COVID-19で全数検査した場合と事前診断で抽出検査した場合はどう違うのか

COVID-19による罹病判定

前提:

COVID-19に感染している割合: 0.002

病気にかかっている人が　検診を受けると、8割の確率で陽性と判定される。

病気にかかっていない人が検診を受けると、9割の確率で陰性と判定される。

10000人の集団を考える

- 感染者数　　20人: **入院する必要のある人数**
- 非感染者数　9980人

（1）10000人が無差別に検診をうける場合

- 感染者で　陽性と判定される　　20 * 0.8　=　16人: **正常判定で入院させられる人数**
- 非感染者で陰性と判定される　9980 * 0.9　=　8982人: 正、陰性判定
- 感染者で　陰性と偽判定される　20 – 16　=　4人: 誤、**野放しになっている感染者数**
- 非感染者で陽性と判定される　9980 – 8982 =　998人: 誤、**無駄に使われている病床**

結果: 陽性と判定された人のうち、正しく判定された人はたったの1.6%

20人しか入院の必要がないのに、998床のベッドが無駄遣いされる

感染しているのに野放しにされる人は4人

COVID-19による罹病判定

前提:

COVID-19に感染している割合: **0.002**

病気にかかっている人が　検診を受けると、**1 – 0.002**の確率で陽性と判定される。

病気にかかっていない人が検診を受けると、**1 – 0.002**の確率で陰性と判定される。

10000人の集団を考える

- 感染者数　　　20人: **入院する必要のある人数**
- 非感染者数　9980人

（1'）10000人が無差別に検診をうける場合

- 感染者で　陽性と判定される　　　20 * 0.998 = 　20人: **正常判定で入院させられる人数**
- 非感染者で陰性と判定される　　9980 * 0.998 = 9960人: 正、陰性判定
- 感染者で　陰性と偽判定される　20 – 　20 = 　　0人: 誤、**野放しになっている感染者数**
- 非感染者で陽性と判定される　　9980 – 9960 = 　20人: 誤、**無駄に使われている病床**

結果: 陽性と判定された人のうち、正しく判定された人はやっと50%

検診の精度を0.998まで上げても、半数のベッドが無駄遣いされる

感染しているのに野放しにされる人はいない

COVID-19による罹病判定

（1）10000人が無差別に検診をうける場合

・感染者で　陽性と判定される　　20 * 0.8　=　　16人：正、**正常判定で入院させられる人数**
・非感染者で陰性と判定される　9980 * 0.9　=　8982人：正、陰性判定
・感染者で　陰性と偽判定される　20 – 16　=　　4人：誤、**野放しになっている感染者数**
・非感染者で陽性と判定される　9980 – 8982 =　998人：誤、**無駄に使われている病床**

結果：陽性と判定された人のうち、正しく判定された人はたったの1.6%

20人しか入院の必要がないのに、998床のベッドが無駄遣いされる

感染しているのに野放しにされる人は4人

（2）他の症状や診断（事前診断）で、感染可能性が高い人を100人選択して検診をする場合

前提：事前診断で選択した集団のうち、感染者の割合　10%

・選択した集団のうち感染者　　10人
・選択した集団にとらえられなかった感染者　　10人　**野放しになっている感染者数**

選択した集団のうち：

・感染者で　陽性と判定される　10 * 0.8 =　8人：正、**正常判定で入院させられる人数**
・非感染者で陰性と判定される　90 * 0.9 =　81人：正、陰性判定
・感染者で　陰性と偽判定される　10 – 8　=　2人：誤、**野放しになっている感染者数**
・非感染者で陽性と判定される　90 – 81　=　9人：誤、**無駄に入院している健常者数**

結果：陽性と判定された人のうち、正しく判定された人は47%に増加

無駄遣いされるベッドは9床に減る

感染しているのに野放しにされる人は12人に増える

事前診断で、感染可能性が高い人を100人程度抽出する条件

N人から確率θで x 人を選ぶ場合の期待値

豊田秀樹著、基礎からのベイズ統計学、朝倉書店(2015)

ベルヌイ分布 (Bernoulli distribution)

2値の事象: 確率θで$x = 0$、$1 - \theta$で$x = 1$が出る確率。

$f(x|\theta) = \theta^x(1-\theta)^{1-x}$, $x = 0,1$: 確率質量関数（probability mass function), 確率関数

