

経済価値の保険数理と機械学習

明治大学 総合数理学部

松山直樹

機械学習とアクチュアリー学

武蔵野大学OLIS保険フォーラム

2023.6.10

武蔵野大学有明キャンパス

1. 経済価値ベースの保険数理が求められる背景

伝統的保険数理の前提

- 保険料の計算基礎率は契約時点で固定され保険料は払込期間を通じ不変
- 責任準備金(保険負債の価値)の**計算基礎率は契約時点で固定** (解約返戻金は責任準備金をもとに設定され契約時点で将来の値が確定)
- 代表的な基礎率の特徴
 - **予定死亡率** (標準死亡率) : **年齢構造を持つが時間変化しない**
 - **予定利率** (標準利率) : **期間構造を持たず時間変化しない** ⇒ **金融自由化前の考え方**
- 全て予定基礎率通りに推移すれば、解約時に責任準備金相当を支払うことで**解約は収支に影響ないと仮定**して、一般には解約率は用いない
- **責任準備金では計算基礎率から乖離するリスクを担保できない**ので別途**ソルベンシーマージン** (リスク吸収のための自己資本。VaRに基づく係数から必要量を算出) を積立。
 - $\text{ソルベンシーマージン比率} = 100 \times \text{自己資本等(マージン)} \div \text{リスク総額の半分}$
 - リスク総額は必要資本量 (バリュアットリスク) に基づく係数を用いて計測
 - 法令上200%以上が要請

金融自由化の進展

- 1970年代の国債大量発行前は、金融機関のシンジケート団が発行された国債を全額引き受け、証券会社以外の金融機関は手持ち国債の売却が禁止され、発行後1年以上経過した国債は日銀が買い取ることで、国債の利回りは政府・日銀の規制下にあった
- 1970年代の国債の多量発行で、銀行の負担が大きくなった結果、1977年に発行後1年を経過した国債の市中売買が認められ、国債の流通市場が誕生し、既発国債は自由金利商品となった
- 金融自由化の進展は欧米の方が早く、自由金利商品の拡大で保険会社のよ
うな中間金融機関から顧客の資金が引き出されMMF等の自由金利商品に
向かうディスインターミディエーション（非中間化）が発生し、米国では
多くの保険会社が苦境に陥った
- 「**生命保険業は1980年時点では150年の伝統を誇っていたが、1990年時
点ではたったの10年でしかない**」（Richard M.Todd and Neil Wallace,
FRB-Minneapolis Quarterly Review 1992）⇒**新しい保険数理の必要性**

逆ザヤの発生

- 日本では、逆ザヤ問題（責任準備金の予定利息に利息配当金収入が満たない状況）が深刻化

期間	新契約予定利率%	10年国債応募者利回り%
96.4~99.3	2.75(標準利率)	3.132(96av.)~1.518(98av.)
99.4~01.3	2.00(同上)	1.732(99av.)~1.293(01av.)
01.4~13.3	1.50(同上)	1.751(06av.)~0.721(13av.)
13.4~17.3	1.00(同上)	0.721(13av.)~▲0.031(16av.)
17.4~	0.25(同上)	0.5(23.3)~▲0.09(19av.)

平成生保危機

会社名	日産生命	東邦生命	第百生命	大正生命	千代田生命	協栄生命	東京生命
破綻時点	1997.4	1999.6	2000.5	2000.8	2000.10	2000.10	2001.3
破綻手続	行政手続	行政手続	行政手続	行政手続	更生特例	更生特例	更生特例
債務超過額	3029億円	6500億円	3177億円	365億円	5950億円	6895億円	731億円
資金援助額	2000億円	3663億円	1450億円	267億円	-	-	-
予定利率強制引下後の上限	2.75%	1.50%	1.00%	1.00%	1.50%	1.75%	2.60%

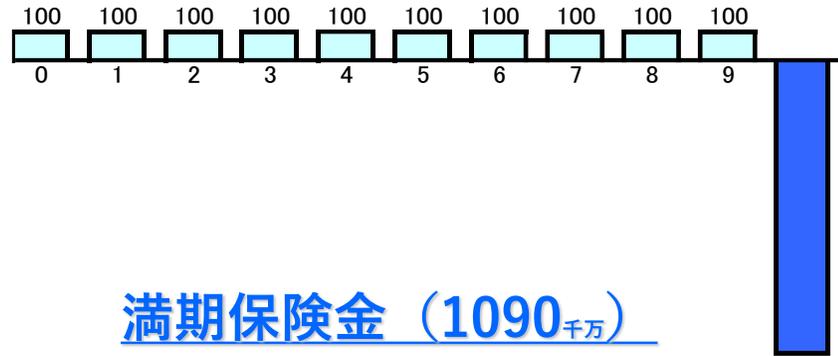
逆ザヤ対策とその限界

- 平成生保危機以降、逆ザヤ解消が保険会社の重要な経営課題となった
- 市場金利低下の環境下でも、新契約予定利率の引き下げと共に、外国債券投資（ただし為替ヘッジコストは別枠処理）などを活用した逆ザヤ解消は進んだ
- しかしながら、2008年、直前まで「逆ザヤ無し、ソルベンシーマージン比率も健全」だった大和生命が破綻し、逆ザヤやソルベンシーマージン比率を目標化したリスク管理の限界が認識されるようになった

(注)法令上200%以上が要請されるが、500%超えが一般的

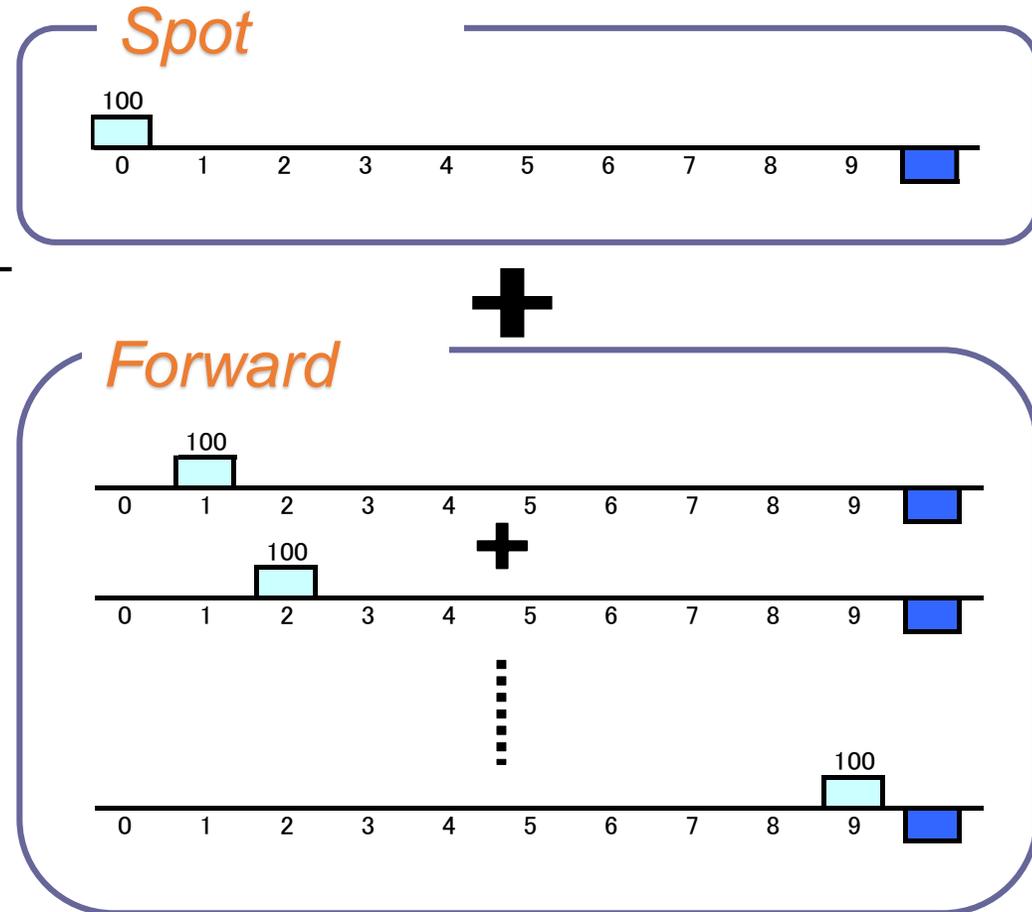
伝統的な平準払保険の金利リスクが手強い理由

平準払保険料 (年P 100_{千万})



満期保険金 (1090_{千万})

伝統的保険は**金利**
フォワードを内在



3つの疑問

- 売却損益等（キャピタルゲイン）を除外して利息配当金収入（インカムゲイン）だけで予定利息という収益目標を達成しようとすることは、資産運用の効率性を損なうのではないか？
- 保険関係収益を勘案すれば、資産運用の目標収益は予定利息よりも下がるので、予定利息を目標化する逆ザヤ回避戦略は過剰な資産運用リスクを取ることになるのではないか？
- 伝統的平準払い保険の将来の払込保険料の利率保証は金利のフォワード契約に相当するから、本来はデリバティブ（金融派生商品）と同等のリスク管理が必要ではないか？

