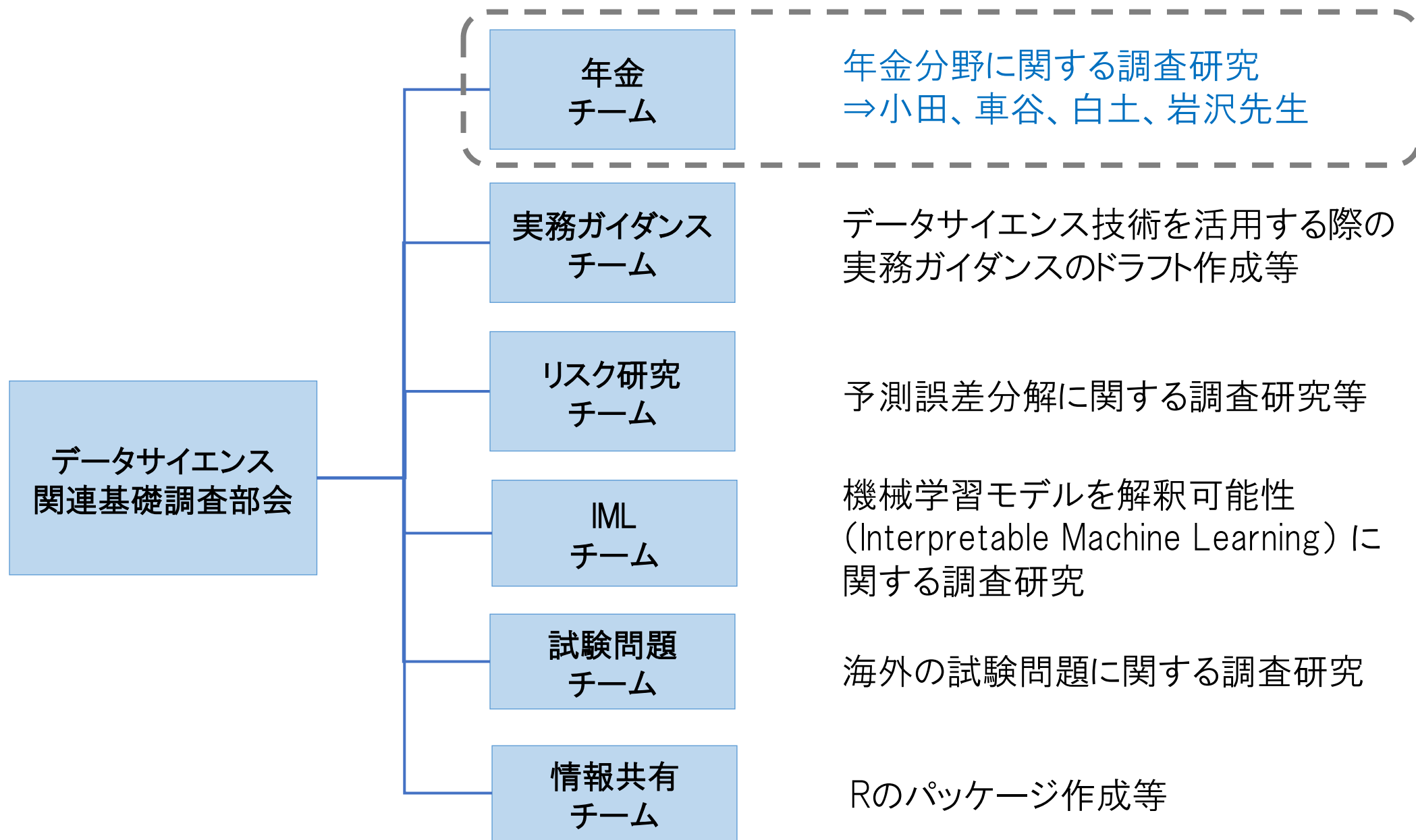


一般化加法モデルを用いた 脱退率推定の改善

三井住友信託銀行 小田直人
りそな銀行、大阪大学MMDS招聘研究員 車谷優樹
2025/11/27

* 本資料において意見に係る部分は講演者の個人的見解であり、所属組織の見解を示すものではありません。

日本アクチュアリー会データサイエンス関連基礎調査部会のチーム編成



本日の概要

1. 現行の脱退率算定方法およびその課題
2. データサイエンスの活用
3. 一般化加法モデルを用いた脱退率算定方法の紹介
4. 数値実験
5. 応用例（説明変数の追加、外部データ脱退率、リスク評価）
6. まとめ
7. Q&A

