

データ解析

生存時間データ解析 (生存関数, ハザード関数)



創域理工学部

Faculty of Science and Technology

東京理科大学
創域理工学部情報計算科学科
安藤宗司

2023年11月30日

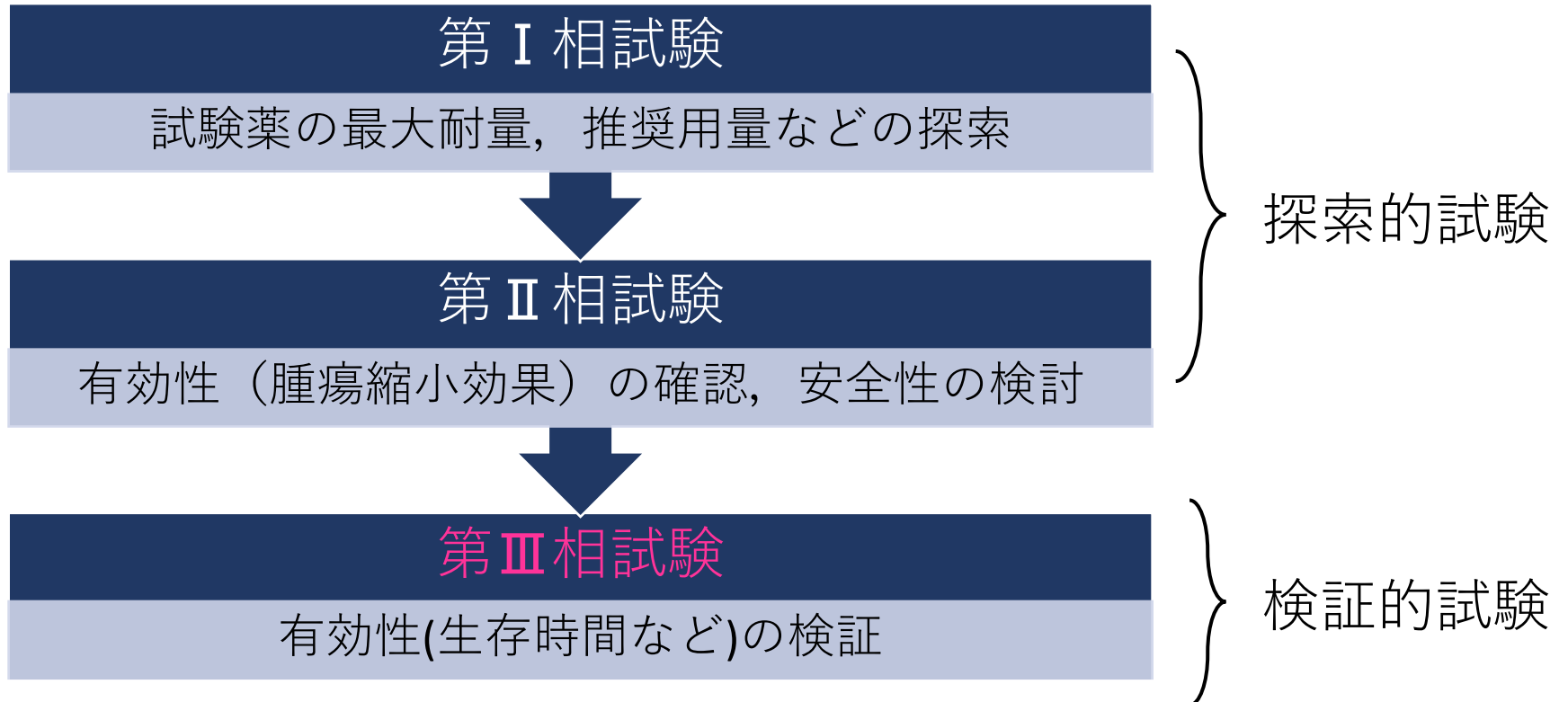
Contents

- がん臨床試験
- 生存時間データの特徴
 - 中途打ち切り
- 生存関数とハザード関数
- 代表的なパラメトリックモデル

がんの臨床試験

□ 新しい治療法（薬の候補など）について，有効性と安全性を科学的に検証する方法

■ 多くの患者さんの理解と協力を得て，実施される



有効性評価項目

□ 真の評価項目

- 全生存期間（Overall survival; OS）
ランダム化からあらゆる原因による死亡までの期間

□ 主な代替評価項目

- 客観的奏効率（Response criteria for solid tumors; RECIST）
抗がん剤投与に伴う腫瘍縮小効果の評価
- 無増悪生存期間（Progression free survival; PFS）
ランダム化からがんの進行または死亡までの期間
- 無病生存期間（Disease free survival; DFS）
ランダム化から客観的な再発または死亡までの期間
- 無増悪期間（Time to progression; TTP）
ランダム化からがんの進行までの期間

RECIST

- 固形がんの治療効果判定のための新ガイドライン

http://www.jcog.jp/doctor/tool/RECISTv11J_20100810.pdf

- 効果判定のために選択された病変を測定する方法
