ベイズ推定

長島 貴之

February 23, 2017

目次

g	データへの対応	母数 (パラメータ)
ベイズ	一期一会に扱う	確率変数でその分布を調べる
頻度	たくさんの中の１つ	母集団固有の唯一値を仮定

確率の定義

例としてサイコロで考える．サイコロを 1 回投げ，偶数の目が出る確率を調べる.(投げる動作を試行，試行によって得られる結果を事象という) 偶数の目が出る確率を求める場合は， 偶数の目が出る事象を A とし，この事象 A の起きる確率 p は

$$p = \frac{\text{偶数の目が出る数}}{\text{起こり得るすべての場合の数}} = \frac{1}{2}$$

一般化すれば

$$p = \frac{\text{事象 A が起こる確率}}{\text{起こり得るすべての場合の数}}$$

確率の記号

P(A) ：事象 A の起こるか確率

確率の記号

2 つの事象 A,B があり，同時に起こる事象この事象が起こる確率 $A \cap B$ この事象が起こる確率

$$P(A \cap B) or P(A, B)$$

と記される (本によってまちまちだが後者多い)

条件付きの確率

ある事象 A がおこった元で事象 B が起こる確率 $P(A \mid B)$
例題
1 組のトランプから 1 枚抜くとする．抜いたカードがハートである事象を A，絵札である事象 B としたとき，条件付き確率 $P(A \mid B)$，$P(B \mid A)$，同時確率 $P(A \cap B)$ を求めよ．

条件付きの確率 2

$P(A \mid B)$ 抜いたカードが ♡ のとき，それが絵札である確率
♡ は 13 通りで，その中で絵柄 3 枚なので $\frac{3}{13}$
$P(B \mid A)$ 抜いたカードが絵札のとき，それが ♡ である確率
絵札が 12 通りで，その中で ♡ は 3 枚なので $\frac{1}{4}$
$P(A \cap B)$ 抜いたカードが ♡ かつ絵札の確率$=\frac{3}{52} = \frac{1}{14}$

条件付きの確率 3

条件付き確率は次の式で表現できる.

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)} \tag{1}$$

例題
1 組のトランプから 1 枚抜くとする．抜いたカードがハートである事象を A，絵札である事象 B としたとき，条件付き確率 $P(A \mid B)$，$P(B \mid A)$ を求めよ.
P(A)=$\frac{13}{52}$,P(B)=$\frac{12}{52}$, 先のページより $P(A \cap B) = \frac{3}{52}$

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)} = \frac{\frac{3}{52}}{\frac{13}{52}} = \frac{3}{13} \tag{2}$$

※ $P(A \mid B)$ も同様なので省略

確率の乗法定理

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)} \tag{3}$$

両辺に P(A) を掛けて

$$P(A \cap B) = P(B \mid A)P(A) \tag{4}$$

導出

乗法定理をを用いれば簡単に導出できる

$$P(A \cap B) = P(B \mid A)P(A) \tag{5}$$

$$P(A \cap B) = P(A \mid B)P(B) \tag{6}$$

以上 より $P(A \mid B)$ について解くと

$$P(A \mid B) = \frac{P(B \mid AP(A)}{P(B)} \tag{7}$$

これが ベイズの定理.

変更

前のページのままだとイメージしづらいので，少し書き換える.

$$P(H \mid D) = \frac{P(D \mid HP(H)}{P(D)} \tag{8}$$

A を原因と仮定 (hypotheis),B を A のもとで得られる結果やデータとして置き換えた.

題 1

男 10 女 7 が一室でパーリーを開いた．男子の喫煙者は 5，女 3 である．部屋に入ったら吸い殻が一本，灰皿の上にあった．この時のタバコを吸った人が女性である確率を求めよ．

題1

男 10 女 7 が一室でパーリーを開いた．男子の喫煙者は５，女３である．部屋に入ったら吸い殻が一本，灰皿の上にあった．この時のタバコを吸った人が女性である確率を求めよ．

H:女性である

題 1

男 10 女 7 が一室でパーリーを開いた．男子の喫煙者は 5，女 3 である．部屋に入ったら吸い殻が一本，灰皿の上にあった．この時のタバコを吸った人が女性である確率を求めよ．

H:女性である
D:(タバコを吸った人すなわち) 喫煙者である．

題1

P(H) はパーティーの中で女性である確率，P(D) は喫煙者である確率，
$P(D \mid H)$ は女性の中で喫煙者である確率

題1

P(H) はパーティーの中で女性である確率，P(D) は喫煙者である確率, $P(D \mid H)$ は女性の中で喫煙者である確率
P(H)=$\frac{7}{17}$,P(D)=$\frac{8}{17}$,$P(D \mid H)$ =$\frac{3}{7}$
式？に代入すると

$$P(H \mid D) = \frac{\frac{7}{17} \times \frac{3}{7}}{\frac{8}{17}} = \frac{3}{8} \tag{9}$$

タバコを吸った人が女性である確率=$\frac{3}{8}$ この図はベン図から容易に求められる.