1. 現行の脱退率算定方法およびその課題

1-1. DB制度、脱退率とは

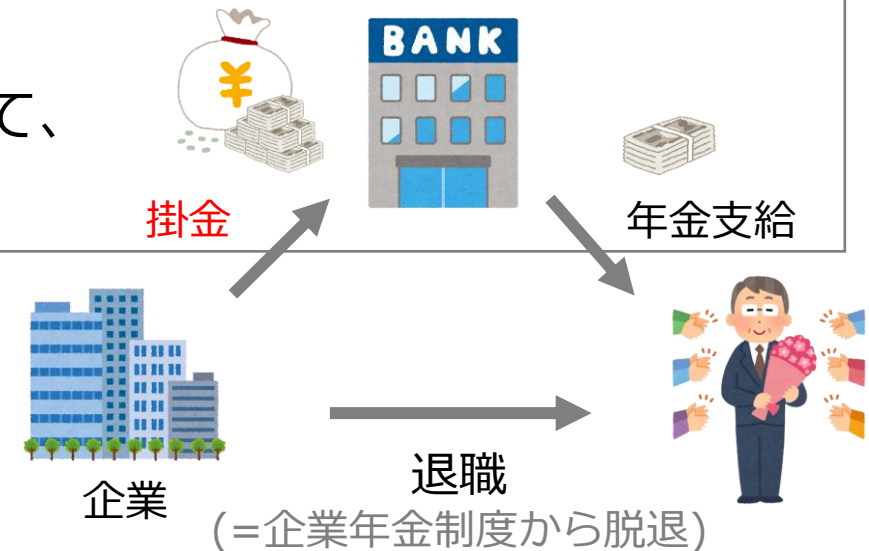
企業年金制度（DB制度）とは？

- 従業員の福利厚生（退職後の生活）のために
公的年金の上乗せとして、**独自の年金制度**を持っている企業がある
- 退職時にもらえるお金は、**長く勤めた者**（=会社への貢献度合いが高い）
ほど高い、という制度が一般的
- 企業年金制度は、**掛金**を毎年（信託銀行等に）積み立てて、
退職者が出たらそこからお金を支払う、という制度



掛金の設定にあたり、次の見積もりが必要になる

✓ **脱退率**（=毎年、どのくらいの従業員が退職するか）



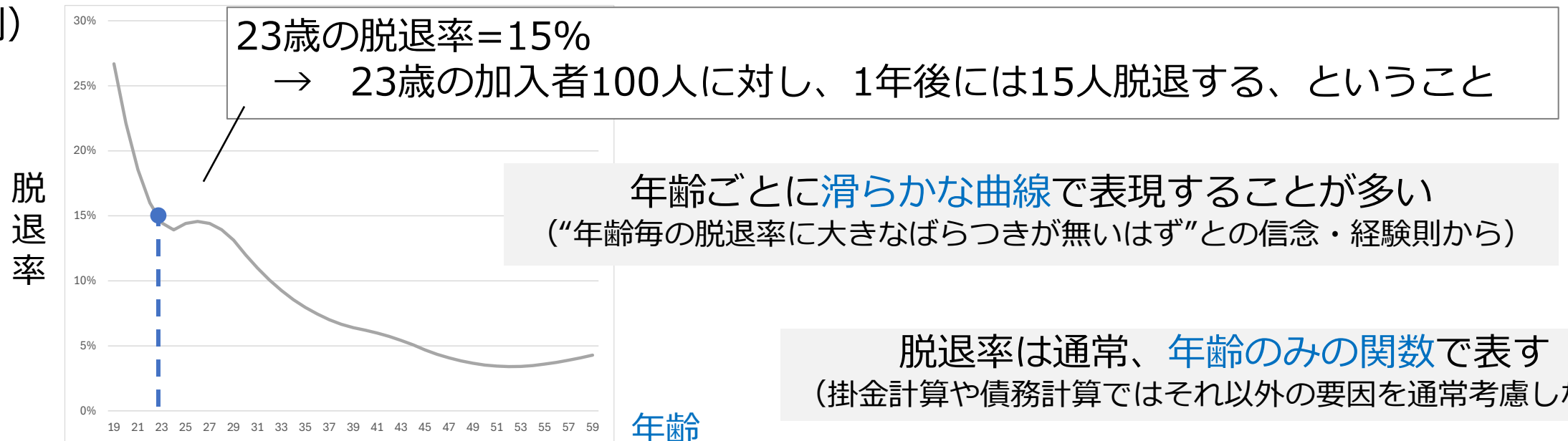
統計的手法を用いて、**脱退率を推定**することは
掛金算定、債務計算に不可欠な計算基礎の1つ