 - CT が最も広く利用可能で、最も再現性に優れた方法
 - 特定の状況（例：体幹部撮影など）においてはMRIも許容

標的病変の評価

□ 完全奏功（Complete Response; CR）

- すべての標的病変の消失。標的病変として選択したすべてのリンパ節病変は、短径で10mm未満に縮小しなくてはならない

□ 部分奏効（Partial Response; PR）

- ベースライン径和に比して、標的病変の径和が 30%以上減少

□ 進行（Progressive Disease; PD）

- 経過中の最小の径和（ベースライン径和が経過中の最小値である場合、これを最小の径和とする）に比して、標的病変の径和が 20%以上増加、かつ、径和が絶対値でも 5 mm以上増加

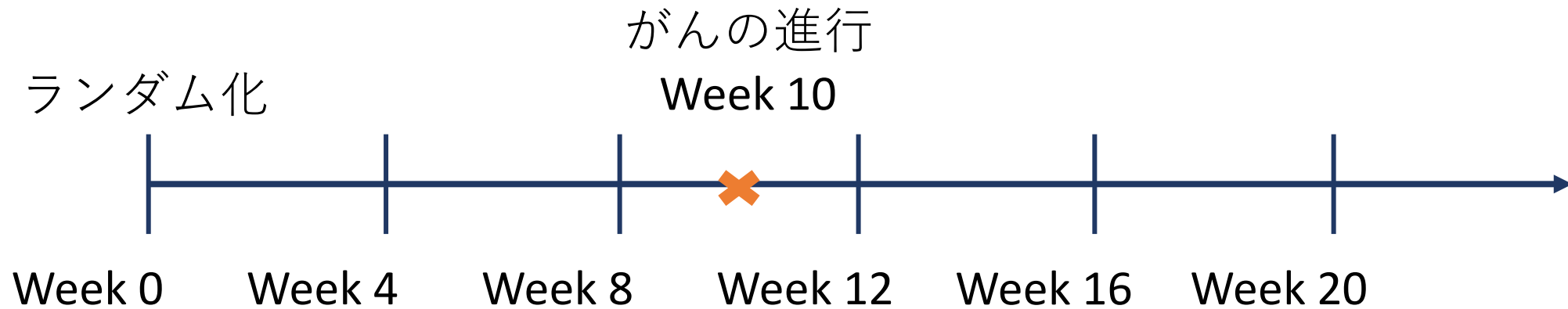
□ 安定（Stable Disease; SD）

- 経過中の最小の径和に比して、PR に相当する縮小がなく PD に相当する増大がない

がんの進行の評価

□ CTは放射線被爆の懸念がある

■ 頻繁に測定できない（検査間隔を短くできない）



真のがんの進行時点と観測されたがんの進行時点に乖離が生じる

Week 10

Week 12

検査間隔の設定

- ヒトを対象とする抗がん剤の評価に関するガイドラインの付録1

<http://cont.o.oo7.jp/36sup26/p275-8.pdf>

- 評価の頻度

- 試験の目的，治療日程および患者の環境を考慮に入れながら，原則試験ごとに判断される必要がある
- できるだけ正確に増悪を評価する必要性と，浸襲的かつ資源集中的な診断法に対する患者の曝露を最小限にする必要性の間でバランスが見出される必要がある

