

## ベイズ理論

### Bayes' Theorem

Thomas Bayes noticed that sometimes the probability of a statistical hypothesis is given before event or evidence is observed (Prior); he showed how to compute the probability of the hypothesis after some observations are made (Posterior).

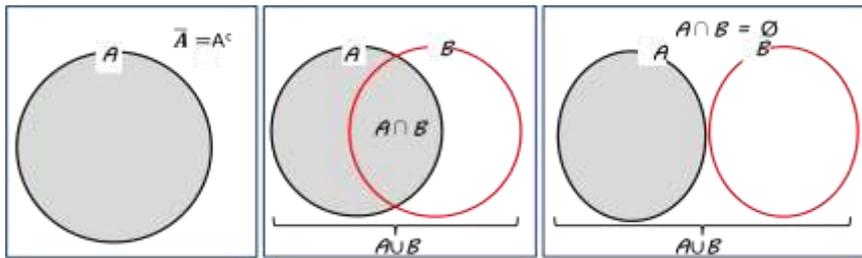
--- Rev. Thomas Bayes

## QUESTION

「先生、子宮頸がんの検診で検査陽性と言われました。癌の可能性は高いですか？」

## THEORY

まずはベン図の復習から。



図左:  $A^c$  は A でない。図中央:  $A \cap B$  は A であり、かつ B である。  $A \cup B$  は A あるいは B である。図右: A と B が同時の起こることはない=相互に排他的: **mutually exclusive**。  
 $A \cup B = A + B$ ,  $A \cap B = \emptyset$  ということ。

## Bayes' Theorem

次に確率を考える。n 回実験を繰り返して m 回 A のイベントが発生する確率は以下。

$$P(A) = \frac{m}{n}$$

A というイベントが既に発生しており、続いて B というイベントも発生する確率は**条件付き確率: conditional probability** と呼ばれ、以下のように表現される。

## $P(B|A)$

まずは全体の中で A+である確率  $P(A)$ が判り、A+ の中で B+ である条件付確率  $P(B|A)$  が判っていれば、 $P(A)$ と  $P(B|A)$ の積を計算することにより、全体の中の  $P(A \cap B)$ である確率を求めることができる。

$$P(A \cap B) = P(A) P(B|A)$$

これは A と B を入れ替えても同じ。

$$P(A \cap B) = P(B) P(A|B)$$

2つの式では  $P(A \cap B)$ が共通なので、

$$\left. \begin{array}{l} P(A \cap B) = P(A) P(B|A) \Rightarrow P(B|A) = \frac{P(A \cap B)}{P(A)} \\ P(A \cap B) = P(B) P(A|B) \end{array} \right\} P(B|A) = \frac{P(B) P(A|B)}{P(A)}$$

という公式ができる。例えば  $A = \text{test +}$ ,  $B = \text{disease +}$  に置き換えると、偽陽性（後述）も含めた  $P(A)$ は検査陽性の確率、 $P(B)$ は検査をする前の病気の確率（対象人口の罹患率: prevalence）、 $P(A|B)$ は病気の人が検査も陽性になる確率= sensitivity。この3つが判っていると、 $P(B|A)$ 、すなわち検査が陽性と判ったあとの病気である確率が判るというわけだ。Bayes' Theorem とは、簡単にいうと  $P(B) =$  事象 A が起きる前の事象 B の確率（事前確率: prior probability）から  $P(B|A) =$  事象 A が起きた後での、事象 B の確率（事後確率, 条件付き確率, posterior probability, conditional probability）が判るという理論である。

## Pap smear

1970年代より子宮頸部細胞診: **Pap smear** がはじまった。この検査により子宮頸がん発症が 64%以上抑制できることが判った<sup>(1)</sup>。子宮頸部を擦過して細胞を採取し、Papanicolaou 染色で癌細胞が混ざっていないか確認するものである。異形度が強い場合、次の検査に進む。そのため、Pap smear が陽性であったからといって生検などで組織を採取し顕微鏡下で確認しても子宮頸がん: **cervical cancer** ではないこともある。