1-1. DB制度、脱退率とは

DB制度での債務計算や掛金計算

- ✓ DB制度の債務や掛金の計算は「現価計算」を元にして行うこととなる
- ✓ 「現価計算」は、将来見込まれるもの（給付や掛金）を基準日まで割戻して現在価値を求めるものであり、その際、
年齢別の「脱退率」（その年齢でDB制度から脱退する確率）を使用する

(例)



実務ではどのように脱退率を計算しているか？

1-2. 現状の取扱い

予定脱退率の取扱い （日本年金数理人会「DB数理実務ガイダンス」）

- ① 算定に用いる過去脱退実績の「算定期間」 （⇒直近3年以上使用など）

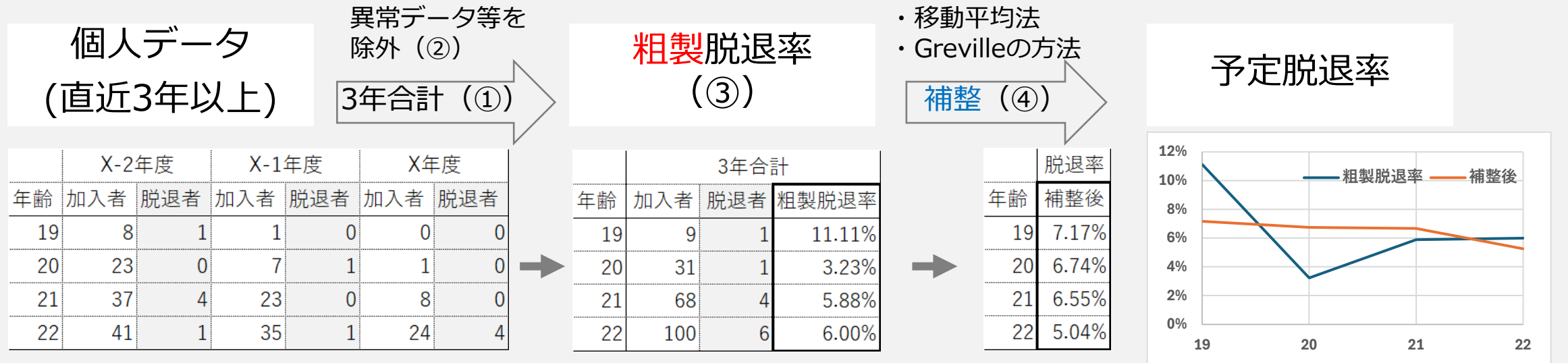
② 過去脱退実績からの「除外データ等の抽出」

③ 過去脱退実績から算出される「粗製脱退率の算定方法」

④ 「補整脱退率の作成方法・粗製脱退率の補整方法」

⑤ 定年延長等の際の「最終年齢までの脱退実績がない場合の取扱い」

⑥ 「異例処理」 （過去実績が使用できない場合等）
- 今回提案する新脱退率算定方法は
③④に関するもの
（他は現行通り取り扱う想定）



1-3. 粗製脱退率の作成

粗製脱退率

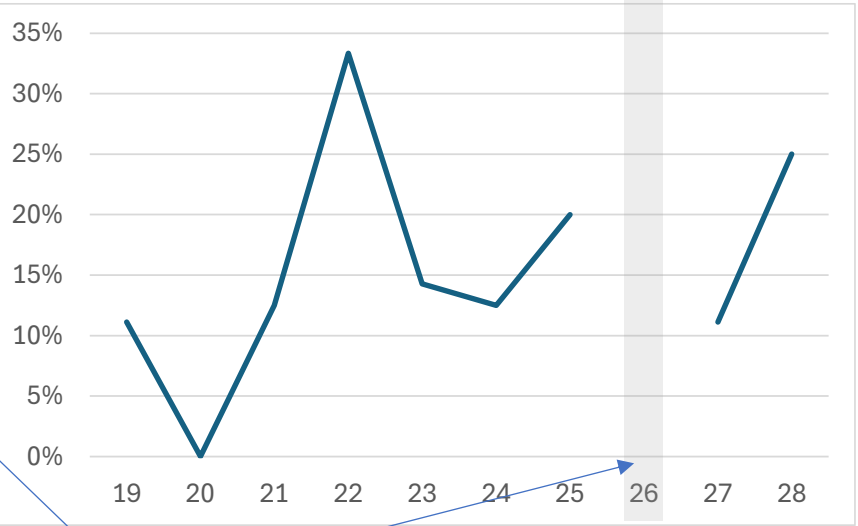
- ① 算定に用いる過去脱退実績の「算定期間」
② 過去脱退実績からの「除外データ等の抽出」
