法と統計学　第 5 回
ベイズの定理 (1)

森　大輔

熊本大学法学部

2017 年

本日の流れ

例題 1-1：乳がん検査 (条件付確率)

　乳がんの早期発見のため、特定の年齢に達した女性は自覚症状がなくても定期的に検査を受けることが推奨されている。あなたが医師として、ある地域でこのような乳がんの検査を実施するとする。この地域で乳がんの検査に参加する 40 歳から 50 歳までの自覚症状のない女性について、以下のことがわかっている。

　これらの女性の 1 人が乳がんである確率は 0.8%である。また乳がんであれば、検査結果が陽性になる確率は 90%である。乳がんでなくても、陽性と出る確率は 5%ある。ある女性の検査結果が陽性と出た。

　この女性が実際に乳がんである確率はどれくらいか？

例題 1-1 の解き方

- $\mathrm{P}(\text{乳がん}) = 0.008$
  - 「これらの女性の 1 人が乳がんである確率は 0.8%」
- $\mathrm{P}(\text{陽性} \mid \text{乳がん}) = 0.90$
  - 「乳がんであれば、検査結果が陽性になる確率は 90%」
- $\mathrm{P}(\text{陽性} \mid \text{乳がんでない}) = 0.05$
  - 「乳がんでなくても、陽性と出る確率は 5%あります」
- $\mathrm{P}(\text{乳がん} \mid \text{陽性}) = \dfrac{\mathrm{P}(\text{乳がん})\times\mathrm{P}(\text{陽性} \mid \text{乳がん})}{\mathrm{P}(\text{乳がん})\times\mathrm{P}(\text{陽性} \mid \text{乳がん})+\mathrm{P}(\text{乳がんでない})\times\mathrm{P}(\text{陽性} \mid \text{乳がんでない})} = \dfrac{0.008\times0.90}{0.008\times0.90+0.992\times0.05} \fallingdotseq 0.13$
  - 「ある女性の検査結果が陽性と出ました。この女性が実際に乳がんである確率」

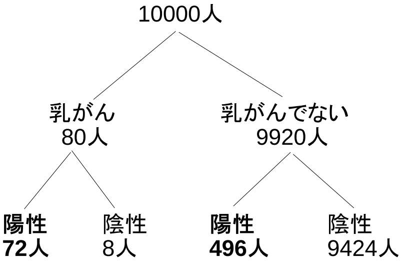
例題 1-2：乳がん検査 (自然頻度)

乳がんの早期発見のため、特定の年齢に達した女性は自覚症状がなくても定期的に検査を受けることが推奨されている。あなたが医師として、ある地域でこのような乳がんの検査を実施するとする。この地域で乳がんの検査に参加する 40 歳から 50 歳までの自覚症状のない女性について、以下のことがわかっている。

これらの女性 1 万人あたり 80 人が乳がんにかかっている。この 80 人の女性のうち、72 人は検査で陽性と出る。乳がんではない 9920 人の女性のうち、496 人はやはり検査結果が陽性になる。

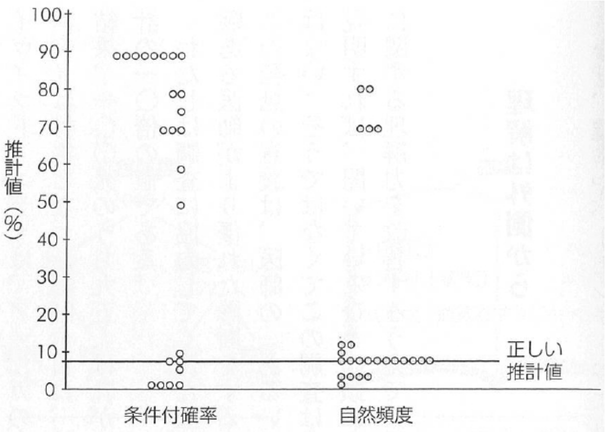
ある女性の検査結果が陽性と出た。この女性が実際に乳がんである確率はどれくらいか？

例題 1-2 の解き方



- 樹形図を書くとわかりやすい
- 検査結果で陽性になる人は全部で $(72 + 496)$ 人
- そのうち、実際に乳がんである人は 72 人
- $\mathrm{P}(\text{乳がん} \mid \text{陽性}) = \frac{72}{72+496} \fallingdotseq 0.13$

条件付確率と自然頻度



乳がんの問題についての医師 48 人の回答 (ギーゲレンツァー 2010:75)

注：例題 1-1 や 1-2 と数字が違う問題 (5%の所が 7%) だったので正答がずれている

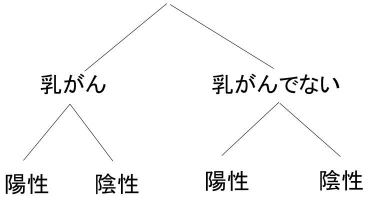
例題 1-1 の解き方 (自然頻度の利用)