⇒ 資産と負債を市場統合的に評価する**経済価値**でのリスク管理

経済価値とは

- 市場価格が利用可能な場合には、現時点での**市場価格と整合的な方法**により導かれ、市場価格が利用可能でない場合には市場整合的な原則、方法、パラメータを用いた資産・負債の**キャッシュフロー(CF)の現在価値**
- **経済価値 = 最良推定(CE) + マージン(MOCE)**
 - $CE = \sum CF_t / (1+r_t)^t$
 - $MOCE = \text{資本コスト率} \times \sum RC_t / (1+r_t)^t$
 - ここで、 CF_t :時点tのCF、 r_t :時点tのリスクフリー金利、 RC_t :時点tの所要資本(リスク量)

規制における経済価値評価の必要性 (IAIS 2007 structure paper)

- 財務諸表上のすべての項目に関連した資産、債務及びリスク・エクスポージャーを、現時点における**経済価値**によって評価することが、保険会社の**財務状況に関する適正かつ信頼できる情報を提供できる唯一の手法**である
- 保険負債固有の不確実性を踏まえると、技術的準備金(保険契約の**経済価値**)は、契約義務を果たすコストの最良推計に加えてリスク・マージンを含む必要がある。リスク・マージンは、**保険会社が保険負債を引き継ぐために求めると期待される額**が技術的準備金となるよう、その額が決められる。

経済価値ベースのソルベンシーマージン 規制の導入

- EUでは、保険会社に対し、資産と負債を経済価値評価する経済価値ベースのソルベンシーマージン基準「Solvency II」を2016年に導入済み。一方で、全世界の国際的に影響の大きい保険会社を対象にICSと呼ばれる経済価値ベースの類似の制度を2025年に導入予定
- 日本では、金融庁の「経済価値ベースのソルベンシー規制等に関する有識者会議」報告書を受けて、2025年に同様の経済価値ベースのソルベンシーマージン規制を導入予定

市場金利と現在価値(割引債券価格)の変化

市場金利	1年後満期100万円の現在価値	損益
2%	98万円	▲1万5千円
1%	99万円	▲5千円
購入時0.5%	99万5千円	0
0.1%	99万9千円	4千円
0%	100万円	5千円

割引債券の残存期間に比例して損益の変化(金利感応度)は大きくなる

一般に保険会社では資産より負債の平均残存期間が長いので、市場金利変化の影響が現行評価と逆方向に働く

評価基準	市場金利	資産（債券）①	保険負債②	純資産①－②
現行評価	↑	↓	－	↓
	↓ 😊	↑	－	↑ 😊
市場整合的評価	↑ 😊	↓	↓ ↓	↑ 😊
	↓	↑	↑ ↑	↓

項目	現行	経済価値
保険負債価値の評価	責任準備金	将来CFの現在推定の割引現在価値(CE) + リスクマージン(MOCE)
死亡率	標準死亡率（契約時固定）	現在推定 + 将来トレンド（可変）
解約率	一般には不使用	現在推定 + 動的解約（可変）
割引利率	標準利率（契約時固定。 期間構造無し）	現在の金利期間構造を参照(可変)
ソルベンシーマージン	自己資本 + 有価証券含み益の一部 等	経済価値ベース資産負債差額 等
リスク計測手法	会計損益の変動リスク	経済価値の変動リスク

標準モデルと内部管理モデル

- 2025年導入予定の経済価値ベースの新規制では、第一の柱として全ての会社が遵守すべき標準モデルが設定され、それにもとづく経済価値ベースのソルベンシー比率の報告が求められるが、会社間の体力格差もあるため一定程度は政策的措置を含む緩和的なモデルとならざるを得ない
- このため、体力のある大手保険会社には、第二・第三の柱を通じて、より高度な経済価値ベースの内部管理が要請されることになる

2. 経済価値評価のための機械学習

経済価値評価（内部管理モデル）のための機械学習のチャレンジ

- 生保の経済価値評価では、超長期のキャッシュフローを割引現在価値で評価（Current Estimate）するためには、死亡率の超長期予測や観測不能な超長期年限の割引率の補外が不可欠（⇒ **従来型のパラメトリックモデルの限界**）
- 経済価値評価のための生保数理モデルのチャレンジ(例)
 - 観測不能年限のイールドカーブ補外（標準モデルのUFR+SW法の非市場整合性、低次元のNSモデルの表現能力不足といった問題の克服）
 - 現実的な動的解約モデルの構築（金利裁定モデルの限界）
 - **現実的な死亡率の超長期トレンドリスク（Longevity Risk）の評価**

伝統的な死亡率予測の考え方

- 死亡率予測で代表的なリレーショナルモデルであるLCモデル (Lee-Carter 1992) は、年齢($x = x_{\min}, \dots, x_{\max}$)、暦年 ($t = t_{\min}, \dots, t_{\max}$)の死亡率 $m_{x,t}$ を次式のように表す。

$$\log(m_{x,t}) = \alpha_x + \beta_x \kappa_t + \varepsilon_{x,t}, \varepsilon_{x,t} \sim N(0, \sigma^2)$$

- LCモデルは**解釈可能性**の高いモデルで、 α_x は『対数死亡率の平均』、 β_x は『 κ_t に対応する年齢別感応度』、 κ_t は『死亡率改善度(トレンド項)』と解釈することができる。 κ_t を時間外挿することで死亡率予測が可能
- Rのパッケージ (StMoMo) で比較的簡単に計算が実行可能

伝統的LC第一段階

- 対数死亡率の年齢別観測平均を求める

$$\alpha_x = \frac{1}{t_{\max} - t_{\min} + 1} \sum_{t=t_{\min}}^{t_{\max}} \log(m_{x,t})$$

- 対数死亡率行列から α_x を引いた残差行列 \tilde{M} を作る

$$\tilde{M} = \begin{bmatrix} \log(m_{x_{\min}, t_{\min}}) - \alpha_{x_{\min}} & \cdots & \log(m_{x_{\min}, t_{\max}}) - \alpha_{x_{\min}} \\ \vdots & \ddots & \vdots \\ \log(m_{x_{\max}, t_{\min}}) - \alpha_{x_{\max}} & \cdots & \log(m_{x_{\max}, t_{\max}}) - \alpha_{x_{\max}} \end{bmatrix}$$

伝統的LC第一段階：次元圧縮

- \tilde{M} をSVDで行列 U, D, V に分解
$$\left\{ \begin{array}{l} \tilde{M} = UDV^T \\ \beta_x = \frac{U_{x,1}}{\sum_x U_{x,1}} \\ \kappa_t = V_{t,1}D_{1,1} \sum_x U_{x,1} \end{array} \right.$$
- パラメータを一意に定めるため以下の識別条件を入れる。
$$\left\{ \begin{array}{l} \sum_{x=x_{\min}}^{x_{\max}} \beta_x = 1 \\ \sum_{t=t_{\min}}^{t_{\max}} \kappa_t = 0 \end{array} \right.$$

