu < \overline{X} + z_{\alpha/2}\times\sqrt{V^2/n}\right) = 1-\alpha.$$

業界の習慣で，しばしば $\alpha = 0.05, 0.01$

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

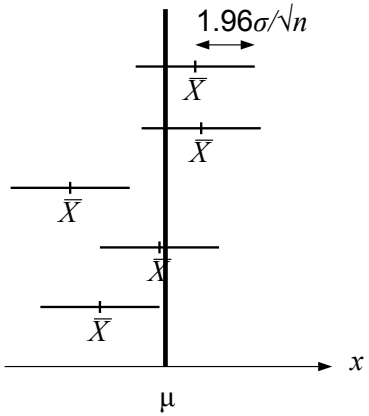
命題 (母平均値の近似的信頼区間 久保川 統計学入門 公式 8.10)

母平均値 μ, 母分散 σ^2 のなにかの分布にしたがう母集団の, 母平均値 μ の 信頼係数 $1-\alpha$ の近似的信頼区間 ($(1-\alpha)$ 信頼区間) は, n が大きいとき, $\overline{X}$ を標本平均値, V^2 を不偏標本分散として,

$$\overline{X} - z_{\alpha/2} \times \sqrt{V^2/n} < \mu < \overline{X} + z_{\alpha/2} \times \sqrt{V^2/n}.$$

$a < \mu < b$ でなく, 閉区間で $\mu \in [a, b]$ とかくこともある.「信頼区間は $[a, b]$」のように答えることもある.

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_α は 久保川 統計学入門 公式 5.9,(8.6)

$z_{0.05/2} = 1.96, z_{0.01/2} = 2.58.$

高校 数学 B では, $z_{0.05/2} = 1.96$ の場合のみ.

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

9 母集団・標本・標本抽出・推定・点推定

10 区間推定
- 大標本の区間推定・近似的信頼区間
- 母比率の区間推定・信頼区間
- 母分散の区間推定・信頼区間とカイ二乗分布

母比率の信頼区間 久保川 統計学入門 例題 8.11

母比率 p は，ベルヌーイ分布 $Ber(p)$ の母平均値 p. 母分散は $p(1-p)$. 標本比率は標本平均値 $\overline{X} = \hat{p} = k/n$.

近似 1 サンプルサイズが大きいとき，中心極限定理により，標本比率は近似的に正規分布 $N(p, p(1-p)/n)$ に従う.

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

近似 2' 母分散 σ^2 を，不偏標本分散 V^2 のかわりに $\hat{p}(1-\hat{p})$ で近似する

母比率の近似的信頼区間 久保川 統計学入門 例題 8.11

$X \sim Ber(p)$ 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,

$$\hat{p} - \Phi^{-1}(1-\alpha/2) \times \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p < \hat{p} - \Phi^{-1}(\alpha/2) \times \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}.$$

$$\hat{p} - z_{\alpha/2} \times \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p < \hat{p} + z_{\alpha/2} \times \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}.$$

L10-Q2

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.

注: 下限, 上限が 0,1 を越えるときは, 近似 1,2 が正確でないことを意味しているが，実用的には 0,1 に直してしまっていい.

ここまで来たよ

9 母集団・標本・標本抽出・推定・点推定

10 区間推定
- 大標本の区間推定・近似的信頼区間
- 母比率の区間推定・信頼区間
- 母分散の区間推定・信頼区間とカイ二乗分布

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?

	の推定値 (点推定)	の信頼区間 (区間推定)
母平均値 μ	標本平均値 $\overline{X} = \frac{1}{n}[X_1 + \cdots]$	左右対称 $\overline{X} - \boxed{\text{正}}\sqrt{} < \mu < \overline{X} - \boxed{\text{負}}\sqrt{}$
母分散 σ^2	不偏標本分散 $V^2 = \frac{1}{n-1}[(X_1 - \overline{X})^2 + \cdots]$	左右対称でない $V^2 \times \boxed{\text{小}} < \sigma^2 < V^2 \times \boxed{\text{大}}$

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 $N(0, 1^2)$
- 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 久保川 統計学入門 §8.1

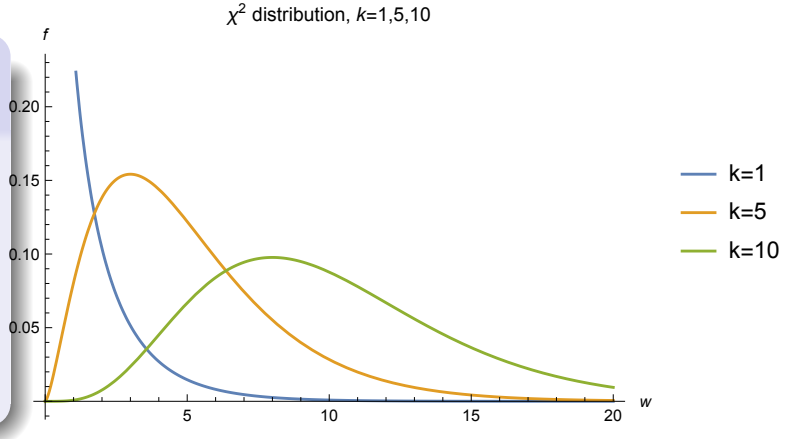
定義 (カイ二乗分布 久保川 統計学入門 定義 8.1(p.170))

$Z_1, \dots, Z_k$, i.i.d. $\sim N(0, 1^2)$ 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

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



久保川 統計学入門 公式 8.2 $W_k \sim \chi_k^2$ に対して,

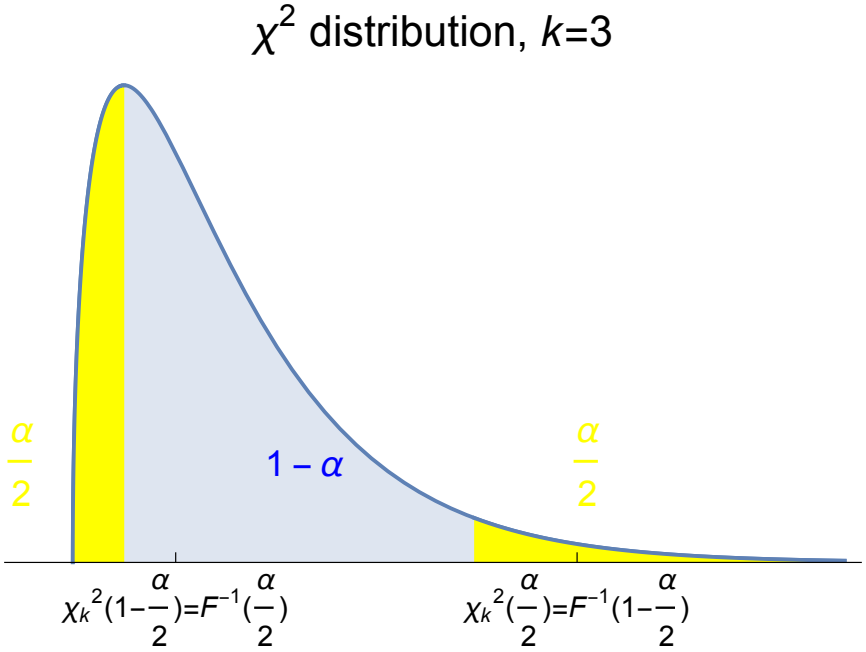
$\mathrm{E}[W_k] = \mathrm{E}[Z_1^2 + \cdots + Z_k^2] = k, \mathrm{Var}(W_k) = 2k, \mathrm{E}[(W_k)^\ell] =$ 簡単じゃない.

カイ二乗分布の確率密度関数と累積分布関数

```
rvx2=stats.chi2(df=k) # k は自由度
```

上側確率 $\alpha/2 = P(W > \chi^2_{k,\alpha/2})$ となる境い目 $\chi^2_{k,\alpha/2} = F_W^{-1}(1-\alpha/2)$.
F はカイ二乗分布の累積分布関数.

久保川 統計学入門 数値表 p.292　久保川 統計学入門 数値表 p.290 と p.292 はフォーマットが違う



$\chi^2_{k,\alpha/2}$ の定義

$$\alpha/2 = P(W > \chi^2_{k,\alpha/2}).$$
$$\chi^2_{k,\alpha/2} = F_W^{-1}(1-\alpha/2).$$

L10-Q3

Quiz(カイ二乗分布)

自由度 2 のカイ二乗分布 χ_2^2 にしたがう確率変数 W を考える．累積分布関数を $F_{\mathrm{W}}(w)$ ，逆関数を $F_{\mathrm{W}}^{-1}(p)$ とする．
次の量を $F_{\mathrm{W}}(), F_{\mathrm{W}}^{-1}()$ で表そう．

1. $P(W < 0.1)$
2. $P(W > 0.9)$
3. $P(W < w) = 0.1$ となる w
4. $P(W > w) = 0.1$ となる w

不偏標本分散のしたがう分布 久保川 統計学入門 §8.2.2

不偏標本分散のしたがう分布

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

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

から定めた

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

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，この比は確率変数 $\frac{W}{n-1}$. $(W \sim \chi^2_{n-1})$

証明じゃないけど説明

$X_i, \dots, X_n$, i.i.d. $\sim N(\mu, \sigma^2)$ に対して,

$$\left(\frac{X_1 - \mu}{\sigma}\right)^2 + \cdots + \left(\frac{X_n - \mu}{\sigma}\right)^2$$

は自由度 n のカイ二乗分布 χ^2_n にしたがう.
μ を $\overline{X}$ に替えた,

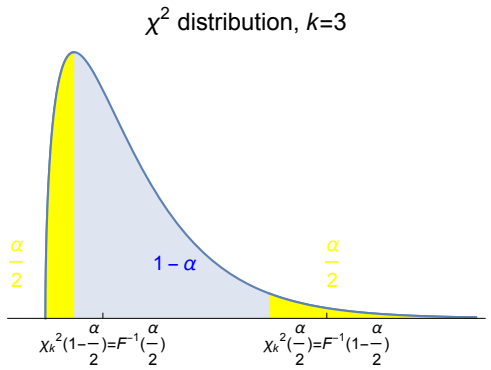
$$W = \left(\frac{X_1 - \overline{X}}{\sigma}\right)^2 + \cdots + \left(\frac{X_n - \overline{X}}{\sigma}\right)^2$$

は自由度は $n-1$ のカイ二乗分布 χ^2_{n-1} にしたがう.

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

母分散の区間推定

$$P\left(F_W^{-1}(\alpha/2) < (n-1)\frac{V^2}{\sigma^2} < F_W^{-1}(1-\alpha/2)\right) = 1-\alpha$$



不等式を σ^2 について解いて次の結果を得る.

正規母集団の母分散の信頼区間 久保川 統計学入門 公式 8.8

$X \sim N(\mu, \sigma^2)$ のサイズ n の標本の不偏標本分散が V^2 のとき, 母分散 σ^2 の信頼係数 $1-\alpha$ の信頼区間は

$$\frac{n-1}{F_W^{-1}(1-\alpha/2)} \times V^2 < \sigma^2 < \frac{n-1}{F_W^{-1}(\alpha/2)} \times V^2$$
$$\frac{n-1}{\chi^2_{n-1,\alpha/2}} \times V^2 < \sigma^2 < \frac{n-1}{\chi^2_{n-1,1-\alpha/2}} \times V^2$$

n 大という条件なし，近似なしだけど，X が正規分布にしたがう（正規母集団 久保川 統計学入門 §8.1）という超強い条件下.

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

L10-Q4

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 80g, 不偏標本分散は $72\mathrm{g}^2$ だった.
母分散 σ^2 を信頼係数 $1-\alpha=0.95$ で区間推定しよう.

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 L11(2025-06-30 Mon)

最終更新: Time-stamp: "2025-07-06 Sun 14:36 JST hig"

今日の目標

- カイ二乗分布を説明できる 久保川 統計学入門 §8.1
- 母分散を区間推定できる 久保川 統計学入門 §8.2
- t 分布を説明できる 久保川 統計学入門 §8.1

L10-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. $\Phi^{-1}(1 - \frac{0.05}{2}) = \Phi^{-1}(0.975) = \texttt{.ppf(0.975)} = 1.96$.