　これらの女性の 1 人が乳がんである確率は 0.8%である。また乳がんであれば、検査結果が陽性になる確率は 90%である。乳がんでなくても、陽性と出る確率は 5%ある。ある女性の検査結果が陽性と出た。

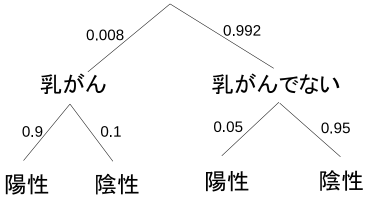
　この女性が実際に乳がんである確率はどれくらいか？

例題 1-1 の解き方 (自然頻度の利用)

- まず、樹形図を書いて起こりうる結果の分類を行なう

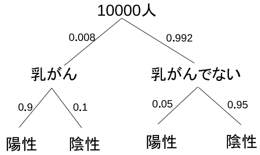


- 問題に従って、それぞれの枝の確率を書く

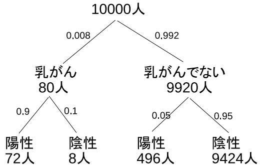


例題 1-1 の解き方 (自然頻度の利用)

- 一番上の人数を決める (何人でもよいが、最後の結果が小数や分数にならないようにするとわかりやすい)

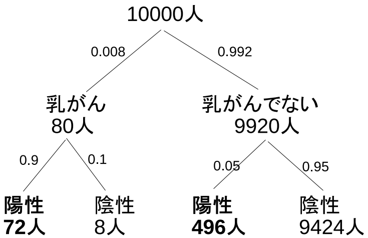


- 枝の確率を掛けて、それぞれの部分の人数を計算



例題 1-1 の解き方 (自然頻度の利用)

- いま求めるのは、陽性になった人の中で、本当に乳がんだった人の割合なので、太字の部分に注目して計算する



- $\mathrm{P}(\text{乳がん} \mid \text{陽性}) = \frac{72}{72+496} \fallingdotseq 0.13$

偽陽性と偽陰性

	乳がん	乳がんでない
陽性		偽陽性 (第一種の過誤)
陰性	偽陰性 (第二種の過誤)	

- 検査が誤っている場合は 2 つある
  1. 本当は乳がんでないのに、検査で陽性になる
     - これを偽陽性という
     - 別名で第一種の過誤 (タイプ 1 エラー)
  2. 本当は乳がんなのに、検査で陰性になる
     - これを偽陰性という
     - 別名で第二種の過誤 (タイプ 2 エラー)

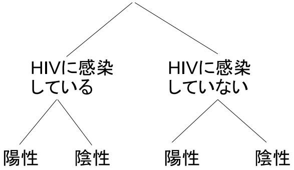
ベイズの定理の練習問題 1：HIV 検査

　HIV 検査について調べてみたところ、次のような情報が見つかったとする。特にリスクの高い行動をとっていない男性の 0.01%が、HIV に感染している。こうした男性が HIV に感染していれば、検査結果が陽性になる確率は 99.9%である。感染していなければ、検査結果が陰性になる確率は 99.9%である。

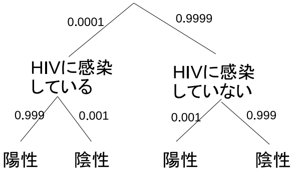
　普通の男性である太郎さんが HIV 検査を受けてみたところ、検査結果は陽性となった。太郎さんが実際に HIV に感染している確率はどれくらいか。

練習問題 1 の解き方の復習

- まず、樹形図を書いて起こりうる結果の分類を行なう

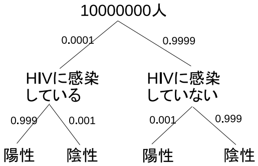


- 問題に従って、それぞれの枝の確率を書く

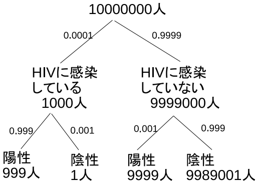


練習問題 1 の解き方の復習

- 一番上の人数を決める (何人でもよいが、最後の結果が小数や分数にならないようにするとわかりやすい)

