標本抽出と推定

樋口さぶろお

龍谷大学理工学部数理情報学科

確率統計☆演習 I　L09(2014-12-05 Fri)

今日の目標

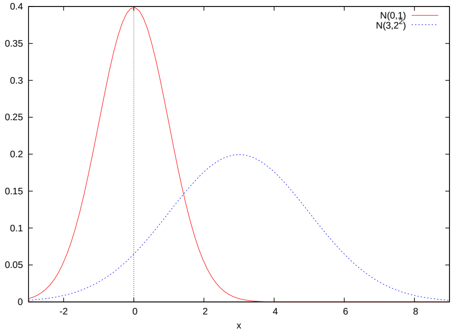
- 母集団, 標本抽出, 推定の定義を説明できる
- 標本から母平均, 母分散を推定できる

http://hig3.net

L09-Q1

Quiz 解答:正規分布の確率密度関数の拡大縮小平行移動

$\mu = 3, \sigma^2 = 2^2$. $\mathrm{N}(3, 2^2)$ の確率密度関数は, $f(x) = \frac{1}{\sqrt{2\pi \cdot 2^2}} \mathrm{e}^{-\frac{(x-3)^2}{2 \cdot 2^2}}$



L09-Q2

Quiz 解答:正規分布の確率

1. 標準正規分布に従う Z が $Z \geq \frac{5-3}{2}$ となる確率だから, 標準正規分布の表をひいて, $Q(1.00) = 0.1587$. (これは図からも求められる $(1-0.6827)/2 = 0.1587$).
2. 標準正規分布に従う Z が $-1 \leq Z \leq 2$ となる確率だから, 標準正規分布の表をひいて, $1 - Q(2.00) - Q(1.00) = 0.8186$. (これは図からも求められる. $0.6827/2 + 0.9545/2 = 0.8186$)

L09-Q3
Quiz 解答:正規分布の確率

1. $\int_{0.5}^{0.7} \phi(x)\,\mathrm{d}x = \Phi(0.7) - \Phi(0.5) = (1 - \Phi(0.7)) - (1 - \Phi(0.5)) = Q(0.5) - Q(0.7) = 0.3085 - 0.2420 = 0.0665.$
2. $z = \frac{y}{2}$ と変数変換すると, 標準正規分布に従う Z が $0.25 \leq Z \leq 0.35$ を満たす確率. よって,
   $\Phi(0.25) - \Phi(0.35) = Q(0.35) - Q(0.25) = 0.4013 - 0.3632 = 0.0381.$
3. $z = \frac{z-3}{2}$ と変数変換すると, 標準正規分布に従う Z が $0.5 \leq Z \leq 0.7$ を満たす確率. よって, 1. と同じで $0.0665.$

ここまで来たよ

1. 略解:正規分布
2. 標本抽出と推定
   - 独立同分布の母平均値と母分散
   - 母集団と標本
   - 母平均値・母分散の推定

独立同分布の母平均値と母分散

確率変数 $X_1, X_2, \ldots, X_n$ が, 独立で, すべて同じ確率分布に従う (同じ確率密度関数 $f(x)$ を持つ) とする. 正規分布でなくてよい.

独立事象 → 高校数学

これを $X_1, \ldots, X_n$ は $\boxed{}$, 独立同分布に従う, という

母平均値 $\mathrm{E}[X_i] = \mu$, 母分散 $\mathrm{V}[X_i] = \sigma^2$ とする.

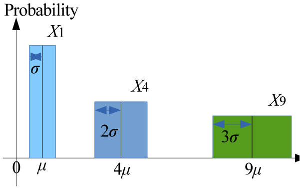
新しい確率変数 Y_n を定義: $Y_n = X_1 + \cdots + X_n$.

$$\mathrm{E}[Y_n] = \sum_{i=1}^{n} \mathrm{E}[X_i] = n\mu.$$

上は実は独立じゃなくても OK

$$\mathrm{V}[Y_n] = \sum_{i=1}^{n} \mathrm{V}[X_i] = n\sigma^2.$$

Y_n の確率密度関数はこんな感じ?



本当は長方形じゃない. まて中心極限定理.

ここまで来たよ

母集団と標本 (1) 有限母集団

AKB48 の身長ふたたび

- AKB48 メンバー全員 (=　　　) の身長の母平均値 $\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$ を求めたい!
- メンバー全員分のデータがあれば公式使うだけ
- しかし, データ非公開になった今, 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 74 枚集めないで何とかすませたい.

$\rightsquigarrow$ 握手会参加券がゲットできて質問できたメンバー 5 人の答え (=　　　) から　　　したい.

5 人を '無作為に' 選ぶ (=　　　)

母集団サイズ=　　, 標本サイズ=　　, 標本の数=　　.

母集団と標本 (2) 離散 or 連続型確率変数

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

- 賞金の母平均値 $\mathrm{E}[X] = \sum_{k=1}^{m} x_k f(x_k)$ を求めたい.
- しかし, 中を見ることはできない.
- $+\infty$ 回くじ買わず, 何とかすませたい.

