ベイズ推定・標本サイズの決定

樋口さぶろお http://hig3.net

龍谷大学理工学部数理情報学科

生活の中の統計技術 L12(2019-01-07 Mon)

最終更新: Time-stamp: "2019-01-07 Mon 10:20 JST hig"

今日の目標

- 2×2 分割表に基づいてベイズ推定ができる
- 平均値, 比率の区間推定において, 求められる精度に応じて標本サイズを決められる

L12-Q1

Quiz 解答:無相関な二元分割表

	右利き	右利きでない
早生まれ	2	5
早生まれでない	6	15

L12-Q2

期待度数は

	右利き	右利きでない	計
早生まれ	$\frac{4}{3}$	$\frac{2}{3}$	2
早生まれでない	$\frac{8}{3}$	$\frac{4}{3}$	4
計	4	2	6

$$\chi^2 = \frac{(1-\frac{4}{3})^2}{\frac{4}{3}} + \frac{(1-\frac{2}{3})^2}{\frac{2}{3}} + \frac{(3-\frac{8}{3})^2}{\frac{8}{3}} + \frac{(1-\frac{4}{3})^2}{\frac{4}{3}} = \frac{3}{8}.$$

L12-Q3

Quiz 解答:ピアソンの χ^2

左から, 大きい, 大きい (同じ値), 中程度, zero.

ここまで来たよ

ベイズ推定 I

L12-Q1

Quiz(ベイズ推定)

抽選用の袋に 100 個の色つきボールが入っている．ボールを割ると, 20 個は「あたり」,80 個は「はずれ」の紙が入っている．
「あたり」のボールのうち 10%は赤, 90%は白に塗られている．
「はずれ」のボールのうち 70%は赤, 30%は白に塗られている．

1. 無作為にボールを取り出すとき, 「あたり」である確率 (事前確率) を求めよう．
2. 無作為にボールを取り出すとき, 白に塗られている確率を求めよう．
3. 無作為にボールを取り出したところ, 赤に塗られたボールだった．このとき, 「あたり」である確率 (事後確率) を求めよう．

条件付き確率

赤であるという条件のもとでの, あたりの確率

$$P(\text{あたり} \mid \text{赤}) = \frac{P(\text{あたり}, \text{赤})}{P(\text{赤})}$$

理論的にはベイズの公式から

$P(A)$: A の起きる確率
$P(A, B)$: A かつ B の起きる確率
$P(A|B)$: B が起きるという条件の下で A の起きる確率

ベイズの公式

$P(A, B) = P(A|B)P(B) = P(B|A)P(A)$

$\rightsquigarrow$

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

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

$P(A|B)$ を $P(B|A)$ (と $P(A)$) で書き表す式, およびその逆の式.

L12-Q2

Quiz(ベイズ推定)

ある国の国民 (必要なら 1000 人と思ってもよい) の, 病気 D の感染率は 1%である.
病気 D にかかっている人に検査 E を行うと, 60%が陽性, 40%が陰性になる.
病気 D にかかっていない人に検査 E を行うと, 10%が陽性, 90%が陰性になる.

1. 特に情報がないとき, 無作為に選んだ国民が, 病気 D にかかっている確率 (事前確率) を求めよう.
2. 無作為に選んだ国民が, 検査 E で陽性になったとき, 病気 D にかかっている確率 (事後確率) を求めよう.
3. 無作為に選んだ国民が, 検査 E で陰性になったとき, 病気 D にかかっていない確率 (事後確率) を求めよう.

ベイズ推定, ベイズ的な考え方

$\boxed{\text{事後確率 } P(A|B)} \longleftarrow \boxed{\text{事前確率 } P(A)}$

$\uparrow$

$\boxed{\text{情報 } B}$

主観確率　　ベイズの定理=ベイズの公式 (+ニュアンス?)

L12-Q3

Quiz(ベイズの公式)

外見で区別できない, 甘い品種 1 と渋い品種 2 の柿がある.
甘い品種 1 は, 確率 0.95 で赤に, 確率 0.05 で黄色になる.
渋い品種 2 は, 確率 0.125 で赤に, 確率 0.875 で黄色になる.
確率変数 X, Y を用いて, 甘い品種 1 を $X = 1$, 渋い品種 2 を $X = 2$, 赤を $Y = 10$, 黄色を $Y = 20$ と表現する.

1. 問題文から $P(Y = y|X = x)$ を読み取ろう.
2. かごの柿の $1/5$ が甘い柿であるとする. 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. ベイズの公式を使って, 取り出した赤い柿が甘い確率 $P(X = 1|Y = 10)$ を求めよう.
3. かごの柿の $1/5$ が渋い柿であると考えている. いま, 無作為に 1 個の柿を取りだしたところ, 黄色い柿だった. ベイズの公式を使って, 取り出した黄色い柿が渋い確率を求めよう.