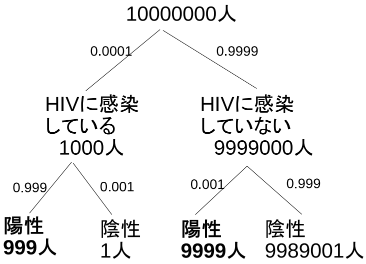


- 枝の確率を掛けて、それぞれの部分の人数を計算



練習問題 1 の解き方の復習

- いま求めるのは、陽性になった人の中で、本当に HIV に感染している人の割合なので、太字の部分に注目して計算する



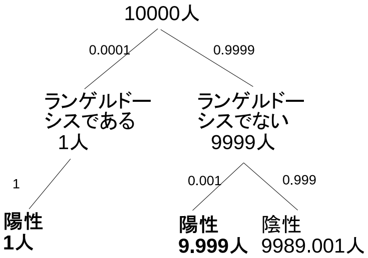
- $\mathrm{P}(\text{HIV に感染している} \mid \text{陽性}) = \frac{999}{999+9999} \fallingdotseq 0.09$

DVD(ハードナッツ) の例：ランゲルドーシス

- 伴田刑事は、ランゲルドーシス (架空の病気) だと診断
- ランゲルドーシスは、1 万人に 1 人がかかる難病
- 検査の精度は 99.9%
- 伴田刑事が本当にランゲルドーシスである確率は？
  - 伴田刑事「間違っている可能性は 0.1%しかない」と悲観
  - くるみ「伴田さんが病気の確率は、99.9%じゃなくて、10 人のうちの 1 人、せいぜい 10%ってことなんです」
- 今まで見てきたベイズの定理から、くるみの方が正しいことがわかる
- ただ、正確な計算をするには一点情報が足りない
  - 検査の「精度」とは何か？ 99.9%とは何の数字か？
  - 本当にランゲルドーシスの人が陽性になる確率か？あるいはランゲルドーシスでない人が陰性になる確率か？

DVD(ハードナッツ) の例：ランゲルドーシス

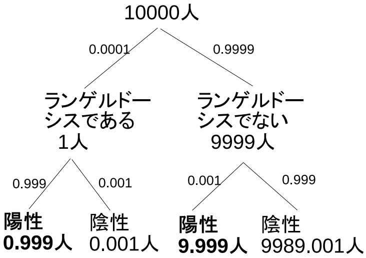
- ランゲルドーシスの人は必ず陽性になり、ランゲルドーシスでない人が陰性になる確率が 99.9%だとすると？
- ここでは数字として出ている 1 万人をあえて使って (小数の人数が出るが) 図を描いてみる



- $\mathrm{P}(\text{ランゲルドーシスである} \mid \text{陽性}) = \frac{1}{1+9.999} \fallingdotseq 0.09$

DVD(ハードナッツ) の例：ランゲルドーシス

- ランゲルドーシスの人が陽性になる確率を 99.9%、ランゲルドーシスでない人が陰性になる確率も 99.9%なら？
  - これでも大差ない結果
- 先ほどの HIV の例とまったく同じ数字の問題



- $\mathrm{P}(\text{ランゲルドーシスである} \mid \text{陽性}) = \frac{0.999}{0.999+9.999} \fallingdotseq 0.09$
- くるみの計算 (約 10%) はだいたい正しかった

練習問題 2：数字の変化

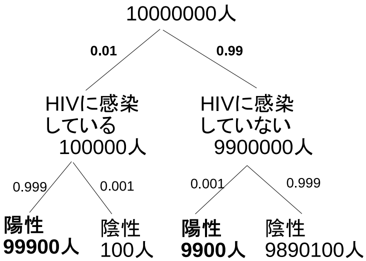
　普通の人々の 0.01%が HIV に感染している。こうした人々が HIV に感染していれば、検査結果が陽性になる確率は 99.9%である。HIV に感染していなければ、検査結果が陰性になる確率は 99.9%である。

この問題を以下のように変えた場合、検査結果が陽性のときに本当に HIV に感染している確率はそれぞれいくつになるか？

(1)「人々の 0.01%」→「人々の 1%」
(2)「陽性になる確率は 99.9%」→「陽性になる確率は 90%」
(3)「陰性になる確率は 99.9%」→「陰性になる確率は 90%」

練習問題 2：数字の変化

- (1) 「人々の 0.01%」→「人々の 1%」



- $\mathrm{P}(\text{HIV に感染している} \mid \text{陽性}) = \frac{99900}{99900+9900} \fallingdotseq 0.91$

参考文献

- マイクル O. フィンケルスタイン (2014)『法統計学入門：法律家のための確率統計の初歩』木鐸社の第 2 章 pp.68–69、第 3 章 pp.70–77.
- Finkelstein, M. O. & B. Levin (2001) *Statisticf for Lawyers* (2nd ed.), Springer の第 1 章、第 2 章、第 3 章.
- ゲルト・ギーゲレンツァー (2010)『リスク・リテラシーが身につく統計的思考法：初歩からベイズ推定まで』早川書房.
- 草野耕一 (2014)「『検察官の誤謬』と『弁護人の誤謬』— ベイズの定理の実践的活用法」論究ジュリスト 10、20–29.
- 市川伸一 (1997)『考えることの科学』中公新書.