

2つの生存関数の比較に必要なサンプルサイズ

M2007MM028 高橋友弥

指導教員：田中豊

1 はじめに

生存分析は、あるイベント（故障、死亡など）が起きるまでの時間に焦点を当てる分析方法であり、工学分野においては機械システムや製品の故障など、医学分野においては、病気の再発や死亡などを対象とした研究分野である。2群の生存関数の比較のための代表的な方法としてログランク検定、2群の比較を含め、いくつかの共変量を含む場合の共変量の効果を検定するための代表的な方法としてCox回帰が知られている。本研究では生存関数の2群比較における検出力の評価あるいはサンプルサイズの設計に焦点を絞り、Rubinstein[6], Freedman[1], Lakatos[4]の提案した方法について検討した。このうちLakatosの方法については理論の検討は行ったが要求される事前情報が多く実用的でないのでシミュレーション研究においては省略し、残りの2つの方法についてのみ検討を行った。

2 サンプルサイズと検出力

生存分析に関連したサンプルサイズと検出力の求め方の一般論をLachin[3]に従って要約しておく。

2.1 一般論：完全データの場合

実験での興味は治療群と対照群の差である。すべての患者がイベントが発生する（以下、イベントの発生を死亡と表す）まで、追跡できると仮定する。検定統計量は H_0 のもとで $N(\mu_0, \Sigma_0^2)$, H_1 のもとで $N(\mu_1, \Sigma_1^2)$ に従うと仮定する。ただし母平均 μ_0, μ_1 は既知、 Σ_0^2, Σ_1^2 は観測値の分散 σ^2 とサンプルサイズ N の関数とする。統計的検定をするとき、第1種の誤りと第2種の誤りを考えなくてはならない。実験計画のサンプルサイズ N は第1種の誤りの確率 α を指定し、第2種の誤りの確率 β が一定の値以下であることを要求する。ここでは片側検定を考え、 $H_0: \mu = \mu_0, H_1: \mu > \mu_0$ （または $H_1: \mu < \mu_0$ ）とする。このとき α, β が指定された値をとるためには次の式が成り立たなければならない。

$$|\mu_1 - \mu_0| = Z_\alpha \Sigma_0 + Z_\beta \Sigma_1 \quad (1)$$

多くの場合

$$\Sigma^2 = \frac{\sigma^2}{N}$$

が成り立つ。式(1)に

$$\Sigma_0^2 = \frac{\sigma_0^2}{N}, \Sigma_1^2 = \frac{\sigma_1^2}{N}$$

を代入すると

$$|\mu_1 - \mu_0| = \frac{Z_\alpha \sigma_0}{\sqrt{N}} + \frac{Z_\beta \sigma_1}{\sqrt{N}} \quad (2)$$

サンプルサイズ N に対して解くと

$$N = \left[\frac{Z_\alpha \sigma_0 + Z_\beta \sigma_1}{\mu_1 - \mu_0} \right]^2 \quad (3)$$

同様に Z_β に対して解くと

$$Z_\beta = \frac{\sqrt{N} |\mu_1 - \mu_0| - Z_\alpha \sigma_0}{\sigma_1} \quad (4)$$

式(3)と(4)からサンプルサイズおよび検出力が計算できるが、これらは式(2)から導かれるので基本的な関係を表す式としては

$$\sqrt{N} |\mu_1 - \mu_0| = Z_\alpha \sigma_0 + Z_\beta \sigma_1 \quad (5)$$

を用いればよい。

2.2 打ち切りありの場合（指数分布の場合）

治療中止、転勤などで追跡不能が生じたり、研究の終了時点で対象患者が生存するなどの理由から生存時間のデータが入手できないとき打ち切り（censoring）が生じたという。実際の臨床試験ですべての患者が死亡するまで追跡するのは不可能である。ここで時間 T で終わることを許して、区間 $[0, T]$ に一定の割合でランダムに患者がエントリーし、エントリーしてから死亡するまでの生存時間は指数分布に従うと仮定する。サンプルサイズ N が十分大きい時、ハザード率 λ の最尤推定量は近似的に正規分布に従う。