PFS (DFS, TTP) の課題

- 正確なイベント（進行）発生時点は観測されない
- 検査間隔が広いほど、観測される**PFS**は長くなる
 - 測定時点を進行時点とすると、正のバイアスが生じる
- 検査間隔を短くすることは困難
 - 放射線被爆の懸念
 - 主治医の判断で検査が実施される

生存時間データ解析

□ 生存時間を指標あるいは応答（目的）として観察・追及する研究分野の統計解析もしくはその方法論の総体

□ 生存時間

■ ある基点からイベント（注目事象）を発生するまでの時間



■ イベント

死亡，病気の進行・再発，喫煙の再開，製品の故障など

生存時間データ解析の基本的な目標

- 個体群の生存時間分布を決定すること
 - 2群以上においては，生存時間分布を比較すること

生存時間分布の推定と検定

- 生存時間に関係すると考えられる変数が利用可能なときに，生存時間への影響を要約・評価すること

Cox回帰分析

生存時間データの特徴

- 多変量解析の多くの手法では，データの分布に正規分布を仮定することが多い
- 生存時間解析の主要な手法では，順位分布を基礎とした，ランクに基づく統計解析法を使用する
- 生存時間データ
 - 観測値が正である
 - 長期にわたって生存する個体が存在する
 - 「中途打ち切り」という観測値を多く含む
 - 正規分布からかけ離れた（とくに右に）歪んだ分布をもつ

中途打ち切り

□ 無情報な打ち切り

- 生存時間分布の推定に影響を与えない打ち切り

- 観察打ち切り

- 研究計画により、観察期間が事前に規定されているために、イベントが発現する前に観察打ち切りになる場合
- 研究計画により、あるイベント数が観測された時点で研究を終了することが規定されている場合

□ 情報をもつ打ち切り

- 生存時間分布の推定に影響を与える打ち切り

- 競合リスクによる観察打ち切り

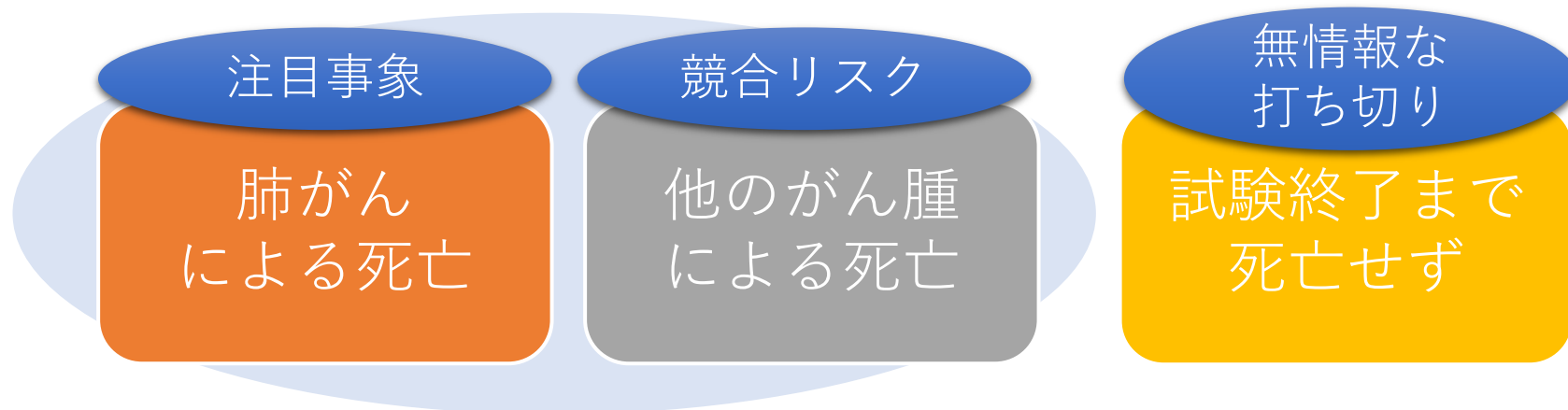
- 注目するイベントよりも先にそれを妨げるイベントが発現し、観察が継続できずに観察打ち切りになる場合