ここまで来たよ

11 2 つのカテゴリ変数の間の関係

12 ベイズ推定・標本サイズの決定
- ベイズ推定
- 検出力と標本サイズ

復習

平均値の区間推定

母集団の平均値 μ の, 信頼係数 $1-\alpha = 0.95, 0.99$ の信頼区間は,

$$\overline{x} - z_{\alpha/2}\frac{s}{\sqrt{N}} < \mu < \overline{x} + z_{\alpha/2}\frac{s}{\sqrt{N}}.$$

$\overline{x}$: 標本の平均値
s: 標本の標準偏差
N: 標本のサイズ
係数 $z_{\alpha/2} = \begin{cases} 1.96 & (1-\alpha = 0.95) \\ 2.58 & (1-\alpha = 0.99) \end{cases}$

L12-Q4

Quiz(母平均値の区間推定 (母分散未知))

あるファミレスのドリンクバーの, ドリンクサーバーの出すコップ 1 杯分のドリンクの体積 (cm^3) は, (未知の) 母平均値 $\mu \mathrm{cm}^3$ と (未知の) 母分散 $\sigma^2(\mathrm{cm}^3)^2$ の正規分布にしたがう.
$n = 6$ 杯いれてみたところ, 体積 (cm^3) は,

$$202, \quad 203, \quad 204, \quad 204, \quad 205, \quad 206$$

だった. ここから標本平均値を求めたところ $\bar{X} = 204\mathrm{cm}^3$ だった. 不偏標本分散を求めたところ $s^2 = \frac{10}{6-1} = 2(\mathrm{cm}^3)^2$ だった.
ドリンクサーバーの出すコップ 1 杯分の体積の母平均値 μ を区間推定して, 信頼係数 $1 - \alpha = 0.99$ の信頼区間を求めよう.
答の整理は不要. 小数や平方根の計算は不要. 分数や平方根が残った形で, 有効数字を考えずに答えてよい.

L12-Q5

Quiz(標本サイズと信頼区間)

学生の (大きな) 母集団の身長の平均値 μ cm を推定したい．標本サイズ $N = 20$ を抽出し，ある信頼係数で区間推定したところ，

$$166 < \mu < 174$$

となった．

1. 信頼区間の長さを 4cm 以下にするには，標本サイズをどのようにとればいいか．
2. 信頼区間の長さを 2cm 以下にするには，標本サイズをどのようにとればいいか．

復習

母比率の信頼区間 (母分散未知)

サイズ N の標本で, 標本の比率が $\hat{p} = y/N$ のとき, 母集団の比率 p の信頼係数 $1-\alpha$ の信頼区間は

$$\frac{y}{N} - z_{\alpha/2} \times \frac{\sqrt{\frac{y}{N}(1-\frac{y}{N})}}{\sqrt{N}} < p < \frac{y}{N} + z_{\alpha/2} \times \frac{\sqrt{\frac{y}{N}(1-\frac{y}{N})}}{\sqrt{N}}$$

係数 $z_{\alpha/2} = \begin{cases} 1.96 & (1-\alpha = 0.95) \\ 2.58 & (1-\alpha = 0.99) \end{cases}$

覚え方: 標本の平均値 $\hat{p} = \frac{y}{N}$, 分散 $\hat{p}(1-\hat{p})$.

L12-Q6

Quiz(母比率の区間推定)

120 人のクラスのうち, 何人が運転免許を持っているか知るために, 10 人に質問したところ 2 人が運転免許を持っていた.

1. クラスで運転免許を持っている人の母比率 p を, 信頼係数 $1-\alpha = 0.95$ で区間推定しよう.
2. クラスの 120 人のうち運転免許を持っている人の人数 m を, 信頼係数 $1-\alpha = 0.95$ で区間推定しよう.

いずれも, 加減乗除平方根の残った未整理な形で答えてよい.

L12-Q7

Quiz(標本サイズと信頼区間)

選挙の出口調査で, 標本サイズ $n = 50$ で候補 A への投票率を推定したところ, $0.41 < p < 0.69$ となった. 当確を出す, すなわち, $0.5 < p$ であることを確信するには, 標本サイズはどのくらい必要か.

L12-Q8

Quiz(標本サイズと信頼区間)

(大きな) 母集団での, ある意見に対する賛成の比率を調べたい.

1. 0.1 以上の精度で求めるには, 標本サイズはどのようにとればいいか. 信頼係数は 0.95 とする.
2. 0.01 以上の精度で求めるには, 標本サイズはどのようにとればいいか. 信頼係数は 0.95 とする.

検出力

統計的仮説検定の言葉で言うと, 検出力 ($1-\beta=$ 第 2 種の過誤が起きない確率) が十分大きくなる標本サイズを探していることに相当.
実験計画法の考え方の一種.

お知らせ

- 2019-01-14 月 は休日
- 2019-01-21 月 2 はたぶん 3-B105 で
- 2019-01-22 火 (講時未定) は補講．期末試験シミュレーション問題演習．この日の出席や提出による加点はありません．
- 2019-01-28 月 2 期末試験
    - 30 ピーナッツ/科目 100 ピーナッツ
    - 60 分
    - 紙は何でも持込可．

期末試験出題計画

毎回の課題を復習することをおすすめします.
大注意:この計画は確定版ではありません. 2018-01-21 月までに精密化・確定します.

- 平均値の区間推定
- 2 群の平均値の差の区間推定
- 比率の区間推定
- ○ 標本サイズの決定
- 統計的仮説検定の意味
- 分散分析 (=多群の平均値の差の検定) の級間平方和と級内平方和と F
- クロス集計表の独立性の指標 χ^2
- ○ ベイズ推定
- ?

電卓は使用しません. Excel の操作についての問題は出題しません.