

# 生命エネルギーモデルを用いた死亡率予測

南優希

早稲田大学大学院基幹理工学研究科  
数学応用数理専攻修士1年

清水 泰隆 (早稲田大学) との共同研究

2019年10月26日  
研究集会「Demographic Change」@早稲田大学

- 生命エネルギーモデルとは
- 非斉次的拡散による SEM
- モデル設定とパラメータ推定
- 死亡率予測
- まとめと今後の課題

# モデル考案の背景

現在、死亡率予測モデルとしては、Lee-Carter モデルが国際的にも標準的な方法とされている。

## Lee-Carter モデル (Lee and Carter, [1])

年齢  $x = 1, \dots, K$ , 時刻  $t = 1, \dots, T$  としたときの中央死亡率  $m_{x,t}$  に対して

$$\log m_{x,t} = \alpha_x + \beta_x \kappa_t + \epsilon_{x,t}.$$

$\alpha_x, \beta_x, \kappa_t$  : パラメータ,  $\epsilon_{x,t}$  : 誤差項

## Lee-Carter モデルの問題点

- パラメータの解釈がしにくい
- 推定したパラメータが一致性を満たさない (Leng and Peng, [2])
- 生まれ年毎の累積死亡率の追跡が面倒である

# 生命エネルギーモデル (Survival Energy Model, SEM)

## 生命エネルギーモデル (伊藤, 清水 [3])

人間に生命エネルギー (Survival Energy, SE) なるものが存在すると仮定し, 誕生から死亡までの SE の推移を, コホート  $c$  毎に確率過程  $\mathbf{X}^c = (X_t^c)_{t \geq 0}$  によってモデリングする.

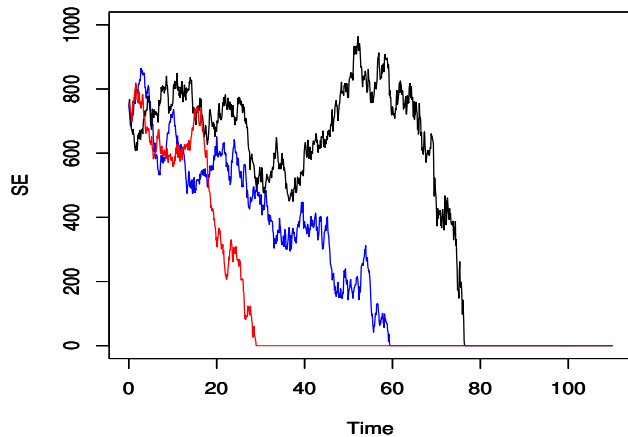
- コホート  $c$  は生まれ年を意味する
- 初期エネルギー  $X_0^c = x_c$  はコホートごとに等しい

## 死亡時刻と死亡率関数

死亡時刻 :  $\tau^c := \inf\{t > 0 : X_t^c < 0\}$

死亡率関数 :  $q_c(t) := \mathbb{P}(\tau^c \leq t), \quad t > 0$

# 生命エネルギーのイメージ図



## ドリフト付きブラウン運動による SEM

$$X_t^c := x_c + \mu_c t + \sigma_c W_t$$

- $x_c > 0$  : 初期エネルギー
- $\mu_c \in \mathbb{R}$  : ドリフト係数,  $\sigma_c > 0$  : 拡散係数
- $W_t$  : 標準ブラウン運動

$$q_c(t) = 1 - \Phi\left(\frac{x_c + \mu_c t}{\sigma_c \sqrt{t}}\right) + e^{-\frac{2\mu_c x_c}{\sigma_c^2}} \Phi\left(\frac{-x_c + \mu_c t}{\sigma_c \sqrt{t}}\right), \quad t > 0$$

