大数の法則/母集団・標本抽出・推定

樋口さぶろお　https://hig3.net

龍谷大学理工学部数理情報学科

確率統計☆演習 I　L10(2019-12-09 Mon)

最終更新: Time-stamp: "2019-12-15 Sun 10:37 JST hig"

今日の目標

- 大数の法則の内容を説明できる 岩薩林 確率・統計 §4.4
- 母集団, 標本, 標本抽出, 推定の考えが説明できる 岩薩林 確率・統計 §5.1,5.2
- 母平均値, 母期待値, 母分散, 母比率を点推定でき

L08-Q4

Quiz 解答:正規分布の確率 $Z = \frac{X-3}{2}$ とすると, Z は標準正規分布 $\mathrm{N}(0, 1^2)$ にしたがう.

1. $\mathrm{E}[X^2] = \mathrm{V}[X] + \mathrm{E}[X]^2 = 2^2 + 3^2$.
2. $P(X \geq 5) = P(\frac{5-3}{2} \leq Z < +\infty) = \int_1^\infty f(z)\,\mathrm{d}z = I(+\infty) - I(1) = \frac{1}{2} - I(1) = 0.1587$.
3. $P(1 \leq X \leq 7) = P(-1 \leq Z \leq 2) = \int_{-1}^2 f(z)\,\mathrm{d}z = I(2) - I(-1) = I(2) + I(1) = 0.8186$.

L09-Q1

Quiz 解答:二項分布

1. 二項分布 $\mathrm{B}(100, \frac{2}{3})$ に従う確率変数を X とする. $P(X = 40)$ を求めればいいから, ${}_{100}\mathrm{C}_{40} p^{40}(1-p)^{100-40} = \frac{100!}{40!60!}(\frac{2}{3})^{40}(1-\frac{2}{3})^{60}$.
2. $\mathrm{E}[X] = n \times p = \frac{200}{3}$. $\mathrm{V}[X] = n \times p(1-p) = \frac{200}{9}$.

L09-Q2

Quiz 解答:独立同分布の母期待値母分散

1. 3μ.
2. 3μ.
3. 3μ
4. $9\sigma^2$.
5. $3\sigma^2$.
6. $5\sigma^2$.
7. $3(\sigma^2+\mu^2)+3\times 2\times \mu^2$.

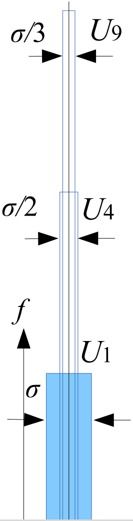
L10-Q3

Quiz 解答:独立同分布にしたがう変数の和

1. A は母平均値が μ, 母分散が $\frac{\sigma}{n}$.
2. B は母平均値が 0, 母分散が σ^2.
3. C は母平均値が 0, 母分散が 1.

二項分布の母平均値，母分散の式の証明 岩薩林 確率・統計例題 4.6

U_n の確率関数は
こんな感じ？



i.i.d にしたがう確率変数の和の $1/n$

新しい確率変数: $U_n = \frac{1}{n}T_n = \frac{1}{n}(X_1 + \cdots + X_n)$

$$\mathrm{E}[U_n] = \mathrm{E}\left[\tfrac{1}{n}T_n\right] = \frac{1}{n} \times n \times \mu.$$

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

ここまで来たよ

9. 二項分布, 独立同分布の和
   - チェビシェフの不等式と大数の法則

10. 大数の法則/母集団・標本抽出・推定
   - 母集団と標本
   - 母平均値・母分散の (点) 推定
   - 母比率とその (点) 推定

チェビシェフの不等式

チェビシェフの不等式 Chebyshev's inequality 岩薩林 確率・統計 §4.3(4.15)

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

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

証明 岩薩林 確率・統計 §4.3
どんな X にも使えて便利な不等式. 意味は…

大数の (弱) 法則アバウト版 岩薩林 確率・統計 §4.4

$X_1, \ldots, X_n$ が独立同分布にしたがい, $\mathrm{E}[X_i] = \mu$, $\mathrm{V}[X_i] = \sigma^2$,
$U_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ のとき, n が十分大きいとき U_n は 'ほぼ μ に等しい'
(U_n が μ から外れる確率はゼロに近づく)

大数の (弱) 法則 岩薩林 確率・統計 §4.4

$X_1, \ldots, X_n$ が独立同分布にしたがい, $\mathrm{E}[X_i] = \mu$, $\mathrm{V}[X_i] = \sigma^2$,$U_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ のとき,