Quiz 解答: 区間推定の性質

1

L10-Q2

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，
$\Phi^{-1}(1-\frac{0.05}{2})=z_{0.05/2}=1.96$ 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」こののりで判定を繰り返すと，20 回に 1 回に間違えるということ）

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

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

命題 (復習：母平均値の近似的信頼区間（大標本）

久保川 統計学入門 公式 8.10 | 確率統計 I(2025)L10)

母平均値 μ, 母分散 σ^2 のなにかの分布にしたがう母集団の, 母平均値 μ の 信頼係数 $1-\alpha$ の近似的信頼区間 ($(1-\alpha)$ 信頼区間) は, n が大きいとき，$\overline{X}$ を標本平均値, V^2 を不偏標本分散として,

$$\overline{X} - z_{\alpha/2} \times \sqrt{V^2/n} < \mu < \overline{X} + z_{\alpha/2} \times \sqrt{V^2/n}.$$

長所

- どんな母集団（分布）でも可

短所

- $\boxed{\text{近似 1}}$ 標本サイズ N が大きい（大標本）ことを使った，中心極限定理に頼る近似
- $\boxed{\text{近似 2}}$ 母分散 σ^2 のかわりに不偏標本分散 V^2 を使う近似

ここまで来たよ

10 区間推定

11 母分散・母平均値の区間推定
- 母分散の区間推定・信頼区間とカイ二乗分布
- 母平均値の区間推定 (正規母集団, 母分散既知，小標本)
- 母平均値の区間推定 (正規母集団, 母分散未知，小標本)

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?

	の推定値 (点推定)	の信頼区間 (区間推定)
母平均値 μ	標本平均値 $\overline{X} = \frac{1}{n}[X_1 + \cdots]$	左右対称 $\overline{X} - \boxed{\text{正}}\sqrt{} < \mu < \overline{X} - \boxed{\text{負}}\sqrt{}$
母分散 σ^2	不偏標本分散 $V^2 = \frac{1}{n-1}[(X_1 - \overline{X})^2 + \cdots]$	左右対称でない $V^2 \times \boxed{\text{小}} < \sigma^2 < V^2 \times \boxed{\text{大}}$

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 $N(0, 1^2)$
- 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

カイ二乗分布 久保川 統計学入門 §8.1

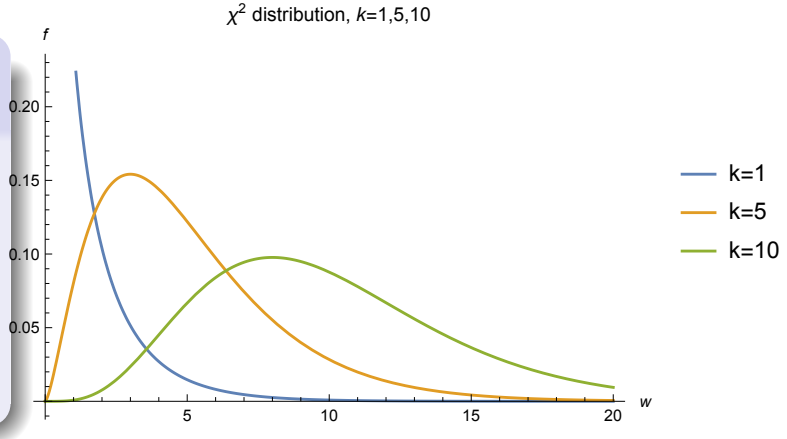
定義 (カイ二乗分布 久保川 統計学入門 定義 8.1(p.170))

$Z_1, \dots, Z_k$, i.i.d. $\sim N(0, 1^2)$ 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

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



久保川 統計学入門 公式 8.2 $W_k \sim \chi_k^2$ に対して,

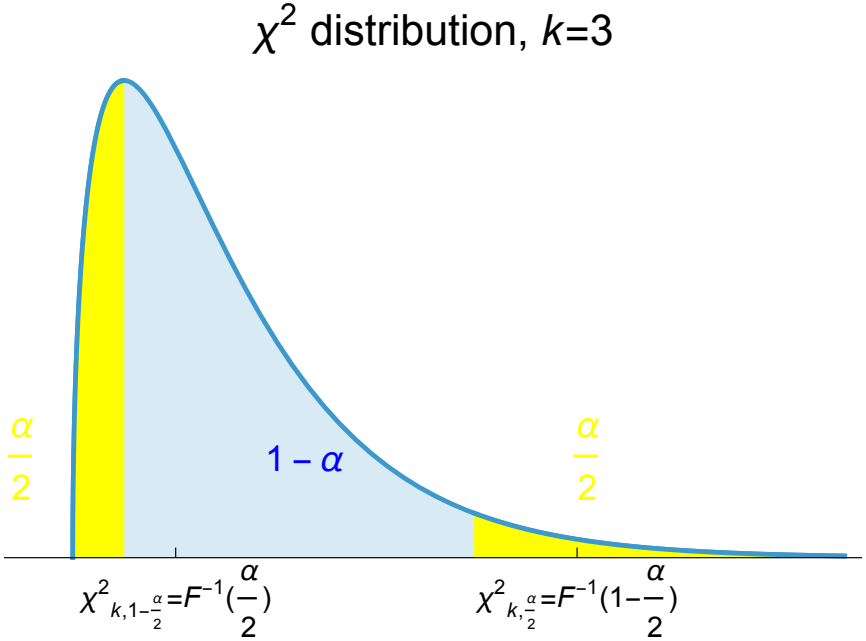
$\mathrm{E}[W_k] = \mathrm{E}[Z_1^2 + \cdots + Z_k^2] = k, \mathrm{Var}(W_k) = 2k, \mathrm{E}[(W_k)^\ell] =$ 簡単じゃない.

カイ二乗分布の確率密度関数と累積分布関数

```
rvx2=stats.chi2(df=k) # k は自由度
```

上側確率 $\alpha/2 = P(W > \chi^2_{k,\alpha/2})$ となる境い目 $\chi^2_{k,\alpha/2} = F^{-1}_{\mathrm{W}k}(1-\alpha/2)$.
F はカイ二乗分布の累積分布関数.

久保川 統計学入門 数値表 p.292　久保川 統計学入門 数値表 p.290 と p.292 はフォーマットが違う



$\chi^2_{k,\alpha/2}$ の定義

$$\alpha/2 = P(W > \chi^2_{k,\alpha/2}).$$

$$\chi^2_{k,\alpha/2} = F^{-1}_{\mathrm{W}k}(1-\alpha/2).$$

分布の再生性

命題 (カイ二乗分布の再生性 久保川 統計学入門 公式 8.2)

$X \sim \chi_n, Y \sim \chi_m$, X, Y 独立のとき，$X + Y \sim \chi_{n+m}$

定義 (再生性)

一般に，1 パラメタ θ を持つ分布の族 $D(\theta)$ があり，
$X \sim D(\theta_1), Y \sim D(\theta_2), Y \sim \chi_m$, X, Y 独立のとき，$X + Y \sim D(\theta_3)$ となるとき，分布の族 $D(\theta)$ は再生性 (reproductive property) を持つという．

例 (二項分布の再生性 久保川 統計学入門 公式 6.21)

$X \sim Bin(n, p), Y \sim Bin(m, p)$, X, Y 独立のとき，
$X + Y \sim Bin(n + m, p)$.

証明：式でも示せるが，コイン投げの意味的には明らか．

L11-Q1

Quiz(カイ二乗分布)

自由度 2 のカイ二乗分布 χ_2^2 にしたがう確率変数 W を考える．累積分布関数を $F_{\mathrm{W}}(w)$ ，逆関数を $F_{\mathrm{W}}^{-1}(p)$ とする．
次の量を $F_{\mathrm{W}}(), F_{\mathrm{W}}^{-1}()$ で表そう．

1. $P(W < 0.1)$
2. $P(W > 0.9)$
3. $P(W < w) = 0.1$ となる w
4. $P(W > w) = 0.1$ となる w

不偏標本分散のしたがう分布 久保川 統計学入門 §8.2.2

不偏標本分散のしたがう分布

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

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

から定めた

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

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，この比は確率変数 $\frac{W}{n-1}$. $(W \sim \chi^2_{n-1})$

証明じゃないけど説明

$X_1, \dots, X_n$, i.i.d. $\sim N(\mu, \sigma^2)$ に対して,

$$\left(\frac{X_1 - \mu}{\sigma}\right)^2 + \cdots + \left(\frac{X_n - \mu}{\sigma}\right)^2$$

は自由度 n のカイ二乗分布 χ^2_n にしたがう.
μ を $\overline{X}$ に替えた,

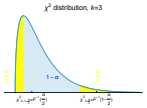
$$W = \left(\frac{X_1 - \overline{X}}{\sigma}\right)^2 + \cdots + \left(\frac{X_n - \overline{X}}{\sigma}\right)^2$$

は自由度は $n-1$ のカイ二乗分布 χ^2_{n-1} にしたがう.

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

母分散の区間推定

$$P\left(F^{-1}_{\mathrm{W}n-1}(\alpha/2) < (n-1)\frac{V^2}{\sigma^2} < F^{-1}_{\mathrm{W}n-1}(1-\alpha/2)\right) = 1-\alpha$$



久保川 統計学入門 図 8.6

不等式を σ^2 について解いて次の結果を得る.

正規母集団の母分散の信頼区間 久保川 統計学入門 公式 8.8

$X \sim N(\mu, \sigma^2)$ のサイズ n の標本の不偏標本分散が V^2 のとき, 母分散 σ^2 の信頼係数 $1-\alpha$ の信頼区間は

$$\frac{n-1}{F^{-1}_{\mathrm{W}n-1}(1-\alpha/2)} \times V^2 < \sigma^2 < \frac{n-1}{F^{-1}_{\mathrm{W}n-1}(\alpha/2)} \times V^2$$

$$\frac{n-1}{\chi^2_{n-1,\alpha/2}} \times V^2 < \sigma^2 < \frac{n-1}{\chi^2_{n-1,1-\alpha/2}} \times V^2$$

n 大という条件なし，近似なしだけど，X が正規分布にしたがう（正規母集団 久保川 統計学入門 §8.1）という超強い条件下.

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

L11-Q2

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 σ^2 を信頼係数 $1-\alpha=0.95$ で区間推定しよう.

久保川 統計学入門 例題 8.9

ここまで来たよ

10 区間推定

11 母分散・母平均値の区間推定
- 母分散の区間推定・信頼区間とカイ二乗分布
- 母平均値の区間推定 (正規母集団, 母分散既知，小標本)
- 母平均値の区間推定 (正規母集団, 母分散未知，小標本)

標本平均値のしたがう分布 (正規母集団, 母分散既知)

久保川 統計学入門 §8.2

$N(\mu, \sigma^2)$ 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.

標本平均値 $\overline{X} = \frac{1}{n}(X_1 + \cdots + X_n)$ は確率変数

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

命題 (正規分布の再生性 久保川 統計学入門 公式 6.21**)**

$X \sim N(\mu_1, \sigma_1^2), Y \sim N(\mu_2, \sigma_2^2)$, X, Y 独立のとき,
$X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

$n \to +\infty$ しなくても最初から正規分布

超強い仮定として，正規母集団を仮定すれば，$\boxed{\text{近似 1}}$ を取り除ける!

