2024 年度春学期　応用数学（解析）　第１０回

第３部・微分方程式に関する話題／　生存時間分布と半減期

ここまで微分方程式のお話をしましたが，テクニカルなことが多く，また「勘」などという内容もあって，数学としては正直あまりおもしろくなかったのではないかと思います。そこで，今回と次回は，微分方程式の実際例への応用についてお話しします。今回は寿命を扱う微分方程式，次回は振動を扱う微分方程式についてです。

生存時間分布とワイブル分布モデル

日本人の平均寿命は世界のトップレベルだそうですが，それはあくまで平均であって，実際には長生きする人も若くして亡くなる人もいます。機械でも同様で，同じ型の機械でも個体によって長持ちするものも早く壊れるものもあります。つまり，「寿命」は偶然のために長短がばらばらになるもので，これを「寿命は**確率変数**である」といいます。では，どのように長短ばらばらか，どのくらいの寿命である確率がいくらくらいかを考えてみましょう。これを寿命の**確率分布**といいます。

寿命，すなわち時刻 0 に誕生した人が死亡する時刻は，確率変数であり，これを T で表します。いま，次のような関数 $l(t)$ を定義します。

$$l(t) = \lim_{\Delta \to 0} \frac{1}{\Delta} P(t < T < t + \Delta | T > t). \tag{1}$$

この関数のうち，条件付き確率 $P(t < T < t + \Delta | T > t)$ は，「$T > t$ であるとわかっているときに，$t < T < t + \Delta$ である確率」，すなわち「時刻 t までは確かに生存している人が，時刻 t 以後 Δ 時間の間に死亡する確率」を意味します[1]。ですから，それを Δ で割ったものは，t 以後の時間あたりの死亡確率密度であり，その $\Delta \to 0$ の極限は「時刻 t までは確かに生存していた人の，次の瞬間の死亡確率密度」ということになります。つまり，この関数 $l(t)$ は「時刻 t までは確かに生存している人が，次の瞬間に死亡する危険の度合」で，これを**ハザード関数**といいます。

さて，この確率変数 T に対して，次の**累積分布関数** $F(t)$ を考えます。

$$F(t) = P(T \leq t) \tag{2}$$

$F(t)$ は「寿命が t 以下である確率」です。さらに

$$S(t) = 1 - F(t) = P(T > t) \tag{3}$$

は「時刻 t の時点でまだ生きている確率」であり，これを**生存関数**と呼びます。ハザード関数が瞬間瞬間での死亡の危険を表現しているのに対して，累積分布関数 $F(t)$ や生存関数 $S(t)$ は「あるひとつの個体が，時刻 t まで時間が過ぎたとき，すでに死んでいる／まだ生きている確率」を表現しています。これを，独立に死亡する十分多くの個体の集団について考えると，大数の法則によって，累積分布関数 $F(t)$ や生存関数 $S(t)$ は「時刻 t まで時間が過ぎたとき，すでに死んでいる／まだ生きている個体の割合」と同等と見なすこともできます。

では，確率変数 T の累積分布関数や生存関数と，ハザード関数の関係を見てみましょう。T の確率密

[1] くわしくは，浅野の講義「統計学」第９回を参照してください。なお，$P()$ は「カッコ内のことが起きる確率」の意味です。

度関数を $f(t)=F'(t)$ として，これまでに出てきた式を用いると

$$\begin{aligned}
l(t) &= \lim_{\Delta\to 0}\frac{1}{\Delta}P(t<T<t+\Delta|T>t) \\
\text{（条件付き確率の定義より）} &= \lim_{\Delta\to 0}\frac{1}{\Delta}\cdot\frac{P\{(t<T<t+\Delta)\text{ and }(T>t)\}}{P(T>t)} \\
\text{（}T>t\text{ は }t<T<t+\Delta\text{に含まれるので）} &= \lim_{\Delta\to 0}\frac{1}{\Delta}\cdot\frac{P(t<T<t+\Delta)}{P(T>t)} \\
\text{（(2) 式と累積分布関数の定義より）} &= \frac{1}{P(T>t)}\lim_{\Delta\to 0}\frac{F(t+\Delta)-F(t)}{\Delta} \\
\text{（微分の定義より）} &= \frac{1}{P(T>t)}F'(t) \\
\text{（確率密度関数の定義と (3) 式より）} &= \frac{f(t)}{S(t)}
\end{aligned} \tag{4}$$

