第1回　生存時間分析（前半）

1　Introduction

生存時間分析は、薬剤投与の効果や疾患による患者の寿命を分析するために用いられる医療統計・生物統計の代表的な手法です。統計検定1級「統計応用」（医薬生物学）でも、ほぼ毎年生存時間分析に関連する問題が出題されています。

この講座では、2回にわたって生存時間分析について解説します。今回は生存時間分析の基本的な用語の確認を行い、生存関数の推定（Kaplan-Meier推定・最尤推定）の紹介、最尤推定に基づく生存関数の差の仮説検定とその導出を解説します。

今回紹介する内容について、統計検定1級「統計応用」（医薬生物学）における過去の出題履歴を確認しましょう。

- 2021年度問1　生存関数のKaplan-Meier推定とその発展
- 2019年度問1　生存関数のKaplan-Meier推定と最尤推定
- 2018年度問1　生存関数の最尤推定と差の仮説検定の導出
- 2014年度問4　生存関数のKaplan-Meier推定と最尤推定
- 2013年度問2　生存関数の最尤推定と差の仮説検定の導出

2　生存時間分析の基礎

薬剤を投与してから効果が切れるまでの時間や、疾患になってから患者が寿命を迎えるまでの時間など、なんらかの基点からイベントが発生するまでの時間のことを**生存時間**といいます。以下では、薬剤の効果を例に生存時間分析に現れるさまざまな概念を解説します。

2.1　生存関数

薬剤を投与してから t 時点後にまだ効果が持続している確率 $S(t) = \mathbb{P}[T > t]$ を**生存関数**といいます。$S(t-) = \mathbb{P}[T \geq t]$ を考えたい場合もありますが、連続型の場合には両者は同じであるため、必要がない限りは同一視して扱います。

生存時間分析の基本的な公式に、生存時間の期待値や分散が生存関数を用いて表せるというものがあります。以下の例題を解いてみましょう。

例題　T を連続型の確率変数とし $T \geq 0$ とします。また生存関数を $S(t)$ と表すものとします。T の期待値と分散が存在するとき、以下の公式が成り立つことを証明してください。

$$\mathbb{E}[T] = \int_0^\infty S(t)dt$$
$$\mathbb{E}[T^2] = 2\int_0^\infty tS(t)dt$$

ただし、T の期待値と分散が存在するなら $\lim_{t\to\infty} tS(t) = 0$ および $\lim_{t\to\infty} t^2S(t) = 0$ が成り立つことは用いてよいものとします。

解答　いずれも部分積分を用いて計算します。期待値から計算してみましょう。以下、$f(t)$ を T の確率密度関数とします。

$$\begin{aligned}\mathbb{E}[T] &= \int_0^\infty tf(t)dt \\ &= [-tS(t)]_0^\infty + \int_0^\infty S(t)dt \\ &= \int_0^\infty S(t)dt\end{aligned}$$

同様に2次モーメントを計算します。

$$\begin{aligned}\mathbb{E}[T^2] &= \int_0^\infty t^2f(t)dt \\ &= \left[-t^2S(t)\right]_0^\infty + 2\int_0^\infty tS(t)dt \\ &= 2\int_0^\infty tS(t)dt\end{aligned}$$

以上で公式を証明することができました。■

なお、生存時間の期待値 $\mathbb{E}[T]$ のことを**平均生存時間**といいます。

2.2　ハザード関数

時刻 t ではまだ投薬の効果が持続している患者さんのうち、時刻 t の瞬間に効果が切れてしまうような患者さんの割合を **ハザード関数** といい、数式で表すと次のようになります。

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{\mathbb{P}[t \leq T < t + \Delta t \mid T \geq t]}{\Delta t}$$

また、時刻 t までのハザードの累計を表す関数を**累積ハザード関数**といい、数式で表すと次のようになります。

$$\Lambda(t) = \int_0^t \lambda(s) ds$$

ハザード関数と累積ハザード関数は、生存時間 T の確率密度関数 $f(t)$ と生存関数 $S(t)$ で表せるという有名な公式があります。以下の例題を確認してみましょう。

例題　T を連続型の確率変数とし、$T \geq 0$ とします。以下の事実を証明してください。

(1)　$\lambda(t) = \dfrac{f(t)}{S(t)}$

(2)　$\Lambda(t) = -\log S(t)$

解答

(1)　条件付き確率の定義から、極限をとっている式の分子は次のように式変形できます。

$$\begin{aligned}
\mathbb{P}[t \leq T < t + \Delta t \mid T \geq t] &= \frac{\mathbb{P}[t \leq T < t + \Delta t, T \geq t]}{\mathbb{P}[T \geq t]} \\
&= \frac{\mathbb{P}[t \leq T < t + \Delta t]}{\mathbb{P}[T \geq t]}
\end{aligned}$$

あとは確率密度関数の定義 $f(t) = \lim_{\Delta t \to 0} \dfrac{\mathbb{P}[t \leq T < t + \Delta t]}{\Delta t}$ と生存関数の定義 $S(t) = \mathbb{P}[T \geq t]$ から