θ: 母数 (parameter): 確率関数に含まれる定数

$E[X] = \theta, V[X] = \theta(1-\theta)$

事前診断 x 回でθ の割合で感染候補者を抽出 => 2項分布 (binomial distribution)

$$f(x|\theta) = {}_nC_x\theta^x(1-\theta)^{1-x} = \frac{n!}{x!(n-x)!}\theta^x(1-\theta)^{1-x}$$

期待値 $E[X] = n\theta$　　分散 $V[X] = n\theta(1-\theta)$

期待値 $E[X] = n\theta = 100$ を抽出するには、$\theta \sim 0.01$が必要

$\theta \sim 0.01$となる事前診断の前提

全体の人数	N
感染者の割合	k_{inf}

感染者のうち、何らかの発症があって事前診断で判定可能な割合	k_{on}
感染者のうち、潜伏期間で健常者と区別できない割合	$1-k_{on}$
事前診断で判定可能な人を感染候補者と判定する確率	α
事前診断で健常者と区別できない人を感染候補者と判定する確率	$1-\beta$

感染者のうち、何らかの発症があって事前診断で感染候補者に抽出される割合	$k_{on}\alpha$
健常者と区別できない人の割合	$(1-k_{on})k_{inf}+(1-k_{inf})$
健常者と区別できない人のうち感染候補者に抽出される割合	$(1-\beta)[(1-k_{on})k_{inf}+(1-k_{inf})]$

事前診断 で抽出される全割合 $\theta = k_{on}\alpha + (1-\beta)[(1-k_{on})k_{inf}+(1-k_{inf})]$

$k_{inf} = 0.002, k_{on} = 0.5, \alpha = 0.9,\ 1-\beta = 0.01$

$\boldsymbol{\theta} = 0.0018 + 0.01[0.5 * 0.002 + 0.998] \sim \mathbf{0.012}$ の割合で抽出されることになる

$N = 10000$

抽出される人数	$x = N\theta = 120$人
全感染者数	$Nk_{inf} = 20$人
感染者で何らかの発症があって抽出される数	$Nk_{inf}k_{on}\alpha = 9$
感染者で健常者と区別できない数	$Nk_{inf}(1-k_{on}) = 10$
感染者で健常者と区別できないが抽出される数	$10(1-\beta) = 0.1$
感染候補者中に捕捉された感染者	9人

=> 11人は野放しにされる: 改善するにはαを上げるしかないが、それならPCR検査をする必要がない

その他の条件での結果

COVID-19による罹病判定

（1）10000人が無差別に検診をうける場合

・感染者で　陽性と判定される　　20 * 0.8　=　16人：正、**正常判定で入院させられる人数**
・非感染者で陰性と判定される　　9980 * 0.9　=　8982人：正、陰性判定
・感染者で　陰性と偽判定される　20 – 16　=　4人：誤、**野放しになっている感染者数**
・非感染者で陽性と判定される　　9980 – 8982 =　998人：誤、**無駄に入院している健常者数**

結果: 陽性と判定された人のうち、正しく判定された人はたったの1.6%

998床のベッドが無駄遣いされる

感染しているのに野放しにされる人は4人

（2）他の症状や診断（事前診断）で、感染可能性が高い人を100人選択して検診をする場合

前提: 事前診断で選択した集団のうち、感染者の割合　15%

・選択した集団のうち感染者　　15人
・選択した集団にとらえられなかった感染者　　5人　**野放しになっている感染者数**

選択した集団のうち:

・感染者で　陽性と判定される　　15 * 0.8 =　12人：正、**正常判定で入院させられる人数**
・非感染者で陰性と判定される　　85 * 0.9 =　77人：正、陰性判定
・感染者で　陰性と偽判定される　15 – 12　=　3人：誤、**野放しになっている感染者数**
・非感染者で陽性と判定される　　85 – 77　=　8人：誤、**無駄に使われている病床**

