

## 最近の癌診断（線虫による診断など）における確率について

2019.09.11 応用数理研究所 佐々木俊久

### 1. 2017年1月に説明した「ベイズ統計入門」の2章 ガンの罹患の例

2019年6月18日 日経産業では、検体数がかなり増加しています。

がん患者 約1400人（2017年時は24人） 健常者 約500人（2017年時は218人）

がん患者をがんと正しく判定できた割合（感度）は84%（2017年時は95.83）,

がんでない人をがんでないと判定できた割合（特異度）は92%（2017年時は94.95）で、

検体数が多くなりどちらの割合も悪くなっています。

記事では、8~9割の確度でがんに反応したとしています。

いろいろな用語が評価として使用されています、これを2章で説明します。

2017年時と同じように、日本の人口 12730万人、がん患者 88万人（がん患者の確率：0.7%）として

がん患者で陽性になる人  $88 \times 0.84 = 73.92$  万人

健常者で陽性になる人  $(12730 - 88) \times (1 - 0.92) = 1011.36$  万人

陽性でがん患者の比率  $73.92 / (73.92 + 1011.36) \times 100 = \underline{6.81\%}$

2017年時より、かなり小さくなりました（2017年時は11.69%）。

#### 別資料

腫瘍マーカーの見方

[https://www.jslm.org/books/guideline/05\\_06/298.pdf](https://www.jslm.org/books/guideline/05_06/298.pdf)

感度80%、特異度80%の検査で、対象者1000名として試算

一般診療所の場合 有病率：1% 陽性的中率 3.8% 誤警告率：96.2%

がん専門病院の場合 有病率：10% 陽性的中率 30.8% 誤警告率：69.2%

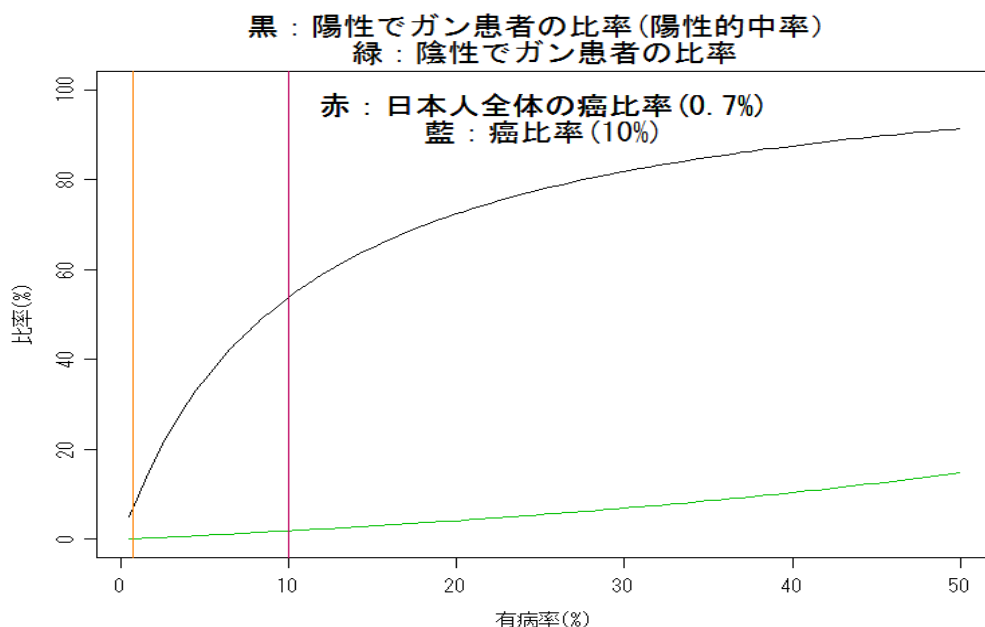
この陽性的中率が、陽性でがんである比率です。

「有病率の低い一般診療所では、陽性的中率はわずか約4%で、96%に誤警告をすることになる。誤警告が大きいと検査を行ったことがムダであっただけでなく、受診者に著しい経済的・精神的負担を負わせることになる。陽性的中率が31%と高いがん専門病院では、検査の妥当性が示される。」と記述されています。