$\Phi$  : 標準正規分布の分布関数

$$\text{平均寿命} : \mathbb{E}[\tau^c] = \int_0^\infty [1 - q_c(t)] dt = \begin{cases} \frac{x_c}{|\mu_c|} & (\mu_c < 0) \\ \infty & (\mu_c \geq 0) \end{cases}$$

$$\text{終身保険} : \bar{A}_x = \int_0^\infty e^{-\delta t} \frac{q'_c(t)}{1 - q_c(t)} dt$$

# 非齊次的拡散過程による SEM

## 非齊次的拡散過程 SEM

時刻  $t \geq 0$  における生命エネルギー  $\mathbf{X}(\theta_c) = (X_t^c)_{t \geq 0}$  が,

$$X_t^c := x_c + \int_0^t U(s, \mu_c) ds + \int_0^t V(s, \sigma_c) dW_s$$

という確率微分方程式に従う.

- $U(s, \mu_c), V(s, \sigma_c) : \mathbb{R}_+ \times \mathbb{R} \rightarrow \mathbb{R}$
- $\theta_c = (x_c, \mu_c, \sigma_c)$

# 非斉次的拡散過程 SEM における死亡率関数

定理 1 (Molini *et al.*, [4])

$M(t, \mu_c) = \int_0^t U(s, \mu_c) ds$ ,  $S(t, \sigma_c) = \frac{1}{2} \int_0^t V^2(s, \sigma_c) ds$  に対して

$$\frac{M(t, \mu_c)}{S(t, \sigma_c)} \equiv \kappa_c, \quad \forall t \geq 0$$

の関係を仮定すると,

$$\begin{aligned} q(t, \theta_c) &:= \mathbb{P}(\tau^c \leq t | \theta_c) \\ &= 1 - \Phi\left(\frac{x_c + M(t, \mu_c)}{\sqrt{2S(t, \sigma_c)}}\right) + \exp(-\kappa_c x_c) \Phi\left(\frac{-x_c + M(t, \mu_c)}{\sqrt{2S(t, \sigma_c)}}\right) \end{aligned}$$

パラメトリック族:  $\mathcal{P}_\Theta = \{q(t, \theta) | \theta \in \Theta\}$ ,  $\Theta \subset \mathbb{R}^m$



# 条件付き経験累積死亡率

Human Mortality Database [5] から男女合計の生命表を入手  
→コホート  $c$  毎の死亡率データになるように加工

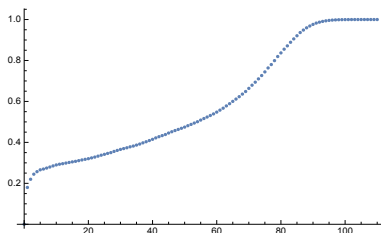


Figure: スウェーデンの 1830 年生まれコホートの経験累積死亡率推移

- 幼少期，青年期における死亡率は不安定
- 高齢期にさしかかると死亡率が急激に増加

# 非斉次的拡散 SEM のパラメトリックモデル

$T > S > 0$  として,  $S$  歳まで生存した部分集団の SE 過程を考え,  $T$  歳でパラメータが変化するとし, 時間依存ドリフト係数として次の 3 つのケースを考える.

①  $U(t, \mu_c) = \alpha_c$

②  $U(t, \mu_c) = \alpha_c + \beta_c(t - T)^{\gamma_c} \mathbf{1}_{\{t > T\}}$

③  $U(t, \mu_c) = \alpha_c + \beta_c \exp(\gamma_c(t - T)) \mathbf{1}_{\{t > T\}}$

$\alpha_c < 0, \beta_c < 0, \gamma_c > 0$  とする.

