制限付き平均生存時間を基にした2つの生存曲線の比較

経営システム工学専攻　農田宏樹

1　はじめに

2 群の生存時間の比較を行う臨床試験において、中心的な問題になってくるものが、異なる治療効果を施した 2 群の生存率の結果に差があるかを比較することである。この問題を考えるうえで用いられる手法の一つに 2 群のカプランマイヤー生存曲線をプロットすることである。次にログランク検定により、全区間において 2 群間で明らかな差があるかを判定することができる。そして、2 群の治療効果にどれ程の差があるかを推定するために、何らかのモデルを設定して、その差を定量的に表現する。このために使われるモデルに比例ハザードモデルがあるが、これはデータ集約のための便利なモデルといえる。比例ハザード性が成り立っているならば、時点によらずハザード比は一定であり、群間のリスクの違いをただ一つのパラメータに要約することができる。ゆえに、2 群の治療効果の差を推定するために、比例ハザード性を前提としたハザード比の推定を行う。

しかし、これらの手法は比例ハザード性が成り立っていることが前提のため、この条件が成り立たないときについて生存率の差を定量的に評価するための別の指針の利用が望ましいとされる。そこで考案されたものが、生存曲線下面積と呼ばれる、リスク集合の生存時間の平均値を利用するという手法である。この手法を実データに用いることで、生存曲線下面積の有用性を示す論文も刊行されている。本稿では、生存曲線下面積を検定に応用し、比例ハザード性が成立していないという条件の付いた様々な設定の下でログランク検定の検出力と生存曲線下面積の差の検定の検出力を比較する。

2　ログランク検定について

2.1　ログランク検定の検定統計量の導出

ログランク検定はまず 2 群の生存時間を個別に検討することから始まる。群を A と B と名付ける。両群全体で r 個の異なる死亡時点 $t_{(1)} < t_{(2)} < \cdots\cdots < t_{(r)}$ があり、時点 $t_{(j)}$ において群 A で d_{1j} 人、群 B で d_{2j} 人死亡したとする。($j = 1, 2 \cdots\cdots r$)。同一の群で同じ時点に 2 人以上の被験者が死亡していなければ、d_{1j} と d_{2j} は 0 か 1 かのいずれかとなる。さらに、時点 $t_{(j)}$ の直前において、群 A のリスク集合が n_{1j} 人、群 B のリスク集合が n_{2j} 人であるとする。結果として、時間 $t_{(j)}$ において $n_j = n_{1j} + n_{2j}$ 人のリスク集合の中から合計 $d_j = d_{1j} + d_{2j}$ 人が死亡したことになる。この状況を表 2.1 に示す.

群	t_j 時の死亡数	t_j を超えて生存	リスク集合
A	d_{1j}	$n_{1j} - d_{1j}$	n_{1j}
B	d_{2j}	$n_{2j} - d_{2j}$	n_{1j}
合計	d_j	$n_j - d_j$	n_j

ここで、2 群の被験者の生存関数に差がないという帰無仮説を考える。この仮説の妥当性を評価する方法の 1 つは、各死亡時点において、観測された死亡数と帰無仮説の下で期待される死亡数の差がどの程度であるかを考えることである。そして差の程度についての情報をまとめる。

表 2.7 の周辺和を固定とみなし、生存と群が独立という帰無仮説が正しいとすると、この 4 つの値は、$t_{(j)}$ における群 A の死亡数である d_{1j} を、0 から、d_j と n_{1j} のいずれか小さい方までの任意の値をとりうる確率変数とみなす。d_j は超幾何分布と呼ばれる分布に従い、群 A の死亡数に関する確率変数が値 d_{1j} となる確率は $A = B$ で表すことができる。超幾何分布に従う確率変数 d_{ij} の平均は $e_{1j} = \frac{n_{1j} d_j}{n_j}$ であり、この e_{1j} が時間 $t_{(j)}$ における群 A の死亡数の期待値となる。時間 $t_{(j)}$ で死亡する確率は群によらないという帰無仮説の下では、t_j における死亡確率は $\frac{d_j}{n_j}$ となることから、この値は直感的に理解しやすいものとなる。n_{1j} をかけると、この e_{1j} は t_j の期待死亡数となる。