競合リスク

□ 注目事象と排他的関係の事象

- 無情報な打ち切り：将来的に注目事象を起こす可能性がある
- 競合リスク：将来的に注目事象を起こす可能性が**ない**

□ 肺がんによる死亡に注目する臨床試験の例



□ 競合リスクを考慮しないと結果に偏りが生じる

(無情報な) 中途打ち切りの種類

□ 右側打ち切り

- イベント発現までの時間について, 「ある期間以上」ということしか観測できないデータ

□ 左側打ち切り

- イベント発現までの時間について, 観察開始時間が測定できないことにより, 「ある期間以下」ということしか観測できないデータ

□ 区間打ち切りデータ

- イベント発現時点を正確に観測できず, 「ある時点から別のある時点までに発生した」という観察打ち切りになった時間の幅としてイベント発現までの時間が得られるデータ

本講義の諸前提

- 臨床研究では右側打ち切りが典型的であるため、本講義での中途打ち切りは「右側打ち切り」に限定する
- 以降では、右側打ち切りを単に「中途打ち切り」をよぶ
- イベントは「あらゆる原因による死亡」とし、全生存期間（Overall survival; OS）を評価項目とする

中途打ち切り機構

□ T_i^*, C_i を, それぞれ第 i 参加者の真の生存時間と中途打ち切り時間を表す確率変数とする

□ 観測できる生存時間の確率変数 T_i

$$T_i = \min(T_i^*, C_i)$$

□ 打ち切りに関する指示変数 I_i

$$I_i = \begin{cases} 1 & (T_i = T_i^* < C_i; \text{イベント}) \\ 0 & (T_i = C_i \leq T_i^*; \text{中途打ち切り}) \end{cases}$$

□ 最も簡単なデータ集合の確率変数の表記

$$(T_i, I_i) \quad (i = 1, \dots, n)$$

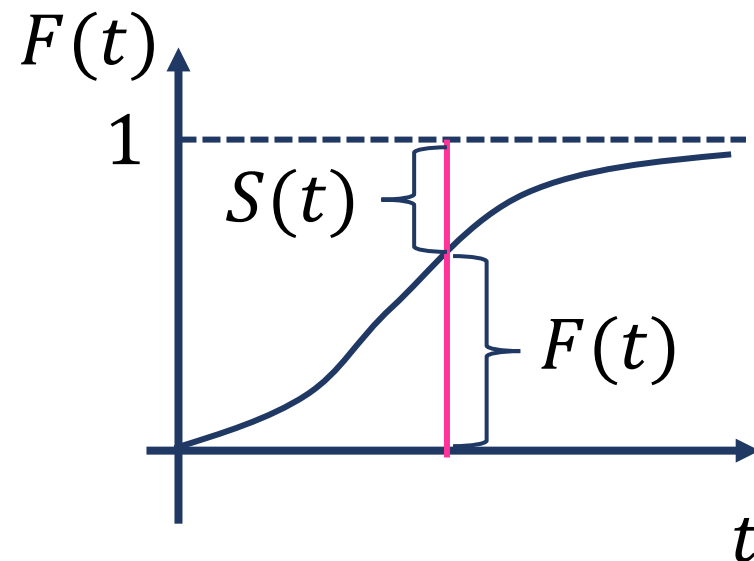
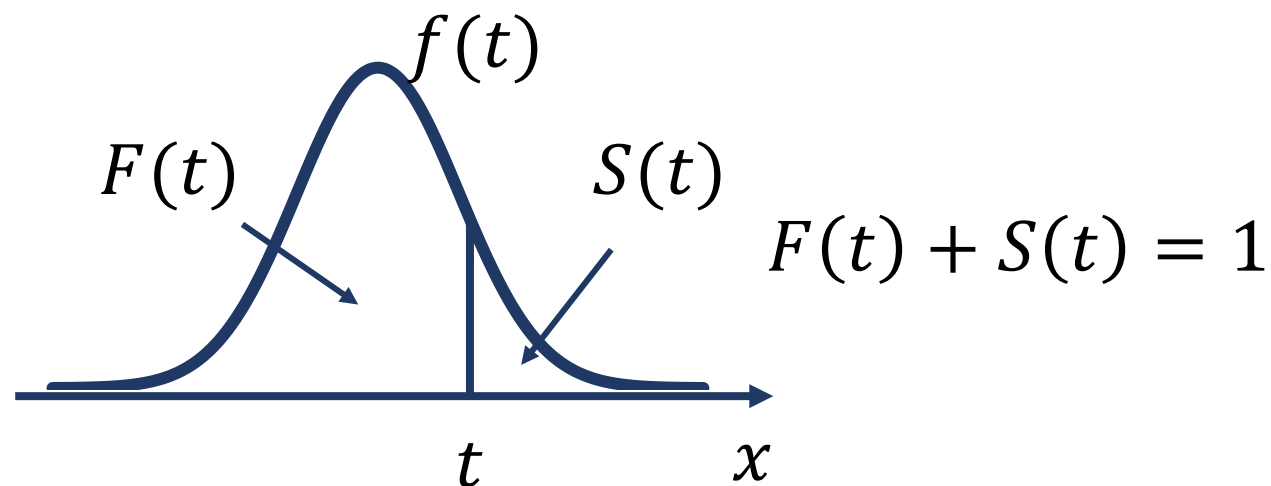
生存時間データの特徴のまとめ