陽性的中率が、どのくらいより大きければ、検査の妥当性があるのか？（佐々木）

今回（日経産業）の有病率と陽性的中率の関係が下図です。

有病率を10%の環境では、陽性的中率は53.8%とかなり大きくなる。



## 2. モデルの精度評価 (分類問題) <sup>1</sup>

下記で表されるような疾病検査を例にとって説明します。

陽性・陰性は予測クラス，疾病あり・なしは正解クラスです。

	疾病あり	疾病なし
陽性	a	b
陰性	c	d

評価指標

### (1) Precision

$$\text{Precision} = a / (a + b)$$

- ・陽性と判定された人のうち，実際に疾病がある人の割合。
- ・Precision(精度，適合率)，Positive Predictive Value(陽性予測値)、陽性的中率と呼ばれる。

### (2) Recall

$$\text{Recall} = a / (a + c)$$

- ・実際に疾病がある人のうち，陽性と判定される人の割合。
- ・Recall(再現率)，Sensitivity(感度)，True Positive Rate(TP,真陽性率)と呼ばれる。

<sup>1</sup> モデルの精度評価 (分類問題)

<http://tkdmah.hatenablog.com/entry/2014/02/22/193008>

### (3) Specificity

$$\text{Specificity} = d / (b + d)$$

- ・実際に疾病がない人のうち、陰性と判定される人の割合。
- ・ Specificity(特異度), True Negative Rate(真陰性率)と呼ばれる。

$$1 - \text{Specificity} = b / (b + d)$$

- ・実際に疾病がない人のうち、陽性と判定される人の割合。
- ・ False Positive Rate(FP,偽陽性率)と呼ばれる。

### (4) F-mesasure

$$F\text{-mesasure} = 2 * \text{Precision} * \text{Recall} / (\text{Precision} + \text{Recall})$$

- ・ Precision と Recall の調和平均。
- ・ F-measure(F 値), F1-measure(F1 値)と呼ばれる。

### (5) Accuracy

$$\text{Accuracy} = (a + d) / (a + b + c + d)$$

- ・ 陽性・陰性が疾病あり・なしを当てる割合。
- ・ Accuracy(正確度)と呼ばれる。

### (6) AUC (詳細は4章)

- ・ AUC(Area Under the Curve)は、陽性・陰性の閾値を変えながら Recall(TP)と(1-Specificity)(FP)をプロットした曲線 (ROC 曲線) の下の面積
- ・ 他の指標と異なり、閾値を一意に特定せずに算出する。

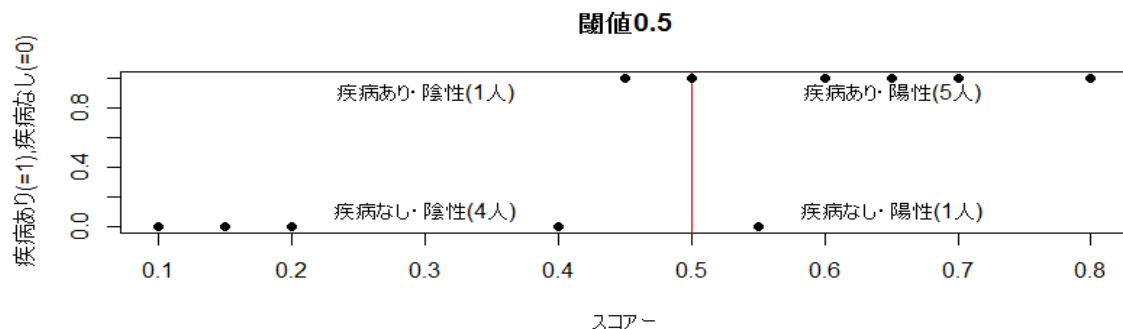
具体例で見ていきましょう。

例えば、クラス1 (癌患者), クラス0 (正常者) を分類する問題で、各人に対して検査した予測スコアが以下のようなになったとします。

大体のモデルでは、「この予測スコアが 0.5 以上ならクラス1とする, 0.5 より小さかったらクラス0とする」という風に閾値 (しきい値) を決めて分類の判断を下します。