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

つまり n 大で U_n は $\mathrm{E}[U_n] = \mu$ に「必ず近い」(確率収束).
証明　$\mathrm{E}[U_n] = \mu, \mathrm{V}[U_n] = \sigma^2/n$ に注意して, U_n に対するチェビシェフの不等式を書くと,

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

$a = \frac{\epsilon}{\frac{\sigma}{\sqrt{n}}}$ とすると, $n \to +\infty$ で

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

これが母平均値・母期待値の直観的意味. 要するに,

ここまで来たよ

9. 二項分布, 独立同分布の和
    - チェビシェフの不等式と大数の法則

10. 大数の法則/母集団・標本抽出・推定
    - 母集団と標本
    - 母平均値・母分散の (点) 推定
    - 母比率とその (点) 推定

母集団と標本 (1) 有限母集団 岩薩林 確率・統計 §§5.1,5.2

某アイドルグループの身長ふたたび

- 某アイドルグループ全員 (→ 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!
  - メンバー 1 名を等確率で選んでくる, という試行を考えると, 確率変数 X の母平均値 $\mu = \mathrm{E}[X]$.
- メンバー全員分のデータがあれば定義の式使うだけ
- 握手会でメンバー 1 人ずつに質問しなければいけないとしたら?
- 握手会参加券 40 枚集めないで何とかすませたい.

$\rightsquigarrow$ 質問できたメンバー 5 人の身長 (=標本)(独立同分布にしたがう確率変数 $X_1, X_2, \cdots, X_5$) から推定したい.
5 人を '無作為に' 選ぶ (=標本抽出する)
母集団サイズ= $\boxed{}$, 標本サイズ= $\boxed{}$, 標本の個数= $\boxed{}$.

母集団と標本 (2) 離散 or 連続型確率変数 岩薩林 確率・統計 §5.1,5.2

賞金額, 個数が謎のスピードくじ (引いて賞金額を見た後で箱に戻す).
賞金額 X は離散型確率変数 $\to$ 無限母集団 (何回でもひけるから).

- 賞金の母平均値 $\mu = \mathrm{E}[X] = \sum_x x \cdot p(x)$ を求めたい.
- くじの中を見れば ($p(x)$ の式を知れば) 定義の式使うだけ.
- しかし, 中を見ることはできない.
- $+\infty$ 回くじを買わず, 何とかすませたい.

$\rightsquigarrow$ 引いた 5 枚のくじの賞金額=標本)(独立同分布にしたがう確率変数 $X_1, X_2, \cdots, X_5$) から推定したい.
5 枚を '無作為に' 選ぶ (=標本抽出する).
母集団サイズ= $\boxed{}$, 標本サイズ= $\boxed{}$, 標本の個数= $\boxed{}$.

母集団・標本抽出・推定 岩薩林 確率・統計例 11(p.115)

- 母集団 population =考えたい集団．どんな分布，母平均値，母分散，などわかっていないことがあるが，全体を調べるわけにはいかない集団.
- 標本 sample (名詞) =母集団から ‘無作為に’ とってきた一部分
- 標本抽出 する sample(動詞)=母集団から ‘無作為に’ とってくる $\rightsquigarrow$ 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

推定には誤差あるかも．標本の選び方ごとに答は違うし．

ここまで来たよ

9. 二項分布, 独立同分布の和
   - チェビシェフの不等式と大数の法則

10. 大数の法則/母集団・標本抽出・推定
    - 母集団と標本
    - 母平均値・母分散の (点) 推定
    - 母比率とその (点) 推定

母平均値の (点) 推定 高校 数学 B 岩薩林 確率・統計 (5.4)p.114

$X_1, X_2, \ldots, X_n$ はサイズ n の標本.
各 X_i $(i = 1, \ldots, n)$ は母平均値 $\mu = \mathrm{E}[X_i]$, 母分散 $\sigma^2 = \mathrm{V}[X_i]$ の独立同分布にしたがう確率変数.

標本平均値

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

が, 母平均値 $\mathrm{E}[X]$ の ‘よい’ 推定値になっている.

母平均値は μ はひとつに定まっているが, 標本平均値 $\overline{X}_{(n)}$ は確率変数であり, 試行=標本抽出のたびにかわる ($\overline{X}_{(n)}$ は確率分布をもつ)

標本期待値

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

が, $\mathrm{E}[g(X)]$ の ‘よい’ 推定値になっている.