ベイズの展開公式

データ D を得ることができる原因だが普通 1 つではない．原因が n 個あれば H_1,H_2...H_n とかける．原因 H_1 に注目してみる．ベイズの定理の H を H_1 とき置き換える．

$$P(H_1 \mid D) = \frac{P(D \mid H_1)P(H_1)}{P(D)} \tag{10}$$

P(D) は同時確率の和で表現できるので

$$P(D) = P(D, H_1) + P(D, H_2) + \cdots + P(D, H_n) \tag{11}$$

と書ける．

ベイズの展開公式 2

乗法の定理より

$$P(D) = P(D|H_1)P(H_1) + P(D|H_2)P(H_2) + \cdots + P(D|H_n)P(H_n) \quad (12)$$

(乗法定理より) これを一般化すると

$$P(H_i|D) = \frac{P(D|H_i)P(H_i)}{\sum_{i=0}^{n} P(D|H_i) P(H_i)} \quad (13)$$

用語の確認

- $P(H|D)$:事後確率... データ D が原因 h_i から得られる確率
- $P(D|H)$:尤度... 原因 h_i のもとでデータ D が得られる確率

用語の確認

- $P(H|D)$:事後確率... データ D が原因 h_i から得られる確率
- $P(D|H)$:尤度... 原因 h_i のもとでデータ D が得られる確率
- P(H):事前確率... データ D を得る前の原因 h_i の確かしらさ

用語の確認

- $P(H|D)$:事後確率... データ D が原因 h_i から得られる確率
- $P(D|H)$:尤度... 原因 h_i のもとでデータ D が得られる確率
- P(H):事前確率... データ D を得る前の原因 h_i の確かしらさ

ベイズ理論の計算ステップ

1. モデル化し，尤度を導出
2. 事前確率を設定
3. ベイズに公式を用いて事後確率を算出

理由不十分の原則

形が同じの 2 つの壷，赤い壷，青い壷，が置いてある．赤い壷には白玉 2 個，赤玉 3 個が入っていて，青い壷には，白玉 4 個，赤玉 8 個入っている．壷を一つを選択し，玉を一つを取り出したら，白玉だった．このときの赤い壷から取り出された確率を求めよ．
H_r:赤い壷から取り出す
H_b:青い壷から取り出す
D:取り出した玉が白である．

理由不十分の原則

ベイズの展開公式

$$P(H_r|D) = \frac{P(D|H_r)P(H_r)}{P(D|H_r)P(H_r) + P(D|H_b)P(H_b)} \tag{14}$$

尤度を算出:赤い壺から取り出された玉が白である確率
$P(H_r|D) = \frac{2}{5}$
事前確率の設定
2つの壺がどの割合で割合で選ばれるかの乗法がないので，事前確率は等確率になるように設定 $P(H_r) = P(H_b) = \frac{1}{2}$
今までのことを踏まえて

$$P(H_r|D) = \frac{\frac{2}{5} \times \frac{1}{2}}{\frac{2}{5} \times \frac{1}{2} + \frac{1}{3} \times \frac{1}{2}} = \frac{6}{11} \tag{15}$$

理由不十分の原則

壺を選ぶ確率は主観確率になっている．これが頻度主義に敬遠されている理由.

ベイズ更新

A 社の壺には，水晶玉とガラス玉が 4:1 の割合で入っている．B 社の壺には，水晶玉とガラス玉が 2:3 の割合で入っている．壺の見た目は全く同じで壺にはたくさん玉が入っているとする．いま，A 社製か B 社製か不明の壺があり，続けて 3 回玉を取り出したら，順に 水晶玉，水晶玉，ガラス玉であった．この壺が A 社製の確率を求めよ．H_a:A 社製の壺から取り出す
H_b:B 社製の壺から取り出す
D_1=S:取り出した玉が水晶玉
D_2=G:取り出した玉がガラス玉