$\rightsquigarrow$ 引いた 5 枚のくじの賞金額 (= $\boxed{}$) から $\boxed{}$ したい.

5 枚を '無作為に' 選ぶ (= $\boxed{}$).

母集団サイズ= $\boxed{}$, 標本サイズ= $\boxed{}$, 標本の数= $\boxed{}$.

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．
- 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 sampling =母集団から ‘無作為に’ とってくること
- 推定 estimation =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こと

推定には誤差あるかも．もともと，標本の選び方ごとに答えは違うし．

ここまで来たよ

母平均値の推定

以下, X_i $(i = 1, \ldots, n)$ はサイズ n のサンプル. X_i $(i = 1, \ldots, n)$ は母平均値 μ, 母分散 σ^2 の独立同分布にしたがう確率変数.

標本平均値

$$\text{標本平均値 } \bar{X} = \frac{1}{n}[X_1 + \cdots + X_n]$$

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 $\mathrm{E}[X]$ はひとつに定まっているが, 標本平均値 $\bar{X}$ は, 確率変数で有り, 試行=標本抽出のたびにかわる (それ自体が確率分布をもつ)

$$\mathrm{E}[\bar{X}] = \mu, \qquad \text{不偏 ‘よい’}$$

$$\mathrm{V}[\bar{X}] = \frac{\sigma^2}{n}, \qquad n \to \infty \text{ で正確}$$

なぜなら

母分散の推定

(不偏) 標本分散

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

が, 母分散の ‘よい’ 推定値になっている.
ここで, $\bar{X}$ は母平均値でなく, 上のように計算した標本平均値.

なぜ $n-1$? だって… こうするとちょうど不偏 $\mathrm{E}[S^2] = \sigma^2$.
おぼえ方 (不偏) 標本分散は…

- $n = 1$ のとき, $\frac{0}{0}$ で定義されない.
- $n = 2$ のときに, $\boxed{}$.

$\mathrm{E}[S^2] = \sigma^2$ を $n = 2$ のときに確認

$$\begin{aligned}
\text{左辺} =& \frac{1}{2-1}\mathrm{E}[(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2] \\
=& \mathrm{E}[X_1^2 + X_2^2 - 2(X_1 + X_2)\bar{X} + 2\bar{X}^2] \\
=& \mathrm{E}[X_1^2 + X_2^2 - 2\bar{X}^2] \\
=& \mathrm{E}[X_1^2] + \mathrm{E}[X_2^2] - 2\mathrm{E}[\bar{X}^2]
\end{aligned}$$

ここで,
$\sigma^2 = \mathrm{V}[X_1] = \mathrm{E}[X_1^2] - (\mathrm{E}[X_1])^2 = \mathrm{E}[X_1^2] - \mu^2$,
$\frac{\sigma^2}{2} = \mathrm{V}[\bar{X}] = \mathrm{E}[\bar{X}^2] - (\mathrm{E}[\bar{X}])^2 = \mathrm{E}[\bar{X}] - \mu^2$,
より,

$$\begin{aligned}
\cdots =& (\mu^2 + \sigma^2) + (\mu^2 + \sigma^2) - 2(\mu^2 + \tfrac{\sigma^2}{2}) \\
=& \sigma^2 \\
=& \text{右辺}
\end{aligned}$$

L09-Q1

Quiz(推定)

瀬田学舎の龍大生の通学時間の分布を知るために, 無作為に 5 人を選んで質問したところ次のようだった. 母平均値, 母分散, 母標準偏差を推定しよう.

10 分, 20 分, 30 分, 30 分, 110 分.

龍大生でなく, $\boxed{}$ だったら? $\boxed{}$ だったら?
$\boxed{}$ だったら?

標本抽出と推定の実験

`http://hig3.net`
→ (左上) 統計データ収集
学籍番号と X のみ入力.

各チームで

- サンプル数=1
- サンプルサイズ=チーム人数

のサンプルを作って, 母平均値と母分散を推定しよう.

L09-Q2

Quiz(推定)

ある確率分布に従うスピードくじを 10 回ひいたところ, 賞金は,
0 円, 0 円, 0 円, 0 円, 0 円, 0 円, 10 円, 10 円, 30 円, 100 円
だった. 確率分布の母平均値と母分散と母標準偏差を推定しよう.

L09-Q3

Quiz(母平均値, 母分散の点推定)

フライドチキン屋さんのフライドチキンの在庫 (=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.

$117\text{g}, \quad 109\text{g}, \quad 109\text{g}, \quad 119\text{g}, \quad 100\text{g}, \quad 112\text{g}.$

1. 重さの母平均値を点推定しよう.
2. 重さの母分散を点推定しよう.

連絡

- 2014-11-17 から チューターは月火水木昼 (1-614).
- ~~2014-12-03~~ →2014-12-17 水 4 数理情報学科特別講義
- $\boxed{\text{2014-12-12 金 2 休講}}$　しか〜し, 来年度の 3 年次必修科目 学外実習・総合演習 履修説明会. 2 年生は全員出席必須. 1-542.
- $\boxed{\text{いつか補講} \times 2}$