L10-Q1

Quiz(母平均値, 母分散, 母比率の点推定)

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

117g, 109g, 109g, 119g, 100g, 112g.

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

よい推定量がもつ性質

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

母平均値の推定量 $\overline{X}_{(n)}$ は不偏性を持つ 岩薩林 確率・統計 p.113

「標本平均値 $\overline{X}_{(n)}$」の不偏性

母ナントカの推定量の母平均値 = 母ナントカ

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

一致性も持つ大数の法則から 岩薩林 確率・統計 p.143

母分散の (点) 推定 高校 数学 B 岩薩林 確率・統計 (5.11) の V (p.122)

不偏標本分散

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

が, 母分散の 'よい' 推定値になっている.
ここで, $\overline{X}$ は母平均値でなく, 上の標本平均値 ($\overline{X}_{(n)}$ の略記).

$n-1$ の理由 こうするとちょうど不偏: $\mathrm{E}[S^2] = \sigma^2$.
直観的理由 $\overline{X}$ は X_i の重心だから, $(X_i - \overline{X})^2$ は $(X_i - \mu)^2$ より小さくなりがち ($\frac{n-1}{n}$ 倍) なので修正.

おぼえ方 不偏標本分散は $n = 1$ のとき, $\boxed{}$

$\mathrm{E}[S^2] = \sigma^2$ を $n = 2$ のときに確認 (証明). $\overline{X} = \frac{1}{2}(X_1 + X_2)$.

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

ここまで来たよ

9. 二項分布, 独立同分布の和
   - チェビシェフの不等式と大数の法則
10. 大数の法則/母集団・標本抽出・推定
   - 母集団と標本
   - 母平均値・母分散の (点) 推定
   - 母比率とその (点) 推定

比率=ratio 岩薩林 確率・統計 p.107

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

- $X \sim$ ある分布, $Y = \mathbf{1}_{[\cdots \text{である}]}(X)$, たとえば $X > 10$ なら $Y = 1$ とか.
- 母集団=日本国民, 国民 x の血液型が A である $Y = 1$.

母比率 岩薩林 確率・統計 p.107

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

$$\text{母集団の「…である」母比率 } p = \frac{\text{「…である」データ } x \text{ の個数}}{\text{母集団サイズ}} = \mathrm{E}[Y]$$

やりたいこと:母比率の推定

ベルヌーイ分布の p (母比率) を標本から推定したい!

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

母比率の (点) 推定 岩薩林 確率・統計 p.115

標本比率 岩薩林 確率・統計 p.115

標本のデータ n 個中 k 個が「…」であるとき,

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

が「…」の母比率 p のよい推定値になっている.

母比率 p の推定=母平均値 $\mathrm{E}[Y]$ の推定

サイズ n の標本中 k 個が「…である」とき,

$$\begin{aligned} \text{母平均値 } \mathrm{E}[Y] \text{ の推定値 } =& \text{標本平均値} \overline{Y} \\ =& \frac{1}{n}[\underbrace{1+\cdots+1}_{k}+\underbrace{0+\cdots+0}_{n-k}] \\ =& \frac{k}{n} = \hat{p}. \end{aligned}$$

連絡

- 2020-01-20 月 1 テスト B1 実施. 45 分, PC 受験の予定. 教科書の p.xii–xv のコピーを配布. 他の持込不可.
- 2020-01-27 月 1(定期試験期間) テスト A2,B2 実施. A1,B1 と同等の評価内容.
- 2020-02-15 土 学力認定試験 2 年生も受験可能 `https://wiki.math.ryukoku.ac.jp` (全学認証)
- 成績の 70 ピーナッツ分$=\min(\max(A1, A2), \max(B1, B2))$.
- 予習復習問題を, 期限後も (再/初) 受験できるようにしてます. ピーナッツにはカウントしないけど, テスト準備に活用してね.
- 2019-12-16 は区間推定. §§6.1,7.1,7.3 読んできてね.
- 2019-12-16 の Trial 予告
- 2019-12-16 までの (短い作文の) レポート課題 `https://manaba.ryukoku.ac.jp`
- オフィスアワー木 6(1-539) 金昼 (1-542), Math ラウンジ (1-536/538)

学籍番号での LINE 公式アカウント登録

龍大 manaba course