$$\phi(\lambda) = \frac{\lambda^3 T}{\lambda T - 1 + e^{-\lambda T}} \quad (6)$$

とおき、 Q_e, Q_c はサンプル比で $Q_e + Q_c = 1$ また $\bar{\lambda} = Q_e \lambda_e + Q_c \lambda_c$ とすると

$$\sigma_0^2 = \phi(\bar{\lambda})(Q_e^{-1} + Q_c^{-1})$$

$$\sigma_1^2 = \phi(\lambda_e)Q_e^{-1} + \phi(\lambda_c)Q_c^{-1} \text{ が成り立つ。式(5)に}$$

$$|\mu_1 - \mu_0| = |\lambda_e - \lambda_c|$$

$$\sigma_0^2, \sigma_1^2 \text{ を代入すると}$$

$$\begin{aligned} \sqrt{N} |\lambda_e - \lambda_c| &= Z_\alpha \sqrt{\phi(\bar{\lambda})(Q_e^{-1} + Q_c^{-1})} \\ &\quad + Z_\beta \sqrt{\phi(\lambda_e)Q_e^{-1} + \phi(\lambda_c)Q_c^{-1}} \end{aligned}$$

を導くことができ、サンプルサイズ N あるいは検出力 Z_β について解くことができる。通常 $Q_e = Q_c, \sigma_1^2 > \sigma_0^2$ であるから、このことを利用すると上の式がもっと簡単になり

$$\sqrt{N} |\mu_e - \mu_c| = (Z_\alpha + Z_\beta) \sqrt{\phi(\lambda_e)Q_e^{-1} + \phi(\lambda_c)Q_c^{-1}} \quad (7)$$

となる。

3 ログランク検定

ログランク検定は、死亡というイベントが発生した地点 t ごとに、グループと生存状況を示す 2×2 分割表を構成し、その結果を Mantel - Haenszel(1959) の考え方で総合して群間の差を検定しようとするアイデアがもとになっている。

A 群, B 群をひとまとめにして, 相異なる死亡時点が k 個あったとする. 時点 t_i での A 群, B 群の死亡数を o_{ai}, o_{bi} , 時刻 t_i の直前の時点 $t_i - 0$ での生存数を n_{ai}, n_{bi} とする. これらの集合をリスク集合と呼び, n_{ai}, n_{bi} をリスク集合の大きさと言う. 時間 t_i において, 両群をプールした死亡率 q_i と, 各群の期待死亡数 e_{ai}, e_{bi} は

$$q_i = \frac{o_{ai} + o_{bi}}{n_{ai} + n_{bi}} \quad (8)$$

$$e_{ai} = n_{ai}q_i, e_{bi} = n_{bi}q_i \quad (9)$$

で表され, 全ての死亡時点について両群の期待死亡数と観察死亡数の合計を

$$E_a = \sum e_{ai}, E_b = \sum e_{bi}$$

$$O_a = \sum o_{ai}, O_b = \sum o_{bi} \quad (10)$$

のようにとる. そうするとログランク検定は

$$\chi_0^2 = \frac{(O_a - E_a)^2}{E_a} + \frac{(O_b - E_b)^2}{E_b} \quad (11)$$

となり, H_0 が正しい時, χ_0^2 が自由度 1 の χ^2 分布に従う.

4 Rubinstein の方法

T 年の試験期間の間に患者はポアソン過程に従ってランダムに試験に参加し, 治療群と対照群にランダムに振り分けるとする.

試験参加後の生存時間はハザード率 λ_e, λ_c の指数分布に従うと仮定する. 参加した患者は試験期間終了後 τ 年間治療を継続する.

$$H_0 : \theta = \frac{\lambda_c}{\lambda_e} = 1, H_1 : \theta > 1$$

追跡不能もランダムに発生し, 追跡時間は生存時間とは独立にパラメータ ϕ_e, ϕ_c の指数分布に従うとする. 死亡より追跡不能の方が先に発生すれば打ち切りとなる.

D_c : 治療群の観測された死亡数

D_e : 対照群の観測された死亡数

N : 各群のサンプルサイズ

Z_α : 有意水準 α に対応する $N(0,1)$ の分位点

Z_β : 検出力 β に対応する $N(0,1)$ の分位点

$\lambda_i^* = \lambda_i + \phi_i$

$i = c, e$ である.