1970年代、アメリカ 44州、306の検査室のデータを集めたところ、子宮頸がんの診断がついた人のうち pap smear が陽性だった人の割合は 0.8375 であった:  $P(\text{test positive} \mid \text{cancer})$ <sup>(2)</sup>。この割合は**感度: sensitivity\***と呼ばれる。ということは子宮頸がんの診断を受けた人の中に pap smear が陰性だった人:  $P(\text{test negative} \mid \text{cancer})$  が  $1-0.8375=0.1625$  いたということだ。この割合は**偽陰性: false negative\***と呼ばれる。一方で、子宮頸がんではないのに pap smear が陽性だった人の割合は 0.1864 だった:  $P(\text{test positive} \mid \text{no cancer})$ 。この割合は**偽陽性: false positive\***と呼ばれる。ということは癌ではなかった人で pap smear も陰性だった人の割合は  $1-0.1864=0.8136$  ということ:  $P(\text{test negative} \mid \text{no cancer})$ 。この割合は**特異度: specificity\***と呼ばれる。整理すると以下のようになる。

\*sensitivity, specificity は病理検査などの **golden standard** に対して検査でどれだけ最終診断に近づけるかを示すものである。例えばインフルエンザの迅速診断の **golden standard** は PCR あるいは培養によるウイルス検査である。

	disease (+)	disease (-)
test (+)	a	c
test (-)	b	d
	a+b	c+d

$$\text{sensitivity} = \frac{a}{a+b} \quad \text{specificity} = \frac{c}{c+d}$$

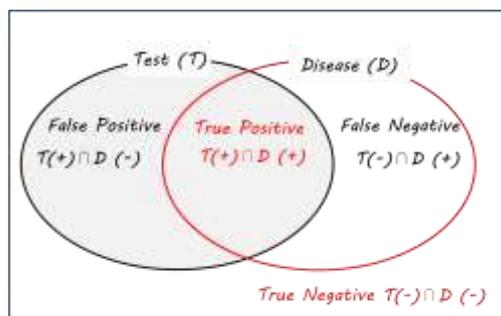
$$\text{False negative} = 1 - \text{sensitivity} \quad \text{False positive} = 1 - \text{specificity}$$

簡便のため pap smear test 陽性を T(+), 陰性を T(-), 子宮頸がん(=disease) 有りを D(+), 無しを D(-)とする。もしも Pap smear が陽性であった場合、「先生、私は子宮頸がんなのですか?」と聞かれるだろう。これに答えるためには  $P(\text{cancer} | \text{test positive}) = P(D+|T+)$ を知る必要があるが、これは sensitivity の分母、分子の関係が逆転している点にご注意いただきたい。ここで Bayes' Theorem を使う。

$$P(\text{cancer} | \text{test positive}) = \frac{P(\text{cancer}) P(\text{test positive} | \text{cancer})}{P(\text{test positive})}$$

$$P(D+ | T+) = \frac{P(D+) P(T+ | D+)}{P(T+)}$$

$P(\text{cancer})=P(D+)$ は所謂罹患率: prevalence で 0.000083 とする。 $P(\text{test positive} | \text{cancer})=P(T+|D+)$ は sensitivity で 0.8375。 $P(\text{test positive}) P(T+)$ は 図の true positive と false positive を合わせたものである。



検査を受けた人の中でPap smearが陽性である確率:  $P(\text{Test}+) \Rightarrow P(T+)$   
 検査を受けた人の中で「子宮頸がん」である確率:  $P(\text{Disease}+) \Rightarrow P(D+)$

$$\text{True positive/total} = P(T+ \cap D+) = P(D+)P(T+|D+) = \text{prevalence} * \text{sensitivity}$$

$$\text{False positive/total} = P(T+ \cap D-) = P(D-)P(T+|D-) = \text{prevalence} * (1 - \text{specificity})$$

$$P(D+ | T+) = \frac{\text{prevalence} * \text{sensitivity}}{\text{prevalence} * \text{sensitivity} + (1 - \text{prevalence})(1 - \text{specificity})}$$

結果は 0.000373。100 万回の検査陽性者の中で本当に子宮頸がんなのは 373 回だけ。

これを陽性的中率: **positive predictive value** という。逆に  $P(D|T-)$  陰性的中率: **negative predictive value** も似た計算式が適用できる。

$$P(D|T-) = \frac{(1 - prevalence) * specificity}{1 - \{prevalence * sensitivity + (1 - prevalence)(1 - specificity)\}}$$

結果は **0.999983** である。ということは **100** 万回の検査で子宮頸がんでないのは **999,983** 回ということ。このスクリーニング検査は世界中で行われているが、数値を見る限り、あまり効率的とはいえない。