③ 過去脱退実績から算出される「粗製脱退率の算定方法」
④ 「補整脱退率の作成方法・粗製脱退率の補整方法」
⑤ 定年延長等の際の「最終年齢までの脱退実績がない場合の取扱い」
⑥ 「異例処理」（過去実績が使用できない場合等）

$$x \text{ 歳の粗製脱退率} = \frac{x \text{ 歳での脱退者数}}{x \text{ 歳での加入者数}}$$

(分子分母は3年合計)

(例)

年齢	加入者	脱退者	粗製脱退率
19	9	1	11.11%
20	12	0	0.00%
21	8	1	12.50%
22	6	2	33.33%
23	7	1	14.29%
24	8	1	12.50%
25	10	2	20.00%
26	0	0	#DIV/0!
27	9	1	11.11%
28	8	2	25.00%



特に小規模集団の場合、粗製脱退率の年齢ごとのブレは大きくなる傾向にある
→ブレを抑制する観点から、「算定年齢幅の拡大」を行う場合がある(次頁)

課題 1 : 分母が0人の場合、この式で計算できない
(仮に0と設定してもそれが実態を表しているとは言えない)

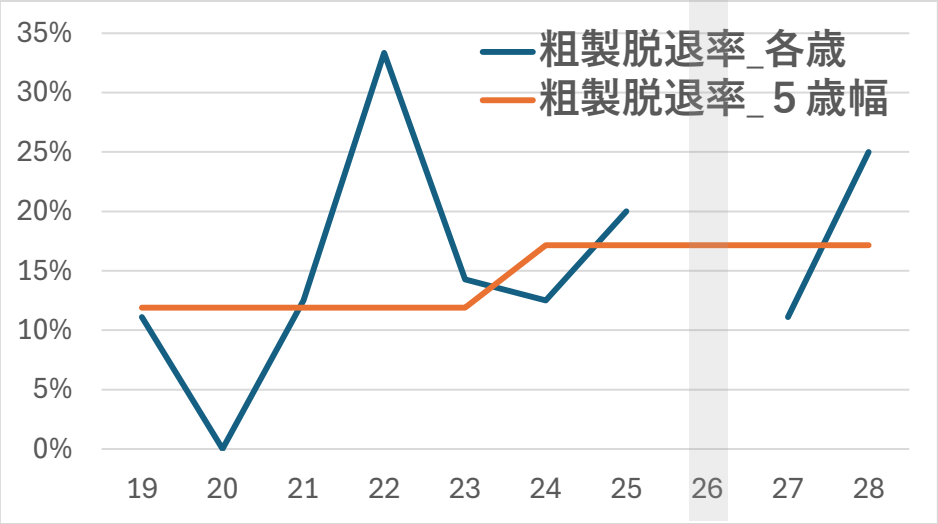


1-3. 粗製脱退率の作成

年齢ごとのばらつきを抑える観点から、各歳ごとでなく**算定年齢幅を拡大**（例えば**5歳幅**）し、人数規模を大きくした上で粗製脱退率を作成することもある

(例)