- 生存時間データには中途打ち切り観測値が含まれる
 - 通常の要約統計量である算術平均だと生存時間を過少推定してしまう
- 中途打ち切りデータにも対応した要約統計量を導出するために、生存関数とハザード関数を導入する

生存関数

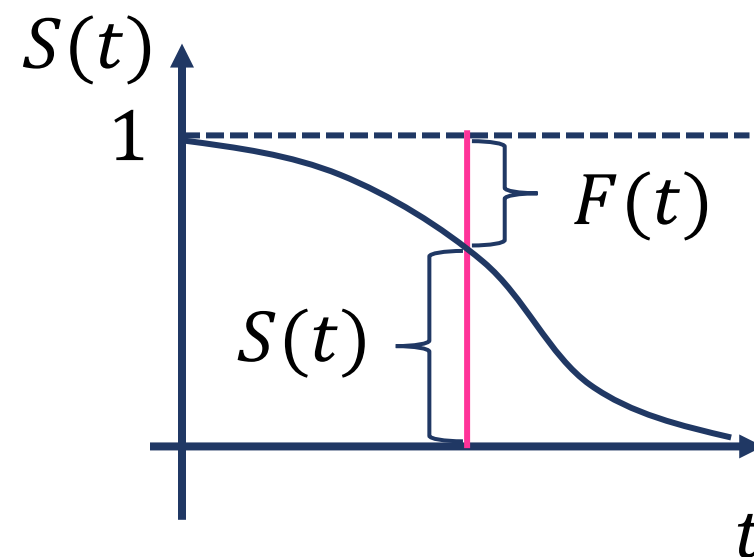
- 個体の生存時間を表す確率変数 T
- 累積分布関数 $F(t) = \Pr(T \leq t) = \int_0^t f(u)du$
 - 確率密度関数 $f(x)$
 - ある個体の生存時間が t 以下の確率
- 生存関数 $S(t) = \Pr(T > t) = 1 - F(t)$
 - 少なくとも時点 t で生存している確率
- $S(t)$ によって表される T の分布 Lifetime distribution
- $F(t)$ によって表される T の分布 Failure time distribution

生存関数と累積分布関数の関係



$$f(t) = \frac{dF(t)}{dt} = \frac{d}{dt}(1 - S(t)) = -\frac{dS(t)}{dt}$$

$$\begin{aligned} f(t) &= \lim_{\delta t \rightarrow 0} \frac{F(t + \delta t) - F(t)}{\delta t} \\ &= \lim_{\delta t \rightarrow 0} \frac{\Pr(t < T \leq t + \delta t)}{\delta t} \end{aligned}$$



