生存曲線解析と ROC 解析：資料

芳賀昭弘 *

1　生存分析（教科書にはない内容です）

医療では、データを一斉に生成し、観測するということはまず不可能です。症例の発症、診断のタイミング、研究対象として症例を登録するという時期もバラバラになることが多くあります。また症例の発生に長い時間を要するものもあります。

そのように、対象の登録時点が異なるものであっても、また、事象の発生（イベントの発生、アウトカム、等々という）が一部の症例でしか起きていなくても、解析可能であるような方法が望まれます。その代表がカプラン・マイヤー（Kaplan-Meiyer: KM）法です。

KM 法に必要なデータ · · · **経過時間と打ち切り**

経過時間とは、症例の登録時点からアウトカム発生までの時間 or 打ち切りまでの時間（症例登録の時点としては、手術日、治療終了日などをとることが多い）

打ち切りとは、期間内のアウトカム発生の有無以外に、様々な理由で研究対象から除かれる症例があり、それを打ち切りという（例えば、患者が来なくなった、別の要因で亡くなった、等）

1.1　生存曲線のプロット

表１：生存率データと手計算による解析例

経過時間 [day]	打ち切り	i	n_i	e_i	q_i	p_i	s_i
301	0	1	9	1	1/9	8/9	8/9
565	1						
638	0	2	7	2	2/7	5/7	$8/9 \times 5/7$
638	0						
729	1						
866	0	3	4	1	1/4	3/4	$8/9 \times 5/7 \times 3/4$
868	1						
1196	1						
1319	1						

ここで i はイベント発生の順番、n_i はアウトカムが発生していない数、e_i は各期間のアウトカム発生症例

* Electronic address: haga@tokushima-u.ac.jp

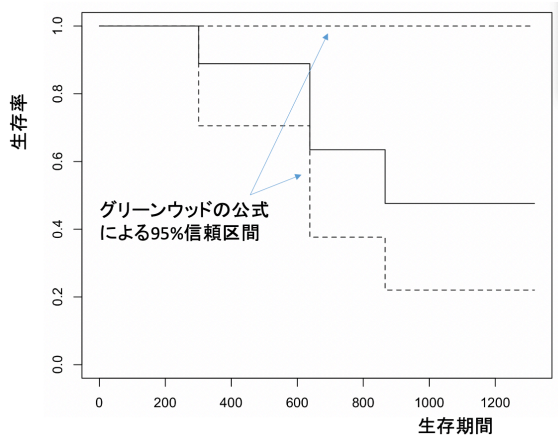

図 1　表 1 から得られる KM plot の例 （横軸は生存期間, 縦軸は累積生存率）

数、q_i はアウトカム発生率、p_i はアウトカムが発生しない率、s_i は累積生存率です。経過時間に対する累積生存率のプロットが、生存曲線となります。経過時間と打ち切りのデータがあれば、生存曲線を描くことができます。

生存曲線における代表値として、平均生存期間や中央値生存期間があります。平均生存期間は、KM 生存曲線の下の面積（生存関数の積分）として求められ、中央値生存期間は、生存確率が 50% となる時点を指します。統計解析の実務では、打ち切りデータが多い場合は中央値生存期間を使い、データが完全な場合は平均生存期間を報告するのが一般的です。

図はある二つの群の生存曲線を描いたものです。これは縦軸が累積生存率で横軸が時間となっており、この図がカプラン・マイヤー生存曲線と言われるものです。みなさんは上の表のデータを使ってカプラン・マイヤー生存曲線を描いてみてください。

生存曲線の信頼区間は、グリーンウッドの公式と呼ばれている次の表式で計算されています。

$$95\% \text{ 信頼区間} = s_i \pm 1.96 SE(s_i), \tag{1}$$

$$SE(s_i) = s_i \sqrt{\sum \frac{e_i}{n_i(n_i - e_i)}} \tag{2}$$

和はその時点までのデータを取ります。

1.2　２つの生存曲線の比較

２つの患者群に分けたとき、片方が一方よりも生存曲線の振る舞いが異なっているということを示すのに χ^2 検定が使えます。この χ^2 検定は、イベントが起きたごとに実施することができますが、これを全てのイベントで解析できるように考えると、別な統計検定量（超幾何分布）が必要となってきます。超幾何分布を使った生存曲線の差の検定はログランク検定（Logrank test）と呼ばれています。さらに、生存時間に応じた解析症例数の変化を考慮した一般化ウィルコクソン検定（Genaralized Wilcoxon test）と呼ばれるものもよく利用されます。

KM 法は、時間経過でイベントが発生するようなデータの解析に使えます（装置の故障率などにも使えますね)。さらに、複数の原因（因子）による事象の発生の影響を調べる方法もあります（比例ハザードモデル)。皆さんも将来、使える場面に遭遇するかもしれません。卒業研究で使用する機会もあるでしょう。生存率や故