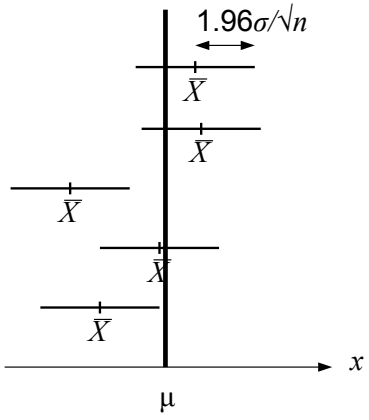
$\boxed{\text{近似 2}}$ (σ^2 を V^2 で置き換えちゃう) は使わずに書いてみる

命題 (母平均値 (正規母集団, 母分散既知) の信頼区間 久保川 統計学入門 公式 8.6(1))

$N(\mu, \sigma^2)$ 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,

$$\overline{X} - z_{\alpha/2} \times \sqrt{\sigma^2/n} < \mu < \overline{X} + z_{\alpha/2} \times \sqrt{\sigma^2/n}.$$

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_α は $z_{\frac{0.05}{2}} = 1.96, z_{\frac{0.01}{2}} = 2.58.$

ここまで来たよ

10 区間推定

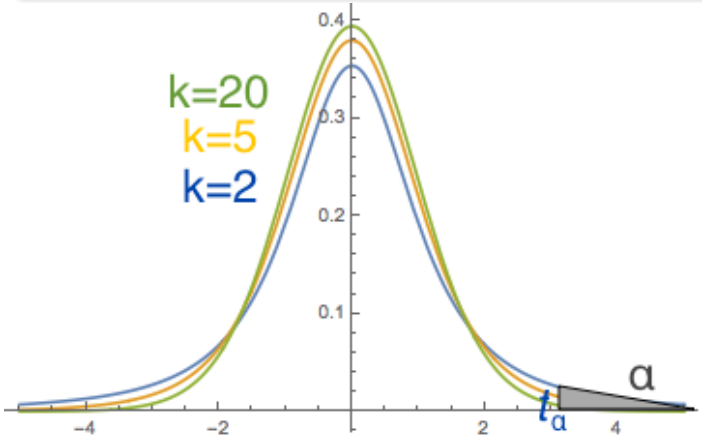
11 母分散・母平均値の区間推定
- 母分散の区間推定・信頼区間とカイ二乗分布
- 母平均値の区間推定 (正規母集団, 母分散既知，小標本)
- 母平均値の区間推定 (正規母集団, 母分散未知，小標本)

t 分布

あとは，近似 2 をなんとかしたい．

定義 (t 分布 久保川 統計学入門 定義 8.3 **)**

$Z \sim N(0, 1^2)$, $W \sim \chi^2_k$, Z と W が独立,
のとき連続型確率変数 $T = \frac{Z}{\sqrt{W/k}}$ 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, $N(0, 1^2)$ より低く広い.
自由度 $k \to +\infty$ で $N(0, 1^2)$ に一致する

t 分布の確率密度関数と累積分布関数

久保川 統計学入門 p.171

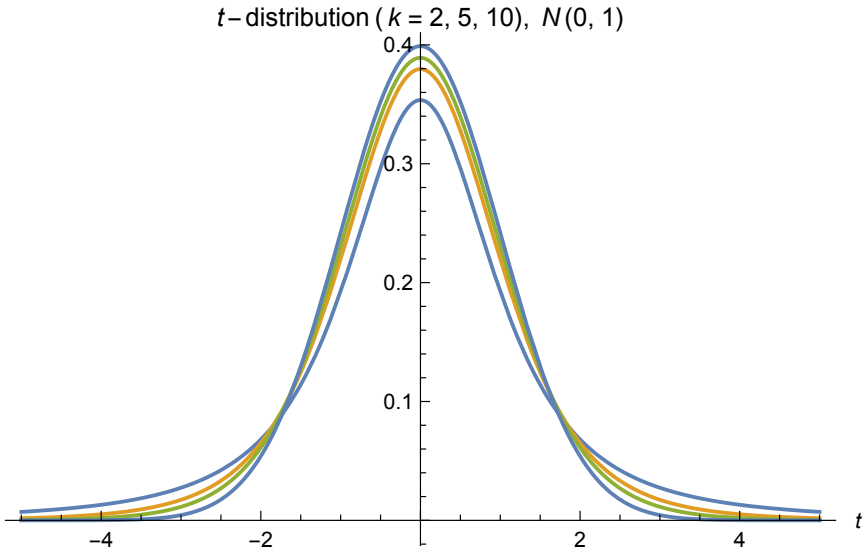
Python scipy.stats.t(df=k)

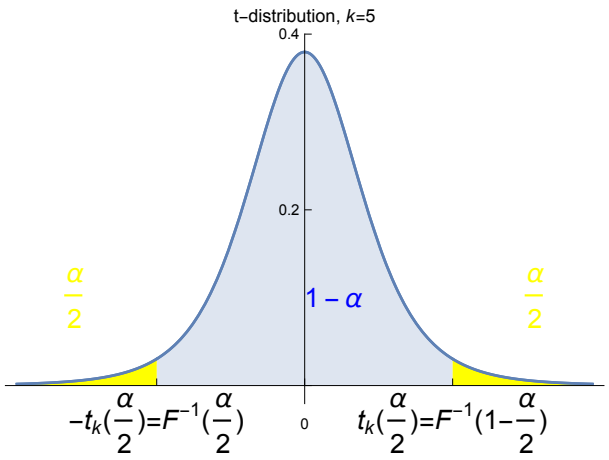
t 分布の確率密度関数は偶関数なので, $t_{k,1-\alpha} = -t_{k,\alpha}$.

久保川 統計学入門 数値表 p.291 上側確率 $\alpha = P(T > t_{k,\alpha})$ となる $t_{k,\alpha}$ の値.

久保川 統計学入門 数値 p.291 と p.290 はフォーマットが違う

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





L11-Q3

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 $Z \sim N(0, 1^2)$ 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_{t}^{-1} を使って答え, さらに数値にしよう.

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)

母分散既知の区間推定では, 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 V^2 (確率変数) を使った, 標準化もどき
$T = \frac{\overline{X}-\mu}{\sqrt{V^2/n}} = \frac{\frac{\overline{X}-\mu}{\sqrt{\sigma^2/n}}}{\sqrt{\frac{(n-1)V^2}{\sigma^2}/(n-1)}}$ でやっちゃいたい. 分子 $\sim N(0,1^2)$, 分母
$= \sqrt{W/(n-1)}, W \sim \chi^2_{n-1}$，実は W と X は独立，より, $T \sim \mathrm{t}_{n-1}$.

命題 (T の分布 久保川 統計学入門 公式 8.4 **)**

母集団 $N(\mu, \sigma^2)$ から, サイズ n の標本 $X_1, \cdots, X_n$ を取り出したとき,

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

は, 自由度 $n-1$ の Student の t 分布にしたがう.

母平均値の信頼区間 (母分散未知) 久保川 統計学入門 公式 8.6(2)

(母分散未知の) 正規分布 $N(\mu, ?^2)$ 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

$$\overline{X} - t_{n-1,\alpha/2} \times \sqrt{V^2/n} < \mu < \overline{X} + t_{n-1,\alpha/2} \times \sqrt{V^2/n}.$$

ただし, $\overline{X}$: 標本平均値, V^2: 不偏標本分散, n: 標本サイズ, $t_{n-1,\alpha/2}$: 自由度 $n-1$ の t 分布の上側確率が $\alpha/2$ となる点.

母分散既知 確率統計 I(2025)L10 と比べて,

$\sigma^2 \rightsquigarrow V^2$

$z_{\alpha/2} \rightsquigarrow t_{n-1,\alpha/2}$

L11-Q4

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 Xg 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.

母平均値の区間推定・線形回帰モデル

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(2025-07-07 Mon)

最終更新: Time-stamp: "2025-07-07 Mon 06:45 JST hig"

今日の目標

- t 分布を説明できる 久保川 統計学入門 §8.1
- 母平均値の区間推定ができる（正規母集団，小標本，母分散未知） 久保川 統計学入門 §8.2
- 線形回帰モデルを説明できる 久保川 統計学入門 §12.1

L11-Q1

Quiz 解答: カイ二乗分布

1. $F_{\mathrm{W}}(0.1)$. Python では `stats.chi2(df=2).cdf(0.1)`.
2. $1 - F_{\mathrm{W}}(0.9)$. Python では `1-stats.chi2(df=2).cdf(0.9)`.
3. $F_{\mathrm{W}}^{-1}(0.1)$. Python では `stats.chi2(df=2).ppf(0.1)`.
4. $F_{\mathrm{W}}^{-1}(1 - 0.1)$. Python では `stats.chi2(df=2).ppf(1-0.1)`.

L11-Q2

Quiz 解答: 母分散の区間推定

標本サイズは $n = 9$, 自由度は $9 - 1$.
不偏標本分散を V^2 とすると，$(9-1)\frac{V^2}{\sigma^2}$ が自由度 $9-1$ のカイ二乗分布に従うことから，母分散 σ^2 の信頼係数 $1 - \alpha = 0.95$ の信頼区間は,

$$
\begin{aligned}
\frac{n-1}{F_{\mathrm{W8}}^{-1}(1-\frac{\alpha}{2})} \times V^2 <& \sigma^2 < \frac{n-1}{F_{\mathrm{W8}}^{-1}(\frac{\alpha}{2})} \times V^2 \\
\frac{n-1}{\chi^2_{n-1,\frac{\alpha}{2}}} \times V^2 <& \sigma^2 < \frac{n-1}{\chi^2_{n-1,1-\frac{\alpha}{2}}} \times V^2 \\
\frac{8}{17.53} \times 72 <& \sigma^2 < \frac{8}{2.180} \times 72 \\
32.85 <& \sigma^2 < 264.2 \quad [\mathrm{g}^2]
\end{aligned}
$$

ここまで来たよ

11 母分散の区間推定

12 母平均値の区間推定・線形回帰モデル
- 母平均値の区間推定 (正規母集団, 母分散既知，小標本)
- t 分布
- 母平均値の区間推定 (正規母集団, 母分散未知，小標本)
- 線形回帰モデルとしての回帰分析

復習：大標本からの母平均値の近似的信頼区間 確率統計 I(2025)L10

未知（母集団）母平均値 μ, 母分散 σ^2
↑ 推定
既知 標本 $X_1, \dots, X_n \rightsquigarrow$ 標本平均値 $\overline{X}$, 不偏標本分散 V^2.

近似 1 サンプルサイズ n が大きいとき，中心極限定理により，近似的に．

$$\overline{X} \sim N(\mu, \sigma^2/n)$$

$$\text{標準化} \quad \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0, 1^2)$$

近似 2 σ のかわりに $\sqrt{V^2}$ を使っても近似的になりたつ．

$$\frac{\overline{X}-\mu}{\sqrt{V^2/n}} \sim N(0, 1^2)$$

命題 (母平均値の近似的信頼区間 久保川 統計学入門 公式 8.10)

母平均値 μ, 母分散 σ^2 のなにかの分布にしたがう母集団の, 母平均値 μ の 信頼係数 $1-\alpha$ の近似的信頼区間 ($(1-\alpha)$ 信頼区間) は, n が大きいとき，$\overline{X}$ を標本平均値, V^2 を不偏標本分散として,

$$\overline{X} - z_{\alpha/2} \times \sqrt{V^2/n} < \mu < \overline{X} + z_{\alpha/2} \times \sqrt{V^2/n}.$$

標本平均値のしたがう分布 (正規母集団, 母分散既知) 久保川 統計学入門 §8.2

$N(\mu, \sigma^2)$ 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.
標本平均値 $\overline{X} = \frac{1}{n}(X_1 + \cdots + X_n)$ は確率変数

命題 (正規分布の再生性 久保川 統計学入門 公式 6.21)

$X \sim N(\mu_1, \sigma_1^2), Y \sim N(\mu_2, \sigma_2^2)$, X, Y 独立のとき,
$X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

$\to +\infty$ しなくても $n = 2$ から正規分布
$\overline{X} \sim N(\mu, \sigma^2/n)$. $Z = \frac{\overline{X} - \mu}{\sqrt{\sigma^2/n}} \sim N(0, 1^2)$.

超強い仮定として，正規母集団を仮定すれば，$\boxed{\text{近似 1}}$ を取り除ける!

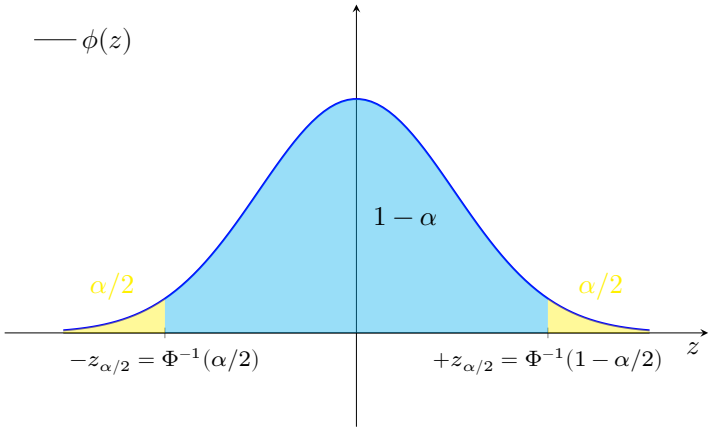
σ^2 はなぜかわかってるとして，$\boxed{\text{近似 2}}$ (σ^2 を V^2 で置き換えちゃう) なしで書いてみる

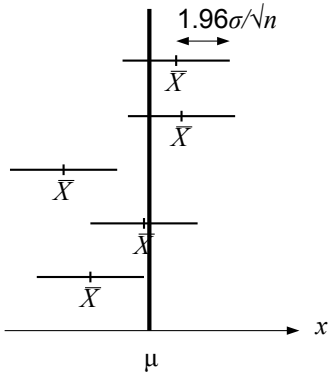
命題 (母平均値 (正規母集団, 母分散既知) の信頼区間 久保川 統計学入門 公式 8.6(1))

$N(\mu, \sigma^2)$ 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,

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

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.