時間依存拡散係数は,  $\kappa_c < 0$  として

$$V^2(t, \sigma_c) = \frac{2}{\kappa_c} U(t, \mu_c), \quad \forall t \geq 0.$$

# パラメータの推定

- 真の条件付死亡率 :  $q_c(t|S) := \mathbb{P}(\tau^c \leq t | \tau^c > S)$
- 経験条件付死亡率 :  $\hat{q}_c(t|S) := \frac{\frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{S < \tau_i^c \leq t\}}}{\frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{\tau_i^c > S\}}}$

$$\sup_{t>0} |\hat{q}_c(t|S) - q_c(t|S)| \xrightarrow{P} 0, \quad n \rightarrow \infty \quad (\text{Glivenko - Cantelli})$$

- パラメトリック族 :

$$\mathcal{P}_\Theta = \{q(t, \theta) | \theta = (x, \alpha, \beta, \gamma, \kappa) \in \Theta\}, \quad \Theta \subset \mathbb{R}^5: \text{有界開集合}$$

- 真値の存在を仮定 :

$$\exists \theta_0 \in \Theta \quad \text{s.t.} \quad q(t, \theta_0) = q_c(t), \quad \text{a.e.}$$

## 最小二乗推定量 (LSE)

$0 < t_0 < t_1 < \dots < t_d$  に対して,

$$\hat{\theta}_n := \arg \min_{\theta \in \Theta} \sum_{i=1}^d |q(t_i, \theta | S) - \hat{q}_c(t_i | S)|^2$$

ただし,  $q(t, \theta | S) := \frac{q(t, \theta) - q(S, \theta)}{1 - q(S, \theta)}$

- 変化点  $T$  以前で  $x_c, \alpha_c, \kappa_c$  を,  $T$  以降で  $\beta_c, \gamma_c$  を推定
- $x_c, \alpha_c, \kappa_c$  は識別性がないため,  $x_c$  の値を固定

## 定理 2

$q(t, \cdot)$  は各  $t$  に対して  $C^1(\Theta)$  級であるとする. さらに, 次の識別性条件を仮定する:

$$q(t_i, \theta | S) = q(t_i, \theta_0 | S) \text{ for } i = 1, 2, \dots, d \Rightarrow \theta = \theta_0.$$

このとき, LSE  $\hat{\theta}_n$  は弱一貫性を持つ:

$$\hat{\theta}_n \xrightarrow{P} \theta_0, \quad n \rightarrow \infty.$$

## 定理 3

$q(t, \cdot)$  は各  $t$  に対して  $C^2(\Theta)$  級であり、識別性があるとする。さらに、 $\Theta$  に凸性を仮定し、 $m = \dim(\Theta)$  とする。このとき、

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} R_d^{-1} Q_d \cdot N_d(0, \Sigma), \quad n \rightarrow \infty$$

が成り立つ。ただし、

$$Q_d = (\partial_{\theta} q(t_1, \theta | S), \dots, \partial_{\theta} q(t_d, \theta | S)) \in \mathbb{R}^m \otimes \mathbb{R}^d,$$

$$R_d = \left( \sum_{i=1}^N \partial_{\theta_i} \partial_{\theta_j} q(t_i, \theta | S) \right)_{1 \leq i, j \leq m} \in \mathbb{R}^m \otimes \mathbb{R}^m,$$

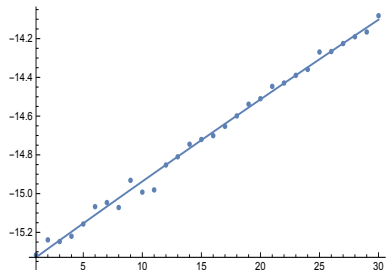
$$\Sigma = (p_c(t_i \wedge t_j | S) - p_c(t_i | S)p_c(t_j | S))_{1 \leq i, j \leq d}, \quad p_c(\cdot) := 1 - q_c(\cdot).$$

# オランダにおける死亡率予測

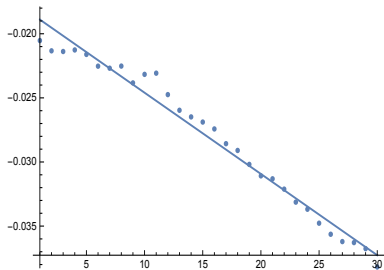
1841 年生まれから 1870 年生まれの死亡率データを用いて、1880 年生まれ、1890 年生まれ、1900 年生まれの死亡率予測を行った。