障率などのような解析にはこのような方法があるということは覚えておいてください。

2　ROC 解析（教科書にはない内容です）

KM 法を使った生存解析と並んで医療分野でよく使われる解析が、Receiver Operating Characteristic (ROC) curve、ROC 曲線による解析です。ROC 解析の例を考えるため、「健康・病気」などのような 2 つの値（二値）をとる事象を考えます。そしてこれを検診の「白血球数」から予測することを考えてください。白血球数がある基準範囲内であれば健康、外れるようであれば病気と予測するとしましょう。検査結果で病気が疑われるケースを**陽性**、疑われないケースを**陰性**と言います。この陽性と陰性を分ける点（今の例では白血球の数[*1]）を**カットオフ値**（**閾値**）と呼ばれます。この値を極端にとれば、被検者の多くを陽性とすることもできますし、逆にその多くを陰性とすることもできます。被検者の多くを陽性とするようなカットオフ値では、実際に病気でない人を病気と判定することで（これを**偽陽性**という[*2]）さらなる検査を要請することになり資源の無駄や被検者の負担を増やしてしまいます。逆に陰性に偏るようなカットオフ値では、偽陽性は少なくなりますが、本当は病気なのに誤って健康と判定する（これを**偽陰性**という[*3]）ことを避けることができなくなります。適切なカットオフ値を決めることは重要であることがわかります。また、「白血球数」に加えて「血小板数」も病気の予測に用いた場合、予測の精度は増すでしょうか？このように二値推論の複数のモデルを比較することでより良い予測モデルを定量的に評価したいことがあります。カットオフ値による予測性能の変化の振る舞いや推論モデルの比較に使われるのが Receiver Operating Characteristic (ROC) curve（ROC 曲線）です。実際起こった現象と予測の結果を比較する場合、表 2 の 2×2 分割表（**混同行列**）を使うとわかりやすくなります。ROC 曲線は、閾値を変えながら混同行列を作成し、偽陽性率を横軸に陽性率をプロットすると描くことができます。まずは表 2 に示されている混同行列の形と指標の定義を頭に入れましょう

表 2：混同行列

		実際	
		真	偽
予測	真	真陽性	偽陽性
	偽	偽陰性	真陰性

- **感度** Sensitivity：真陽性/(真陽性 + 偽陰性)（実際が真の中で、予測が真の割合）
- **特異度** Specificity：真陰性/(偽陽性 + 真陰性)（実際が偽の中で、予測が偽の割合）
- **陽性的中率** Positive Predictive Value：真陽性/(真陽性 + 偽陽性)（予測が真のうち、実際に真の割合）
- **陰性的中率** Negative Predictive Value：真陰性/(偽陰性 + 真陰性)（予測が偽のうち、実際に偽の割合）
- **オッズ比** Odds Ratio：(真陽性/偽陰性)/(偽陽性/真陰性)
- **相対危険度** Risk Ratio：真陽性/(真陽性 + 偽陽性) / 偽陰性/(偽陰性 + 真陰性)

2.1　ROC 曲線の描画

横軸を偽陽性率、縦軸を陽性率としてプロットしたものが ROC 曲線です。模試の点数によって、実際の試験の合格者はどのくらい正確に予測できたのか、ということをこの ROC 曲線から評価します。ROC 曲線を積分したもの（曲線の下の領域の面積）を Area Under the Curve (AUC) といい、モデルの予測性能を評価

[*1] もう少し詳しく述べると、検査による白血球数を x、日本人成年男性の中央値を m として $|x - m| < c$ を基準範囲とすると、この c が日本人成年男性のカットオフ値となります

[*2] 第一種の過誤

[*3] 第二種の過誤

表３：模試の点数による実際の試験の合格者の予測

模試の点数	合格 or 不合格	陽性率*	偽陽性率**
80	合	1/3	0/2
70	不	1/3	1/2
60	合	2/3	1/2
20	合	3/3	1/2
10	不	3/3	2/2

* 陽性率：模試の点数に閾値を設定したとき、真の合格者のうち何名を検知できたか

** 偽陽性率：模試の点数に閾値を設定したとき、真の不合格者のうち何名を誤って合格者としてしまうか

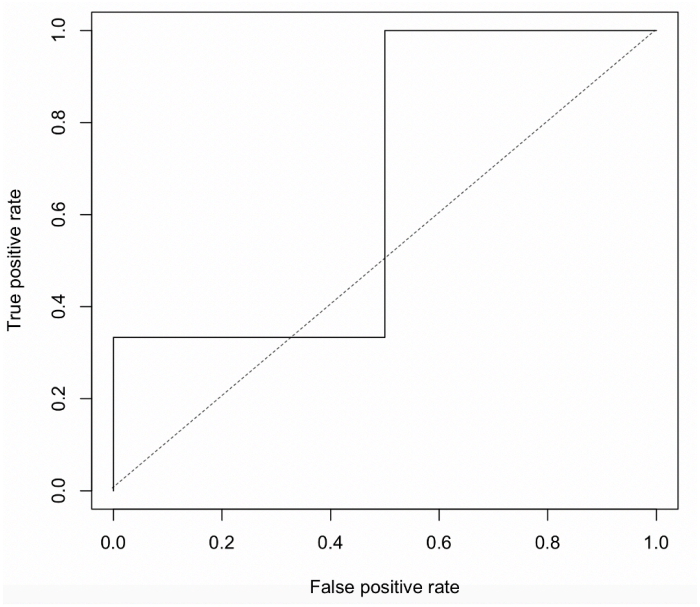

図２　表３から得られる ROC 曲線の例 （横軸は偽陽性率, 縦軸は陽性率）

するときに使います（今の例では、模試の点数そのものが予測モデルです)。また、どの閾値が最も良い予測をすることができたのか、ということに対しては、Balanced Error Rate (BER) という指標がよく使われます。BER は（偽陽性率 + 偽陰性率）/2 で与えられます。これは誤分類の確率を意味しており、BER を最小にする閾値が最も良い性能を与える閾値であるという考え方に則っています。[*4]

[*4] “がん”ではない人を “がん”と判定してしまうことと、“がん”である人を “がん”ではないと判定してしまうこととでは、どちらが深刻でしょう？ BER は偽陽性率と偽陰性率の平均値を与えていますが、モデルの良し悪しの評価には、両者に重み（例えば偽陽性率:偽陰性率=1:1000 など）を掛けて加える重み付き平均をとる損失関数を用意する、といった手法がよいと考えられます。