

生存時間研究における「比較」のデザインと解析

杉本知之*・後藤昌司[†]

生存時間研究の比較において、デザインの多くは用いる統計解析法に依存し、また、用いる統計解析法は比較の目標に応じて変化する。比較の方式において、帰無仮説と予想される対立仮説が立てられ、用いる統計解析法は、これらの仮説の設定に対して最強力な手法が採用される。簡単に考えられる状況において、よく知られた強力な検定はFisherの直接法やログランク検定などが挙げられる。現実の状況はかなり複雑なことが多く、これら共変量調整としてロジスティック・モデルや比例ハザード・モデルの導入により対処ができる。そのようなときでもやはり簡単に考えられる仮説の設定の状況を脱出しておいならず、このようなときの割付けの仕方もあまり明確ではない。他の場合でも、予想される対立仮説の族が上記の検定が用いられるような簡単な状況でないときは、その検定はもはや最強力ではない。本報告では、既存の検定方式がある複雑な状況でどのように検出力が変化するかを議論し、二三の他の検定方式と比較し、デザイン方法についても議論する。通常、メカニズムがわかりにくいため実験デザインの状況で用いにくと思われる二三のノンパラメトリック検定のデザインについても議論する。まずは、簡単な状況から出発する。

例 1. ログランク検定 生存時間研究において様々な断片的な情報の同等性もしくは差を検出することに関心がある。ただし、比例ハザード・モデルではそのような断片的な差よりもむしろ「生存時間分布全体としての差はどうか」という比較を行える利点をもっている。標準治療群を指示する固定共変量を $z_i = 0 (i = 1; \dots; n_1)$ とし、同様に試験治療群の共変量を $z_i = 1 (i = 1; \dots; n_2)$ とする。この2群に比例ハザード・モデル

$$\hat{t}(t; z) = \hat{t}_0(t) \exp(\hat{a}z) \quad (1)$$

を想定して非劣性/同等性を検証することを考える。ここに、 $\hat{t}_0(t)$ は潜在基礎ハザード関数、 $\hat{a}$ は比例性のパラメータである。 $\hat{a} > 0$ のときは試験治療群のリスクは高く、 $\hat{a} < 0$ のときはリスクが少ないと考えられる。そのため、帰無仮説として

$$H_0 : \hat{a} \in B_0$$

をおく。ここに、 B_0 は比例ハザード・モデル上での許容度である。 $\hat{a} = 0$ の検定ときはこの比較はログランク統計量になる。生存時間データ $t_i; \delta_i; z_i; i = 1; \dots; n = n_1 + n_2$ が得られたとき、 $\hat{a}$ を推測するためのCoxの部分尤度は

$$L_p(\hat{a}) = \prod_{i=1}^n \frac{\exp(\hat{a}z_i)}{\sum_{j=1}^n \exp(\hat{a}z_j)} = \prod_{j=1}^n \frac{y_j(t_i) \exp(\hat{a}z_j)}{\sum_{j=1}^n y_j(t_i) \exp(\hat{a}z_j)} \quad (2)$$

*大阪大学 大学院基礎工学研究科情報数理系専攻 〒560-8531 豊中市待兼山町 1-3 (Tel. 06-6850-6486) (Tel. 06-6850-5688)

である．ここに， $y_j(t) = I(t_j \leq t)$ である．

$\hat{\alpha} = B_0$ の一点分布を仮定してその是非を検定する．近似的に $\hat{\alpha} \sim N(\alpha; \sigma_{\hat{\alpha}}^2)$ を利用して $H_0: \alpha = B_0 > 0$ のもとで， $\hat{\alpha} - B_0 \sim N(\alpha - B_0; \sigma_{\hat{\alpha}}^2)$ であるので，検定統計量

$$t_{\hat{\alpha}} = (\hat{\alpha} - B_0)/\sigma_{\hat{\alpha}} \sim N(0; 1)$$

を得る． B_0 の有意義な値を現象的側面，検出力の観点から決定する．

B_0 の設計：薬剤の場合であれば，「試験薬の有効率が 0:1 以上も劣る」のに非劣性/同等性の仮説を採択されてしまうというタイプ II エラー（ベータは比例性のパラメータ α と紛らわしくなるのでここではタイプ II エラーもしくは error-II とする）がほとんどないように $\hat{\alpha}_0 = \hat{\alpha}0:1$ が設計され，かつ，「試験薬の有効率に全く差がない」という本命の対立仮説には error-II = 0:1 や error-II = 0:2 などが使用されていることが多いと思われる．また，生存率に議論を移したときでも「調査期間の中で生存率が最大で 0.1 以上も劣る」のに非劣性/同等性の対立仮説 $\hat{\alpha} = 0$ が採用されてしまうというタイプ II エラーはもちろんほとんどないようにしたい．ここでは，治療により治癒率は変化しないという仮説を少なくとも仮定する．そのために，ここでの議論に通常の比例ハザード・モデルを仮定した．