年齢	各歳幅			5歳幅		
	加入者	脱退者	粗製脱退率	加入者	脱退者	粗製脱退率
19	9	1	11.11%	42	5	11.90%
20	12	0	0.00%	42	5	11.90%
21	8	1	12.50%	42	5	11.90%
22	6	2	33.33%	42	5	11.90%
23	7	1	14.29%	42	5	11.90%
24	8	1	12.50%	35	6	17.14%
25	10	2	20.00%	35	6	17.14%
26	0	0	#DIV/0!	35	6	17.14%
27	9	1	11.11%	35	6	17.14%
28	8	2	25.00%	35	6	17.14%



DBに関する数理実務ガイダンス

各年齢集団が小集団の場合、各歳ごとでなく**算定年齢幅を拡大**し、人数規模を大きくした上で粗製脱退率を作成することができる（例示）**5歳幅毎**の粗製脱退率

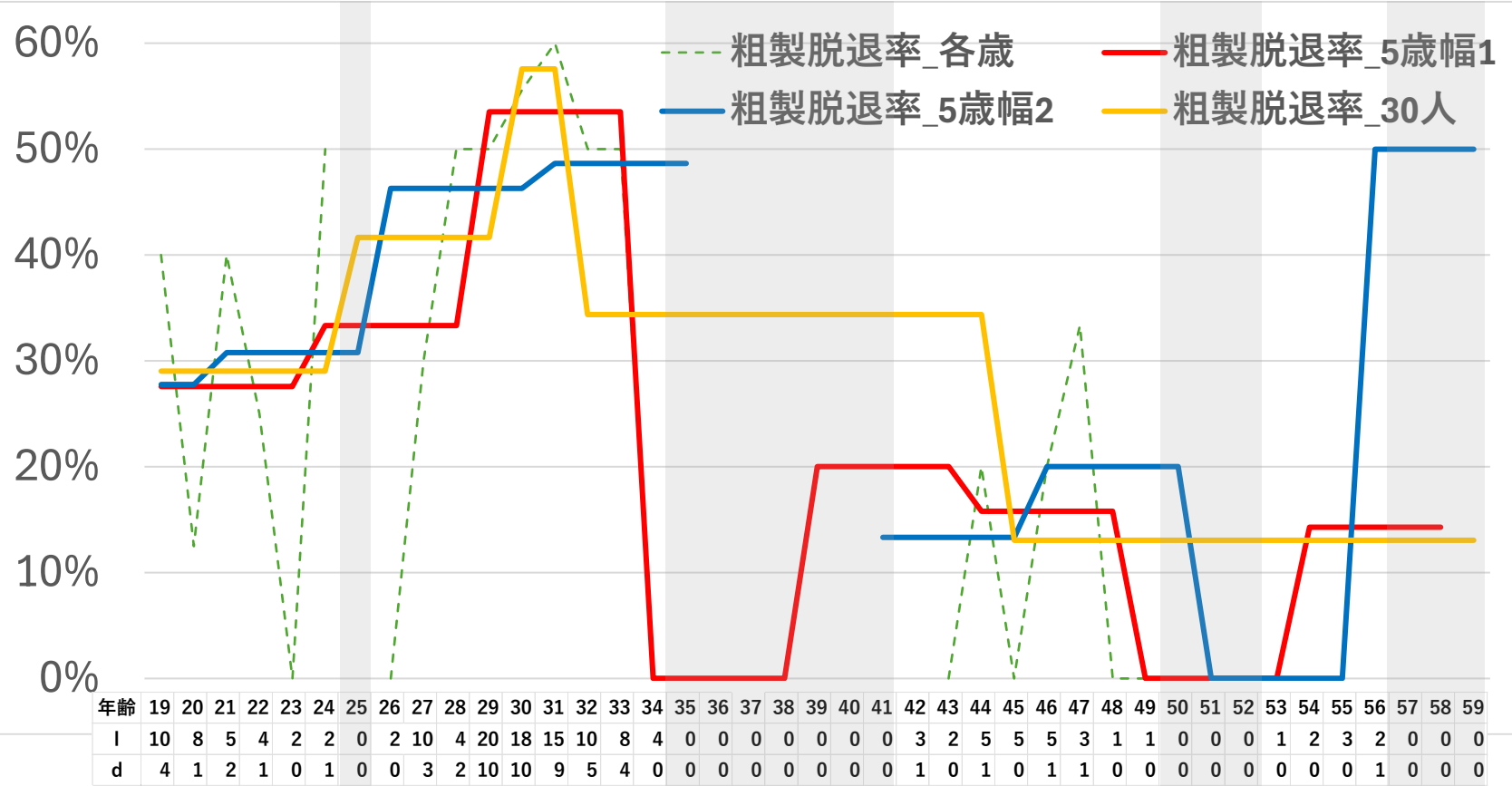
厚生年金基金実務基準第1号 代行保険料率の計算に関する細目

年齢毎の粗脱退率の分母となる人数が算定年度幅を通算して30人以上となる時その脱退率は有意と判定し実績値を用いるが、30人未満の場合にはその年齢の前後に同年齢だけ、**分母となる人数が30人以上となるまで算定年齢幅を拡大**(中略)し、該当する年齢幅の分母・分子をそれぞれ合算して脱退率を算定する *1

* 1 代行保険料率の計算では、予定脱退率の平滑化（補整）は行われない

1-3. 粗製脱退率の作成

具体例：算定年齢幅の拡大方法による粗製脱退率の違い



- ✓ **5歳幅1**：19歳から5歳幅
5歳幅2：21歳から5歳幅とする