- ① 30 年分の死亡率データのパラメータ推定
  - スタート年齢は  $S = 20$  とし、変化点は  $T = 60$  とする
  - 初期エネルギーは  $x_c = 1000$  とする
  - 制約条件として、 $\alpha_c < 0$ ,  $\beta_c < 0$ ,  $\gamma_c > 0$ ,  $\kappa_c < 0$  とする
- ② 将来死亡率関数のパラメータ予測
  - (非) 線形回帰によって予測

# 推定したパラメータの推移:case1



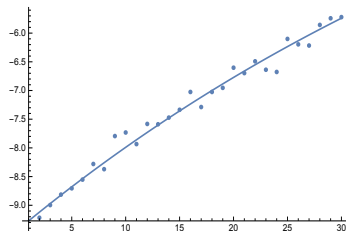
(a)  $\alpha = -15.373 \exp(-0.0028758t)$



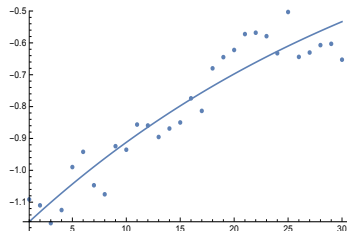
(b)  $\kappa = -0.000633696t - 0.0182603$



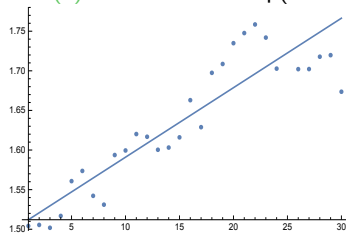
# 推定したパラメータの推移:case2



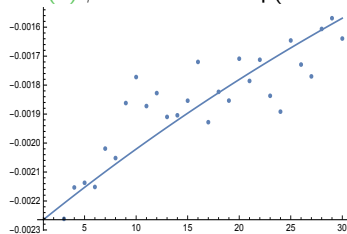
(a)  $\alpha = -9.42654 \exp(-0.016534t)$



(b)  $\beta = -1.19305 \exp(-0.0268324t)$

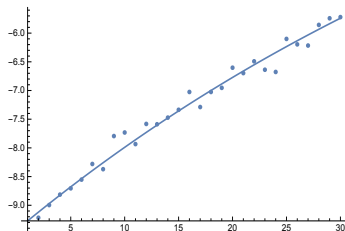


(c)  $\gamma = 0.00877223t + 1.50324$

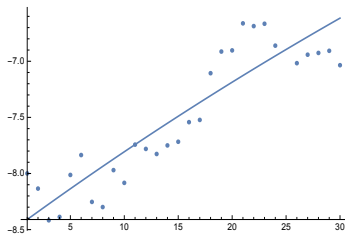


(d)  $\kappa = -0.00229343 \exp(-0.0126517t)$

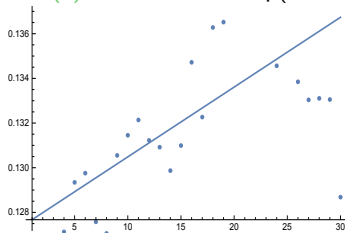
# 推定したパラメータの推移:case3



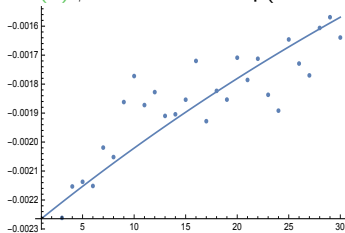
(a)  $\alpha = -9.42654 \exp(-0.016534t)$