ハザード関数

□ 死亡率 $\Pr(t < T \leq t + \delta t \mid T > t)$

■ 時点 t まで生存した個体が $(t, t + \delta]$ で死亡する確率

□ 平均死亡率 $\Pr(t < T \leq t + \delta t \mid T > t) / \delta t$

■ 時点 t まで生存した個体が $(t, t + \delta]$ で死亡する確率の δt 当たりでみた割合

□ ハザード関数 $h(t)$

$$\begin{aligned} h(t) &= \lim_{\delta t \rightarrow 0} \left\{ \frac{\Pr(t < T \leq t + \delta t \mid T > t)}{\delta t} \right\} = \lim_{\delta t \rightarrow 0} \left\{ \frac{\Pr(t < T \leq t + \delta t, T > t)}{\Pr(T > t) \delta t} \right\} \\ &= \frac{1}{\Pr(T > t)} \lim_{\delta t \rightarrow 0} \left\{ \frac{\Pr(t < T \leq t + \delta t)}{\delta t} \right\} = \frac{f(t)}{S(t)} \end{aligned}$$

ハザード関数

- 累積ハザード関数 $H(t) = \int_0^t h(u)du$
- ハザード関数は、確率密度関数、確率関数を1対1変換した量である
- 確率変数 T が離散的な場合と連続的な場合に区別してハザード関数を議論する
 - 生存関数、確率密度関数、確率関数との関係

離散確率変数の場合

□ 離散確率変数 T

■ 離散的に $0(=t_0) < t_1 < t_2 < \dots < t_j$ の値をとり、各時点において確率関数 $p(t_j) = \Pr(T = t_j)$ をもつとする

□ 生存関数 $S(t) = \Pr(T > t) = \sum_{l=j}^J p(t_l) \quad (t_{j-1} \leq t < t_j)$

□ ハザード関数 $h(t)$

$$h(t) = \Pr(T = t_j \mid T > t_{j-1}) = \frac{p(t_j)}{S(t_{j-1})} = \frac{p(t_j)}{\sum_{k=j}^J p(t_k)}$$

□ 累積ハザード関数 $H(t) = \sum_{l=1}^{j-1} h(t_l) \quad (t_{j-1} \leq t < t_j)$

生存関数とハザード関数の関係

$$h(t_j) = \frac{p(t_j)}{S(t_{j-1})} = \frac{S(t_{j-1}) - S(t_j)}{S(t_{j-1})} = 1 - \frac{S(t_j)}{S(t_{j-1})}$$

したがって

$$\begin{aligned} S(t_j) &= S(t_{j-1})(1 - h(t_j)) \\ &= S(t_{j-2})(1 - h(t_j))(1 - h(t_{j-1})) \\ &= \cdots = S(t_0) \prod_{k=1}^j (1 - h(t_k)) = \prod_{k=1}^j (1 - h(t_k)) \end{aligned}$$

カプランマイヤー推定量の形式との関係性を考えるための良い示唆を与える

連続確率変数の場合

□ 生存関数とハザード関数の関係

$$h(t) = \frac{f(t)}{S(t)} = -\frac{d \log S(t)}{dt} \quad \left(\because f(t) = -\frac{dS(t)}{dt}, \frac{d \log S(t)}{dt} = \frac{1}{S(t)} \frac{dS(t)}{dt} \right)$$

$$S(t) = \exp \left(-\int_0^t h(u) du \right) \quad \left(\because \log S(t) = -\int_0^t h(u) du \right)$$

代表的なパラメトリックモデル

- 指数分布
- ワイブル分布
- ガンマ分布
- 対数正規分布
- 対数ロジスティック分布

指数分布

□ 確率密度関数

$$f(t) = \begin{cases} \lambda \exp(-\lambda t) & (t > 0) \\ 0 & (t \leq 0) \end{cases}$$

□ 生存関数

$$S(t) = \int_t^{\infty} \lambda \exp(-\lambda u) du = [-\exp(-\lambda u)]_t^{\infty} = \exp(-\lambda t)$$

□ ハザード関数

$$h(t) = \frac{f(t)}{S(t)} = \frac{\lambda \exp(-\lambda t)}{\exp(-\lambda t)} = \lambda$$

生存時間データの標準的な解析手順

□ 2群（試験群と対照群）の生存関数を比較

■ Step1: 各群の生存関数を推定

- カプランマイヤー法
- 2群生存関数が類似しているか
- 試験群の生存関数が、対照群より完全に上側に存在するか

■ Step2: 観察された生存関数の差が統計的に有意かどうかを確認

- ノンパラメトリック検定
- 比例ハザード性が成り立つ場合、ログランク検定

■ Step3: 治療効果の推定

- Cox比例ハザードモデル（ハザード比の推定）