- ✓ 5歳幅1の場合、最も高年齢では54歳～58歳、59歳、となり均等に年齢を割り当てられない
- ✓ 5歳幅2の場合、依然、**加入者数0の集団が存在するという問題が残っている**
- ✓ 算定年齢が「30人」を超えるように定めると、**33歳以降2つの値しか取っておらず、年齢ごとの実績の違いを十分に反映できない**

課題2：小規模集団の場合、算定年齢幅の拡大方法により結果が大きく変わり得る



1-4. 粗製脱退率の補整

補整に関する取扱い

DBに関する数理実務ガイダンス（抜粋）

(A) 統計資料の偶発的な要素や変動を排除し、当該制度の脱退傾向を示す率として、凹凸の少ない滑らかな曲線となるように粗製脱退率を補整することができる。

(例示) ・移動平均法による補整
・グレヴィルの補整式による方法

今回提案する新脱退率算定方法は
(A)に関するもの

(B) 特異な年齢における脱退についても、年金財政の健全性を考慮し、補整を行う。

(C) 実績脱退率の水準、傾向、安定性又は将来の動向（見込み）等を考慮して、年金財政の健全性の観点から必要と認める場合は、脱退率の割（増）掛け等により補整することができる。

(D) 年齢別に異なる率により割（増）掛けを行うことができる。

(E) 掛金率が低下する方向への割（増）掛けは原則として行わないこととする。

- ① 算定に用いる過去脱退実績の「算定期間」
- ② 過去脱退実績からの「除外データ等の抽出」
- ③ 過去脱退実績から算出される「粗製脱退率の算定方法」
- ④ 「補整脱退率の作成方法・粗製脱退率の補整方法」
- ⑤ 定年延長等の際の「最終年齢までの脱退実績がない場合の取扱い」
- ⑥ 「異例処理」（過去実績が使用できない場合等）