ここまで来たよ

11 母分散の区間推定

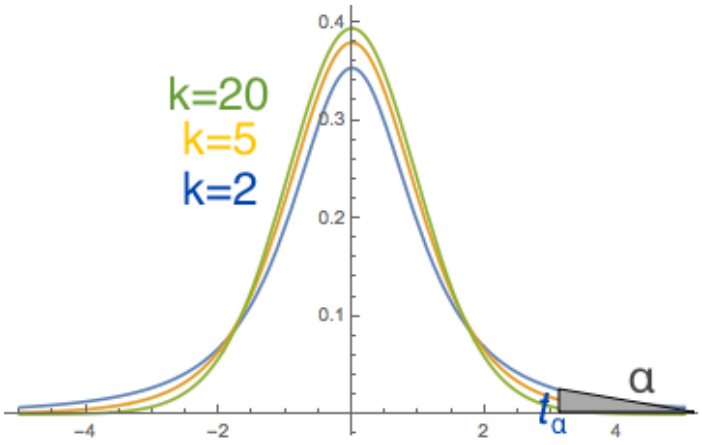
12 母平均値の区間推定・線形回帰モデル
- 母平均値の区間推定 (正規母集団, 母分散既知，小標本)
- t 分布
- 母平均値の区間推定 (正規母集団, 母分散未知，小標本)
- 線形回帰モデルとしての回帰分析

t 分布

あとは，$\boxed{\text{近似 2}}$をなんとかしたい．

定義 (t 分布 久保川 統計学入門 定義 8.3 **)**

$Z \sim N(0, 1^2)$, $W \sim \chi_k^2$, Z と W が独立,
のとき連続型確率変数 $T = \frac{Z}{\sqrt{W/k}}$ 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，$N(0, 1^2)$ より低く広い．
自由度 $k \to +\infty$ で $N(0, 1^2)$ 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

久保川 統計学入門 p.171

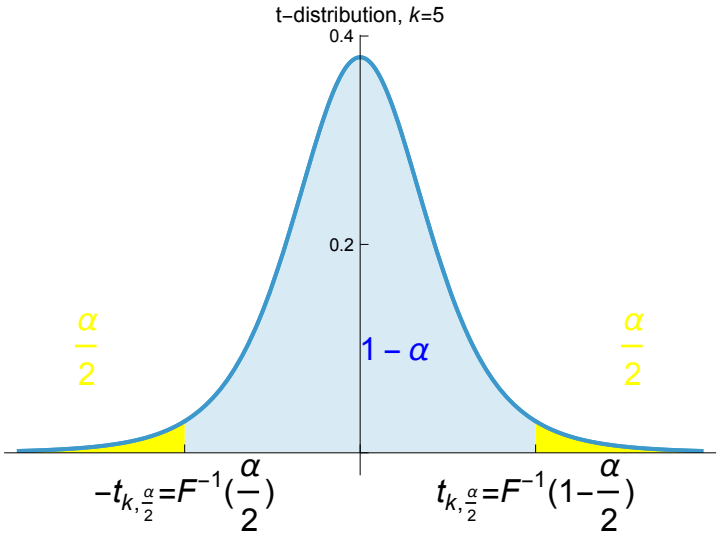
Python scipy.stats.t(df=k)

久保川 統計学入門 数値表 p.291 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, $t_{k,1-\alpha} = -t_{k,\alpha}$.

久保川 統計学入門 数値 p.291 と p.290 はフォーマットが違う

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



久保川 統計学入門 図 8.5

L12-Q1

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 $Z \sim N(0, 1^2)$ 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_{t}^{-1} を使って答え，さらに数値にしよう．

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．

ここまで来たよ

11 母分散の区間推定

12 母平均値の区間推定・線形回帰モデル
- 母平均値の区間推定 (正規母集団, 母分散既知，小標本)
- t 分布
- 母平均値の区間推定 (正規母集団, 母分散未知，小標本)
- 線形回帰モデルとしての回帰分析

母平均値の区間推定 (正規母集団, 母分散未知)

母分散既知の区間推定では, 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 V^2 (確率変数) を使った, 標準化もどき
$T = \frac{\overline{X}-\mu}{\sqrt{V^2/n}} = \frac{\frac{\overline{X}-\mu}{\sqrt{\sigma^2/n}}}{\sqrt{\frac{(n-1)V^2}{\sigma^2}/(n-1)}}$ でやっちゃいたい. 分子 $\sim N(0,1^2)$, 分母
$= \sqrt{W/(n-1)}, W \sim \chi^2_{n-1}$，実は W と X は独立，より, $T \sim \mathrm{t}_{n-1}$.

命題 (T の分布 久保川 統計学入門 公式 8.4 **)**

母集団 $N(\mu,\sigma^2)$ から, サイズ n の標本 $X_1,\cdots,X_n$ を取り出したとき,

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

は, 自由度 $n-1$ の Student の t 分布にしたがう.

標準化に σ^2 のかわりに V^2 を使うと，標準正規分布のかわりに t_k になる

母平均値の信頼区間 (母分散未知) 久保川 統計学入門 公式 8.6(2)

(母分散未知の) 正規分布 $N(\mu, ?^2)$ 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

$$\overline{X} - t_{n-1,\alpha/2} \times \sqrt{V^2/n} < \mu < \overline{X} + t_{n-1,\alpha/2} \times \sqrt{V^2/n}.$$

ただし, $\overline{X}$: 標本平均値, S^2: 不偏標本分散, n: 標本サイズ, $t_{n-1,\alpha/2}$: 自由度 $n-1$ の t 分布の上側確率が $\alpha/2$ となる点.

母分散既知 確率統計 I(2025)L10 と比べて,

$\sigma^2 \rightsquigarrow S^2$

$z_{\alpha/2} \rightsquigarrow t_{n-1,\alpha/2}$

L12-Q2

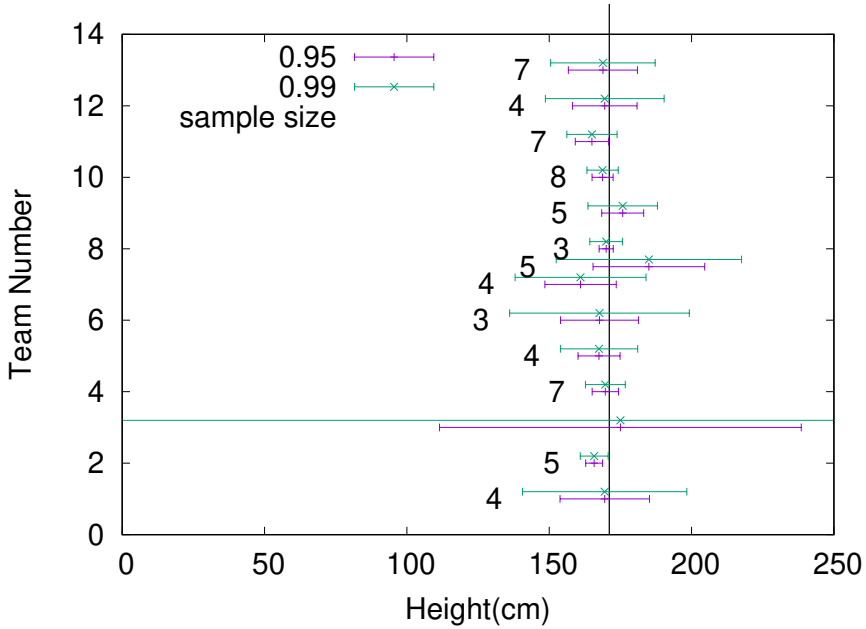
Quiz(母平均値の区間推定 (母分散未知))

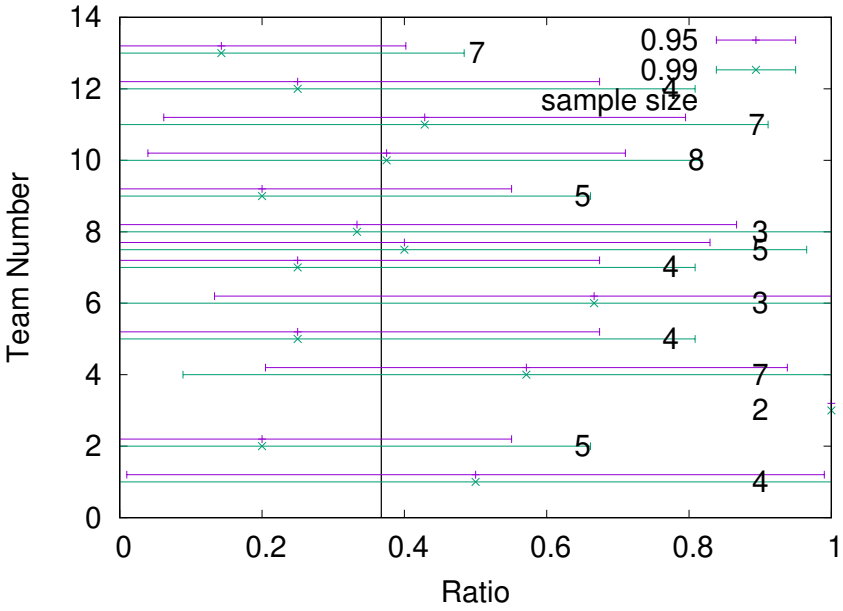
あるドーナツ製造マシンが製造するドーナツの重さ Xg は, 正規分布にしたがう確率変数である
製造された 4 個のドーナツの重さを測定したところ, 次のようだった.
$51\mathrm{g}, 52\mathrm{g}, 47\mathrm{g}, 50\mathrm{g}$.

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.

久保川 統計学入門 例題 8.7

クラス 2017 の身長の母平均値/「滋賀県」の母比率の区間推定





注: 母比率が 0.5 のとき, 信頼区間の幅は最大で $2 \times 1.96 \times 0.5/\sqrt{n}$.

ここまで来たよ

11 母分散の区間推定

12 母平均値の区間推定・線形回帰モデル
- 母平均値の区間推定 (正規母集団, 母分散既知，小標本)
- t 分布
- 母平均値の区間推定 (正規母集団, 母分散未知，小標本)
- 線形回帰モデルとしての回帰分析

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, u: 連続型確率変数, σ^2, β_0: 定数=パラメタ=母数.

$$Y = \beta_0 + u, \qquad u \sim N(0, \sigma^2)$$

u: 誤差, ノイズ. 小文字だけど確率変数.
(隠された) パラメタ β_0, σ^2 を, サイズ n 個の Y の標本から推定したい.

$$\begin{aligned} \mathrm{E}[Y] =& \beta_0 + \mathrm{E}[u] = \beta_0, \\ \mathrm{Var}(Y) =& \quad \mathrm{Var}(u) = \sigma^2. \end{aligned}$$

$u \sim N(0, \sigma^2) \rightsquigarrow Y = u + \beta_0 \sim N(\beta_0, \sigma^2)$. ここしばらくやってた, 正規母集団, 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.

(確率変数でない) 変数 x に依存する確率変数 Y

ドーナツ製造機の作るドーナツの重さ Y は, 温度 x による場合，線形単回帰モデルを仮定できる.

定義 (線形単回帰モデル 久保川 統計学入門 (12.1))

Y, u: 連続型確率定数, β_0, β_1: 回帰係数, σ^2: 定数=パラメタ

$$Y = \beta_0 + \beta_1 \cdot x + u, \qquad u \sim N(0, \sigma^2)$$

あれ？ データ分析 久保川 統計学入門 §2.2 で見た？ ここでは (x, Y) は平等ではない.

x: 説明変数 (独立変数) ここでは確率変数でない

Y: 被説明変数目的変数 (従属変数) ここでは確率変数

$\beta_0, \beta_1, \sigma^2$ は，未知の母ナントカ．データ (x_i, Y_i) から推定したい.