結果: 陽性と判定された人のうち、正しく判定された人は60%に増加

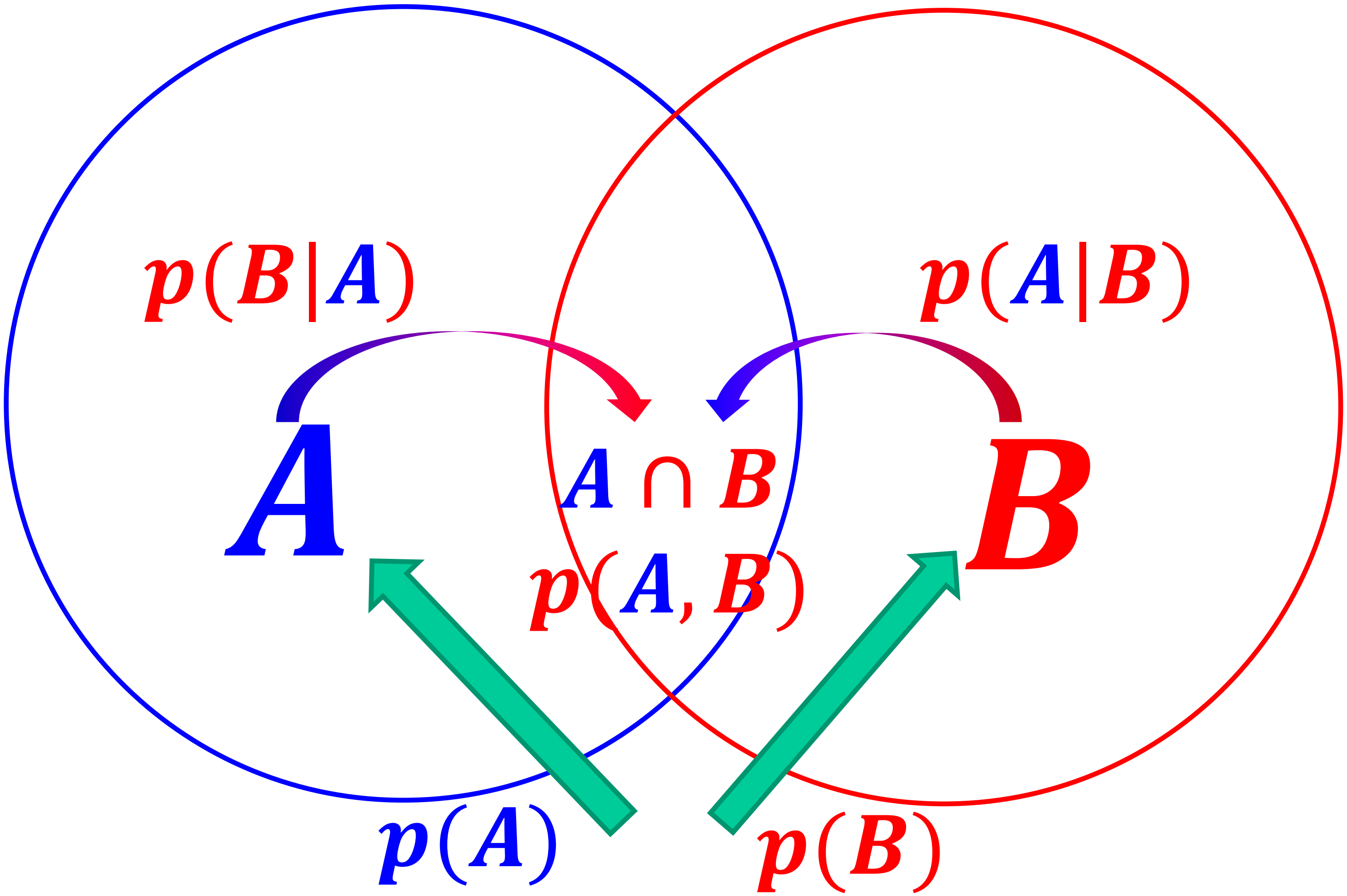
無駄遣いされるベッドは8床に減る

感染しているのに野放しにされる人は5人

ベイズの定理からの計算

確率に関するベイズの定理

$$p(A,B) = p(B|A)p(A) \quad = p(A|B)p(B)$$



ベイズの定理: $p(A|B) = p(B|A)p(A)/p(B)$

事後確率　　事前確率

確率に関するベイズの定理

豊田秀樹著、基礎からのベイズ統計学、朝倉書店(2015)

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

A, B: 事象 (event)

$p(A)$: 事象Aが観察される確率

$p(A,B)$: 同時確率 (joint probability)。事象AとBが同時に観察される確率

$\sum_{i,j} p(A_i, B_j) = 1$

$\sum_i \ p(A_i, B_j) = p(B_j)$: 周辺確率 (marginal probability)

$p(B|A)$: 条件付き確率 (conditional probability)。

事象Aが観察されたという条件の下で、Bが観察される確率

$p(B_j|A_i) = \frac{p(A_i,B_j)}{p(A_i)}$

$\sum_j \ p(B_j|A_i) = 1$

乗法の定理 (multiplication theorem of probability): $p(A_i, B_j) = p(B_j|A_i)\,p(A_i) = p(A_i|A_j)\,p(B_j)$