伝統的LC第二段階：時系列のあてはめ

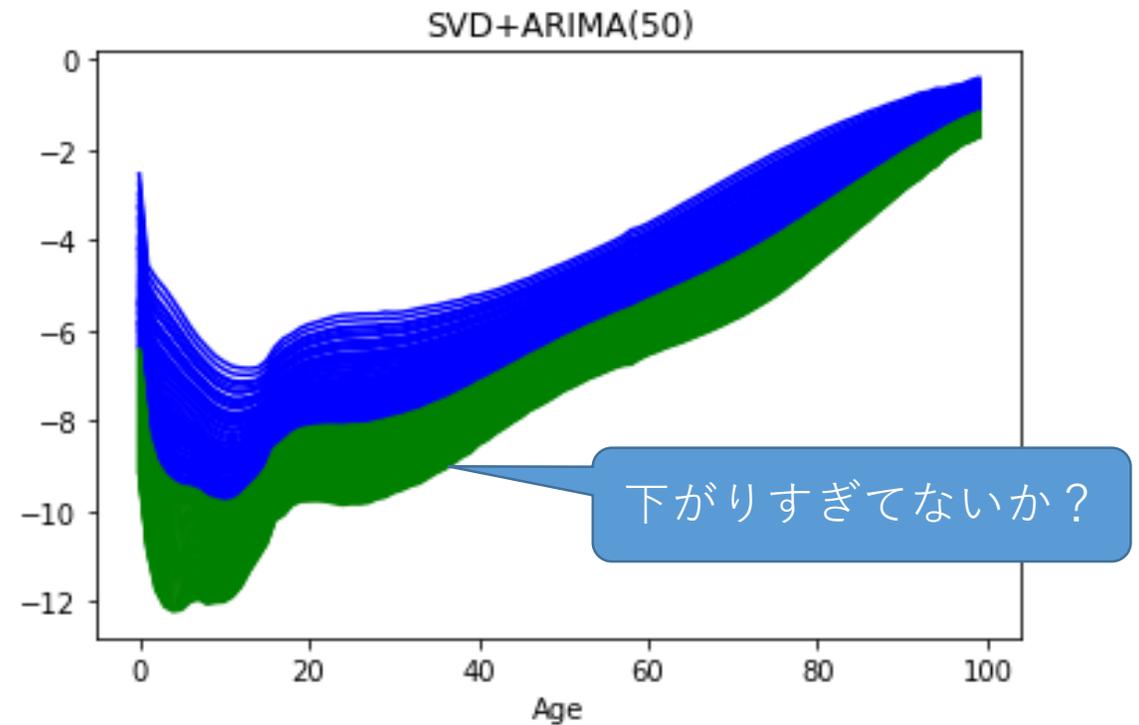
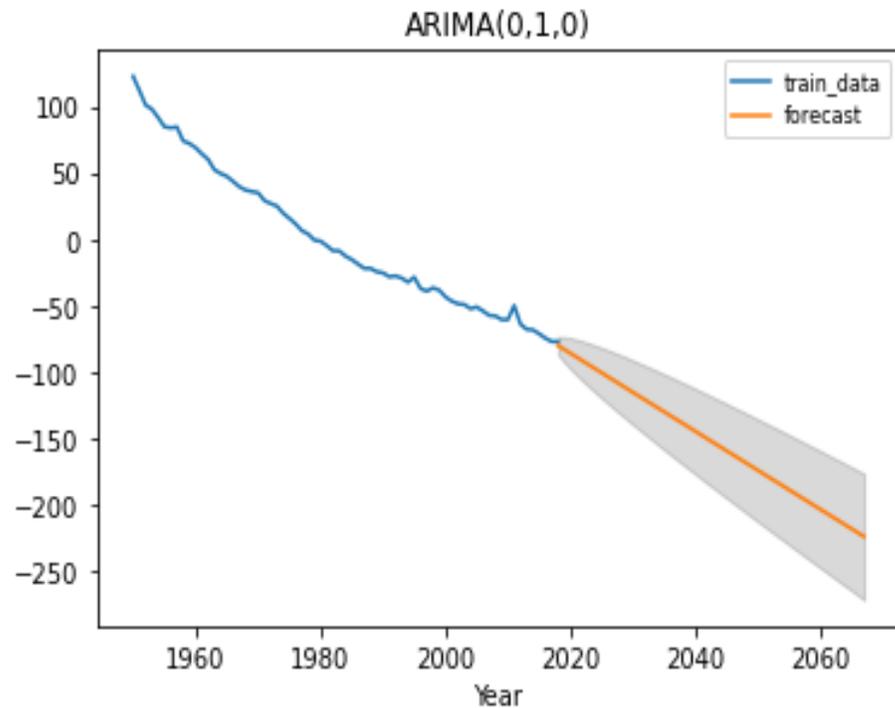
- 第一段階で得られた κ_t に時系列モデルを適用して外挿（二段階推定！）
- ARIMA(p, d, q)

$$\kappa_t - \kappa_{t-d} = \delta + \epsilon_t + \sum_{i=1}^p \phi_i \kappa_{t-i} + \sum_{j=1}^q \theta_j \epsilon_{t-j}$$

- 実際にはARIMA(0,1,0)が選ばれることが多い

$$\kappa_t = \kappa_{t-1} + \delta + \epsilon_t$$

LCの予測では死亡率が直線的に低下
(50年予測の κ_t (左)と対数死亡率(右))



LCモデルの限界への対応

- 二段階推定によるパラメータ間の不整合性の問題

⇒ ベイズ状態空間(BSS)モデルによる **一段階推定** への拡張

モデル構造の柔軟性は高く区間推定も可能だが、多くの場合MCMC（マルコフ連鎖モンテカルロ）が必要で計算コスト大。このため一般に計算粒度が粗く予測精度を追求することが難しい。

- 非線形表現能力の問題

⇒ Neural Network (NN) への拡張

万能近似定理(Cybenco 1989：コンパクトなサポートを持つ如何なる連続関数も連続なシグモイド活性化関数を持つ二層のNNで一様に近似可能)が示すように、NNは**高い非線形表現能力**を持ち、予測精度が得やすいが、一般にモデルのseed頑健性と解釈可能性に難がある

ベイズ状態空間(BSS)アプローチ

- LCモデルのBSSモデルへの直接的翻訳 (Pedroza (2006))

観測方程式: $x_t = \alpha + \kappa_t \beta + \varepsilon_t, \varepsilon_t \sim iid N(0, \sigma_\varepsilon^2 I)$

状態方程式: $\kappa_t = \kappa_{t-1} + \mu + \epsilon_t, \epsilon_t \sim iid N(0, \sigma_\epsilon^2)$

$\alpha, \beta, \sigma_\varepsilon, \mu, \sigma_\epsilon$ の同時分布を推定 (事後分布を得るために計算コストの大きいMCMCを使用)

$$x_t = (\log m_{t,0}, \log m_{t,1}, \dots, \log m_{t,n})^T; \alpha = (\alpha_0, \alpha_1, \dots, \alpha_n)^T; \beta = (\beta_0, \beta_1, \dots, \beta_n)^T$$

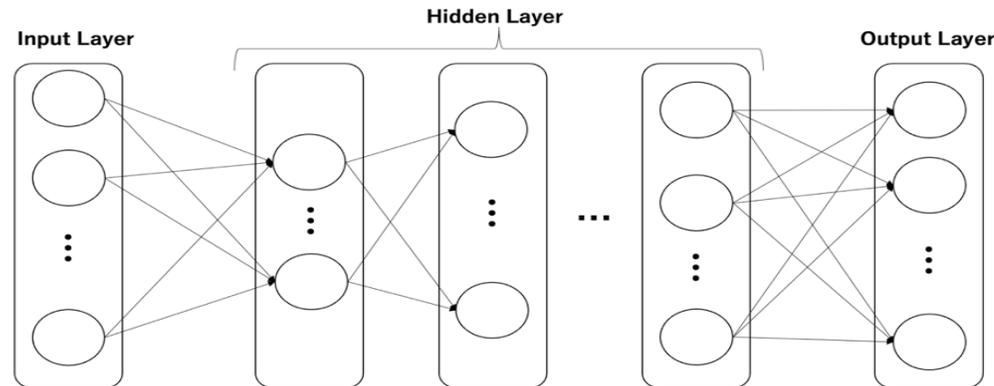
ニューラルネットワーク(NN)アプローチ

- 近年、死亡率予測において、**広い非線形クラスへのモデルの拡張**のために、様々なNNの活用が提案されている。
- NNは、脳にある『シナプス』と『ニューロン』を基にしているモデルであり、NNは一般的に3層で構成され、各層にニューロンが存在する。
- それらは『入力層、隠れ層(中間層)、出力層』と呼ばれており、ネットワーク内の各ニューロンは接続されているシナプスを介して加重情報を取得し、入力信号の加重和(線形和)を変換する活性化関数を使用して出力を返す。
- 特に、中間層上に存在する潜在変数は、データを入力層で入力し抽出された特徴量を表している。

最も基本的なNN (FNN)

- 入力層から出力層までの層の数が n であるとき、 $(i + 1)$ 層目における d_{i+1} 次元ユニットの出力 y_{i+1} ($i = 1, 2, \dots, n - 1$)は y_i 層目における d_i 次元ユニットの出力 y_i 、活性化関数 ϕ_i 、重み行列 $W_i \in \mathbb{R}^{d_{i+1} \times d_i}$ 、バイアスベクトル $\mathbf{b}_i \in \mathbb{R}^{d_{i+1}}$ を用いて次のように計算する

$$y_{i+1} = \phi_i(W_i y_i + \mathbf{b}_i)$$



活性化関数 ϕ

- Sigmoid: $1/(1+e^{-x})$
- Hyperbolic tangent: $\tanh(x) = (e^x - e^{-x}) / (e^x + e^{-x})$
- ReLU: x ($x > 0$); 0 ($x \leq 0$)
- LeakyReLU: x ($x > 0$); αx ($x \leq 0$) α は微少な正数
- Softmax: $e^{x_i} / \sum_{j=1}^n e^{x_j}$ n はノード数

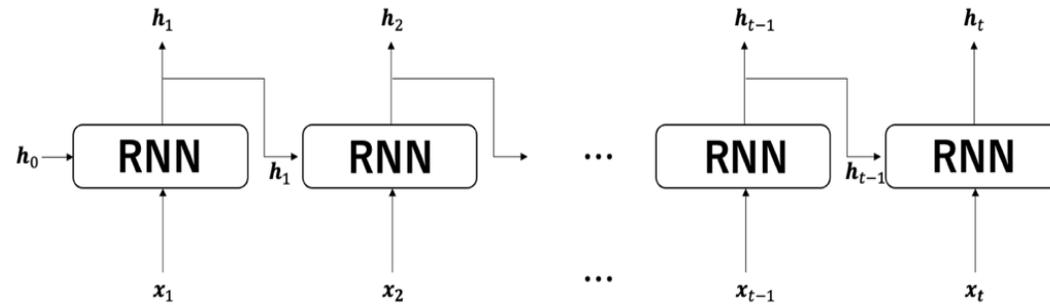
CNN(畳み込みNN)