が得られます。ここで，(3) 式より

$$S'(t)=(1-F(t))'=-F'(t)=-f(t) \tag{5}$$

ですから，これを (4) 式に代入すると

$$l(t)=-\frac{S'(t)}{S(t)} \tag{6}$$

という微分方程式が得られます。ここで，時刻 0，すなわち誕生の瞬間に生存している確率は 1 ですから，$S(0)=1$ です。これを初期条件として (6) 式の微分方程式を解くと，

$$\begin{aligned}
l(t) &= -\frac{S'(t)}{S(t)} \\
&= -\frac{d}{dt}(\log S(t)) \\
-\int_0^t l(u)du &= \log S(t)+C \qquad \text{（}C\text{ は積分定数）} \\
S(t) &= \exp\left(-\int_0^t l(u)du\right) \qquad \text{（}t=0\text{ のとき }S(0)=1\text{ なので，}C=0\text{）}
\end{aligned} \tag{7}$$

のように，ハザード関数と生存関数との関係が得られます。

さて，ここでハザード関数 $l(t)$ を

$$l(t)=\lambda p(\lambda t)^{p-1} \tag{8}$$

とおいて（λ, p は正の実数），(7) 式に代入すると，

$$\begin{aligned}
S(t) &= \exp\left(-\int_0^t \lambda p(\lambda u)^{p-1}du\right) \\
&= \exp\left(-\left[(\lambda u)^p\right]_{u=0}^{u=t}\right)=\exp\left(-(\lambda t)^p\right) \\
F(t) &= 1-S(t)=1-\exp\left(-(\lambda t)^p\right)
\end{aligned} \tag{9}$$

のような累積分布関数 $F(t)$ が得られます。このような累積分布関数をもつ確率分布を**ワイブル分布**とよびます。

ワイブル分布の累積分布関数には λ と p の 2 つのパラメータがあります。λ が大きいとハザード関数が全ての時刻で大きくなるわけですから，全ての時刻を通じて死亡する危険が大きくなり，全体として寿命が短いことになります。一方，p については，

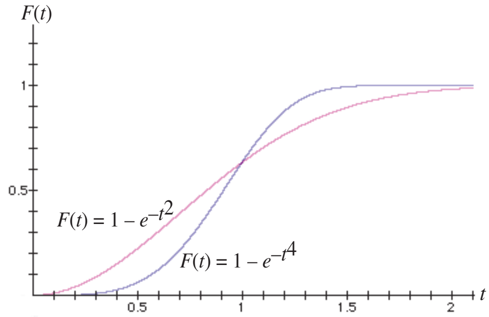

図 1: ワイブル分布における，パラメータと累積分布関数

- $p > 1$ のときは (8) 式右辺の指数部が正になり，ハザード関数は単調増加になります。これは，誕生から時間がたつにつれて，死亡する危険（すなわち，死亡率や故障率）が大きくなることを示しています。例えば，機械を使い続けていると，あちこちにガタが来て，だんだんと壊れやすくなるという場合（**摩耗故障**）がこれに相当します。
- $0 < p < 1$ のときは，時間がたつにつれて死亡する危険が減ることになります。これは，機械で言えば，初期不良が多く，時間がたつと安定してきて故障が減少してゆくという場合（**初期故障**）に相当します。

