2019年6月5日　統計数理研究所　オープンハウス

# Kaplan–Meier推定量に基づく一標本検定のサンプルサイズ設計

長島 健悟　医療健康データ科学研究センター 特任准教授

## 【研究背景】

- 生存時間を評価指標とした単群試験は, 抗がん剤の臨床研究のデザインとしてよく用いられる
- このデザインでは, 特定の時点$t$の生存率$S(t)$が許容限界の値$S_0(t)$より大きいか否かの仮説検定を用いた評価がよく行われる

$$\begin{cases} H_0 : S(t) - S_0(t) \le 0 \\ H_1 : S(t) - S_0(t) > 0 \end{cases}$$

- Kaplan–Meier推定量$\hat{S}(t)$は$S(t)$のノンパラメトリック推定量であり, $\hat{S}(t)$及びその変換$g\{\hat{S}(t)\}$が漸近正規性を持つことを利用した推測が行われる
- 変換には恒等変換: $g_1(x) = x$, 対数変換: $g_2(x) = \log(x)$, 補二重対数変換: $g_3(x) = \log\{-\log(x)\}$, ロジット変換: $g_4(x) = \log\{x(1-x)^{-1}\}$, 逆正弦平方根変換: $g_5(x) = \arcsin\sqrt{x}$などが用いられる
- しかし, $g$に恒等変換や対数変換を用いた場合の正規近似の精度は悪く, 仮説検定の第一種の過誤の確率を名義の値付近に制御できないことが知られている [1]
- $\hat{S}(t)$の性質 (漸近正規性など) がよく知られているためか, $g\{\hat{S}(t)\}$に基づくサンプルサイズ設計法の研究はなく, Web上で計算可能な対数変換に基づいた方法 [2] が多くの臨床研究で用いられてきた

## 【研究目的】

- 対数変換は近似精度が悪いため，検討を進めたところ, 既存の方法 [2] は妥当でない計算式に基づくことがわかった
- 本研究では妥当なサンプルサイズ計算式を提案し，適切な変換$g$についても検討する

## 【Kaplan–Meier推定量】

- $T_i\ (i = 1, \dots, n)$が分布$F(t)$に従う生存時間確率変数, $U_i$がある特定の分布に従う打切時間確率変数, $X_i = \min\{T_i, U_i\}$が観測されると仮定
- Kaplan–Meier推定量とその漸近分散の推定量$\hat{\sigma}^2(t)$は

$$\hat{S}(t) = \prod_{s<t} \left\{1 - \frac{\Delta \bar{N}(s)}{\bar{Y}(s)}\right\},\ \hat{\sigma}^2(t) = \hat{S}(t)^2 \int_0^t \frac{1}{\bar{Y}(s)\{\bar{Y}(s) - \Delta \bar{N}(s)\}} \mathrm{d}\bar{N}(s).$$

ただし，$N_i(t) = I(X_i \le t, \delta_i = 1)$は計数過程，$\delta_i = I(T_i < U_i)$，$Y_i(t) = I(X_i \ge t)$はat risk過程，$\bar{N}(t) = \sum_{i=1}^n N_i(t)$，$\bar{Y}(t) = \sum_{i=1}^n Y_i(t)$, $\Delta X(t) = X(t) - X(t-)$, $X(t-) = \lim_{h \downarrow 0} X(t-h)$.

- $\sqrt{n}\{\hat{S}(t) - S(t)\} \to_L -S(t)L(t)$で$\sigma^2(t) = S^2(t)\mathrm{Var}[L(t)]$であり, $L(t)$は$\mathrm{E}[L(t)] = 0$で$\mathrm{Var}[L(t)] = \int_0^t \{\Pr[U > s]S(s)\}^{-1}\mathrm{d}\Lambda(s)$のブラウン運動.
- $\sqrt{n}[g\{\hat{S}(t)\} - g\{S(t)\}] \to_L g'\{S(t)\}\{-S(t)L(t)\}$であり，漸近分散は$\tau^2(t) = \mathrm{Var}(\sqrt{n}[g\{\hat{S}(t)\} - g\{S(t)\}] = g'\{S(t)\}\sigma^2(t)$.

## 【提案するサンプルサイズ設計の方法 [3]】

- 仮説$H_0 : \epsilon \le 0$, $H_1 : \epsilon > 0$の検定に基づく($\epsilon = g\{S_1(t)\} - g\{S_0(t)\}$)
- 有意水準$\alpha$で$Z = \frac{g\{\hat{S}(t)\} - g\{S_0(t)\}}{\hat{\tau}(t)/\sqrt{n}} > z_{1-\alpha}$なら$H_0$を棄却; $H_0$の下で$Z \to_L N(0,1)$である事に基づく. $z_p = \Phi^{-1}(p)$, $\hat{\tau}(t) = g'\{\hat{S}(t)\}\hat{\sigma}(t)$.
- $H_1$の下では$n$が大きいとき$Z$が$N(\epsilon\sqrt{n}/\tau_1, 1)$で近似できるため,

$$\Pr(Z > z_{1-\alpha} \,|\, H_1) \approx \Phi\left(-z_{1-\alpha} + \frac{\epsilon\sqrt{n}}{\tau_1}\right).$$

ただし, $\tau_j^2 = g'\{S_j(t)\}^2\sigma_j^2(t)$.