β_1: 回帰係数 (x を 1 だけ変えたときの y の変化量)

ノイズ・誤差 $u = Y - (\beta_0 + \beta_1 \cdot x) \sim N(0, \sigma^2)$ は縦向きのみ.

x_i の値を変えて，Y_i の実現値を測定 $\rightsquigarrow$ 独立だけど n 個の同一分布でない！

線形ではないけど

最小 2 乗法

直線 $y = \beta_0 + \beta_1 x$ からの，実現値 (x_i, y_i) の縦のずれの 2 乗

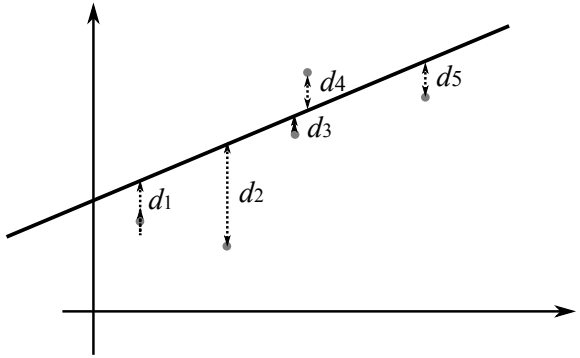
$$S(\beta_0, \beta_1) = \sum_{i=1}^{n} (y_i - (\beta_1 x_i + \beta_0))^2.$$

これが最小になるという条件で決める $\hat{\beta}_0, \hat{\beta}_1$ を，最小二乗推定量という．

$\frac{\partial S}{\partial \alpha}(\hat{\beta}_0, \hat{\beta}_1) = \frac{\partial S}{\partial \beta}(\hat{\beta}_0, \hat{\beta}_1) = 0$

微積分 II

u, Y が確率変数なので，$\hat{\beta}_0, \hat{\beta}_1$ も確率変数．

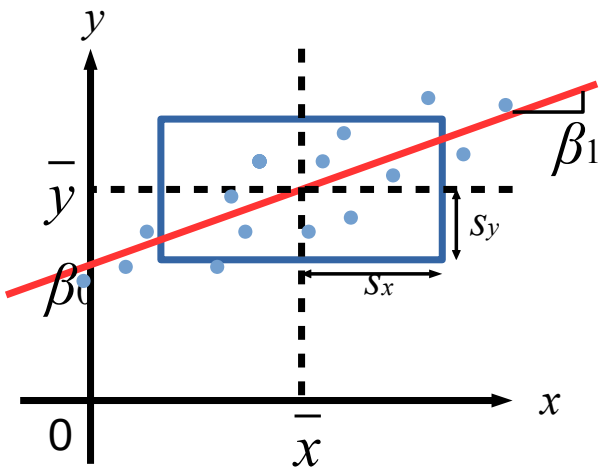


物理実験

回帰直線の公式再訪データ分析

$$\hat{\beta}_1 = \frac{S_{xy}}{S_x^2}$$

$$\hat{\beta}_0 = (\text{ 点 } (\overline{x}, \overline{Y}) \text{ を通るような値})$$



命題 (回帰直線 久保川 統計学入門 公式 2.8 **)**

x_i, Y_i $(i = 1, \ldots, n)$ の平均値を $\overline{x}, \overline{y}$, 標本標準偏差を S_x, S_y, 標本共分散を S_{xy} とする．このとき回帰直線は,

$$y = \frac{S_{xy}}{S_x^2} \times (x - \overline{x}) + \overline{y} = \hat{\beta}_0 + \hat{\beta}_1 \cdot x.$$

定義

- 予測値 ($\neq$ 推定量) $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i = \overline{y} + \hat{\beta}_1(x_i - \overline{y})$.
- 残差 $e_i = y_i - \hat{y}_i =$ 実測値 $-$ 予測値. これも確率変数.
- 残差平方和 $\mathrm{RSS} = \sum_i e_i^2$.

σ^2 の推定量 $\hat{\sigma}^2 = \frac{\sum_i e_i^2}{n-2}$.
自由度は $n-2$. β_0, β_1 の 2 個分ずるして e_i^2 を小さくできるから，その分を引いておく．$\hat{\beta}_1$ が最重要な推定量

$\hat{\beta}_1$ の母平均値と母分散

$$
\begin{aligned}
\hat{\beta}_1 =& \frac{1}{S_x^2}\frac{1}{n}\sum_i (x_i - \overline{x})(\beta_1(x_i - \overline{x}) + u_i), \\
\mathrm{E}[\hat{\beta}_1] =& \beta_1, \\
\mathrm{Var}(\hat{\beta}_1) =& \left(\frac{1}{nS_x^2}\right)^2 \mathrm{Var}(\sum_i (x_i - \overline{x})u_i) = \cdots = \frac{\sigma^2}{nS_x^2}.
\end{aligned}
$$

命題 (久保川 統計学入門 公式 12.1)

- $\hat{\beta}_1$ は β_1 の不偏推定量: $N(\beta_1, \frac{\sigma^2}{nS_x^2})$.
- $\hat{\sigma}^2$ は何倍かするとカイ二乗分布にしたがう：$\frac{(n-2)\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-2}$

$\hat{\beta}_1$ の精度を上げるには

標準偏差を小さくするには，n を大きくするか，S_x^2 を大きくするか．どちらも実験計画により実現できる．

さらなる問 久保川 統計学入門 p.232, 公式 12.2

- β_1 の信頼区間は？　確率統計 II,III
- $\beta_1 \neq 0$ なのか？ Y と x は「関係」あるのか？　統計的仮説検定 確率統計 I(2025)L13

統計的仮説検定・母平均値の両側 t 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 L13(2025-07-14 Mon)

最終更新: Time-stamp: "2025-07-15 Tue 07:00 JST hig"

今日の目標

- 統計的仮説検定を説明できる 久保川 統計学入門 SS9.1,9.2
- 母平均値の両側 t 検定ができる 久保川 統計学入門 §9.4.1

L12-Q1

Quiz 解答:t 分布の確率と $t_{k,\alpha}$

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

L12-Q2

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 $V^2 = \frac{1}{4-1} \cdot 14\mathrm{g}^2$.

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

2.

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

ここまで来たよ

12 母平均値の区間推定・線形回帰モデル

13 統計的仮説検定・母平均値の両側 t 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) 久保川 統計学入門 §9.1

点推定　μ は値 xx と推定 久保川 統計学入門 §7.2

区間推定　μ は値 yy と値 zz の間と推定 (信頼係数 $1-\alpha$ で) 久保川 統計学入門 §8

仮説検定　μ は値 xx と差がある，と, (とても確信のある $(1-\alpha)$ ときだけ) 断言 (有意水準 α で) =見逃し多いけど発色したら正しい血痕試験紙 久保川 統計学入門 §9

あるドーナツ製造機は, 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).

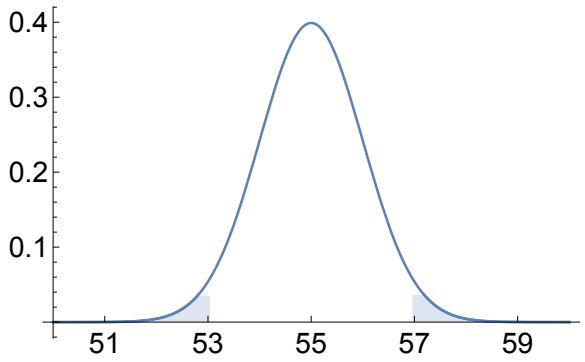
ある学習法を使ってるある生徒の, 毎日のテストでの 1 か月の平均点は 63 点．自分が別の学習法で教えた 5 日間の平均点は… 自分の方法は優れていると言いたい.

検定はだいたいこんな考え方 久保川 統計学入門 §9.2

確率変数 X は, 正規分布 $N(55, 2^2)$ 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. → 標本平均値 $\overline{X} = \frac{1}{4}(X_1 + \cdots + X_4) = 57$.
$\overline{X} \sim N(55, 2^2/4)$.
確率 α(=有意水準) でしか起きないくらいまれな, 帰無仮説から離れたことが起きたら (検定統計量 $\overline{X}$ の実現値が棄却域に含まれるなら) おかしいと判定する (帰無仮説を棄却する)
棄却域は, $\overline{X}$ の数直線上の領域のこと（不等式で指定される），その棄却域とそれ以外の領域（受容域）との境い目の値を棄却限界という.



帰無仮説と対立仮説 久保川 統計学入門 §9.2

母集団，母ナントカに対する仮説（標本に対するものではない）

- 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)
帰無仮説を受容する accept　「帰無仮説を棄却できない」
実験失敗 (背理法始めたけど矛盾導けなかった). 何も言えない (発色しな

ここまで来たよ

12 母平均値の区間推定・線形回帰モデル

13 統計的仮説検定・母平均値の両側 t 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

L13-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.
$50\mathrm{g}, 50\mathrm{g}, 51\mathrm{g}, 46\mathrm{g}, 48\mathrm{g}$.
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

久保川 統計学入門 p.189

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).

統計的仮説検定の一般的手順

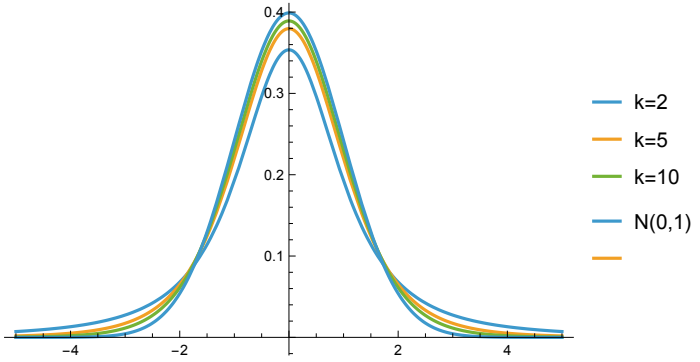
1. 「有意水準 $\alpha = \cdots$ で」「…検定を行う」(2,3 を名前で予告する)
2. 「帰無仮説を…とする」「対立仮説を…とする」
3. 「帰無仮説のもとでナントカ検定統計量 Y は …分布にしたがう」
4. 「この標本に対してナントカ検定統計量 Y の実現値は $y = \cdots$ である」
5. （棄却限界を計算して)「棄却域は…」
6. 「…より帰無仮説を棄却する/受容する」「よって（問に即した対立仮説の内容）と結論する/とは結論できない」

(6)「結論」の基本形は，(対立仮説) と結論する/結論できない，です．問と対立仮説の間に翻訳がはさまってる（例：変化した）ときは，問に対する答にします．単に (対立仮説) というだけでなく，「結論」というワードを入れるのは，「心から信じてるわけではないが，統計的仮説検定の手順に従うとこうなる」というニュアンスをいれるため

復習：t 分布 確率統計 I(2025)L12

定義 (t 分布 久保川 統計学入門 定義 8.3)

$Z \sim N(0, 1^2)$, $W \sim \chi_k^2$, Z と W が独立,
のとき連続型確率変数 $T = \frac{Z}{\sqrt{W/k}}$ 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.



命題 (T の分布 久保川 統計学入門 公式 8.4)

母集団 $N(\mu, \sigma^2)$ から, サイズ n の標本 $X_1, \cdots, X_n$ を取り出したとき,

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

は, 自由度 $n-1$ の Student の t 分布にしたがう.