(b)  $\beta = -8.47987 \exp(-0.00827468t)$

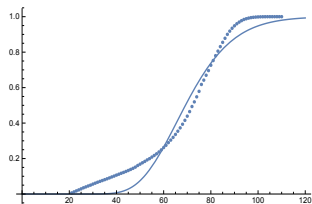


(c)  $\gamma = 0.000312531t + 0.12736$

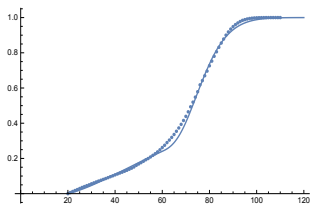


(d)  $\kappa = -0.00229343 \exp(-0.0126517t)$

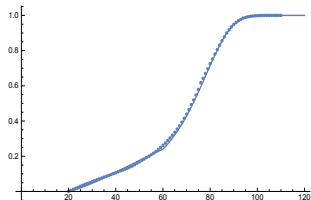
# 推定死亡率関数と実データとの比較



(a) case1



(b) case2



(c) case3

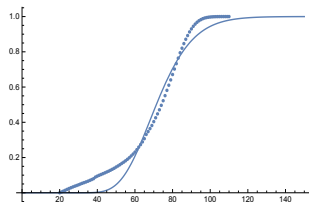
Figure: 1870 年生まれコホートの推定死亡率関数 (実線) と実データ (点線)

# 推定死亡率関数と実データとの比較

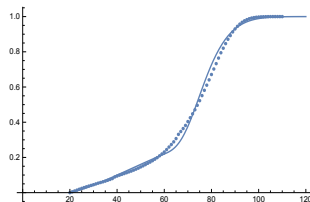
Table: 推定死亡率関数と実データとの誤差の二乗和の平均

ケース 1	ケース 2	ケース 3
0.427944	0.0220319	0.00671647

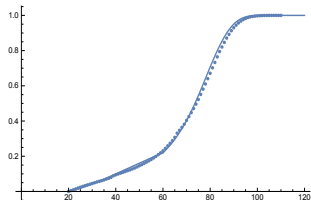
# 死亡率予測の結果：10年後



(a) case1



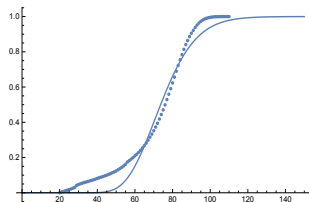
(b) case2



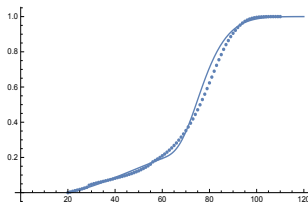
(c) case3

Figure: 1880 年生まれコホートの予測死亡率関数 (実線) と実データ (点線)

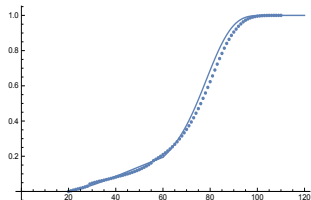
# 死亡率予測の結果：20年後



(a) case1



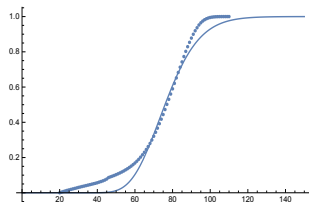
(b) case2



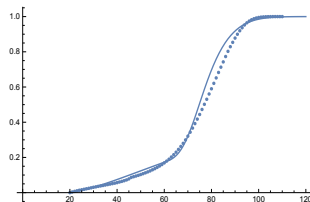
(c) case3

Figure: 1890 年生まれコホートの予測死亡率関数 (実線) と実データ (点線)

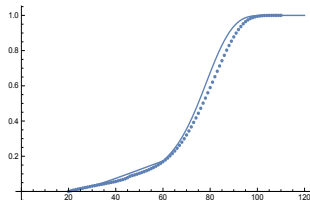
# 死亡率予測の結果：30年後



(a) case1



(b) case2



(c) case3

Figure: 1900 年生まれコホートの予測死亡率関数 (実線) と実データ (点線)