(B)～(E)は、(A)の平滑化補整以外の各種補整に関する記載だが、主な内容は、

「健全性」を考慮し、掛金が低下しない方向、債務が減少しない方向への補整を検討するというもの
→ 新脱退率算定方法でも、必要に応じて実施することを想定

1-4. 粗製脱退率の補整

(A)：平滑化補整に関する記載

DBに関する数理実務ガイダンス（抜粋）

(A) 統計資料の偶発的な要素や変動を排除し、当該制度の脱退傾向を示す率として、凹凸の少ない滑らかな曲線となるように粗製脱退率を補整することができる。

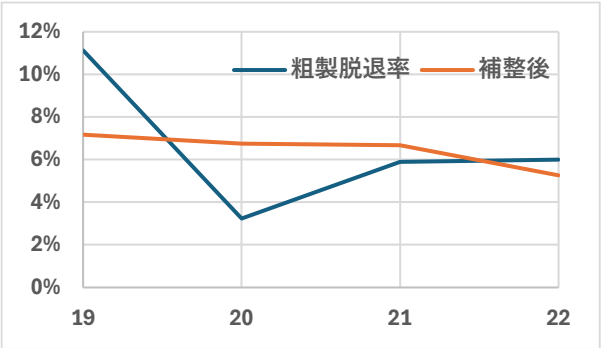
- (例示)
- ・移動平均法による補整
 - ・グレヴィルの補整式による方法

- ✓ 現行記載では移動平均法とGrevilleの方法の2種類が例示されている（詳細は次頁以降で説明）
- ✓ いずれの方法も粗製脱退率を作成した上でそれを補整する形
- ✓ なお、今回提案する新脱退率算定方法は、粗製脱退率を補整するのではなく、粗製脱退率を経由せずに脱退実績から直接求める形になるので、その点で現行の2方法とは異なる

1-4. 粗製脱退率の補整

粗製脱退率の補整

- ✓ 粗製脱退率そのものを脱退率とする場合、年齢ごとのばらつきが大きくなることが多い
- ✓ “年齢毎の脱退率に大きなばらつきが無いはず”（年齢別脱退率を表すグラフの形状は滑らか）との信念・経験則（今まで問題が生じていない）から粗製脱退率を補整することが一般的
- ✓ 移動平均法は、各年齢でその前後の粗製脱退率の単純平均を取るもの
- ✓ Grevilleの方法は、重み付きでの平均（重みの係数は年齢に拠らない固定の値）を取るもの



一般的な補整方法	補整式 $\{c_x\} \rightarrow \{q_x\}$
$2n + 1$ 点移動平均法	$q_x = \sum_{t=-n}^n \frac{1}{2n + 1} c_{x+t}$
m 次 $2n + 1$ 項式 Greville法	$q_x = \sum_{t=-n}^n L_t c_{x+t} \text{ *1}$

c_x : 粗製脱退率
 q_x : 補整脱退率

(参考)
 $m = 3$ のときの L_t

$t \setminus n$	2	3	4	5	6
0	0.559440	0.412588	0.331140	0.277944	0.240058
1	0.293706	0.293706	0.266557	0.238693	0.214337
2	-0.073426	0.058741	0.118470	0.141268	0.147356
3	—	-0.058741	-0.009873	0.035723	0.065492
4	—	—	-0.040724	-0.026792	0.000000
5	—	—	—	-0.027864	-0.027864
6	—	—	—	—	-0.019350

* 1 $L_{-t} = L_t$ を満たす。厳密には外挿値を計算するための係数も必要になるが、ここでは触れない

1-4. 粗製脱退率の補整

移動平均法

色々なバージョンはあるが、例えば、5点移動平均法は以下の算式のとおり

$$q_x = \frac{1}{5}(c_{x-2} + c_{x-1} + c_x + c_{x+1} + c_{x+2})$$

- c_x : 年齢 x における粗製脱退率
 - q_x : 年齢 x における平滑化された脱退率
-
- ✓ 移動平均法は、上記のように単純平均を取るものだが、各年齢で上記の対応を行うことにより、脱退率の曲線は滑らかになる
 - ✓ 一方、前後の年齢の粗製脱退率について単純に平均を取るとは適切かといった問題が生じる
 - ✓ 例えば、1歳上の粗製脱退率と2歳上の粗製脱退率が平均を取る時に同じ重みで良いのかといった論点があり、Grevilleの方法がその解決策の一つになる
 - ✓ 実務では移動平均法を連続で2回取する場合もある

1-4. 粗製脱退率の補整

Grevilleの補整法

- ✓ どのように重みを付けるかを考える際、
 1. 乖離度（出来るだけ粗製脱退率から乖離しないようにすること）
 2. 平滑度（出来るだけ滑らかにすること）の最小化を両立するようにするという考え方があり、Grevilleの方法はその一つ
- ✓ 移動平均法と同様、Grevilleの方法にも色々とバージョンはあるが、例えば「3次5項方式」では、以下のように重みが付いたものとなる

$$q_x = -0.0734c_{x-2} + 0.2937c_{x-1} + 0.5594c_x + 0.2937c_{x+1} - 0.0734c_{x+2}$$

- ✓ 「3次」は3次多項式近似を意味し、「5項」は推定年齢の前後5歳の年齢を使用することを意味する
- ✓ Grevilleの方法は、次数：3次、項数：もっと大きいものが使用されることが多い