次の手順は、各死亡時点における個別の 2 × 2 表を併合して、死亡時点全体における d_{1j} の観測値の期待値からの逸脱の指標を得ることである。最も単純な方法は、2 群のすべての死亡時点数 r について、差 $d_{1j} - e_{1j}$ を足し合わせることである。結果として得られるスコ

アは $U_L = \sum_{j=1}^{r}(d_{1j} - e_{1j})$ である。さらに、死亡時点は、互いに独立なので、U_L の分散は d_{ij} の分散の和となる。ここで、d_{1j} は、超幾何分布に従うので、d_{1j} の分散は $v_{1j} = \frac{n_{1j}n_{2j}d_j(n_j-d_j)}{n_j^2(n_j-1)}$ であり、これより U_L の分散は、$var(U_L) = \sum_{j=1}^{r} v_{1j} = V_L$ と書ける。さらに、死亡時点数が少なくなければ、U_L が近似的に正規分布に従うことが示される。すると、$\frac{U_L}{\sqrt{V_L}}$ は、平均 0、分散 1 の正規分布に従う。また、標準正規分布に従う確率変数の平方は、自由度 1 のカイ 2 乗分布に従う。

3 制限付き平均生存時間

生存時間解析において興味のある基本量の一つに、イベント発生までの平均生存時間 (Mean Survival Time) μ がある。この μ は数式としてどのように表すことができるかを考える。まず $\hat{\mu}$ は期待値の定義より、確率密度関数を用いて $\int_0^\infty tf(t)dt$ と書ける。これに部分積分を施すことで

$$\hat{\mu} = \int_0^\tau \hat{S}(t)dt$$

を導くことができる。ここで、τ は最大観測時間か、調査研究者が前もって選択する最大時間のいずれかである。平均生存時間の分散は以下のように書ける。

$$\hat{V} = \sum_{j=1}^{D}[\int_{t_j}^{\tau} \hat{S}(t)dt]^2 \frac{d_j}{n_j(n_j-d_j)}$$

また、2 つの生存曲線の差がないという帰無仮説の下で n を十分大きくすると以下の統計量は自由度 1 のカイ 2 乗分布に従う。

$$\frac{(\hat{\mu}_1-\hat{\mu}_2)^2}{\frac{\hat{V}_1}{m}+\frac{\hat{V}_2}{n}}$$

4 Simulation Study

4.1 Simulation の手順とパラメータの設定

まず、比例ハザード性が成り立つ場合は生存関数が指数分布に従っているとする。ただし、ワイブル分布において shape パラメータを共に 1 にすれば比例ハザード性が成立し、かつ指数分布と同値になるため今回は shape パラメータがともに 1 であるという条件の下で scale パラメータとタイプ 1 打ち切りの時間を変化させた場合の検出力への影響を検討する。サンプルサイズは両群で 200 とし、第 1 群は shape を 1 に、scale を 30 に固定し、第 2 群はサンプルサイズと shape パラメータは同じであるが、scale を 3 ずつ変化させていく。また、打ち切り時間は最大 90 から始めて 5 ずつ下げて変化させる。

比例ハザード性が成り立たない場合は、生存関数が早期に開く場合と後期に開く場合の 2 種類について分けて考える。早期に開く場合には生存関数が従う分布として確率変数が正の値をとる連続分布の中から、カイ 2 乗分布、F 分布、ガンマ分布、ワイブル分布、対数正規分布、ロジスティック分布、一般化パレート分布のいずれかに従うと仮定し、生存曲線が早期に開く設定が成立するようなパラメータを決めたうえでシミュレーションを行い、後期に開く場合に関しても同様確率変数が正の値をとる連続分布の中から、カイ 2 乗分布、F 分布、ガンマ分布、ワイブル分布、対数正規分布、ロジスティック分布、一般化パレート分布のいずれかに従うとし、生存曲線が後期に開く設定が成立するようなパラメータを決めたうえでシミュレーションを行う。