### Likelihood ratio

子宮頸がんのスクリーニング検査で見つかる癌の頻度は見方によっては極めて低い。では、この検査の評価を数値で表すことはできないだろうか? **尤度比: Likelihood ratio (LR)** がある。

$$likelihood\ ratio\ (+) = \frac{sensitivity}{1 - specificity} \quad likelihood\ ratio\ (-) = \frac{(1 - sensitivity)}{specificity}$$

**LR(+)**が **10** を超えていれば余計な検査は省略して生検などの最終診断のための検査に進む。逆に **LR(-)**が **0.1** を下回れば、疑っている病気を否定できる。これ以上無駄な検査をする必要がない。**Pap smear** で陽性、陰性のときの **LR** はそれぞれ **4.5**、**0.2** である。ということは生検 (侵襲的最终検査) の前に診断を煮詰めるための **2 次検査** が必要であろう。

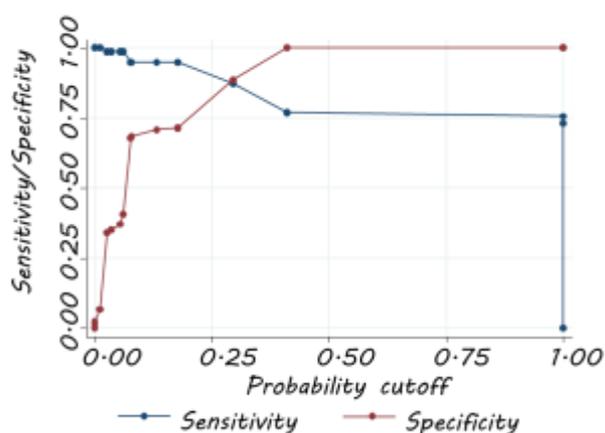
### Sensitivity & Specificity

胃液が食道に逆流して発生する逆流性食道炎が増えている。この病態が慢性化すると下部食道の粘膜が扁平上皮から円柱上皮に変化する。このような病態をバレット食道

と呼んでいるが、特に 3cm 以上のバレット食道に成長すると、この中に **specialized intestinal epithelium (SIM)** が発生することがある。これを放置すればやがて腺癌に進展し得る。そのため食道内視鏡で SIM を早期発見早期切除することが重要になってくる。診断精度を上げるために最近開発された狭帯域光観察 **Narrow Band Imaging; NBI** を SIM の発見に使ってみた<sup>(3)</sup>。246 病変を対象に、NBI で病変部を粘膜の微細な表面構造(**fine mucosal pattern: FMP**)と毛細血管(**capillary pattern: CP**)のパターンの組み合わせで 0.00 から 1.00 までスコアリングを行った。その後、全ての病変を生検して SIM の有無を確認している。この研究で明らかにしたい点は、生検すべきスコアのカットオフポイントを決めることである。生検は侵襲的な検査で食道粘膜に穴があかないとも限らない。よって不必要な検査は避けたいからだ。逆に、どのような病変を生検すれば効率的に SIM を発見できるかを知りたい。

まずはスコアのカットオフポイントを 0.000 から 1.000 に向けて少しずつあげていったとき、それぞれのカットオフポイントにおける **sensitivity** と **specificity** を調べてみた。

Cutoff point	Sensitivity	Specificity	Sum
0-000	100-0	0-0	100-0
0-001	100-0	6-6	106-6
0-025	98-7	33-9	132-7
0-050	98-7	36-9	135-6
0-075	94-9	67-9	162-7
0-100	84-9	70-8	155-7
0-200	87-2	88-7	175-9
0-300	76-9	100-0	176-9
0-400	76-9	100-0	176-9
0-500	75-6	100-0	175-6
0-600	75-6	100-0	175-6
0-700	75-6	100-0	175-6
0-800	75-6	100-0	175-6
0-900	75-6	100-0	175-6
0-990	73-1	100-0	173-1
1-000	0-0	100-0	100-0