- CNN(Convolutional NN)は後述のRNNと同じく時系列との親和性が高い
- CNNは、第1段階で重みを共有した畳み込みを行い、第2段階では畳み込みによって得られた値を活性化関数によって非線形変換し、最後にプーリング関数(ウィンドウ領域の平均/最大/最小)によって値を出力する
- Perla et al.(2021)は1D-CNNによるLCモデルの拡張を提案
- 1D-CNNでは、カーネルサイズ m ・フィルタ数 J の畳み込みフィルタ $W^{i,j} \in \mathbb{R}_{d \times m}$ ($j = 1, \dots, J$)を用いて、第 i 層のデータ $y^i \in \mathbb{R}_{d \times T}$ を以下で変換

$$y_{j,k}^{i+1} = \phi \left(\sum_{s=1}^m \sum_{l=1}^d W_{l,s}^{i,j} y_{l,k+s-1}^i + b^{i,j} \right);$$
$$k = 1, \dots, T + 1 - m; y^{i+1} \in \mathbb{R}_{J \times (T+1-m)}.$$

RNN（再帰型NN）

- 時系列データと親和性の高いNN
- 時刻 t の入力データ $\mathbf{x}_t \in \mathbb{R}^d$ 、RNN層の k 次元ユニットである時、時刻 t におけるRNN層の出力 \mathbf{h}_t は時刻 $(t - 1)$ のRNN層の出力 \mathbf{h}_{t-1} と活性化関数 ϕ 、入力に対応する重み行列 $W \in \mathbb{R}^{k \times d}$ 、時刻 $(t - 1)$ の出力結果に対応する重み行列 $U \in \mathbb{R}^{k \times k}$ 、バイアスベクトル $\mathbf{b} \in \mathbb{R}^k$ を用いて、次のように計算する



LSTM

- 代表的なRNN：期時系列データを扱う上での勾配消失問題に対応できるLSTM(Long short term memory)
- 時刻 t における入力値ベクトルを $\mathbf{x}_t \in \mathbb{R}^d$ 、LSTMのユニット数を k とするとき、忘却ゲート出力 \mathbf{f}_t 、入力ゲート出力 \mathbf{i}_t 、出力ゲートの出力 \mathbf{o}_t 、入力情報からCECを計算する \mathbf{g}_t はシグモイド関数 σ と活性化関数 ϕ を用いて次のように計算される

$$\mathbf{f}_t = \sigma(W_f^T \mathbf{x}_t + U_f^T \mathbf{h}_{t-1} + \mathbf{b}_f)$$

$$\mathbf{o}_t = \sigma(W_o^T \mathbf{x}_t + U_o^T \mathbf{h}_{t-1} + \mathbf{b}_o)$$

$$\mathbf{i}_t = \sigma(W_i^T \mathbf{x}_t + U_i^T \mathbf{h}_{t-1} + \mathbf{b}_i)$$

$$\mathbf{g}_t = \phi(W_g^T \mathbf{x}_t + U_g^T \mathbf{h}_{t-1} + \mathbf{b}_g)$$

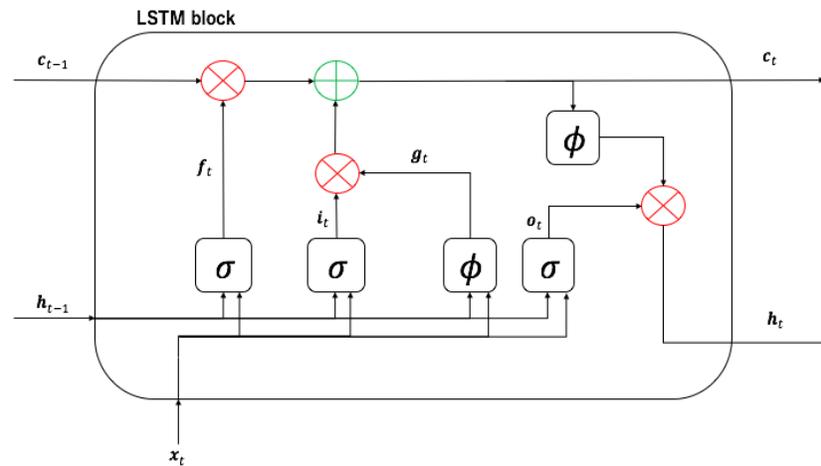
- ただし、時刻 $(t-1)$ の出力結果： \mathbf{h}_{t-1} 、各ゲートの入力値に対応する重み行列： $W_f, W_o, W_i, W_g \in \mathbb{R}^{k \times d}$ 、時刻 $(t-1)$ の出力結果に対応する重み行列： $U_f, U_o, U_i, U_g \in \mathbb{R}^{k \times k}$ 、バイアスベクトル： $\mathbf{b}_f, \mathbf{b}_o, \mathbf{b}_i, \mathbf{b}_g \in \mathbb{R}^k$

LSTM

- 時刻 t におけるLSTMブロックの出力結果 \mathbf{h}_t は各ゲートの出力結果を用いて、次のように計算することができる

$$\mathbf{c}_t = \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{g}_t \odot \mathbf{i}_t$$
$$\mathbf{h}_t = \phi(\mathbf{c}_t) \odot \mathbf{o}_t$$

ただし、 \odot は要素ごとの積（アダマール積）で、 \mathbf{c}_t はLSTMブロックのCEC（Constant Error Carousel）とよばれる勾配消失防止の仕組み



LSTMによるLC拡張①

(Nigri et al. 2019)

- SVDで得られたLCの (κ_t) 列にラグ τ のLSTMをあてはめて次式のように逐次アルゴリズムで外挿したものをLCモデルにあてはめ (二段階推定)

$$\bullet \left\{ \begin{array}{l} \kappa_t = f_{\text{LSTM}}(\kappa_{t-1}, \dots, \kappa_{t-\tau}) \\ \kappa_{t+1} = f_{\text{LSTM}}(\kappa_t, \dots, \kappa_{t-\tau+1}) \\ \vdots \\ \kappa_{t+T} = f_{\text{LSTM}}(\kappa_{t+T-1}, \dots, \kappa_{t-\tau+T}) \end{array} \right.$$

LSTMによるLC拡張② (Perla. et al.2021)

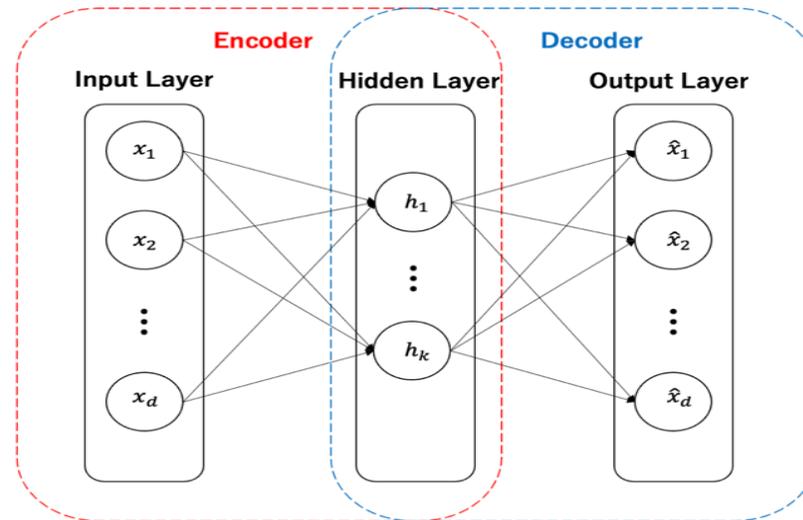
- 複数母集団の死亡率データに対してLSTMを利用した一段階推定を達成
- 年齢 $x(x_{\min}, \dots, x_{\max})$ 、歴年 $t(t_{\min}, \dots, t_{\max})$ 、国 $i(\text{JPN, UK, USA, } \dots)$ の死亡率を $m_{x,t}^{(i)}$ 、対数死亡率 $\log(m_{x,t}^{(i)})$ を訓練データの最大値と最小値を利用して $[0,1]$ 正規化したものを $\tilde{m}_{x,t}^{(i)}$ とする
- 時刻 t でLSTMに入力するベクトルとそれに応答するベクトル $\mathbf{u}_t^{(i)}, \mathbf{s}_t^{(i)}$ を次のように定義。

$$\mathbf{u}_t^{(i)} := \begin{pmatrix} m_{x_{\min},t}^{(i)} & \dots & m_{x_{\max},t}^{(i)} \end{pmatrix}$$
$$\mathbf{s}_t^{(i)} := \begin{pmatrix} \tilde{m}_{x_{\min},t}^{(i)} & \dots & \tilde{m}_{x_{\max},t}^{(i)} \end{pmatrix}$$