全確率の公式 (law of total probability):$p(B_j) = \sum_i \ p(A_i, B_j)p(A_i)$

ベイズの定理 (Bayes' theorem): $p(A_i|B_j) = \frac{p(B_j|A_i)p(A_i)}{p(B_j)} = \frac{p(B_j|A_i)p(A_i)}{\sum_i \ p(B_j|A_i)p(A_i)}$

$p(A_i)$: 事前確率 (prior probability)

$p(A_i|B_j)$: 事後確率 (posterior probability)

$p(A_i|B_j) = p(B_j|A_i)$: AとBは互いに独立である(independent)

=> $p(A_i, B_j) = p(A_i)p(B_j)$

ベイズの定理の応用: COVID-19による罹病判定

豊田秀樹著、基礎からのベイズ統計学、朝倉書店(2015)

https://qiita.com/oki_mebarun/items/ee345ba45c7d3752b54c

2020/3/14 イタリア　感染者数 145,637人、死者 5,436人、人口 6000万人

COVID-19 (A) に感染する割合: 0.0024 (事前確率 p(A))

病気Aにかかっている人が　検診Bを受けると、8割の確率で陽性になる。

病気Aにかかっていない人が検診Bを受けると、9割の確率で陰性になる。

検診Bによって陽性と判定された場合、その受信者が病気Aにかかっている確率 (事後確率) を求めよ。

条件付確率: 病気にかかっている人の判定確率

正常診断: $p(\text{陽性}|\text{病気}) = 0.8,\ p(\text{陰性}|\text{病気でない}) = 0.9$　　**正常に判定する確率は8割以上だが・・・**

誤診　　: $p(\text{陰性}|\text{病気}) = 0.2,\ p(\text{陽性}|\text{病気でない}) = 0.1$

事後確率 $p(\text{病気}|\text{陽性}) = p(A|B) = \frac{p(B|A)p(A)}{p(B)} = \frac{p(B|A)p(A)}{\sum_A p(B|A)p(A)}$

$$= \frac{p(\text{陽性}|\text{病気})p(\text{病気})}{p(\text{陽性}|\text{病気})p(\text{病気}) + p(\text{陽性}|\text{病気でない})p(\text{病気でない})}$$

事前確率:

(1) $p(\text{病気}) = 0.0024$

$\boldsymbol{p}(\text{陽性}) = p(\text{陽性}|\text{病気})p(\text{病気}) + p(\text{陽性}|\text{病気でない})p(\text{病気でない})$

$= \mathbf{0.8 * 0.0024 + 0.1 * (1 - 0.0024) = 0.1017}$

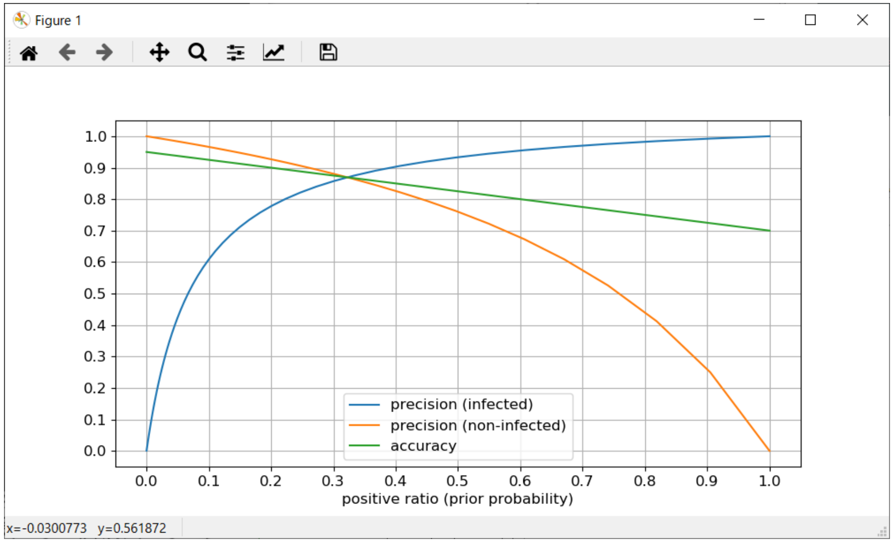
=> $\boldsymbol{p}(\text{病気}|\text{陽性}) = \mathbf{0.018}$　　**陽性と判断されても、98%は病気ではない**