正規分布にしたがう母集団の母平均値の両側 t 検定

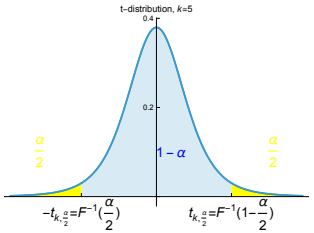
久保川 統計学入門 公式 9.4

母平均値の両側 t 検定 久保川 統計学入門

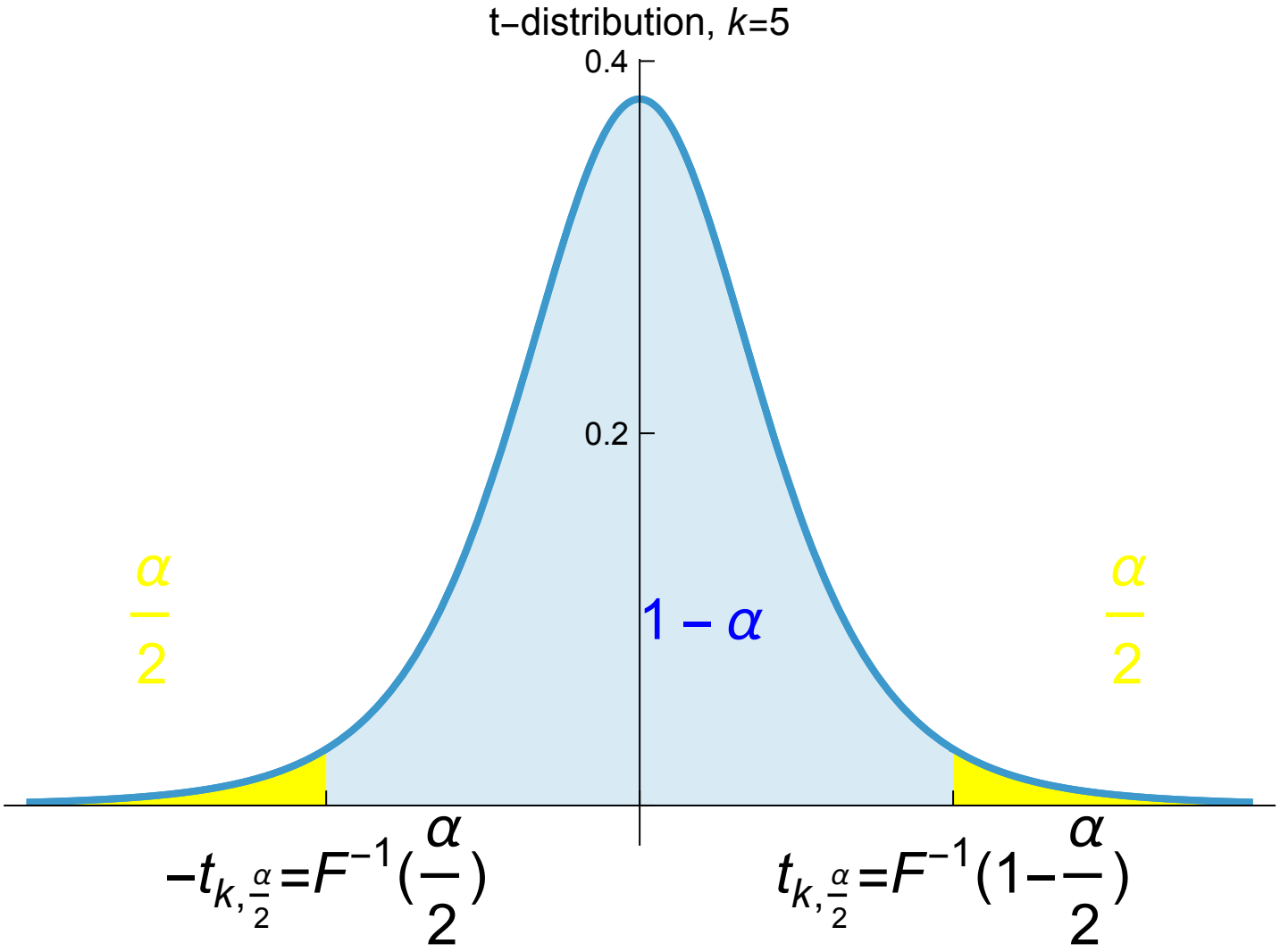
前提 母集団が正規分布 $N(\mu, \sigma^2)$ にしたがう，μ, σ^2 は不明.

1. 有意水準 α 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, 標本サイズ n の標本から計算する検定統計量 $T = \frac{\overline{X}-\mu_0}{\sqrt{V^2/n}}$ は自由度 $n-1$ の t 分布にしたがう.
4. 統計検定量 T の実現値 t を標本平均値 $\overline{X}$, 不偏標本分散 V^2 の実現値から計算すると $t = \cdots$.
5. 棄却限界は $C = F^{-1}(1-\frac{\alpha}{2}) = t_{n-1,\frac{\alpha}{2}}$, 棄却域は $|t| > C$ である.
6. 実現値 t は棄却域に含まれる/含まれないので (結論) 帰無仮説を棄却する/受容する, (対立仮説) 母平均値は $\mu \neq \mu_0$ と結論する/できない.

棄却域は t 軸の部分集合 $(-\infty, -C) \cup (C, +\infty)$. 受容域は $[-C, +C]$. ここで C は，$P(t \in$ 棄却域$) = \alpha$ であるように, $C = F^{-1}(1-\frac{\alpha}{2}) = t_{n-1,\frac{\alpha}{2}}$ と決めた.



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


t−distribution, k=5
0.4
0.2
$\frac{\alpha}{2}$
$1-\alpha$
$\frac{\alpha}{2}$
$-t_{k,\frac{\alpha}{2}}=F^{-1}(\frac{\alpha}{2})$
$t_{k,\frac{\alpha}{2}}=F^{-1}(1-\frac{\alpha}{2})$

L13-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 $\alpha = 0.05$ 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「$\mu = 55$」とする. 対立仮説を「$\mu \neq 55$」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 V^2 とするとき, 検定統計量
$$T = \frac{\bar{X} - 55}{\sqrt{V^2/5}}$$
は, 帰無仮説のもとで, 自由度 $5-1$ の t 分布に従う.
4. 検定統計量 T の実現値
$t = \frac{\overline{X}-55}{\sqrt{V^2/5}} = \cdots = \frac{49-55}{\sqrt{\frac{1}{5}\frac{16}{5-1}}} = -3\sqrt{5} = -6.708.$

5. 棄却限界は
$F^{-1}(1-0.05/2) = t_{4,0.05/2}$ =`rvt.ppf(1-0.05/2)`$= 2.776$ より，棄却域は $|t| > 2.776$.
棄却域は，「$t < F^{-1}(0.05/2)$ または $F^{-1}(1-0.05/2) < t$」とも書ける．
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 55g と異なる，と結論する．

(注: 帰無仮説の棄却のことを，業界用語「有意」significant で表現する人もいる．検定は有意である，母平均値 μ は 55g 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．

久保川 統計学入門 例題 9.5, 基本問題問 1,3(p.200)

L13-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.

母平均値・母分散・母比率・適合度・独立性の検定

	両側	片側
母平均値	**両側 t 検定** 久保川 統計学入門 公式 9.4	**片側 t 検定** 久保川 統計学入門 公式 9.4
母分散	**両側カイ二乗検定** 久保川 統計学入門 公式 9.7	**片側カイ二乗検定** 久保川 統計学入門 公式 9.7
母比率	**両側二項検定** 久保川 統計学入門 (9.7)	**片側二項検定** 久保川 統計学入門 例題 9.10
適合度	NA	**片側カイ二乗検定** 久保川 統計学入門 公式 11.2
独立性	NA	**片側カイ二乗検定** 久保川 統計学入門 公式 11.4
母平均値の差	**両側 t 検定** 対応なし (等分散, Welch) 久保川 統計学入門 公式 10.7, あり 久保川 統計学入門 公式 10.5	**片側 t 検定** 対応なし (等分散, Welch), あり
母分散の比	**両側 F 検定** 久保川 統計学入門 公式 10.3	**片側 F 検定**

ふつうじゃない trial L13 非参照テストではなく, 2025-07-28 月 20:00 までにアップロードするレポート，したがって，2025-07-21 月 1 には trial はない.

- 練習問題 L13-01, L13-02 の両方に合格（今回に限って満点でなくても合格することがある）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.

片側検定・母比率の検定

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(2025-07-21 Mon)

最終更新: Time-stamp: "2025-07-23 Wed 07:46 JST hig"

今日の目標

- 母比率の片側検定ができる 久保川 統計学入門 §9.5
- p 値を使った検定ができる 久保川 統計学入門 §9.3
- 第 1 種, 第 2 種の過誤, 有意水準と検出力を説明できる 久保川 統計学入門 §9.3

統計的仮説検定の「結論」は，基本形は，(対立仮説) と結論する/結論できない，です．問と対立仮説の間に翻訳がはさまってる（例：変化した）ときは，問に対する答にします．単に (対立仮説) というだけでなく，「結論」というワードを入れるのは，「心から信じてるわけではないが，統計的仮説検定の手順に従うとこうなる」というニュアンスをいれるためです．

L13-Q1

解答は 確率統計 I(2025)L13 に掲載

L13-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 $n = 4$ の標本の標本平均値を $\bar{X}$, 不偏標本分散を V^2 とすると, 検定統計量
$$T = \frac{\bar{X} - 196}{\sqrt{V^2/4}}$$
は, 帰無仮説のもとで, 自由度 $4 - 1$ の t 分布に従う.
4. この標本の実現値は $\bar{X} = 200, V^2 = \frac{224}{4-1} = 74.7$. よって, 検定統計量の実現値は，$t = \frac{200-196}{\sqrt{\frac{224}{3}/4}} = 0.92582$.
5. 棄却限界は $C = F^{-1}(0.975) = t_{3,0.05/2}$. 棄却域は $|t| > C$ すなわち, $|t| > 3.182$.
6. $|0.92582| < 3.182$ であり, 実現値は棄却域に含まれないので, 帰無仮説を受容する．来店客数の母平均値が変化したとは結論できない. (注: 検定は有意でなかった, 母平均値 μ と 196g の間には有意差がない, 有意な標本でないなど).

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 Y の実現値は $y = \cdots$ である」
5. （棄却限界を計算して)「棄却域は…」
6. 「…より帰無仮説を棄却する/受容する」「よって（問に即した対立仮説の内容）と結論する/とは結論できない」

母平均値・母分散・母比率・適合度・独立性の検定

	両側	片側
母平均値	**両側 t 検定** 久保川 統計学入門 公式 9.4	**片側 t 検定** 久保川 統計学入門 公式 9.4
母分散	**両側カイ二乗検定** 久保川 統計学入門 公式 9.7	**片側カイ二乗検定** 久保川 統計学入門 公式 9.7
母比率	**両側二項検定** 久保川 統計学入門 (9.7)	**片側二項検定** 久保川 統計学入門 例題 9.10
適合度	NA	**片側カイ二乗検定** 久保川 統計学入門 公式 11.2
独立性	NA	**片側カイ二乗検定** 久保川 統計学入門 公式 11.4
母平均値の差	**両側 t 検定** 対応なし (等分散, Welch) 久保川 統計学入門 公式 10.7, あり 久保川 統計学入門 公式 10.5	**片側 t 検定** 対応なし (等分散, Welch), あり
母分散の比	**両側 F 検定** 久保川 統計学入門 公式 10.3	**片側 F 検定**

ここまで来たよ

13 統計的仮説検定・母平均値の両側 t 検定

14 片側検定・母比率の検定
- 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 $Bin(n,p)$ の p を, 今日は r と書きます

例 (母比率の検定のアイデア 久保川 統計学入門 §9.1)

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 1 回投げたときの表 (1) 裏 (0) $X_i \sim Bin(1,r)$. (無限母集団はベルヌイ分布)
$Bin(1,r)$ のサイズ n のサンプルに対して合計 $X = \sum_{i=1}^{n} X_i$ は
$X \sim Bin(n,r)$. (二項分布).
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(2025)L13

$$P(X = 15) = \frac{50!}{15!(50-15)!}\left(\frac{1}{5}\right)^{15}\left(1 - \frac{1}{5}\right)^{50-15} = 0.0299 \quad \text{稀じゃん}$$

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 $50 \cdot \frac{1}{5} = 10$ 回だって稀じゃん?

$$P(X = 10) = \frac{50!}{10!(50-10)!}\left(\frac{1}{5}\right)^{10}\left(1 - \frac{1}{5}\right)^{50-10} = 0.1398$$

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? …計算機で直接計算する手もある正確確率検定

正規近似による計算 確率統計 I(2025)L08

$\mathrm{E}[X] = 50 \cdot \frac{1}{5} = 10, \mathrm{Var}(X) = 50 \cdot \frac{1}{5} \cdot \frac{4}{5} = 8.$
サンプルサイズ 50 大と思うと，中心極限定理より，近似的に $X \sim N(10, 8)$. $Z = \frac{X-10}{\sqrt{8}} \sim N(0, 1^2)$.