Table: 予測死亡率関数と実データとの誤差の二乗和

	ケース 1	ケース 2	ケース 3
10 年後	0.330259	0.0433915	0.0201326
20 年後	0.329853	0.0822591	0.0653867
30 年後	0.256442	0.135588	0.125765

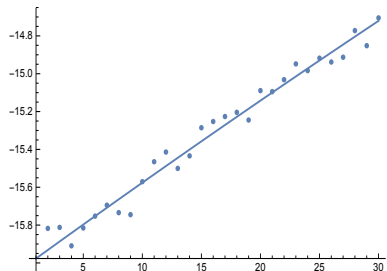


# スウェーデンにおける死亡率予測

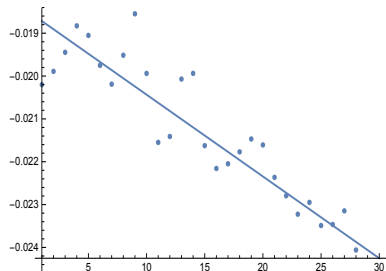
1801 年生まれから 1830 年生まれの死亡率データを用いて, 1840 年生まれ, 1850 年生まれ, ..., 1900 年生まれの死亡率予測を行った.

- ① 30 年分の死亡率データのパラメータ推定
  - スタート年齢は  $S = 20$  とし, 変化点は  $T = 50$  とする
  - 初期エネルギーは  $x_c = 1000$  とする
  - 制約条件として,  $\alpha_c < 0$ ,  $\beta_c < 0$ ,  $\gamma_c > 0$ ,  $\kappa_c < 0$  とする
- ② 将来死亡率関数のパラメータ予測
  - (非) 線形回帰によって予測

# 推定したパラメータの推移:case1

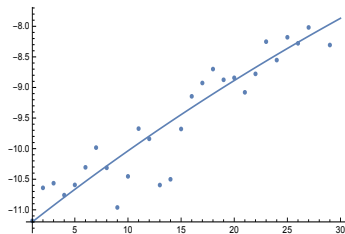


(a)  $\alpha = -16.0218 \exp(-0.00282317t)$

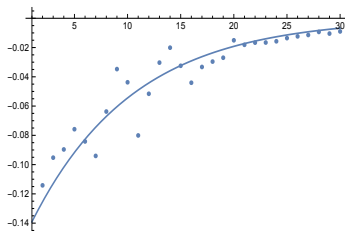


(b)  $\kappa = -0.000190898t - 0.0185255$

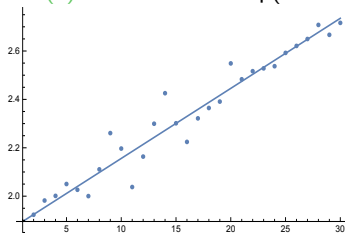
# 推定したパラメータの推移:case2



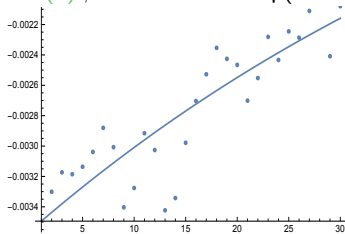
(a)  $\alpha = -11.3345 \exp(-0.0121655t)$