図 1 は，時間 t と，累積分布関数 $F(t)$ つまり「寿命が t 以下である確率」を， $\lambda = 1, p = 2$ の場合と，$\lambda = 1, p = 4$ の場合について表したものです。どちらも単調増加で，機械でいえば，時間がたつにつれて壊れやすくなる摩耗故障の状況を表しています。しかし， $p = 4$ の場合は $p = 2$ の場合にくらべて，はじめは故障が少ないですが途中から急激に故障が増加することがわかります。

ワイブル分布は，実際の寿命データにワイブル分布を仮定し，λ と p を推測するという方法によって，寿命データ解析に広く用いられています。これを簡便に行うのが，**ワイブルプロット**とよばれるものです。これは，製品の寿命を解析するため，その製品のたくさんの個体の耐久試験を行なって，横軸に時刻 t の対数，縦軸に「『時刻 t の時点で故障せずに稼働している個体の割合（すなわち $S(t)$）の逆数』の，対数の対数」をプロットし，それらの点の並びを近似する直線の傾きを求めると，パラメータ p がわかるというものです。

このようなことがなぜできるのかは，次の計算をしてみるとわかります。(9) 式から

$$\frac{1}{S(t)} = \exp\left((\lambda t)^p\right) \tag{10}$$

となりますから，この式の両辺の対数を 2 回とると

$$\begin{aligned}\log\left\{\log\left(\frac{1}{S(t)}\right)\right\} &= \log\left\{\log\left(\exp\left((\lambda t)^p\right)\right)\right\} \\ &= \log\left\{(\lambda t)^p\right\} \\ &= p(\log t + \log\lambda)\end{aligned} \tag{11}$$

となります。そこで，$Y = \log(\log(1/S(t))), X = \log t$ とおくと (11) 式は $Y = p(X + \log\lambda)$ という直線の式となり，その傾きがパラメータ p となります。ワイブルプロットを簡単に描けるように，縦横の軸に $S(t)$ と t を上の Y, X のように変換した目盛りを刻んだ，「ワイブル確率紙」も広く用いられています。

指数分布

ワイブル分布で $p=1$ の場合を考えると，(8) 式からハザード関数は $l(t)=\lambda$ と時刻に依存しない定数になりますから，「誕生からいくら時間がたっても，死亡する危険は一定」「機械の故障率は，時間に関係なく一定（**偶発故障**）」ということになります。このとき (9) 式から

$$F(t)=1-e^{-\lambda t} \tag{12}$$

となります。このような累積分布関数をもつ確率分布を**指数分布**とよびます。指数分布は，「どの時刻においても，その時点で生存している個体のうちその後一定時間内に死亡するものの割合は，常に同じ」という場合の寿命の分布を表します。

この例としては，原子核の崩壊のモデルが有名です。原子核は，どの時刻においても，その時点で存在する原子核のうち一定の割合が崩壊すると考えられているので，ある時刻に残っている原子核の割合，すなわち生存関数 $S(t)$ は指数分布で表現でき，(12) 式から

$$S(t)=e^{-\lambda t} \tag{13}$$

となります。

ここで，時刻 t に存在した原子核の数が半分になる時刻 t' を考えてみましょう。このとき

$$S(t')=\frac{1}{2}S(t) \tag{14}$$

がなりたちますから，(13) 式から

$$e^{-\lambda t'}=\frac{1}{2}e^{-\lambda t} \tag{15}$$

となり，両辺の対数をとると

$$\begin{aligned} -\lambda t' &= -\log 2-\lambda t \\ t'-t &= \frac{\log 2}{\lambda} \end{aligned} \tag{16}$$

となります。このことは，$t'-t$ すなわち「時刻 t に存在する原子核の数が半分になるまでの時間」は，t によらず常に一定であることを意味します。この時間を**半減期**とよびます。

問題

ある原子核は指数分布にしたがって崩壊し，半減期が 2 年であるとします。この原子核が多数あるとき，いまから 1 年以内に崩壊する原子核の割合はいくらですか。

参考文献

柳川堯，統計数学，近代科学社，ISBN978-4764-910140, 1990.