生存率の関数的な最大差：生存率の差は一般的にメジアンあたりでもっとも大きくひらくことが多い．ここでは比例性の許容度 B_0 の値により実際にどの程度の生存率の差がもたらされるのかをみる． $S_1(t)$ を標準治療群の生存時間関数とすると試験治療群の生存時間関数は $S_2(t) = S_1(t)^{\exp(\hat{\alpha})}$ である．このとき，2 群の生存率の差を $\hat{\epsilon}_b = S_2(t) - S_1(t)$ とする．各 $\hat{\alpha} = 0:0; 0:05; 0:025$ 刻みで生存率の差をプロットしていくと，いずれの場合も生存率 0.5 くらいでもっとも生存率変動し，生存率の最大差が 0.1 くらいになる $\hat{\alpha}$ の値は 0.275 であった．したがって，倫理的な議論もあるであろうが，比例性の許容度として $B_0 = 0:275$ を一つの目安とする．

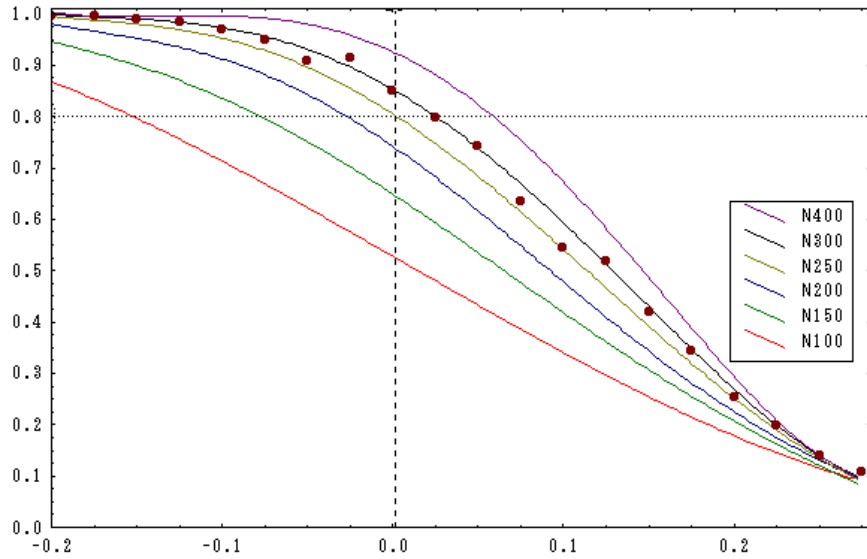
簡単な観測機構での標本サイズと検出力の調査：Cox モデルの使用のため，潜在基礎生存時間関数に対するある特定の仮定は全く必要ない．したがって，規定するのは有意水準 error-I と中途打ちきり機構である．以下の例では，有意水準を 0.1 とし，中途打ち切りなしにおいて検出力を計算した． $B_0 = 0:275$ とする．

図 1 の五つの曲線は試験群と対照群で等標本サイズ ($n_1 = n_2$) とし， $n_1 + n_2 = 100; 150; 200; 300; 400$ に分け，error-I = 0:1 のもとで検出力と真の比例性パラメータ α との関係を描いたものである．ここでは，中途打ち切りは発生させていない．対立仮説 $\hat{\alpha} = 0$ で検出力を約 0.85 を得るためには $n_1 = 150; n_2 = 150$ 程度が必要である．

例 2. 重みづけログランク検定 2 標本において，一般化順位検定等計量は，重みづけログランク統計量 $Z_n(\hat{A})$ の形式で，所与の重み関数 $\hat{W}(\hat{A})$ に対して，同順位なしのとき

$$Z_n = \frac{1}{\sigma^2} \sum_{i=1}^X \hat{W}(t_i) \frac{f\hat{y}_2(t_i)I(z_i = 0) - \hat{y}_1(t_i)I(z_i = 1)g}{\hat{y}(t_i)} \quad (3)$$

$$\text{ここに, } \sigma^2 = \sum_{i=1}^X \hat{W}(t_i)^2 \frac{\hat{y}_1(t_i)\hat{y}_2(t_i)}{\hat{y}(t_i)^2}$$



縦軸：検出力 (1- α error-II), 横軸：真の $\hat{a}$. 中途打ち切りなし.

図 1. 標本サイズごとの検出力と比例性パラメータの関係

と表わされる. ここに, $\tilde{y}(t) = \prod_{i=1}^n y_i(t)$, $\tilde{y}_k(t) = \prod_{i=1}^n I(z_i = k - 1) y_i(t)$; $k = 1, 2$ である. 2 標本に差がないという帰無仮説のもとで, Z_n は漸近的に $N(0, 1)$ に従う.