(b)  $\beta = -0.154447 \exp(-0.104273t)$

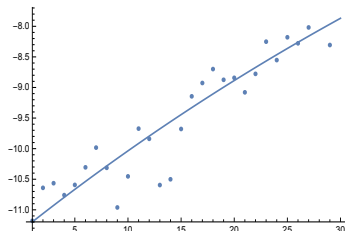


(c)  $\gamma = 0.0289585t + 1.86641$

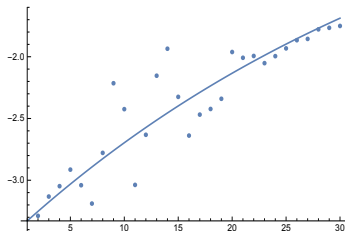


(d)  $\kappa = -0.00355286 \exp(-0.0166043t)$

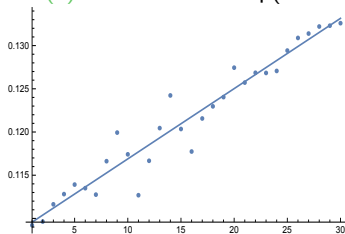
# 推定したパラメータの推移:case3



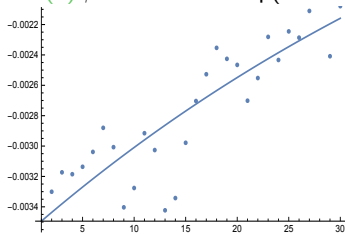
(a)  $\alpha = -11.3345 \exp(-0.0121655t)$



(b)  $\beta = -3.40801 \exp(-0.0234189t)$

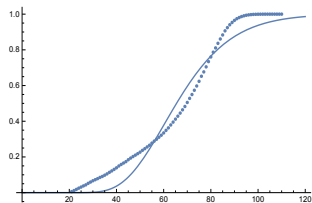


(c)  $\gamma = 0.000813139t + 0.108758$

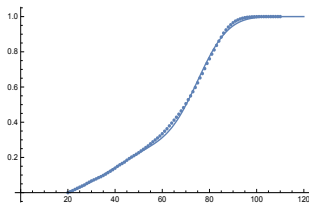


(d)  $\kappa = -0.00355286 \exp(-0.0166043t)$

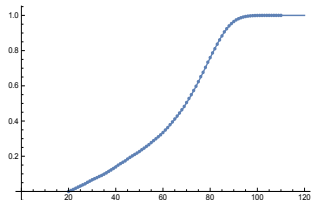
# 推定死亡率関数と実データとの比較



(a) case1



(b) case2



(c) case3

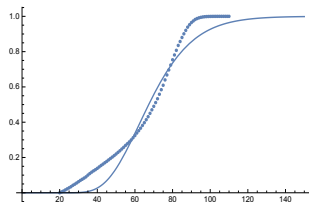
Figure: 1830 年生まれコホートの推定死亡率関数 (実線) と実データ (点線)

# 推定死亡率関数と実データとの比較

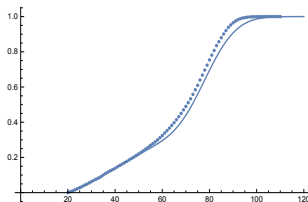
Table: 推定死亡率関数と実データとの誤差の二乗和の平均

ケース 1	ケース 2	ケース 3
0.347101	0.0103527	0.000823361

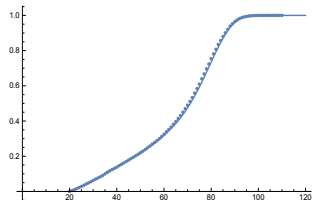
# 死亡率予測の結果：10年後



(a) case1



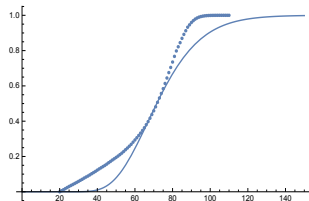
(b) case2



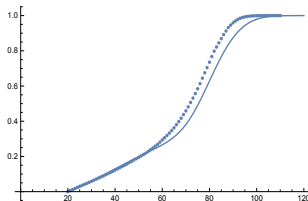
(c) case3

Figure: 1840 年生まれコホートの予測死亡率関数 (実線) と実データ (点線)