$$\begin{aligned} P(15 \leq X) =& P(\tfrac{15-10}{\sqrt{8}} \leq Z < +\infty) \\ =& \Phi(\infty) - \Phi(\tfrac{15-10}{\sqrt{8}}) \\ =& 1 - \mathsf{cdf}(\tfrac{15-10}{\sqrt{8}}) = 1 - 0.96145 = 0.0385. \end{aligned}$$

確率統計 I(2024)L09

$P(X \geq 15) = 0.0385$ (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. そのときのぎりぎりの z が棄却限界，それより遠くが棄却域

母比率の検定 IV

上で出てきた Z は, 分子分母を $n = 50$ で割ると, 帰無仮説 $r = r_0 = \frac{1}{5}$, 標本比率の実現値 $\hat{r} = \frac{3}{5}$ により,

$$\frac{15-10}{\sqrt{8}} = \frac{\frac{15}{50} - \frac{10}{50}}{\sqrt{\frac{1}{50}\frac{1}{5}\frac{4}{5}}} = \frac{\hat{r} - r_0}{\sqrt{\frac{1}{n} \cdot r_0 \cdot (1 - r_0)}}$$

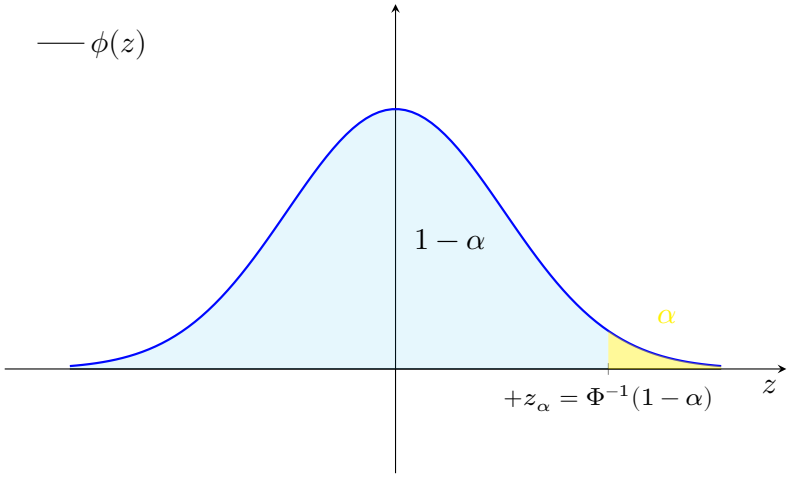
という, 母比率の区間推定の導出に似た形になる.
ただし, 平方根の中にあるのは帰無仮説の母比率 $r_0 = \frac{1}{5}$, 標本比率は $\hat{r}$.

久保川 統計学入門 公式 9.9+式 (9.7)

復習 確率統計 I(2025)L10

母比率 r は，ベルヌーイ分布 $Ber(r)$ の母平均値 r. 母分散は $r(1-r)$.
標本比率は標本平均値 $\overline{X} = \hat{r} = k/n$.

$\boxed{\text{近似 1}}$ サンプルサイズが大きいとき，中心極限定理により，標本比率は近似的に正規分布 $N(r, r(1-r)/n)$ に従う.



定義 (標準正規分布の z_α 久保川 統計学入門 公式 5.9,(8.5), 図 8.4 確率統計 I(2025)L10**)**

$Z \sim N(0, 1^2)$ のとき，上側確率 $P(Z \geq z_\alpha) = \alpha$ となる点を z_α と定める.
Φ を累積分布関数，Φ^{-1} をその逆関数とするとき，
$P(+z_\alpha < Z < +\infty) = 1 - \Phi(z_\alpha) = \alpha$, $z_\alpha = \Phi^{-1}(1-\alpha)$.

母比率の片側 (二項) 検定 (の正規近似)

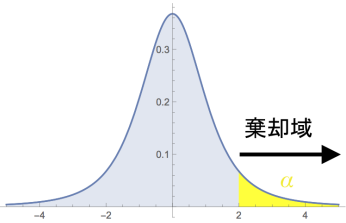
久保川 統計学入門 公式 9.9+式 (9.7)

母比率の片側 (二項) 検定 (の正規近似)

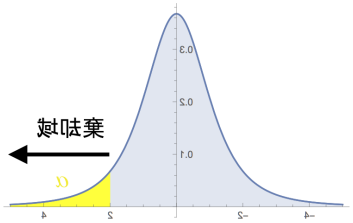
前提 $X \sim Bin(n, r)$.

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，サイズ n の標本から計算する検定統計量 $Z = \frac{X - nr_0}{\sqrt{nr_0(1-r_0)}} = \frac{\hat{r} - r_0}{\sqrt{r_0(1-r_0)/n}}$ は近似的に標準正規分布 $N(0, 1^2)$ にしたがう (n 大のとき). $\hat{r} = \frac{X}{n}$: 標本比率.
4. 標本に対して検定統計量 Z の実現値 z は…
5. 棄却限界は $\Phi^{-1}(1-\alpha)$ (or $\Phi^{-1}(\alpha) = -\Phi^{-1}(1-\alpha)$) 棄却域は $z > \Phi^{-1}(1-\alpha)$ (or $z < \Phi^{-1}(\alpha)$).
6. 「(…) より帰無仮説を棄却する/受容する. 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



L14-Q1(上側片側)

TA Prob and Sol: 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.

略解

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 $X \sim Bin(100, \frac{1}{10})$. 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 $N(0, 1^2)$ 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 $z = \frac{5}{3} = 1.667$.
5. 棄却限界は $z_{0.05} = \Phi^{-1}(1 - 0.05) =$ scipy.stats.norm(loc=0,scale=1).ppf$(1 - 0.05) = 1.645$. 棄却域は $z > z_{0.05}$
6. 実現値 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.

p 値による場合

5. p 値は $p = P(Z > \frac{5}{3}) = \Phi(+\infty) - \Phi(\frac{5}{3}) =$ $1 -$ 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．

棄却域に含まれない場合．p 値が α より大きい場合．

6. 帰無仮説を受容する．母比率 $r > \frac{1}{10}$ とは結論できない.」

久保川 統計学入門 例題 9.11

L14-Q2 (下側片側)

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.

久保川 統計学入門 問 1, 問 5(p.200)

ここまで来たよ

13 統計的仮説検定・母平均値の両側 t 検定

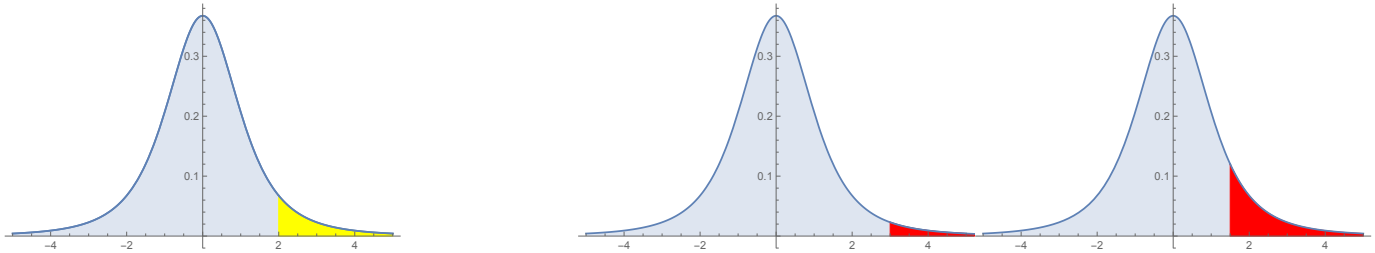
14 片側検定・母比率の検定
- 母比率の検定=二項検定 (の正規近似)
- p 値=有意確率
- 仮説検定での第 1 種・第 2 種の過誤と混同行列

p 値=有意確率による棄却する/しない判定 久保川 統計学入門 §9.3

検定の Step6 では, 検定統計量の実現値 z, t 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 (α で決まる) に入っているか?
$\Updownarrow$
p 値=有意確率 $P(X \geq 15) = P(z \geq \frac{15-30}{\sqrt{8}})$ が有意水準 α より小さいか?

標本の p 値 (p-value)=有意確率 久保川 統計学入門 (9.3)(9.4)

帰無仮説のもとで, 検定統計量がこの標本より極端な値をとる確率
$p < \alpha \Leftrightarrow$ 帰無仮説棄却



母比率の検定の上側片側検定のとき p 値 $= 1 - \Phi(z)$.
棄却域の考え方での棄却: $z > \Phi^{-1}(1-\alpha)$.　$\Phi(z)$ は増加関数だから…
p 値の考え方での棄却: $\Phi(z) > 1 - \alpha$ つまり $p = 1 - \Phi(z) < \alpha$ だから…

母比率の片側 (二項) 検定 (の正規近似) in p 値

久保川 統計学入門 §8.2

母比率の片側 (二項) 検定 (の正規近似)

前提 $X \sim Bin(n, r)$.

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，サイズ n の標本から計算する検定統計量 $Z = \frac{X - nr_0}{\sqrt{nr_0(1-r_0)}} = \frac{\hat{r} - r_0}{\sqrt{r_0(1-r_0)/n}}$ は近似的に標準正規分布 $N(0, 1^2)$ にしたがう (n 大のとき). $\hat{r} = \frac{X}{n}$: 標本比率.
4. 標本に対して検定統計量 Z の実現値 z は…
5. ~~棄却限界は $\Phi^{-1}(1-\alpha)$. 棄却域は $z > \Phi^{-1}(1-\alpha)$ (or $z < \Phi^{-1}(\alpha)$).~~ 実現値 z の p 値は $p = \Phi(\infty) - \Phi(z)$. (or $p = \Phi(z) - \Phi(-\infty)$)
6. 「$p < \alpha$ より帰無仮説を棄却する (受容する). よって母比率 $r > r_0$ (or $r < r_0$) と結論する/できない」

ここまで来たよ

13 統計的仮説検定・母平均値の両側 t 検定

14 片側検定・母比率の検定
- 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

仮説検定での第 1 種・第 2 種の過誤と混同行列

	帰無仮説を棄却, 有意である	帰無仮説を受容, 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, その上で β をなるべく小さくするように試みる習慣.
検定とは, 発色 (検出) したら (まれな誤検出を除いて) ほぼ確かなだけど, 見逃すことはある異常検査薬のようなもの.
α 誤検出率, 有意水準, 危険率
β: 見逃し率 (検出力:$1-\beta$)
類似: 混同行列

機械学習

ふつうじゃない trial L13 非参照テストではなく, 2025-07-28 月 20:00 までにアップロードするレポート，したがって，2025-07-21 月 1 には trial はない．

- 練習問題 L13-01, L13-02 の両方に合格（今回に限って満点でなくても合格することがある）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.
2025-08-04 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
災害や交通機関などにより全学的に定期試験が中止・延期になるときは同様に中止・延期.
成績評価
考慮対象になるかもしれない欠席の届は 2025-07-30 水までに提出.