ベイズの定理の応用: COVID-19による罹病判定

Quiita**ベイズの定理で理解する新型コロナウイルスへのPCR検査の検査対象を絞る意義**
https://qiita.com/oki_mebarun/items/ee345ba45c7d3752b54c

感　度: $\mathrm{RC} = p(\text{陽性}|\text{病気})$
特異度: $\mathrm{SP} = p(\text{陰性}|\text{病気でない})$
適合率: $\mathrm{PC} = p(\text{陽性}|\text{病気})$
正解率: $\mathrm{AC} = p(\text{陽性}|\text{病気}) + p(\text{陰性}|\text{病気でない})$
偽陽性率: $\mathrm{FP} = p(\text{陽性}|\text{病気でない}) = 1 - p(\text{陰性}|\text{病気でない})$
偽陰性率: $\mathrm{FN} = p(\text{陰性}|\text{病気}) = 1 - p(\text{陽性}|\text{病気})$

前提条件
再現率（＝感度）RCは0.7
特異度SPは0.95
事前確率p(病気)およびp(病気でない)をパラメータとする。



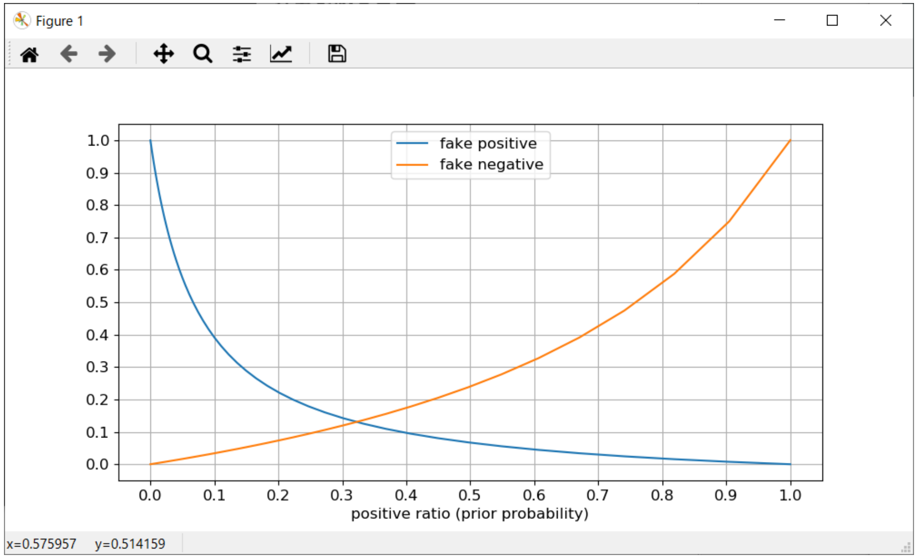
covid19.py
検査陽性の適合率PC(T): 事前確率P(罹患=T)が低いとかなり悪化し、
P(罹患=T)=0.1のとき、検査陽性でも罹患している確率は60%程度。
検査陽性の適合率PC(F): 事前確率P(罹患=T)が高いとかなり悪化し、
P(罹患=T)=0.9のとき、検査陰性でも罹患していない確率は25%程度。
正解率(AC)は、特異度と感度を線形補間する形状であり、両方とも値が高いに越したことはない。

ベイズの定理の応用: COVID-19による罹病判定

Quiita**ベイズの定理で理解する新型コロナウイルスへのPCR検査の検査対象を絞る意義**
https://qiita.com/oki_mebarun/items/ee345ba45c7d3752b54c

※以下は造語
嘘陽性率：検査陽性でも罹患していない確率。　p(病気でない|陽性)
嘘陰性率：検査陰性でも罹患している確率。　p(病気|陰性)
⇔偽陽性率: FP $= p$(陽性|病気でない)、偽陰性率: FN $= p$(陰性|病気)

前提条件
再現率（＝感度）RCは0.7
特異度SPは0.95



covid19.py
嘘陽性率は検査陽性の適合率PC(T)と、嘘陰性率は検査陰性の適合率PC(F)と逆の関係。
嘘陽性率: 事前確率P(罹患=T)が低いとかなり悪化し、
　　P(罹患=T)=0.1のとき、検査陽性でも罹患していない確率は40%程度である。
嘘陰性率: 事前確率P(罹患=T)が高いとかなり悪化し、
　　P(罹患=T)=0.9のとき、検査陰性でも罹患している確率は75%程度である。