ベイズ更新 2

データは 3 つの事象が連続になっている．またそれを取り込むためには??
⇒ ベイズ更新
今回求めるものは

$$P(H_a|S) = \frac{P(S|H_a)P(H_a)}{P(S|H_a)P(H_a) + P(S|H_b)P(H_b)} \tag{16}$$

$$P(H_b|S) = \frac{P(S|H_b)P(H_b)}{P(S|H_b)P(H_b) + P(S|H_b)P(H_b)} \tag{17}$$

$$P(H_a|G) = \frac{P(G|H_a)P(H_a)}{P(G|H_a)P(H_a) + P(G|H_b)P(H_b)} \tag{18}$$

$$P(H_b|G) = \frac{P(G|H_b)P(H_b)}{P(G|H_b)P(H_b) + P(G|H_b)P(H_b)} \tag{19}$$

ベイズ更新 3

尤度の算出

$P(S|H_a) = 0.8 = 4/5, P(G|H_a) = 0.2, P(S|H_b) = 0.4 = 2/5, P(G|H_b) = 0.6$

一回目の玉の取り出し
事前確率の設定
2 つの壺の選択各率は不明なので，理由不十分の原則をもちいて，

$$P(H_a) = P(H_b) = 0.5$$

一回目は水晶玉なので

$$P(|H_a)|S) = 2/3, P(H_b|S) = 1/3$$

ベイズ更新 4

二回目の取り出し
二回目の取り出しの取り出しときのどうするか？ここでベイズ更新を使う．一回目の事後確率を二回目の分析する際の事前確率として使う．(ベイズ更新の使う理由を説明)

$$P(H_a) = 2/3, P(H_b) = 1/3$$

二回目の事後確率は

$$P(H_a|S) = 0.8, P(H_b|S) = 0.2$$

となる．

ベイズ更新5

三回目の取り出しも同様に,

$$P(H_a) = 0.8, P(H_b) = 0.2$$

三回目の事後確率は

$$P(H_a|G) = 4/7 \fallingdotseq (0.571)$$

となる.

平均，分散

平均
: データの平均値

分散
: 平均値からの散らばり具合

標準偏差
: 標準偏差は分散の正の平方根（ルート）で定義されます.

確率変数，確率分布

確率変数

コイントスという確率変数 X を考える. 表は x=1, 裏は x=0 としますと，x=X(事象) は

$$0 = X(裏), 1 = X(表)$$

左辺の従属変数の値 x を実現値という.

確率分布

確率密度の実現値と実現値に付与された確率との対応表

実現値 x	0	1
確率 p	1/2	1/2

連続的な確率変数と確率密度変数

確率変数がサイコロの目のように離散的な値をとるならば，表形式で確率分布を示せる．しかし連続的な値を取るものは表形式で示せない．
$\Rightarrow$ 確率密度関数

ベイズ統計学における母数の扱い

従来の統計学は母数は定数として扱われます. その定数で規定された確率分布ベイズ理論では原因 H としてこの母数を扱う.

$$P(\theta_i|D) = \frac{P(D|\theta_i)P(\theta_i)}{\sum_{i=0}^{n} P\left(D|\theta_i\right) P\left(\theta_i\right)} \tag{20}$$

母数が連続的な値を取るときのベイズ統計

シンプルにするために

$$P(D) = \sum_{i=0}^{n} P\left(D|\theta_i\right) P\left(\theta_i\right) \tag{21}$$

$$\begin{aligned} \text{事前確率 } P(\theta_i) &\Rightarrow \text{事前分布} \pi(\theta) && (22)\\ \text{尤度 } P(D|\theta_i) &\Rightarrow \text{尤度 } f(D|\theta) && (23)\\ \text{事後確率 } P(\theta_i|D) &\Rightarrow \text{事後分布} \pi(\theta|D) && (24) \end{aligned}$$

$$\pi(\theta|D) = \frac{f(D|\theta)\pi(\theta)}{P(D)} \tag{25}$$

データを得られた段階では P(D) は確定をしているので，定数になる.

母数が連続的な値を取るときのベイズ統計2

$$\pi(\theta|D) = kf(D|\theta)\pi(\theta) \tag{26}$$

k は定数で P(D) の逆数
上記の式からわかるように事後分布は尤度と事前分布に比例している

$$\pi(\theta|D) \propto f(D|\theta)\pi(\theta) \tag{27}$$

ベイズ統計例題

ある工場から作られるチョコレート菓子の内容量は正規分布に従い分散は 1^2 であり，製品の 1 つを選んで調べたら，その内容量は 101g であった．この時のこの工場から作られる製品内容量の「平均値 μ を求めよ」

ベイズ統計例題

$$\pi(\theta|D)) = kf(D|\theta)\pi(\theta) \tag{28}$$

母数 θ には平均値がデータ D には内容量 x=101 が対応します.

$$\pi(\theta|x = 101)) = kf(x = 101|\mu)\pi(\mu) \tag{29}$$

尤度の算出
平均値 μ の正規分布に従うでーたから, x=101 というデータが取り出される確率密度を表す. すなわち, 母数 μ をもつ確率密度関数と一致します

$$f(x = 101|\mu) = \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{(101-\mu)^2}{2}}\right) \tag{30}$$

ベイズ統計例題

事前分布の意味
データ得る前に, どの母数を持つ分布が起こりやすいか (選ばれやすいか) を表現するもの．ここでは平均値についてのじょうほうがないので理由不十分の原則より，一様分布を仮定します.

$$\pi(\mu) = 1$$

後は計算するだけ

$$\pi(\theta|x = 101) \propto \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{(101-\mu)^2}{2}}\right) \tag{31}$$