このとき最尤推定量 $\hat{\theta}$ の $\log \hat{\theta}$ は漸近的に正規分布に従い有意水準 α の検定の棄却域は

$$\log \hat{\theta} > \hat{\sigma} Z_\alpha \quad (12)$$

となる. $P[\log \hat{\theta} > \hat{\sigma} Z_\alpha] = 1 - \beta$ とすると, 次の式が成り立つ.

$$\frac{(\log \theta)^2}{(Z_\alpha + Z_\beta)^2} = (E[D_c])^{-1} + (E[D_e])^{-1} \quad (13)$$

ただし $E(D_i)$ は次の式を満たさなければならない.

$$E(D_i) = \frac{NT\lambda_i}{2\lambda_i^*} \left(1 - \frac{e^{-\lambda_i^* \tau}}{\lambda_i^* T} [1 - e^{-\lambda_i^* T}] \right) \quad (14)$$

$i = c, e$

式 (13) と式 (14) より

$$\sum \left(\frac{2(\lambda_i^*)^2}{n\lambda_i} \right) \left(\frac{1}{\lambda_i^* T - e^{-\lambda_i^* \tau} (1 - e^{-\lambda_i^* T})} \right) = \frac{(\log \theta)^2}{(Z_\alpha + Z_\beta)^2} \quad (15)$$

式 (15) を T または Z_β について解き, サンプルサイズ N または検出力 $1 - \beta$ を求める. 式 (15) は非線形関数を含む式であるので T を求めるためには反復的な数値解法を利用する必要がある. 初期値を

$$T = \left(\sum \frac{2(\lambda_i^*)^2 (Z_\alpha + Z_\beta)^2}{n\lambda_i \log \theta} \right)^2 \quad (16)$$

としておいて Newton-Raphson を用いて T を解く. T が求まるとサンプルサイズは

$$N = nT \quad (17)$$

で求めることができる.

5 Freedman の方法

生存時間は指数分布に従うと仮定する.

P_c : ある時点の治療群の治療の生存率

P_e : 同じ地点の対照群の治療の生存率とする.

生存時間分布が指数分布のとき,

$$P_i = e^{-\lambda_i t}, i = c, e$$

指数分布の場合どの時点で 2 つの群のハザード率を取っても比は同じで, 2 つの群のハザード比 θ は

$$\theta = \frac{\log_e(P_e)}{\log_e(P_c)} = \frac{\lambda_e}{\lambda_c} \quad (18)$$

であらわせる.

ϕ_i を i 番目の死亡時点のリスク集合の大きさの比とすると, 真のハザード比が θ である場合のログランク統計量の期待値 (E) と分散 (V) は,

$$E = \sum \left[\frac{\frac{\phi_i \theta}{1 + \phi_i \theta} - \frac{\phi_i}{1 + \phi_i}}{\sqrt{\sum \frac{\phi_i}{(1 + \phi_i)^2}}} \right] \quad (19)$$

$$V = \sum \left[\frac{\frac{\phi_i \theta}{1 + \phi_i \theta^2}}{\sum \frac{\phi_i}{(1 + \phi_i)^2}} \right] \quad (20)$$

となり, $\phi_i=1$ の時これらの式は以下のように簡単な式になる.

$$E = \frac{d^{\frac{1}{2}}(\theta - 1)}{\theta + 1} \quad (21)$$

$$V = \frac{4\theta}{(\theta + 1)^2} \quad (22)$$

両群の総死亡数 d を平均 E , 分散 V を持った正規変数として扱うと

$$d = \frac{(\theta + 1)^2}{(\theta - 1)^2} \left[Z_\alpha + \frac{2Z_\beta\sqrt{\theta}}{\theta + 1} \right]^2 \quad (23)$$

が導かれ, Z_β の係数はほぼ

$$2 \frac{\sqrt{\theta}}{\theta + 1} = 1 \quad (24)$$

になることを考慮すると次の式が得られる.

$$d = (Z_\alpha + Z_\beta)^2 \left(\frac{1 + \theta}{1 - \theta} \right)^2 \quad (25)$$

d が求まるとサンプルサイズは

$$N = \frac{2d}{2 - P_c - P_e} \quad (26)$$

で求められる. これは打ち切りなしの場合である. 脱落率 w の打ち切りがある場合は単純にそれだけ上積みして

$$N = \frac{2d}{2 - P_c - P_e} \frac{1}{1 - w} \quad (27)$$

とすればよい.