4.2 Simulation の結果 ＜比例ハザード性が成立する場合＞

ここでは、比例ハザード性が成り立っているパターンについての結果と考察を行う。打ち切り時間をどのように設定しても、そして 2 群間での scale パラメータの差をどのように設定しても、ログランク検定の検出力と RMST の差の検定の検出力は多少の誤差はあるにせよ、ほぼログランク検定の検出力と遜色はないという結果となった。

表 1: 2 群のパラメータの差が 12 のときの検出力

打切り時間	ログランク	RMST の差
90	0.894	0.873
85	0.908	0.888
80	0.896	0.874
75	0.890	0.865
70	0.865	0.834
65	0.874	0.843
60	0.871	0.834
55	0.834	0.813
50	0.857	0.806

4.3 Simulation の結果 ＜比例ハザード性が成立しない場合＞

次に、比例ハザード性が不成立の下では、生存曲線が早期 に開く場合に RMST の差の検出力はログランク検定よりも検出力が高いことが確認された。その中で

も生存曲線が早期に開き、かつ全時間にわたり 2 曲線が離れている場合はログランク検定、RMST の差の検定共に検出力が高いことが確認されたが、一方で生存曲線が早期に開くが、終盤に差が知縮まり、打ち切りの直前にはほぼ同じ位置に 2 曲線が存在するパターンの検出力はログランク検定の検出力は開きがあるにもかかわらず検出力が低く RMST の検出力は高く現れた。

逆に生存曲線が遅れて開く場合には RMST の差の検出力はログランク検定よりも低いことが確認 された。

従う分布	ログランク	RMST
ワイブル分布	0.685	0.996
対数正規分布	0.713	0.940
ガンマ分布	0.525	0.818
カイ 2 乗分布	0.734	0.866
対数ロジスティック分布	0.198	0.936

従う分布	ログランク	RMST
対数正規分布	0.807	0.504
対数ロジスティック分布	0.943	0.704
F 分布	0.956	0.754
一般化パレート分布	0.978	0.407

5　課題対処のための検定の構築

前節より、比例ハザード性が成り立たない条件のうち生存曲線の開きが後期におこる設定において RMST の差の検定の検出力はあまり高くないという結果を得た。同じような条件でログランク検定を行うと検出力が RMST の差の検定の検出力よりも高く現れる。このとき、検出力という側面のみで考えればログランク検定を活用することが正しいといえるが、前々章でも述べたようにハザード比要約を行う際に数学的に不整合な結果を得ることになる。従ってこの節では、生存曲線の開きが後期におこる設定に対する比例ハザードモデルベースによらない検定について考える。このことを考慮するうえでの基礎は Uno,Takeuchi の論文より見ることができる。これらの論文の中で、2 群の差を定量的に評価する指標として、t-year survival probability というものがある。これは、適切な時間 t_0 を解析を行う前に予め選択し、その時点における生存率の差を推定するということである。カプランマイヤー推定量自体がノンパラメトリックに推定する統計量であるため求まる差 t-year survival probability もまたモデルの仮定は存在しない。これを数式で表現すると、$S_1(t_0) - S_2(t_0)$ となる。またこの分散 $V(S_1(t_0) - S_2(t_0))$ をグリーンウッドの公式を用いて求め、比をとった式

$$Z = \frac{S_1(t_0)-S_2(t_0)}{V(S_1(t_0)-S_2(t_0))}$$

これが検定統計量となり、帰無仮説が真であるとき大標本の場合において、標準正規分布に従う。

帰無仮説は　$H_0{:}S_1(t_0) - S_2(t_0) = 0$　で、対立仮説は　$H_1{:}S_1(t_0) - S_2(t_0) \neq 0$　とする。