- 多変量時系列の入力に対応する時系列モデル： $\mathbf{s}_t^{(i)} = f_{\text{LSTM}}(\mathbf{u}_{t-1}^{(i)}, \dots, \mathbf{u}_{t-\tau}^{(i)})$ に変換($\mathbf{s}_t^{(i)} \Leftrightarrow \mathbf{u}_t^{(i)}$)とLSTMの逐次アルゴリズムを適用して予測
- 同論文では複数のNNを比較しCNN(畳み込みネットワーク)を推奨

NNによる次元削減：AE(Autoencoder)

- AE(Autoencoder)とは、一般的に入力層の次元より低い次元の隠れ層を用い、入力層と隠れ層を経由し、より低い次元で入力データを再構成するように学習させる教師無し学習であり、隠れ層で非線形な特徴を抽出し、多次元データを低次元で特徴づけることが可能
- 下図は、最も単純な3層AEであり、前半のEncoder部分では、入力層の次元 d を隠れ層の次元 k ($k < d$)に圧縮しており、後半のDecoder部分で、圧縮された次元 k から出力層の次元 d に復元



AEによるLC拡張：NNAモデル(Hainaut 2018)

- LCのSVDに替えてAEによる次元削減を行うモデル（二段階推定）。過学習防止の為に疎結合なネットワーク構造を採用
- 入力データ $X(t) := M_{(x_{\min} \sim x_{\max}), t} = \begin{bmatrix} \log(m_{x_{\min}, t}) - \alpha_{x_{\min}} \\ \vdots \\ \log(m_{x_{\max}, t}) - \alpha_{x_{\max}} \end{bmatrix}$
- 再構成データを $\hat{X}(t)$ 、 d 次元潜在変数を $\kappa_t^{nn} = (\kappa_t^{nn,1}, \dots, \kappa_t^{nn,d})$
- AE構造：
$$\begin{cases} \kappa_t^{nn} = f^{(enc)}(X(t)) \\ \hat{X}(t) = f^{(dec)}(\kappa_t^{nn}) \end{cases}$$
- ARIMA(0,1,0)で κ_t^{nn} を外挿し、 $\log m_t = \alpha + f^{(dec)}(\kappa_t^{nn})$ で予測

既存研究(NNによるLCモデル非線形拡張)

■ 二段階推定

- Hainaut(2018) : 従来のSVDの代わりに疎結合Autoencoder(AE)を利用
- Nigri et al.(2019) : RWの代わりにLong Short-Term Memory(LSTM)で外挿
- Scognamiglio(2022) : 疎結合Deep feed forward NN(DFNN)でLCのパラメータを推定



- モデルに**解釈可能性**あり (リスク管理に向く)
- パラメータ間の**不整合性**あり

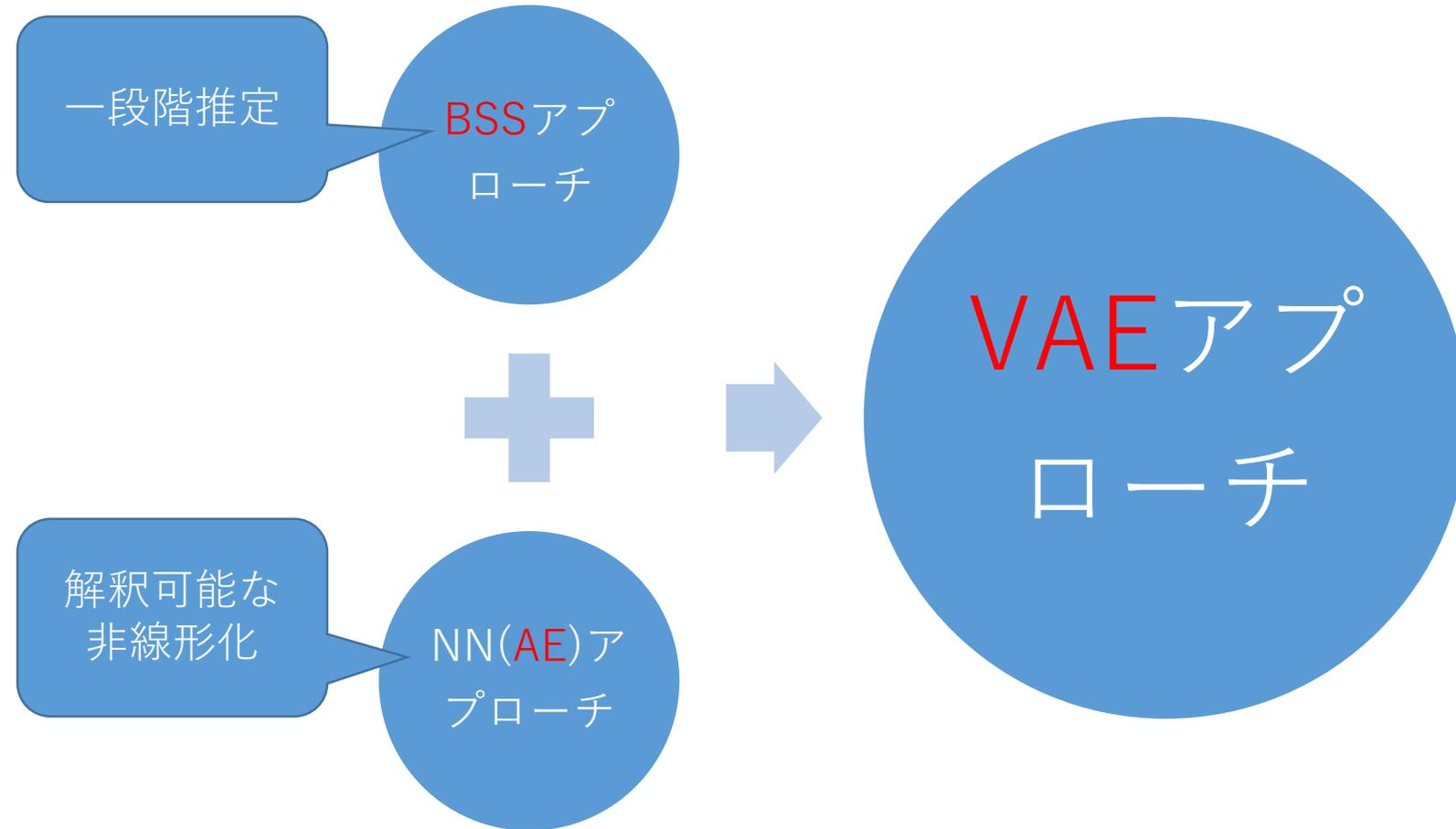
■ 一段階推定

- Wüthrich & Richman(2019) : DFNNを利用したLC拡張
- Perla et.al(2021) : 1D-Convolutional NN(CNN)やLSTM等を利用したLC拡張
- Schnürch & Korn(2022) : 区間推定を考慮した2D-CNNを利用したLC拡張



- モデルに**解釈可能性**なし (リスク管理に不向き)
- パラメータ間の**不整合性**なし (高精度)

NNアプローチとBSSアプローチの融合: 解釈可能な一段階推定の実現 (MM2022)



1. VAEのしくみ(Kingma & Welling 2013)

- VAEは一般的なAEと異なり潜在変数に確率分布を仮定
- VAEの目的は、高コストのMCMC計算を回避して複雑なデータの生成分布 $p_{\theta}(x)$ を求めることであるが、その過程でデータの次元削減された潜在表現を確率分布として獲得するネットワークパラメータを学習
- ベクトル値データセット $X = \{x_i\}_{i=1}^T$ を生成する潜在変数 $Z = \{z_i\}_{i=1}^T$ が事前分布 $p_{\theta}(z_t)$ から生成され、データ x_t は $p_{\theta}(x_t|z_t)$ により生成されるとする。ここで、 $p_{\theta}(z_t)$ と $p_{\theta}(x_t|z_t)$ は、 $p_{\theta}(z_t)$ の分布パラメータと生成パラメータ θ について微分可能と仮定
- 一般に複雑な多次元事後分布 $p_{\theta}(z_t|x_t)$ を直接推定することは難しいため、変分パラメータ φ をもつ $q_{\varphi}(z_t|x_t)$ で変分近似を行う

確率的エンコーダ/デコーダ, 平均場近似

- VAEの変分近似では因子分解可能な分布での平均場近似を行う

$$q_{\varphi}(Z|X) = \prod_{t=1}^T q_{\varphi}(z_t|x_t)$$

- ここで、事後分布を近似する分布 $q_{\varphi}(z_t|x_t)$ をエンコーダと考え、生成分布 $p_{\theta}(x_t|z_t)$ をデコーダと考えることでこのアルゴリズムはAEと見なせるので、 $q_{\varphi}(z_t|x_t)$ を確率的エンコーダと呼び、生成分布 $p_{\theta}(x_t|z_t)$ を確率的デコーダと呼ぶ

ELBO(変分下限)最大化

- パラメータ φ と θ は ELBO (evidence lower bound) 最大化で学習

$$\text{ELBO} := \int q_{\varphi}(Z|X) \log \frac{p_{\theta}(Z,X)}{q_{\varphi}(Z|X)} dZ$$