6 例題とシミュレーションによる精度の検討

Rubinstein と Freedman の方法について, それらの方法の検出力に対する推定精度を検討するため, ハザード比, サンプルサイズのいろいろな条件のもとでの検出力の推定値と, モンテカルロ法で生成したデータにログランク検定を適用したときの経験検出力の比較を行った.

6.1 打ち切りなしの場合

Rubinstein の方法で λ_c と λ_e は Freedman の P_c, P_e から \log を取って計算した.

仮定として θ を推定しても P_c, P_e は一意に決まらないがここでは

$$P_c + P_e = 1$$

を満たすようにしている.

また Rubinstein の式 (15) を

$$Z_\beta = \sqrt{\left(\sum \frac{2(\lambda_i^*)^2}{n\lambda_i} \frac{1}{\lambda_i^* T - e^{-\lambda_i^* \tau} (1 - e^{-\lambda_i^* T})} \frac{1}{(\log \theta)^2} \right)^{-1} - Z_\alpha}$$

Freedman の式 (24) を

$$Z_\beta = \sqrt{d} \sqrt{\left(\frac{1 + \theta}{1 - \theta} \right)^{-2} - Z_\alpha} \quad (28)$$

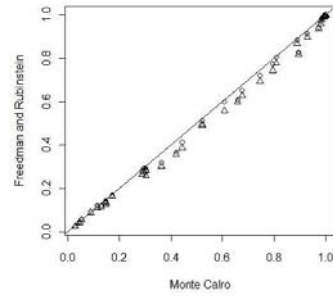


図 1 打ち切りなしの検出力のプロット

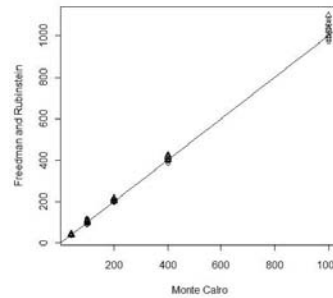


図 2 打ち切りなしのサンプルサイズのプロット

として変形して検出力を計算する.

$\theta=1.33$ の時, $P_c = 0.45, P_e = 0.55$.

$\theta=1.5$ の時, $P_c = 0.43, P_e = 0.57$.

$\theta=2$ の時, $P_c = 0.382, P_e = 0.618$.

$\theta=3$ の時, $P_c = 0.3172, P_e = 0.6828$ と指定した.

まず Rubinstein の T の反復のプロセスは

回数	$n = 20, 1 - \beta = 0.8$	$n = 20, 1 - \beta = 0.9$
1	9.7156	18.6412
2	15.1347	27.2361
3	21.3096	33.8633
4	25.4926	36.0403
5	26.5821	36.1982
6	26.6345	36.1989
7	26.6346	36.1989
8	26.6346	36.1989

表 1 $\phi = 0$ の時の T の計算

収束判定を $|T_{n-1} - T_n| < 10^{-5}$ とおいて Newton-Raphson 法を用いて T を計算したところ表 1 のように 7 回までには収束した. 得られた検出力とサンプルサイズの散布図を描くと図 1, 2 になる.

6.2 打ち切りありの場合

追跡不能のパラメータ $\phi_i = \lambda_c/2, i = c, e$ の打ち切りのある状態を考える. 脱落率 w は指数分布に従うパラメー

タ ϕ, λ_c を与えて 10000 組のデータを生成し i 番目の個体の生存時間, 追跡時間を X_i, Y_i としたとき

$X_i < Y_i$ の時打ち切りなし,

$X_i > Y_i$ の時打ち切りあり.