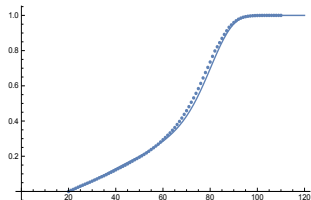
# 死亡率予測の結果：30年後



(a) case1



(b) case2

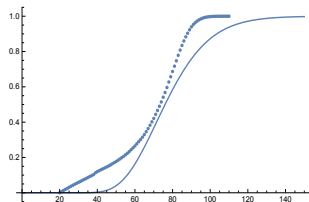


(c) case3

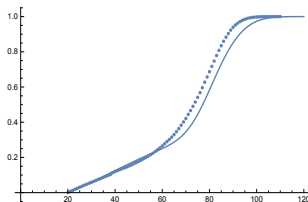
Figure: 1890 年生まれコホートの予測死亡率関数 (実線) と実データ (点線)



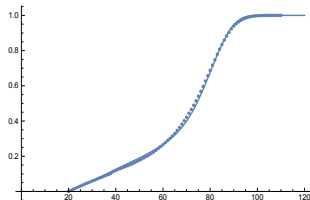
# 死亡率予測の結果：50年後



(a) case1



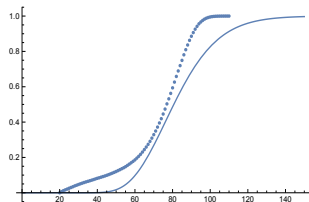
(b) case2



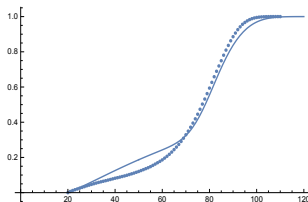
(c) case3

Figure: 1900 年生まれコホートの予測死亡率関数 (実線) と実データ (点線)

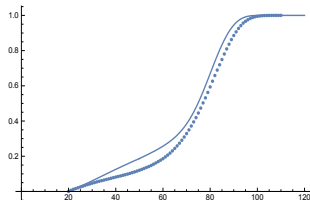
# 死亡率予測の結果：70年後



(a) case1



(b) case2



(c) case3

Figure: 1900 年生まれコホートの予測死亡率関数 (実線) と実データ (点線)

# 死亡率予測の結果

Table: 予測死亡率関数と実データとの誤差の二乗和

	ケース 1	ケース 2	ケース 3
10 年後	0.457803	0.101353	0.0064886
30 年後	0.670725	0.257575	0.0257175
50 年後	1.0998	0.254276	0.00915939
70 年後	1.1654	0.143401	0.239306

# まとめと今後の課題

- SEM は LCM に比べ以下の点で優れている
  - パラメータの解釈が容易である
  - コホート毎の累積死亡率を追いやすい
  - 推定したパラメータに一致性や漸近正規性がある
- データが十分にあれば、非常に高い精度での死亡率予測が可能
- モデルの拡張性が高い
  
- データが少ない国での死亡率予測
- 男女別の死亡率予測
- コホートや変化点を取り入れた死亡率モデルの先行研究との比較

- [1] Lee, R. D. and Carter, L. (1992). Modeling and Forecasting U.S.Mortality. *Journal of the American Statistical Association*, Vol.87, No.419, 659-675.
- [2] Leng, X and Peng, L. (2016). Inference pitfalls in Lee-Carter model for forecasting mortality. *Insurance: Mathematics and Economics*, **70**, 58-65.
- [3] 伊藤龍之介, 清水泰隆. (2019). 生命エネルギー仮説に基づく構造アプローチとコホート別死亡率推定. JARIP 会報 大会プロシーディングス特集号, Vol.6, 17-30, 2019.3.
- [4] Molini, A.; Talkner, P.; Katul, G. G. and Porporato, A. (2011). First passage time statistics of Brownian motion with purely time dependent drift and diffusion. *Physica A* 390 1841-1852.
- [5] Human Mortality Database: <https://www.mortality.org/>