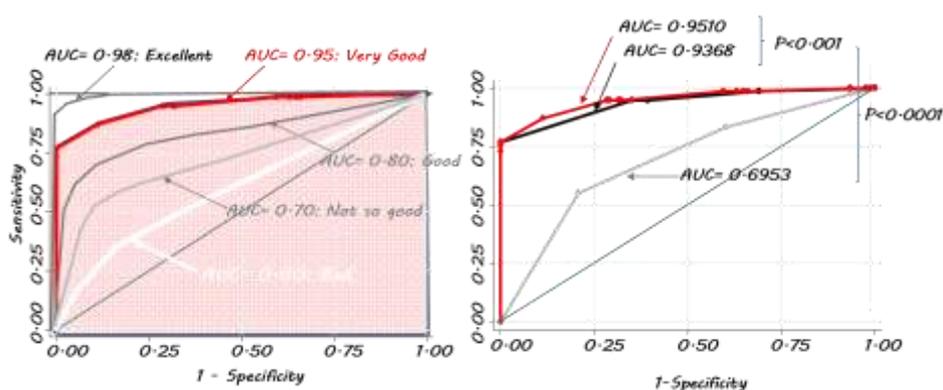


**sensitivity** を高くすると (カットオフポイントを低くセットすると)、SIM 病変検知の洩れが少なくなるが、SIM がないのに生検という侵襲的検査を受ける患者さんの数が

増える。一方、**specificity** を上げると生検検体中の **SIM** の割合が増えるが、生検をしなかった患者さんの中にも実は生検をしたら **SIM** が発見されるケースが増えてくる。検査の性格上 **sensitivity** と **specificity** のどちらを重要視するかで、カットオフポイントを決めるべきである。しかし、**sensitivity** と **specificity** を加えた値が最大のカットオフポイントを選ぶのも 1 つの考えだ。この例ではスコア **0.3** 以上で生検するのがよいと思う。**Specificity** が **100%** であると、**LR** は無限大となり、この検査の陽性で最終診断に進んでよいことを示唆している。

### Receiver operator characteristic (ROC) curve

新しい診断法が優れているかは **ROC curve** で評価することが多い。先の表を縦軸に **sensitivity**, 横軸に **1-specificity** をとると図左赤線のようなになる。



全体の面積を 1 としたとき赤線の下面積: **area under curve (AUC)** は **0.95** と計算できる (図左赤のバックグラウンド)。この値が **1** に近いほど左上の隅に曲線の凸部が近づき、よい診断法ということになる。逆に凸部が弱まり対角線に近づくと精度が悪くなる。**AUC** を比較することにより、検査の精度を統計学的に比較することもできる (図右)。赤線、黒線の違いは僅かに見えるが、統計学的には有意差ありだ。

## EXAMPLE AND EXERCISE

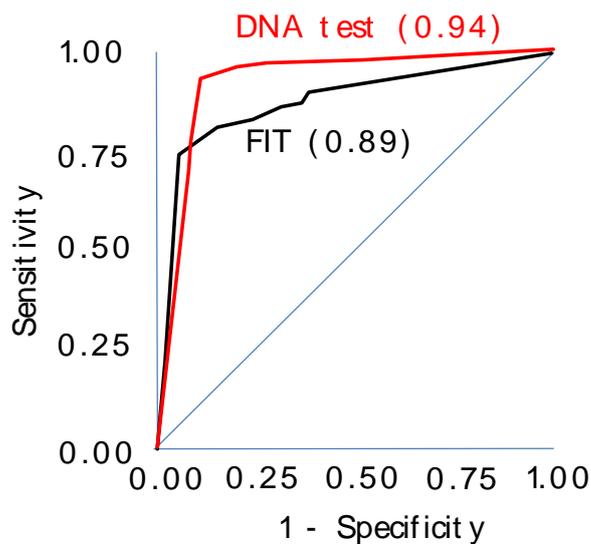
### Exercise 1. Colorectal cancer screening by DNA test <sup>(2)</sup>

大腸直腸がん colorectal cancer (CRC)のスクリーニング検査として正確かつ非侵襲性の検査が期待される。North Carolina の研究チームは便を使って KRAS 遺伝子変異, NDRG4 および BMP3 の異常メチル化を測定する DNA test 、そしてヘモグロビンの免疫染色 fecal immunochemical test (FIT)の CRC screening test を開発した。9,989 人を対象に大腸内視鏡検査を施行したところ、65 人(0.7%)に CRC を発見した(golden standard)。DNA test, FIT の sensitivity はそれぞれ 92.3%, 73.8%、specificity は 86.6%, 94.9%であった。DNA test は FIT と比較して感度に優れているが、特異度に劣り、偽陽性が多かった。

Q1. DNA test, FIT それぞれの PPV, NPV を算出せよ。

Q2. DNA test, FIT それぞれの LR(+), LR(-)を算出せよ。

Q3. ROC curves から DNA test と FIT どちらの検査法が優れているか？



Q4. Q2 と Q3 の回答に違いがあれば、それはどこに要因があるだろうか？

A1.