- ELBO 最大化 \Rightarrow 近似誤差最小化 + 再構成誤差最小化 + 正則化

- $\text{ELBO} = \log p_{\theta}(X) - D_{KL}[q_{\varphi}(Z|X) || p_{\theta}(Z|X)]$

- $\Rightarrow q_{\varphi}(Z|X)$ の $p_{\theta}(Z|X)$ の近似誤差 (D_{KL} : KL (Kullback-Leibler) 距離) 最小化

$$D_{KL}[q_{\varphi}(Z|X) || p_{\theta}(Z|X)] = \int q_{\varphi}(Z|X) \log \frac{q_{\varphi}(Z|X)}{p_{\theta}(Z|X)} dZ.$$

- $\text{ELBO} = \int q_{\varphi}(Z|X) \log p_{\theta}(X|Z) dZ - D_{KL}[q_{\varphi}(Z|X) || p_{\theta}(Z)]$

- $\Rightarrow \log p_{\theta}(X|Z)$ はデコーダ再構成精度が高い ($\doteq 1$) と 0、低い (< 1) と一

- \Rightarrow 近似事後分布 $q_{\varphi}(Z|X)$ と事前分布 $p_{\theta}(Z)$ の KL 距離を罰則とする正則化

Reparameterization trick

- ELBO計算におけるサンプリング近似

$$\int q_{\varphi}(z_t|x_t) \log p_{\theta}(x_t|z_t) dz_t \approx \frac{1}{L} \sum_{l=1}^L \log p_{\theta}(x_t|z_{l,t})$$

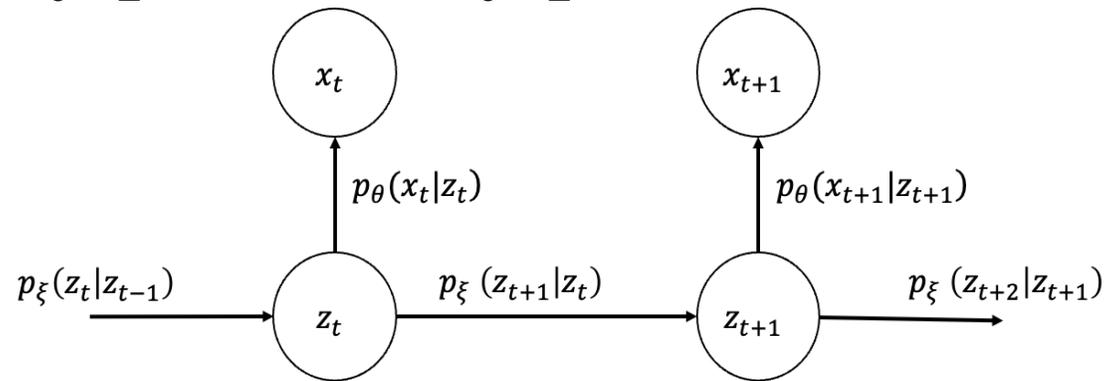
- $z_t \sim q_{\varphi}(z_t|x_t)$ のサンプル生成において、勾配降下法を適用可能にするため、reparameterization trickを導入

$$z_t = g_{\varphi}(x_t, \epsilon_t) = \mu_t + \sigma_t \epsilon_t$$
$$(\mu_t, \sigma_t) = f_{\varphi}^{enc}(x_t), \epsilon_t \sim iid N(0, 1)$$

2. BSS形式でのLCモデルの非線形拡張

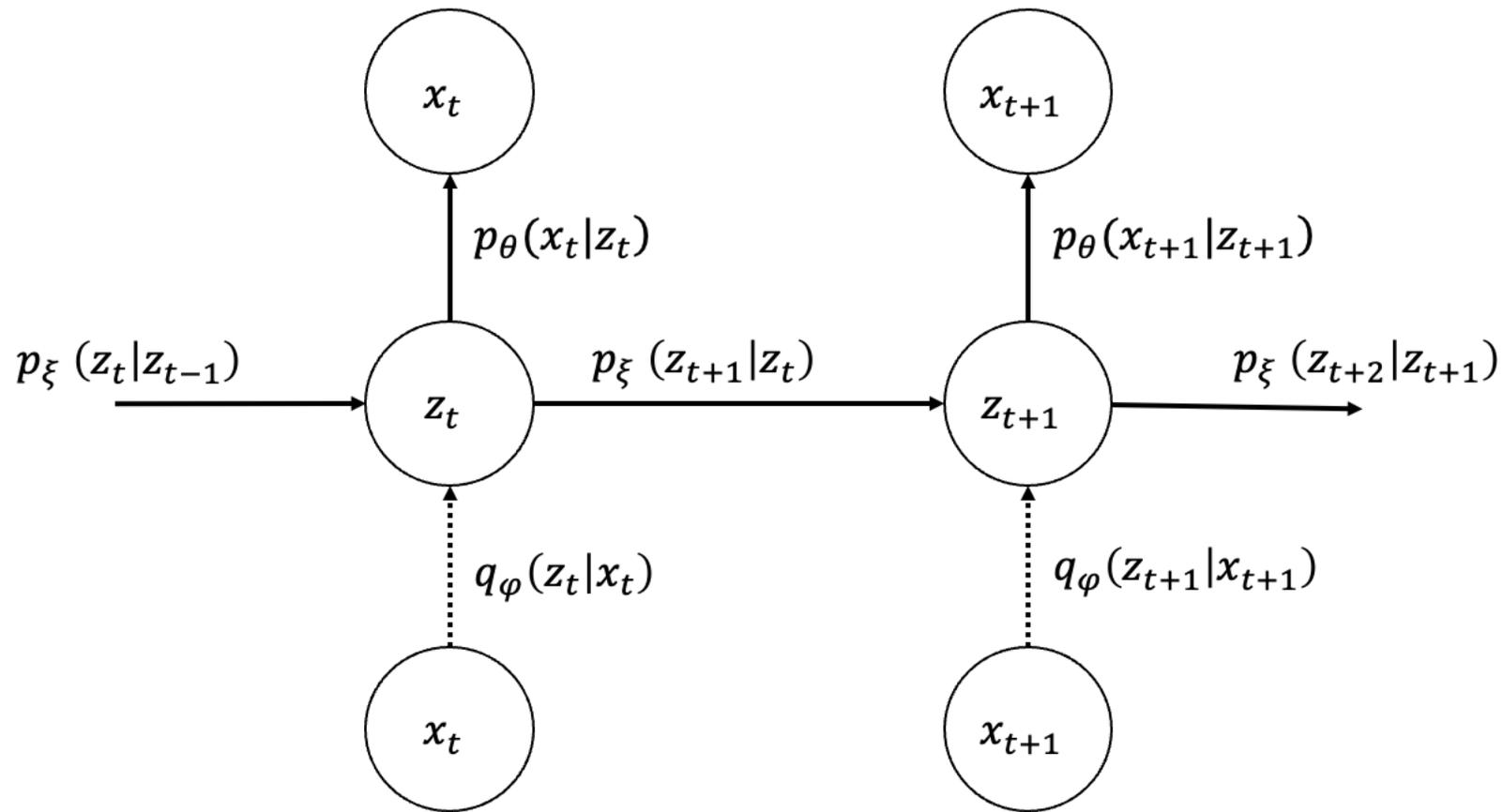
- LCモデルをBSS形式で一般的に記述すると、年齢別対数死亡率データ $\{x_t\}_{t=1}^T$ と潜在変数 $\{z_t\}_{t=1}^T$ について

$$p_{\theta, \xi} (x_1, x_2, \dots, x_T, z_1, z_2, \dots, z_T)$$
$$= \prod_{t=1}^T p_{\theta}(x_t | z_t) \prod_{t=1}^{T-1} p_{\xi}(z_{t+1} | z_t) p_{\xi}(z_1)$$