これで脱落率 w は

$$w = \frac{|X_i > Y_i|}{10000}$$

で判定した. これを Monte Carlo, Freedman, Rubinstein の方法で解析した. Freedman の方法で P_1 と P_2 は Rubinstein の λ_c, λ_e から exp を取って計算した. Monte carlo 法で人工データは指数分布に従う生存時間, 打ち切りをサンプルサイズ分作った. この人工データを 1000 セット作りログランク検定をした. 次にログランク検定の p 値 $< \alpha$ の値となる割合が検出力である. Rubinstein の方法で T を求めるために反復法を使った. これは式 (15) から

$$f(t) = \sum \left(\frac{2(\lambda_i^*)^2}{n\lambda_i} \right) \left(\frac{1}{\lambda_i^* T - e^{-\lambda_i^* \tau} (1 - e^{-\lambda_i^* T})} \right) - \frac{(\log \theta)^2}{(Z_\alpha + Z_\beta)^2} = 0$$

となる. また $f(t)$ を微分したものは

$$f'(t) = - \sum \left(\frac{2(\lambda_i^*)^2}{n\lambda_i} \right) \left(\frac{\lambda_i^* (1 - e^{-\lambda_i^* (\tau + T)})}{(\lambda_i^* T - e^{-\lambda_i^* \tau} (1 - e^{-\lambda_i^* T}))^2} \right) \quad (29)$$

これより Newton-Raphson 法は

$$T_1 = T_0 - \frac{f(T_0)}{f'(T_0)} \quad (30)$$

で $T_{n+1} - T_n = 10^{-5}$ の時の T の解. まず Rubinstein の T の反復のプロセスは

回数	$n = 20, 1 - \beta = 0.8$	$n = 20, 1 - \beta = 0.9$
1	22.5507	43.2674
2	32.1481	52.3815
3	38.3834	54.7807
4	39.8902	54.8938
5	39.9536	54.8963
6	39.9537	54.8963
7	39.9537	54.8963
8	39.9537	54.8963

表 2 $\phi = \lambda_c/2$ の時の T の計算

収束判定を $|T_{n-1} - T_n| < 10^{-5}$ とおいて Newton-Raphson 法を用いて T を計算したところ表 2 のように 6 回までには収束した. 検出力とサンプルサイズの散布図を描くと図 3,4 になる.

7 まとめ

打ち切りのある場合でもなしの場合でもどちらも Monte Carlo 法より検出力の値は小さく, サンプルサイズは大きい. 従ってこれらの推定値はサンプルサイズでは大きめに見積もっており, 安全側によった推定になっている. 打ち切りのある場合は Rubinstein の方, 打ち切りのない場合は Freedman の方が精度が高い.

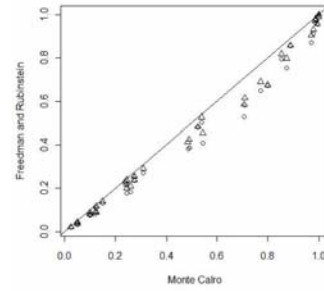


図 3 打ち切りありの検出力のプロット

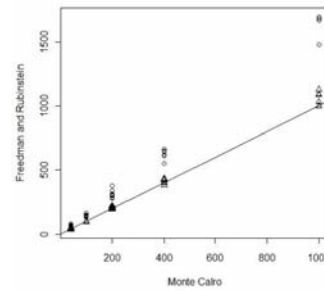


図 4 打ち切りなしのサンプルサイズのプロット

参考文献

- [1] Freedman, L.S. 'Tables of the number of patients required in clinical trials using the logrank test', *Statistics in Medicine*, 1, 121-129 (1982).
- [2] Gross, A.J and Clark, V.A: *Survival Distributions: Reliability Applications in the Biomedical Sciences*. New York: Wiley, (1975)
- [3] Lachin, J.M. 'Introduction to sample size determination and power analysis for clinical trials', *Controlled Clinical Trials*, 2, 93-113 (1981).
- [4] Lakatos, E. 'Sample sizes based on the log-rank statistic in complex clinical trials', *Biometrics*, 44, 229-241 (1988).
- [5] Lakatos, E. and Gordon Lan, K.K. 'A comparison of sample size methods for the logrank statistic', *Statistic in Medicine*, 11, 179-191 (1992)
- [6] Rubinstein, L.V., Gail, M.H. and Santner, T.J. 'Planning the duration of a comparative clinical trial with loss to follow-up and a period of continued observation', *Journal of Chronic Diseases*, 34, 469-479 (1981).
- [7] 金 明哲, R で学ぶデータサイエンス, 森北出版, 183-199 (2007)
- [8] 丹後俊郎・上坂浩之, 臨床試験ハンドブック, 朝倉書店, 172 - 176 (2006)