	P( CRC)	1 - P( CRC)	sensit ivit y	specificit y	P( CRC) *sens	{1-P( CRC) }*(1-spec)	PPV	NPV
DNA test	0.007	0.993	0.923	0.869	0.006461	0.130083	0.047	0.768
FIT	0.007	0.993	0.738	0.949	0.005166	0.050643	0.093	0.901

A2.

	LR (+)	LR (-)
DNA test	7.05	0.09
FIT	14.47	0.28

LR (+)は FIT の方が、LR(-)は DNA test の方が優れている。DNA test が陰性であれば CRC はほぼ否定できる。一方、FIT が陽性であれば CRC の可能性が極めて高い。

A3. DNA test の方が FIT より優れている。図中の数値 0.94, 0.89 はカーブ下の面積でこの数値が 1 に近いほど優れた検査と考えられる。0.90 を超えていれば、かなりよい。

A4. Specificity が高いと LR(+)が、sensitivity が高いと LR(-)が高くなる。ROC の AUC には sensitivity specificity の双方が関与する。総合的にみると DNA test の方が優れていると言えるが、双方の検査を同時に実施すれば、癌の診断精度はさらに高まるであろう。

## SUMMARY

1. Bayes' Theorem を使うことにより、 $P(B) =$  事象 A が起きる前の事象 B の確率(事前確率: prior probability) から  $P(B|A) =$  事象 A が起きた後での、事象 B の確率(事後確率, 条件付き確率, posterior probability, conditional probability) が判る。

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

## 2. 言葉の定義

	disease (+)	disease (-)	t ot al
t est (+)	a	b	a+b
t est (-)	c	d	c+d
t ot al	a+c	b+d	a+b+c+d

sensitivity =  $a/(a+c)$ : 「病気である」ことを検査で正しく言い当てる確率

specificity =  $d/(b+d)$ : 「病気ではない」ことを検査で正しく言い当てる確率

false positive =  $b/(b+d)$  (= 1-specificity): 「病気ではない」のに検査で誤って「病気である」と判断してしまう確率

false negative =  $c/(a+c)$  (= 1-sensitivity): 「病気である」のに検査で誤って「病気ではない」と判断してしまう確率

positive predictive value (PPV) =  $a/(a+b)$ : 検査が陽性で実際病気である確率

negative predictive value (NPV) =  $d/(c+d)$ : 検査が陽性で実際病気である確率

- likelihood ratio (LR) =  $\text{sensitivity}/(1 - \text{specificity})$ : 検査の有用度を評価する指標で、感度を偽陽性で割ることによって得られる。10以上であれば検査法として極めて有用である。
- Receiver operator characteristic (ROC) curve: 新しい診断法が優れているかは縦軸に sensitivity, 横軸に 1-specificity をとった ROC curve で評価する。
- Area under curve (AUC): ROC curve の下の面積を指す。全く診断価値がなければ 0.5, 完全な方法であれば 1 となる。

## MY THOUGHTS

ベイズ理論では事前確率から事後確率を求めるユニークなものである。しかし、事前確率の設定がラフであるため、いくらエビデンスに基づいて事後確率を計算しても、信用できる数値は得られない。日々の診療で使えても、医学研究でエビデンスを創るのには向かない。

## ANSWER TO THE QUESTION

患者さんへは「1 万件の陽性のうち 3~4 件くらいが本当に子宮頸がんです。万が一より若干多いくらい。ということは、1 万件中 9,997 件は陰性ということですから、ほとんどの人はがんではありません。念のため二次検査を受けてみてください。」と説明する。しかし、本当は癌ではないのに (=false positive) 二次検査を受けなくてはならない人がそれだけ多いということだ。一次検査から二次検査の間、受診者の心には葛藤があることだろう。癌のスクリーニング検査では、**likelihood ratio** が高くても罹患率が低いと、なかなか **false positive** の人数を減らすのが難しい。その点、検査を受ける前に説明しておく必要があるだろう。

## REFERENCES

1. La Vecchia C, et al. "Pap" smear and the risk of cervical neoplasia: quantitative estimates from a case-control study. *Lancet*. 1984;2:779-82.
2. Yobs AR, et al. Laboratory reliability of the Papanicolaou smear. *Obstet Gynecol*. 1985;65:235-44.
3. Stata による医療系データ分析入門（第 17 章）. 浦島充佳