- 上式右辺$= 1 - \beta$の解, $n_1 = \left\{\frac{\tau_1(z_{1-\alpha} + z_{1-\beta})}{\epsilon}\right\}^2$を提案法とする

## 【既存法の問題点】

- 既存法 [2] は対数変換に基づいており, $n_2 = \left(\frac{\tau_1 z_{1-\alpha} + \tau_0 z_{1-\beta}}{\epsilon}\right)^2$を用いる
- $n_2$は

$$\Pr(Z > z_{1-\alpha} \,|\, H_1) \approx \Pr\left(\frac{\tau_0 W + \epsilon\sqrt{n}}{\tau_1} > z_{1-\alpha}\right),$$

に基づいている, ただし$W \sim N(0,1)$.

- しかし$(\tau_0 W + \epsilon\sqrt{n})/\tau_1$は$N(\epsilon\sqrt{n}/\tau_1, \tau_0^2/\tau_1^2)$に従うため, 近似は成立せず

## 【シミュレーションによる性能評価】

- 各変換の時の第一種の過誤の確率を評価 ($\alpha = 0.05$)

表1: 第一種の過誤の確率のシミュレーション結果

| Sample size | $S_0(t)$ | 恒等 | 対数 | 二重対数 | ロジット | 逆正弦 |
|---|---|---|---|---|---|---|
| 25 | 0.10 | $\underline{0.009}$ | $\underline{0.098}$ | 0.034 | 0.034 | 0.034 |
| | 0.30 | 0.044 | $\underline{0.098}$ | 0.044 | 0.044 | 0.044 |
| | 0.50 | 0.054 | $\underline{0.115}$ | 0.054 | 0.054 | 0.054 |
| | 0.70 | $\underline{0.091}$ | $\underline{0.091}$ | 0.033 | 0.033 | $\underline{0.091}$ |
| | 0.90 | $\underline{0.072}$ | $\underline{0.072}$ | $\underline{0.072}$ | $\underline{0.072}$ | $\underline{0.072}$ |
| 50 | 0.10 | $\underline{0.025}$ | 0.058 | 0.058 | 0.058 | 0.058 |
| | 0.30 | 0.048 | $\underline{0.085}$ | 0.048 | 0.048 | 0.048 |
| | 0.50 | 0.060 | 0.060 | 0.033 | 0.060 | 0.060 |
| | 0.70 | $\underline{0.080}$ | $\underline{0.080}$ | 0.040 | 0.040 | 0.040 |
| | 0.90 | $\underline{0.112}$ | $\underline{0.112}$ | 0.034 | 0.034 | $\underline{0.112}$ |
| 100 | 0.10 | $\underline{0.020}$ | $\underline{0.072}$ | 0.040 | $\underline{0.072}$ | 0.040 |
| | 0.30 | 0.053 | 0.053 | 0.053 | 0.053 | 0.053 |
| | 0.50 | 0.044 | 0.067 | 0.044 | 0.044 | 0.044 |
| | 0.70 | $\underline{0.075}$ | $\underline{0.075}$ | 0.048 | 0.048 | 0.048 |
| | 0.90 | $\underline{0.117}$ | $\underline{0.117}$ | $\underline{0.024}$ | $\underline{0.024}$ | 0.057 |

- 恒等変換と対数変換は名義の$\alpha$を上回ることが多く, 恒等変換は$S(t)$が0に近いときに過度に保守的
- 他は一部を除いて概ね制御できる, 補二重対数変換はやや保守的

- 提案法と既存法でサンプルサイズ設計をした時の検出力が名義の値$1 - \beta = 0.80$に近いか評価

表2: 提案法と既存法の比較結果

| | $S_0(t)$ | $S_1(t)$ | 提案法 | | | | | 既存法 |
|---|---|---|---|---|---|---|---|---|
| | | | 恒等 | 対数 | 二重対数 | ロジット | 逆正弦 | |
| Power | 0.1 | 0.2 | $\underline{0.862}$ | $\underline{0.738}$ | $\underline{0.760}$ | $\underline{0.768}$ | 0.793 | 0.785 |
| | 0.4 | 0.5 | 0.789 | $\underline{0.762}$ | 0.803 | 0.792 | 0.790 | 0.821 |
| | 0.7 | 0.8 | $\underline{0.756}$ | $\underline{0.719}$ | $\underline{0.857}$ | $\underline{0.845}$ | 0.795 | 0.791 |
| Sample size | 0.1 | 0.2 | 99 | 52 | 75 | 59 | 77 | 71 |
| | 0.4 | 0.5 | 155 | 125 | 166 | 151 | 153 | 144 |
| | 0.7 | 0.8 | 99 | 87 | 142 | 134 | 115 | 106 |

- 提案法は逆正弦平方根変換の場合に最も名義の値に近く, 各変換の傾向は表1と対応 (例：保守的な時$n$が多く検出力大)
- 既存法も名義の値に近いが, 対数変換が第一種の過誤の確率を制御できない点と符合せず. 何故か誤った式が$1 - \beta$に近い$n$を与える.

## 【終わりに】

- 既存法の問題を明らかにし, 妥当なサンプルサイズ設計の方法を提案
- シミュレーション結果から, 逆正弦変換又は補二重対数変換を推奨
- 無変換と対数変換は第一種の過誤の確率を制御できない, 解析する際はソフトウェアの設定確認を (Rのデフォルトは対数変換)
- 提案法Webアプリ公開済; https://nshi.jp/contents/js/onesurvyr/

[1] Borgan Ø and Liestøl K. Scan J Stat 1990; 17(1): 35–41.

[2] Cancer Research and Biostatistics. 2002. https://stattools.crab.org/

[3] Nagashima K, Noma H, Sato Y, Gosho M. Under review.