以上の検定についてもシュミレーションを行った。確率変数が正の値をとる連続分布の中から、F 分布、対数ロジスティック分布、一般化パレート分布のいずれかに従うと仮定し、生存曲線が後期に開く設定が成立するようなパラメータを決めたうえで数値実験を行い、検出力を導く。

従う分布	検出力
対数ロジスティック分布	0.843
F 分布	0.84
一般化パレート分布	0.842

6　まとめ

本論文では中途打ち切りを含む 2 標本問題においてハザード比ベースの推定検定の特徴と問題点について、それに対応する比例ハザード性ベースによらない検定方式を検討した。ログランク検定は比例ハザードモデルと密接に関係しており、比例ハザードモデルに対するスコア検定と同じであること、比例ハザード性が成り立っているときに最強力といえる手法である。一方で比例ハザード性が成り立っていないときにはログランク検定を用いると検出力低下する場合があり、検定統計量のスコアに重みを付けたログランク検定の変法を用いて比例ハザードが成立しないデータを検定してもその後のハザード比要約で不整合が起こるという問題点が考えられる。本稿では、比例ハザード性とは無関係な指標を検定に応用する事の検討を行い、様々な生存時間データのパターンに対しシミュレーションを行った。特に本稿では生存関数がいつ開くかにより分類した。まず生存曲線が早期に開く場合については平均生存時間を用い、それを検定に活用する方法が最適であるという結果が得られた。本稿では検定を中心に検討を行ったが、その後の 2 群の差の定量化を行う時は、ハザード比ではなく各々の群の生存曲線下の面積を求めることで平均生存時間を導き、2 群で差をとることで捉えればよい。これについては Uno(2014) の中

で詳しく検討されている。一方で生存曲線が後期に開く場合については、t-year survival probability という時間を特定し、その時における生存時間の差により検定を行うことの良さを提示した。この手法についても 2 群の差の定量化を行う時は、ハザード比ではなく各々の群の特定された時間における生存率同士の差を求めればよい。ただしこの手法で検定を行う時については事前にどの時間に固定するかを決めなければいけない。ゆえにこの検定は解析を行う前に生存曲線の開きが試験時間の前半では起こらないが一定期間を過ぎた後に開きが起こる可能性を考慮することができる場合に限ることが示唆される。

参考文献

[1] Hajime Uno, Brian Claggett, Lu Tian, Eisuke Inoue, Paul Gallo, Toshio Miyata, Deborah Schrag, Masahiro Takeuchi, Yoshiaki Uyama, Lihui Zhao, Hicham Skali, Scott Solomon, Susanna Jacobus, Michael Hughes, Milton Packer, and Lee-Jen Wei (2014) Moving Beyond the Hazard Ratio in Quantifying the Between-Group Difference in Survival Analysis

[2] David Collett 宮岡悦良 (2013) 医薬統計のための生存時間データ解析

[3] 大橋靖雄　浜田知久馬　魚住龍史 (2016)　生存時間解析　応用編　SAS による生物統計

[4] 大橋靖雄　浜田知久馬　(1995)　生存時間解析　SAS による生物統計

[5] クライン, J. P., メシュベルガー, M. L., (2012) 生存時間解析. 打波守訳，丸善

[6] David Collett 宮岡悦良 (2013) 医薬統計のための生存時間データ解析

[7] 赤澤宏平　柳川尭　(2010) サバイバルデータの解析　イベントヒストリデータ (バイオ統計シリーズ)

[8] 中村剛　(2001) Cox 比例ハザードモデル (医学統計学シリーズ)

[9] Lu Tian Haoda Fu, Stephen J. Ruberg, Hajime Uno, and Lee-Jen Wei (2017) Efficiency of Two Sample Tests via the Restricted Mean Survival Time for Analyzing Event Time Observations

[10] Miki Horiguchi Kyongsun Pak Masaaki Mikami Masahiro Takeuchi (2015) Issues of the hazard ratio estimate and application of the restricted mean survival time to a non-inferiority study

[11] Pepe, M. S. and Fleming, T. R. (1989). Weighted kaplan-meier statistics: a class of distance tests for censored survival data.