条件付き分布，ベイズ推定

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　L15(2025-07-28 Mon)

最終更新: Time-stamp: "2025-07-28 Mon 07:51 JST hig"

今日の目標

- 第 1 種, 第 2 種の過誤, 有意水準と検出力を説明できる 久保川 統計学入門 §9.3
- 適合度のカイ二乗検定ができる 久保川 統計学入門 §11.1
- ベイズ推定できる 久保川 統計学入門 §3.4

L14-Q1(上側片側)
解答は 確率統計 I(2025)L14 に掲載
L14-Q2 (下側片側)

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. 棄却限界は $-z_{0.01} = \Phi^{-1}(0.01) =$ scipy.stats.norm(loc=0,scale=1).ppf$(0.01) = -2.326$. 棄却域は $z < -z_{0.01}$.
6. 検定統計量の実現値 z は棄却域に含まれない．帰無仮説を受容する．母比率 $r < \frac{1}{10}$ と結論できない．

p 値による場合

5. p 値は $p = P(Z < -\frac{5}{3}) = \Phi(-\frac{5}{3}) - \Phi(-\infty) =$ scipy.stats.norm(loc=0,scale=1).cdf$(-\frac{5}{3}) = 0.0475$.
6. $0.0475 = p > \alpha = 0.01$ なので帰無仮説を受容する．よって母比率 $r < \frac{1}{10}$ と結論できない．

棄却域に含まれる場合．p 値が α より小さい場合．

6. 帰無仮説を棄却する．母比率 $r < \frac{1}{10}$ と結論する．

ここまで来たよ

14 片側検定

15 条件付き分布，ベイズ推定
- 検定の第 1 種・第 2 種の過誤・適合度のカイ二乗検定
- 条件付き確率

母平均値・母分散・母比率・適合度・独立性の検定

	両側	片側
母平均値	**両側 t 検定** 久保川 統計学入門 公式 9.4	**片側 t 検定** 久保川 統計学入門 公式 9.4
母分散	**両側カイ二乗検定** 久保川 統計学入門 公式 9.7	**片側カイ二乗検定** 久保川 統計学入門 公式 9.7
母比率	**両側二項検定** 久保川 統計学入門 (9.7)	**片側二項検定** 久保川 統計学入門 例題 9.10
適合度	NA	**片側カイ二乗検定** 久保川 統計学入門 公式 11.2
独立性	NA	**片側カイ二乗検定** 久保川 統計学入門 公式 11.4
母平均値の差	**両側 t 検定** 対応なし (等分散, Welch) 久保川 統計学入門 公式 10.7, あり 久保川 統計学入門 公式 10.5	**片側 t 検定** 対応なし (等分散, Welch), あり
母分散の比	**両側 F 検定** 久保川 統計学入門 公式 10.3	**片側 F 検定**

仮説検定での第 1 種・第 2 種の過誤と混同行列

久保川 統計学入門 §9.3

	帰無仮説を棄却, 有意である	帰無仮説を受容, 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, その上で β をなるべく小さくするように試みる習慣.
検定とは, 発色 (検出) したら (まれな誤検出を除いて) ほぼ確かなだけど, 見逃すことはある異常検査薬のようなもの.
α 誤検出率, 有意水準, 危険率
β: 見逃し率 (検出力:$1-\beta$)
類似: 混同行列

機械学習

カテゴリ変数 久保川 統計学入門 §1.5

量的変数の確率変数 ... 平均や分散に意味がある

- 離散型 確率関数=表 二項分布, ポアソン分布, $\cdots$, x は整数
- 連続型 確率密度関数 正規分布, カイ二乗分布, 指数分布，一様分布, $\cdots$, x は実数

質的変数の確率変数 ... 平均や分散が考えられない
その中でも, 名義変数=カテゴリ (カル) 変数
順序や距離がなくぜんぶが対等. 例: 血液型, 性別, 携帯電話番号, チーム
A 型, B 型などがカテゴリ

2 カテゴリなら, 0,1 のように番号を振って離散型と思える
3 カテゴリ以上なら, 順序や間隔によるので離散型には帰着できない.
あえてやるなら one-hot vector.

カテゴリ分布と多項分布 (multinomial distribution)

久保川 統計学入門 §14.1.3

定義 (k カテゴリのカテゴリ分布 $\mathrm{C}(p_1,\dots,p_k)$)

カテゴリ番号 $i=1,\dots,k$ と，パラメタ $p_1+\cdots+p_k=1$. 略記 $=(p_1,\dots,p_k)$ に対して，

$$\text{確率関数}p(i)=P(X=i)=p_i=\begin{cases} p_1 & (i=1)\\ p_2 & (i=2)\\ \vdots & \\ p_k & (i=k)\end{cases}$$

ベルヌイ分布 $Ber(p_1)$ は $k=2$ の $\mathrm{C}(1-p,p)$.

定義 (k カテゴリの多項分布 $Mult_k(n,p_1,\dots,p_k)$)

確率変数は整数値ベクトル $(X_1,\dots,X_k)$. パラメタ $p_1,\dots,p_k,n$.

$$\text{確率関数}P(X_1=x_1,\dots,X_k=x_k)=f(x_1,\dots,x_k)=\begin{cases} \frac{n!}{x_1!\dots x_k!}p_1^{x_1}\dots p_k^{x_k} & (x_i\ge 0, x_1+\cdots+x_k=n)\\ 0 & (\text{他})\end{cases}$$

カテゴリ分布にしたがう確率変数の n 個の実現値を得て，カテゴリ $1,\dots,k$ をそれぞれ $x_1,\dots,x_k$ 回ずつ得る確率.

	$n=1$	n	ベイズ共役分布	
$k=1$ k	ベルヌイ分布 $Ber(p)$ カテゴリ分布 $\mathrm{C}(p_1,\dots,p_k)$	二項分布 $Bin(n,p)$ 多項分布 $Mult_k(1,p_1,\dots,p_k)$	二項検定 適合度の検定	ベータ分布 ディリクレ分布

適合度基準 久保川 統計学入門 §11.1

カテゴリの個数 $k = 4$ の例

カテゴリ	O 型	A 型	AB 型	B 型
確率 p_i	$p_1 = 0.12$	$p_2 = 0.51$	$p_3 = 0.17$	$p_4 = 0.20$

$$\sum_{i=1}^{k} p_i = 1.$$

カテゴリ分布のサイズ $n = 12$ の標本

番号	血液型
1	B 型
2	O 型
⋮	⋮
12	A 型

$\rightarrow$ 度数分布表は多項分布 $\mathrm{M}(,12)$ の実現値のはず

カテゴリ	O 型	A 型	AB 型	B 型
度数 x_i	$x_1 = 2$	$x_2 = 3$	$x_3 = 6$	$x_4 = 1$

$$\sum_{i=1}^{k} x_i = n = 12.$$

適合度を表す量

カテゴリ i の期待度数=標本サイズ × 母分布の確率=np_i

命題 (ピアソンの適合度基準 Q 久保川 統計学入門 §11.1)

カテゴリ k 個, カテゴリ分布の確率 $p_i (i = 1, \dots, k)$
サイズ n の標本の度数 $x_i (i = 1, \dots, k)$
サイズ n の標本の, 母分布から考えた期待度数 $np_i (i = 1, \dots, k)$

$$\text{ピアソンの適合度基準} Q(x) = \frac{(\text{度数} - \text{期待度数})^2}{\text{期待度数}} \text{の合計} = \sum_{i=1}^{k} \frac{(x_i - np_i)^2}{np_i}$$

が小さいほど, 標本は母分布から予想されるものに近い

命題 (久保川 統計学入門 公式 11.2)

n 大のとき，ピアソンの適合度基準 Q は，自由度 $k-1$ のカイ二乗分布にしたがう．

このことを利用して，帰無仮説「$p_1, \dots, p_k$ のカテゴリ分布にしたがう」対立仮説「したがわない ($\Leftrightarrow p_i$ が 1 個は異なる」で検定が行える 適合度のカイ二乗検定 久保川 統計学入門 §11.1

ここまで来たよ

14 片側検定

15 条件付き分布，ベイズ推定
- 検定の第 1 種・第 2 種の過誤・適合度のカイ二乗検定
- 条件付き確率

復習：同時分布と周辺分布

久保川 統計学入門 §6.1　確率統計 I(2025)L06

X, Y 確率変数

同時分布

$p(x,y) = P(X = x \text{かつ} Y = y) =$

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

周辺分布

X, Y の周辺分布 $p_{\mathrm{X}}(x)$,$p_{\mathrm{Y}}(y)$ は,

$$\text{離散型} p_{\mathrm{X}}(x) = p(x, \bullet) = \sum_y p(x,y), \quad p_{\mathrm{Y}}(y) = p(\bullet, y) = \sum_x p(x,y)$$

要するに 一方を無視した分布．小計．

条件付き確率 久保川 統計学入門 §3.3

定義 (条件付き確率 久保川 統計学入門 定義 3.7, 公式 3.8)

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 久保川 統計学入門 §6.3

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

定義 (離散型確率変数に対する条件付き確率)

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 久保川 統計学入門 定義 6.9)

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

L15-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) 久保川 統計学入門 公式 3.8

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

全確率の法則 (周辺分布との関係) 久保川 統計学入門 §3.4

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

ベイズの定理 久保川 統計学入門 §3.4 I

定理 (ベイズの定理 久保川 統計学入門 公式 3.15)

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

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

$p_{\mathrm{X}|\mathrm{Y}}(x|y)$ 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).

L15-Q2

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．

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

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

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．

ベイズ推定

$$\boxed{\text{事後確率 } p_{\mathrm{X}|\mathrm{Y}}(x|y)} \longleftarrow \boxed{\text{事前確率 } p_{\mathrm{X}}(x)}$$

$$\uparrow$$

$$\boxed{\text{情報 } Y = y}$$

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．

例えばこんな問 I

L15-Q3

Quiz(ベイズの推定)

甘い品種 1 と渋い品種 2 の柿が袋に入っている．
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. 袋の柿の $1/5$ が甘い品種 1 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品種 1 である確率を求めよう.
2. 袋から取り出すの柿の $1/5$ が渋い品種 2 であると考えている．いま, 1 個の柿を取りだしたところ, 黄色い柿だった. この柿が渋い確率を求めよう.

L15-Q4

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 (真陰性).

1. 1 回の検査で陽性となった人が，病気である確率を求めよう.
2. 1 回の検査で陽性となった人が，さらにもう 1 回検査してやはり陽性だった．その人が病気である確率を求めよう.

混同行列 confision matrix

- 検査の性能を表現する数値は 1 通りではない.
- 現実には本当の確率はわからなくて, 人数の表 (混同行列) で記録する
  - 今後出てくる, 標本抽出, 推定の考えが混ざっている

	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
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= $\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}$
- 特異度=真陰性率=specificity= $\frac{\mathrm{TN}}{\mathrm{TN}+\mathrm{FP}}$
- 適合率=precision= $\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}}$
- 正解率=精度=accuracy = $\frac{\mathrm{TP}+\mathrm{TN}}{\text{すべて}}$
- F 値=recall と precision の調和平均

病気に限らない. 例: 猫画像判定 AI の性能, 陽性=画像は猫, 検査で陽性=AI が画像は猫と判定

ふつうじゃない trialL15(締切 2025-08-04 月 20:00)

- trialL15-1 LearnMoodle 練習問題的なものだが，受験は 2 回まで．チェックは使えない
  - ほぼ同内容の予行演習用の練習問題 L15 に合格すると開放されます．
  - 趣旨: 対面でやる機会がない
- trialL15-2 LearnMoodle 作文
  - まずアンケートにご協力お願いします．回答すると開放されます．
  - 学習履歴分析への同意をお願いします

定期試験期間中の任意参加授業内小テスト

用途:LearnMoodle 成績評価方法参照．
2025-08-04 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（注：2024 年度と L08 の内容が異なる）．要学生証, 個人別座席指定．
すべて持込不可．
災害や交通機関などにより全学的に定期試験が中止・延期になるときは同様に中止・延期．

成績評価

考慮対象になるかもしれないものは LearnMoodle 欠席届に 2025-07-30 水までに提出．

メッセージ

- 教科書は売らないで．確率統計 II,III（3 年後期）他で使います．
- 数理と社会（3 年 1Q）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(2025-07-28 Mon)

最終更新: Time-stamp: "2025-07-27 Sun 11:28 JST hig"

今日の目標

-

L15-Q1

Quiz 解答: 条件付き分布

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

L15-Q2

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

L15-Q3

Quiz 解答: ベイズの推定

甘い品種 1:$x=1$, 渋い品種 2:$x=2$, 赤:$y=10$, 黄:$y=20$ と表現する.

1.

$$p_{\mathrm{Y|X}}(y|1)=\begin{cases}0.95 & (y=10)\\ 0.05 & (y=20)\end{cases}$$

$$p_{\mathrm{Y|X}}(y|2)=\begin{cases}0.125 & (y=10)\\ 0.875 & (y=20)\end{cases}$$

$y\backslash x$	1	2
10	0.19	0.10
20	0.01	0.70

$$\begin{aligned}p_{\mathrm{X|Y}}(1|10)=&\frac{p_{\mathrm{Y|X}}(10|1)p_{\mathrm{X}}(1)}{\sum_x p_{\mathrm{Y|X}}(10|x)p_{\mathrm{X}}(x)}\\ =&\frac{0.95\times 0.2}{0.95\times 0.2+0.125\times 0.8}=\frac{19}{29}.\end{aligned}$$

L15-Q4

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

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

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

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