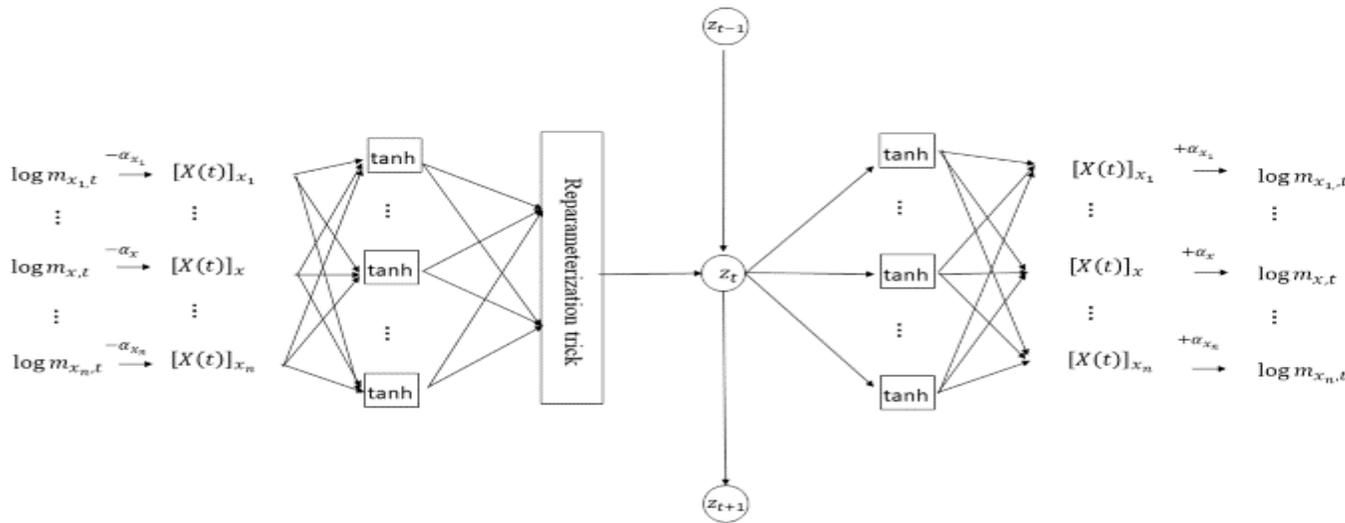
- $p_\xi(z_{t+1}|z_t)$ に伝統的LCと同様にRWを仮定
- LCの「年齢成分 + 年齢成分 × 時系列成分」を（NNを用いて定義される）広いクラスの非線形関数 $f_\theta(z_t)$ に拡張し、モデルを定義する
 - 時点 t の年齢別対数死亡率ベクトル： $x_t = \log m_t = f_\theta(z_t) + \varepsilon_t$,
 - $z_t = \mu_\xi + z_{t-1} + \sigma_\xi \eta_t, t > 1$,
 - $z_1 = z_0 + \sigma_\xi \eta_1$,
 - $\varepsilon_t \sim N(0, \Sigma_\theta); \Sigma_\theta = \text{diag}(\sigma_{\theta_0}^2, \sigma_{\theta_1}^2, \dots, \sigma_{\theta_n}^2); \eta_t \sim N(0, 1)$
 - ここで $(\sigma_{\theta_0}^2, \sigma_{\theta_1}^2, \dots, \sigma_{\theta_n}^2), (\mu_\xi, \sigma_\xi), z_0$ は学習パラメータ
- $p_\xi(z_1)$ が正規分布に従うとき、最大化すべき生成分布の対数尤度は
 - $\log p_\theta(x_t|z_t) = -\log \sqrt{(2\pi)^{n+1} |\Sigma_\theta|} - \frac{1}{2} (x_t - f_\theta(z_t))^T \Sigma_\theta^{-1} (x_t - f_\theta(z_t))$
 - $|\Sigma_\theta| = \sigma_{\theta_0}^2 \sigma_{\theta_1}^2 \dots \sigma_{\theta_n}^2$

3. VAEによるパラメータ推定



ネットワーク構造

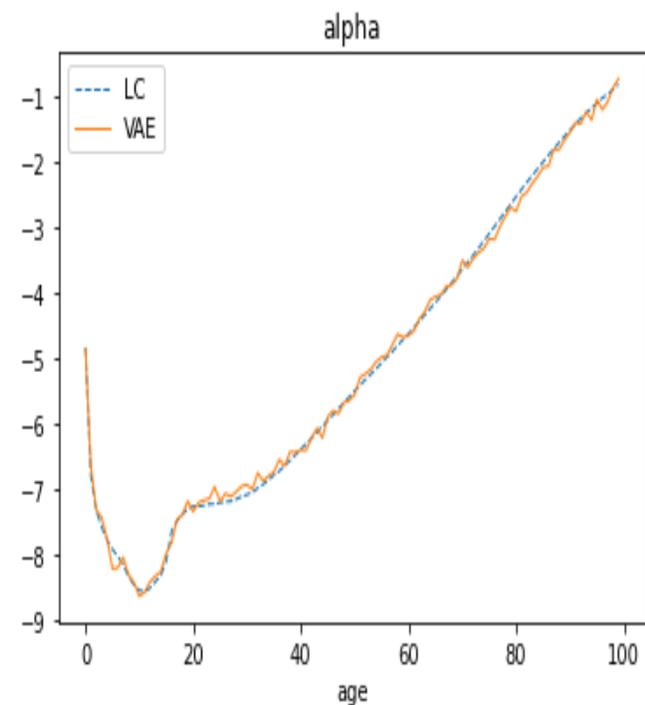
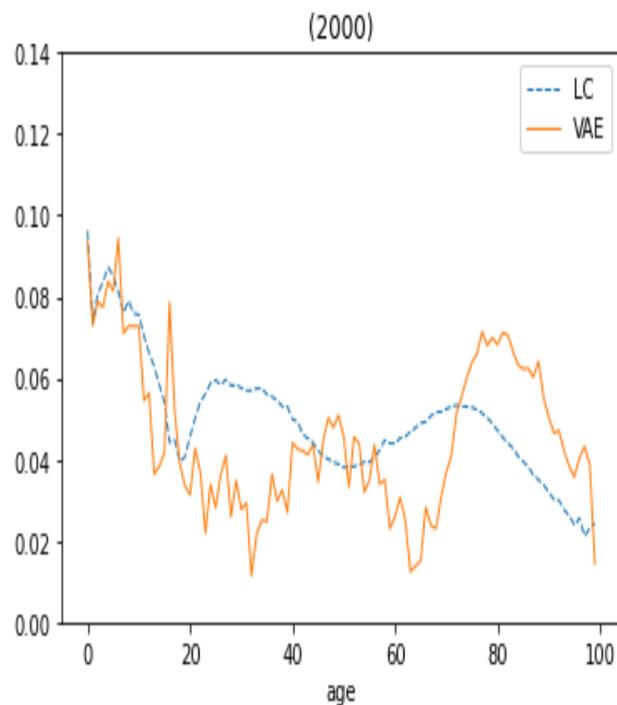
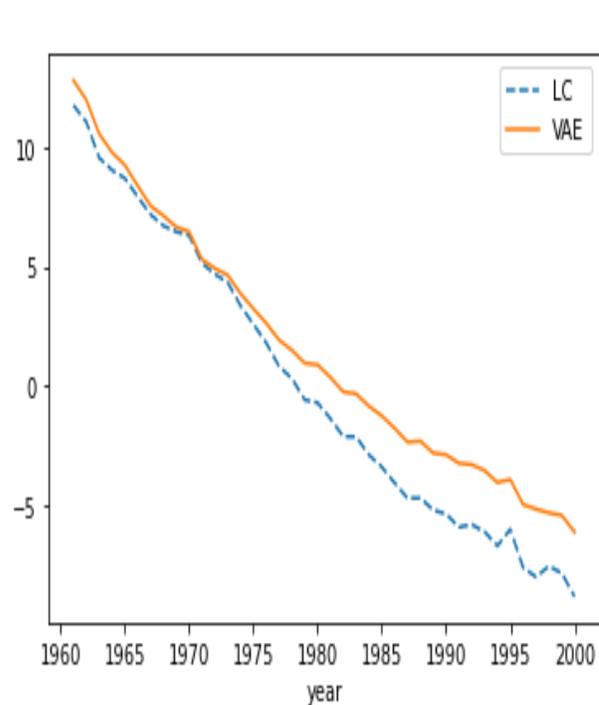
- 3つの隠れ層を持ち、第二隠れ層（潜在変数）の次元（ニューロン数）は1に固定
- エンコーダ部（第一隠れ層）とデコーダ部（第三隠れ層）の活性化関数はtanh + 恒等関数
- エンコーダ部にreparameterization ユニットを含む
- 固定年齢ベクトルの学習パラメータ α をエンコード前に引き、デコード後に足す



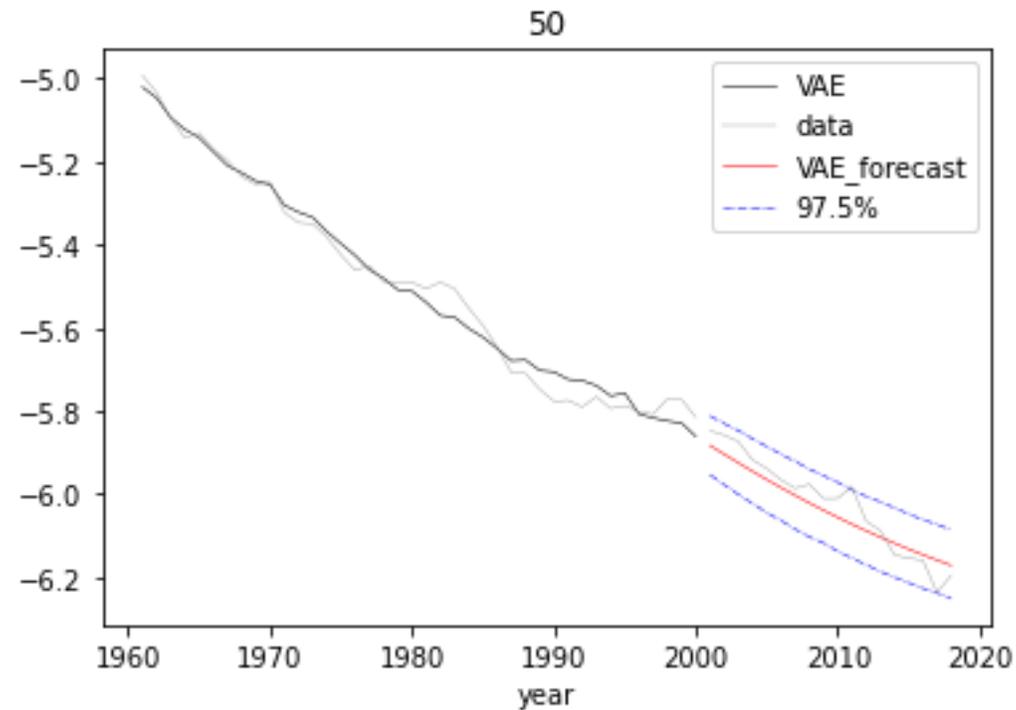
バリデーション(NNの調教)