クラス	1	1	1	0	0	1	0	1	0	0	1
予測スコア	0.80	0.45	0.70	0.10	0.40	0.50	0.15	0.65	0.20	0.55	0.60

以下の図は、閾値 0.5 での陽性・陰性の判定結果です。



この閾値を色々変化させると、それに応じて各評価指標の値も当然変化していきます。

以下の表は、閾値を大きい値から変えた時の各評価指標の値です。

閾値	適合率	感度	特異度	F値	正確度
1.00	NA	0.000	1.000	NA	0.455
0.80	1.000	0.167	1.000	0.286	0.545
0.70	1.000	0.333	1.000	0.500	0.636
0.65	1.000	0.500	1.000	0.667	0.727
0.60	1.000	0.667	1.000	0.800	0.818
0.55	0.800	0.667	0.800	0.727	0.727
0.50	0.833	0.833	0.800	0.833	0.818
0.45	0.857	1.000	0.800	0.923	0.909
0.40	0.750	1.000	0.600	0.857	0.818
0.20	0.667	1.000	0.400	0.800	0.727
0.15	0.600	1.000	0.200	0.750	0.636
0.10	0.545	1.000	0.000	0.706	0.545

このデータからは、F値(F-measure)や正確度(Accuracy)での閾値 0.45 が良さそうです。

(0.40 より大きく 0.45 以下の範囲の閾値は、同じ結果です)

今回の疾病関連では、疾病（クラス 1）が全体の 1%程度であるような、データが不均衡である場合は、Accuracy(正確度)を用いるべきでない。なぜなら、全員に対して陰性と判定することで Accuracy は 99%と非常に高くなるが、1%の疾病の人をすべて見逃すことになるからである。

### 3. 評価指標をどう使い分けるか？

一般に、この評価指標が常に有用である、というようなものはない。何を目標にするかによって、それに応じた評価指標を用いるべきである。

その際の指針としては、以下が役に立つと思う。

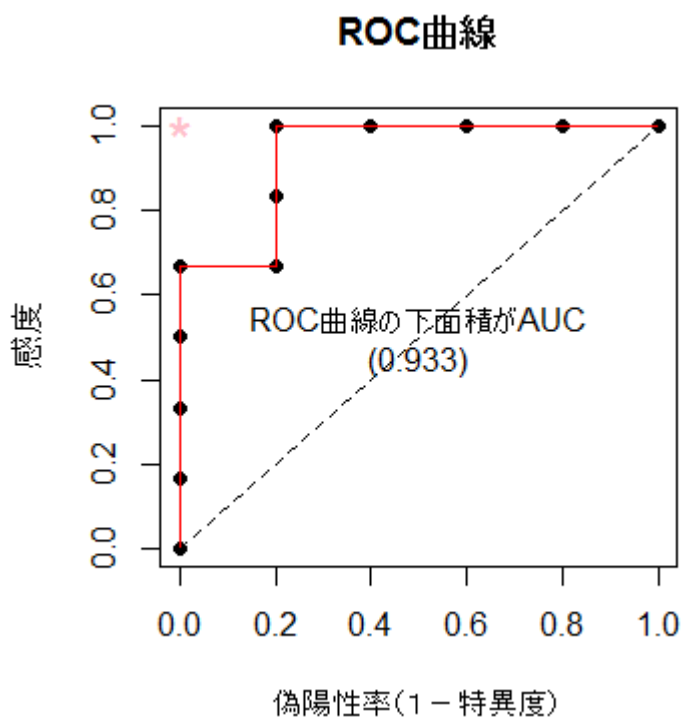
- (1) 特定の閾値を設定せずに評価したい場合 →AUC(後述)
- (2) 特定の閾値を設定して評価したい場合
  - (a) 疾病あり・なしの重要度が同じ場合 →Accuracy(正確度)
  - (b) 疾病ありがより重要である場合
    - (i) 以下の(ii),(iii)の両方に着目したい場合→F-measure(F 値)
    - (ii) 陽性と判定された人のうちの疾病の人の割合が重要である場合→Precision(陽性的中率)
    - (iii) 疾病の人のうち陽性と判定される人の割合が重要である場合→Recall (感度)