Harington-Fleming (1991) 族の $G^{\alpha, \zeta}$ 統計量は $\hat{W}(t) = \hat{S}_n^{\circ}(t) [1 - \hat{S}_n(t)]^{\zeta}$ により与えられる重み関数のクラスである. ここに, $\alpha, \zeta \geq 0$, $\hat{S}_n(t)$ は n 個のデータからの併合標本の Kaplan-Meier 推定値である. これに対し, 併合標本の真の生存時間関数を $S(t)$ とし, $W(t) = S^{\circ}(t) [1 - S(t)]^{\zeta}$ とする. とくに, z を $zW(t)$ におき換えれば, (1) 式より, ハザード・モデル

$$\tilde{r}(t; z) = \tilde{r}_0(t) \exp(\hat{a} z W(t)) \quad (4)$$

を得る. 統計量 Z_n は, (2) からの $\hat{a} = 0$ の Wald 型検定統計量と密接な関係をもつことがわかる. すなわち, 同順位なしのときに,

$$Z_n = \frac{U_n(0)}{\sqrt{A I_n(0)}} \frac{\partial}{\partial \hat{a}} \log L_p(\hat{a}) \bigg|_{\hat{a}=0}$$

である. ここに, $U_n(\hat{a}; z)$ は, $\log L_p(\hat{a})$ の $\hat{a}$ の 1 階微分である評点関数, $I_n(\hat{a}; z)$ は, 2 階微分である観測 Fisher 情報量である.

乱数の生成方法: ここでは, 比例ハザード・モデル (4) 式からの乱数の生成法を手短に述べる. 併合標本の真の生存時間関数 $S(t)$ は, 2 標本の混合分布からの生存時間関数と見なせるため, $S(t) = n_1 S_1(t) / n + n_2 S_2(t) / n$ と書ける. 任意の $S_1(t)$ に対し, $S_1(t) = S_1(t) S_1(t)$ であり, $S_2(t) = S_2(t) S_1(t) \exp[\hat{a} S(t) \circ f_1 \hat{A} S(t) g^{\zeta}]$ である. 二つの式をあわせると関係式

$$S(t) = \frac{n_1}{n} S_1(t) S_1(t) + \frac{n_2}{n} S_2(t) S_1(t) \exp[\hat{a} S(t) \circ f_1 \hat{A} S(t) g^{\zeta}]$$

が得られる. この式は, 現実には, $S(t)$ を階段関数近似により解くことになるであろう.

すなわち, $t \in [t_j; t_{j+1})$ において $S(t)$ が一定であるとするとき, 次の $S(t_{j+1})$ は

$$S(t_{j+1}) = \frac{n_1}{n} S_1(t_j) S_1(t_{j+1}|t_j) + \frac{n_2}{n} S_2(t_j) S_1(t_{j+1}|t_j) \exp[\lambda S(t_j) \theta_1 \Delta S(t_j) g^c] \quad (5)$$

と求められる. したがって, $S(t_j)$ および $S_2(t_j)$ ($j = 1; 2; \dots$) が順次に求められる. 実際の計算では, $S_1(t)$ を特定するが, $S_1(t); S_2(t); S(t)$ の3者の関係は, $S_1(t)$ の形状に依存せずに決定されることがわかる. したがって, (4) 式をもとに, 例1と同様のやり方で, B_0 の設計と簡単な観測機構での標本サイズと検出力の調査を行う.

参考文献

- [1] Cox, D. R. (1972). Regression models and life tables. J. Roy. Statist. Soc., Ser. B34, 187-202.
- [2] Fleming, T. R. & Harrington, D. P. (1991). Counting Processes and Survival Analysis. John Wiley & Sons.
- [3] Goto, M. & Matubara, Y. (1996). The proportional hazard model and its marginals. The Best on Quality [Vol.7] ed. by John D. Hromi, ASQC Quality Press, 127-164.
- [4] Tani, S., Kanba, M. & Goto, M. (1986). Statistical evaluation on equivalence of two treatments. The second Japan-China symposium on statistics, 294-299.
- [5] 後藤昌司 (1996). 生存時間解析における予後因子の探索と評価. 日本医療情報学会・医学知識・診断論理研究会-1, 講演資料集, 1-51.
- [6] 後藤昌司・土屋佳英・松原義弘・辻谷将明・惣田隆生 (医学統計研究会) (1984). 生存時間分布とその応用. MPC [Gross, A. J. & Clark, V. A. (1975). Survival Distribution : Reliability Applications in the Biomedical Science. John Wiley & Sons].
- [7] 杉本知之・後藤昌司 (1999). 退化調整型ハザード・モデルとその推測. 計算機統計学, 12(2), 85-106.
- [8] 畠中・田崎・後藤 (1993). 「同等性」議論の現状と今後の展開. 塩野義解析センター: データ解析の標榜, 16, 95-109.
- [9] 山本・浦狩・後藤 (1992). 臨床評価課程における同等性許容度の評価と実際. 塩野義解析センター: データ解析の標榜, 15, 215-225.
- [10] 吉村 (1995). 統計学は良薬を創るのに役に立つか. 数理科学, 389, 43-49.