- 一般の機械学習アルゴリズムに比べ、NNでは多くのハイパーパラメータ (HP) を決定する必要がある
- このモデルのHPは、第一と第三隠れ層のニューロン数や、learning epoch数など
- 日本の死亡率データ上で、1961-1995で訓練し、バリデーションデータ1996-2000の精度 (10ランダムシード平均) が最も高くなるようHPを特定(⇒neuron:20-1-20,epoch:25000)
- このHPのもとで日本のテストデータ2001-2018上での推定精度 (MSE)を評価すると、LCが3.27、VAE (10ランダムシード平均) が1.50

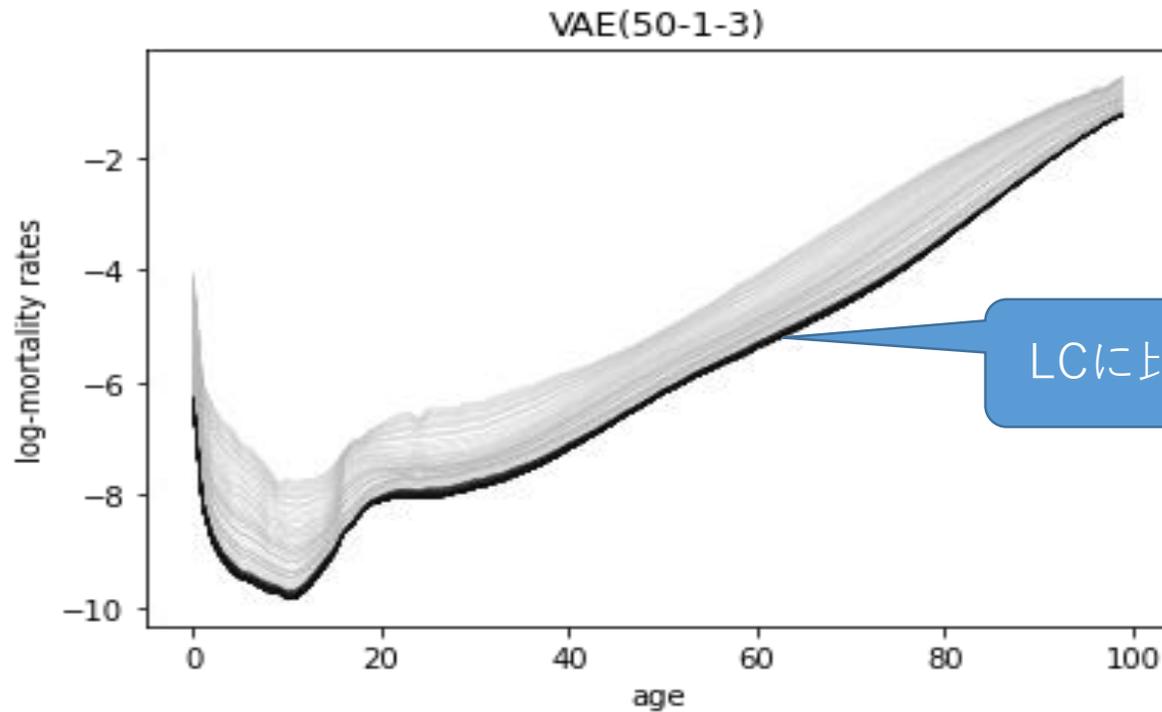
再構成パラメータの解釈性(単一ランダムシードにおける、潜在変数と k 、 $\frac{d}{d\mu_t} f_{\theta}(\mu_t)$ と β 、年齢固定成分と α の比較)



BSSモデルのように予測信頼区間を求め
ることも可能(例：50歳死亡率)



対数死亡率の再構成と50年予測



参考文献 1

- 植村(2007)「生命保険会社の経営破綻要因」保険学雑誌 2007 巻598号,35-52
- 金融庁(2020)「経済価値ベースのソルベンシー規制等に関する有識者会議」報告書
- 金融庁(2022)経済価値ベースのソルベンシー規制等に関する基本的な内容の暫定決定について

参考文献 2

- BROUHNS, N., DENUIT, M. and VERMUNT, J.K. (2002) A Poisson log-bilinear regression approach to the construction of projected lifetables. *Insurance: Mathematics and Economics*, **31**(3), 373-393.
- CAIRNS, A.J., BLAKE, D., DOWD, K. (2006) A two-factor model for stochastic mortality with parameter uncertainty: Theory and calibration. *North American Actuarial Journal of Risk and Insurance*, **73**(4), 687-718.
- CAIRNS, A.J., BLAKE, D., DOWD, K., COUGHLAN, G.D., EPSTEIN, D., ONG, A. and BALEVICH, I. (2009) A quantitative comparison of stochastic mortality models using data from England and Wales and the United States. *North American Actuarial Journal*, **13**(1), 1-53.
- CAIRNS, A.J., BLAKE, D., DOWD, K., COUGHLAN, G.D. and KHALAF-ALLAH, M. (2011) Bayesian stochastic mortality modelling for two populations. *ASTIN Bulletin*, **41**(1), 25-59.
- CYBENKO, G. (1989) Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems*, **2**, 303-314.
- CZADO, C., DELWARDE, A. and DENUIT, M. (2005) Bayesian Poisson log-bilinear mortality projections. *Insurance: Mathematics and Economics*, **36**(3), 260-284.
- FUNG, M.C., PETERS, G.W. and SHEVCHENKO, P.V. (2017) A unified approach to mortality modelling using state-space framework: characterisation, identification, estimation and forecasting. *Annals of Actuarial Science*, **11**(2), 343-389.
- HAINAUT, D. (2018) A neural-network analyzer for mortality forecast. *ASTIN Bulletin*, **48**(2), 481-508.
- KINGMA, D. P. and WELING, M. (2013) Auto-encoding variational Bayes. arXiv:1312.6114.
- KOGURE, A. and KURACHI, Y. (2010) A Bayesian approach to pricing longevity risk based on risk-neutral predictive distributions. *Insurance: Mathematics and Economics*, **46**(1), 162-172.
- LEE, R. D. and CARTER, L. (1992) Modeling and forecasting US mortality. *Journal of the American statistical association*, **87**(419), 659-671.
- LECUN, Y., BOSER, B. E., DENKER, J. S., HENDERSON, D., HOWARD, R. E., HUBBARD, W. E. and JACKEL, L. D. (1990) Handwritten digit recognition with a back-propagation network. *Neural Information Processing Systems 2 (NIPS 1989)*, 396-404.
- MIYATA, A., MATSUYAMA, N. (2022) Extending the Lee-Carter model with variational autoencoder: a fusion of neural network and Bayesian approach *ASTIN Bulletin*, **51**(3), 689-718.
- NIGRI, A., LEVANTESI, S., MARINO, M., SCOGNAMIGLIO, S. and PERLA, F. (2019) A deep learning integrated Lee-Carter model. *Risks*, **7**(1), 33.
- PEDROZA, C. (2006) A Bayesian Forecasting Model: Predicting U.S. Male Mortality, *Biostatistics*, **7**, 530-550.
- PERLA, F., RICHMAN, R., SCOGNAMIGLIO, S. and WÜTHRICH, M. V. (2021) Time-series forecasting of mortality rates using deep learning. *Scandinavian Actuarial Journal*, 1-27.
- RENSHAW, A. E. and HABERMAN, S. (2006) A cohort-based extension to the Lee-Carter model for mortality reduction factors. *Insurance: Mathematics and Economics*, **38**(3), 556-570.
- RICHMAN, R. and WÜTHRICH, M. V. (2021) A neural network extension of the Lee-Carter model to multiple populations. *Annals of Actuarial Science*, **15**(2), 346-366.
- SCHNÜRCH, S., and KORN, R. (2021) Point and interval forecast of death rates using neural networks. *ASTIN Bulletin*, **52**(3), 789-812.
- WANG, CW., ZHANG, J., and ZHU., W. (2021) Neighbouring prediction for mortality. *ASTIN Bulletin*, **51**(3), 689-718.
- WÜTHRICH, M. V., MERZ, M. (2022) Statistical foundations of actuarial learning and its applications. Available at SSRN id=3822407.