### 4. AUC<sup>2</sup>

ROC 曲線は、各スコア（閾値）の感度と特異度から作成します。

<sup>2</sup> ROC 曲線 | 大阪大学腎臓内科・大阪大学医学部

X 軸（横軸）に、偽陽性率（1 - 特異度）（疾病がない人が陽性と判定される人の割合）  
 Y 軸（縦軸）に、感度（疾病がある人が陽性と判定される人の割合）  
 2 章の例で、下図のような黒点を結んだ赤い線が ROC 曲線<sup>3</sup>です。



スコア（閾値）が最低値 0.10 の場合、全員が陽性ですので、感度は 1.0、特異度は 0.0、偽陽性率（1 - 特異度）は 1.0 です。スコア（閾値）が上昇するに従って、感度が低下します。また特異度が上昇するため、偽陽性率（1 - 特異度）は低下します。

スコア値と疾病あり・なしが完全に偶然だった場合、ROC 曲線は右上(1,1)から左下(0,0)への斜点線に近い線になります。一方、理想的にあるスコア値で感度 1.0,特異度 1.0 の場合、左上(0,1)（ピンクの★印）を含む曲線です。

ROC 曲線は、斜点線より左上方に位置しており、ある程度スコア値と疾病あり・なしと関係がありそうです。この関係の度合いを評価するための指標が、ROC 曲線下面積(AUC : area under the curve)であり、0.5 - 1.0 の値をとります。AUC が高ければ高いほど、予測能・診断能が高いと判断できます。

最良な閾値（カットオフ値）の決め方

- (1) 左上（★印(0,1)）からの距離を利用した方法

---

<http://www.med.osaka-u.ac.jp/pub/kid/clinicaljournalclub6.html>

<sup>3</sup> ROC 曲線(Receiver Operating Characteristic Curve, 受信者動作特性曲線)は、もともとレーダーシステムの通信工学理論として開発されたものであり、レーダー信号のノイズの中から敵機の存在を検出するための方法として開発された方法です。

感度と特異度の優れた ROC 曲線は左上隅に近づいていくという事実から、この左上隅との距離が最小となる点の閾値をカットオフ値にする。

## (2) Youden index を用いる方法

(1)の方法とは逆の発想で、最も予測能・診断能の低い  $AUC=0.5$  の斜点線から最も離れた点をカットオフ値にする。すなわち、 $(\text{感度} + \text{特異点} - 1)$  を計算して、その最大値となる点をカットオフ値にする。

## 付録 統計解析パッケージでの分類

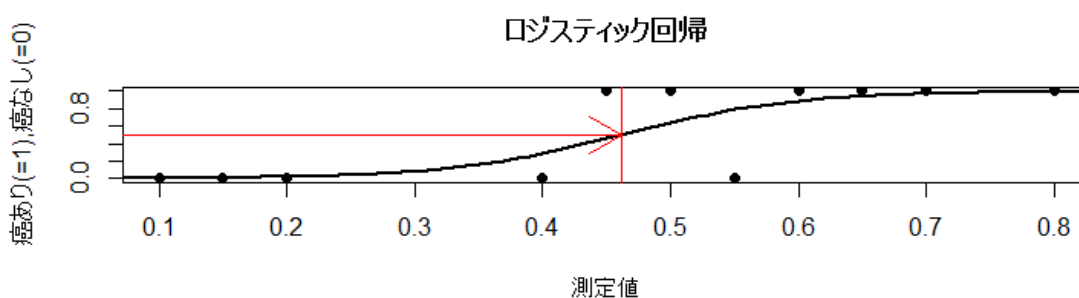
分類の方法は多数あります。今回のような場合にお勧めなのがロジスティック回帰分析です。

### ロジスティック回帰分析

予測する目的変数が 2 値(0 または 1)のとき、量的変数である説明変数を用いた関数を使って判別する方法

特に目的変数の 2 値が不均衡な場合に利用されます。関心のある目的変数を 1 に割り当てます。疾病問題では、疾病ありが 1、クレジット会社の例では不良顧客を 1 とし、小売業の例では優良顧客を 1 とするのが妥当である。

下図はロジスティック回帰の結果図です。



今回は名前の紹介のみ。

以上