課題3（Grevilleの補整法）

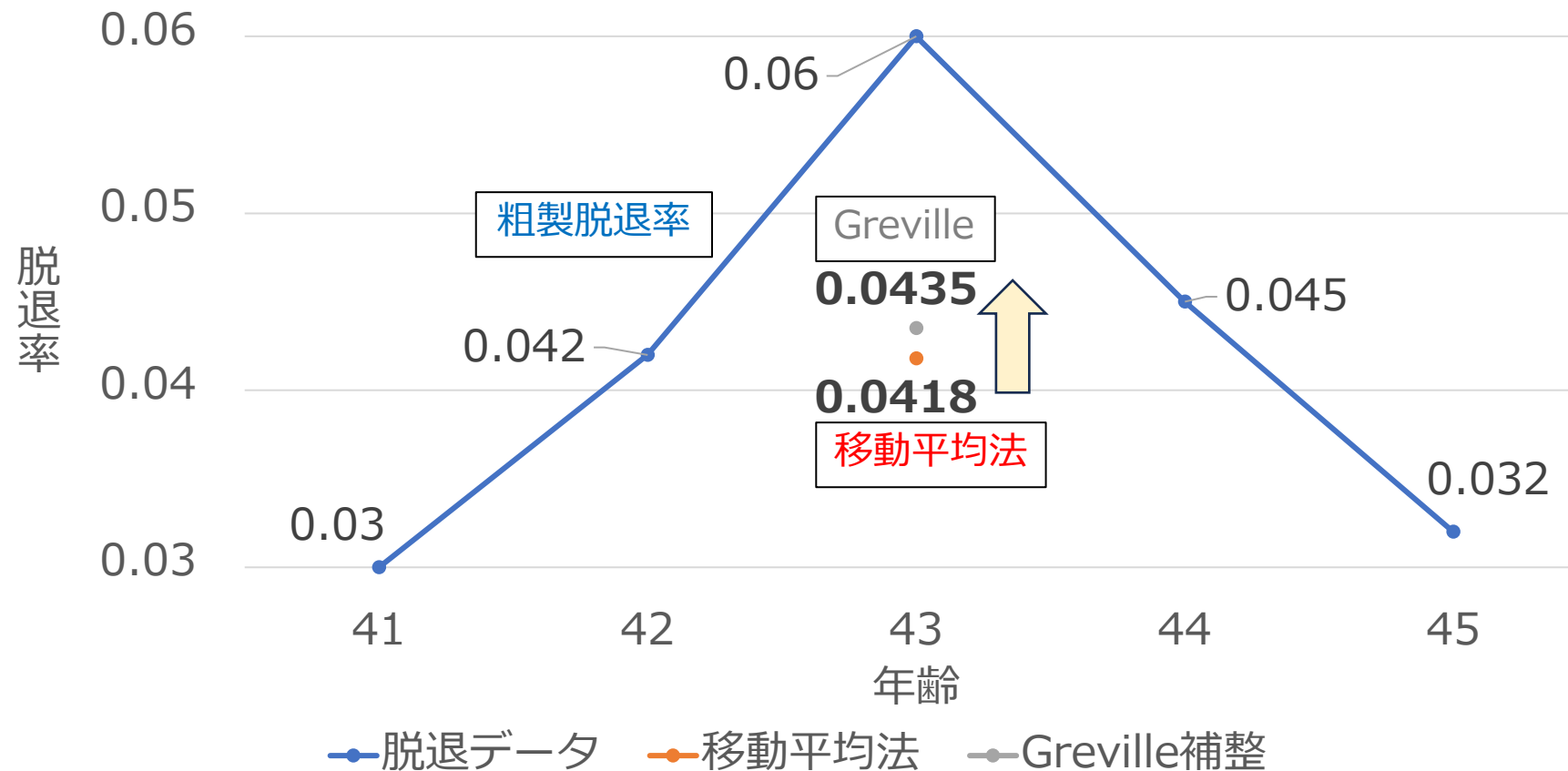
- ✓ 補整結果がマイナスになる場合があり、補整結果をそのまま使用できない場合がある



1-4. 粗製脱退率の補整

$$q_x = -0.0734c_{x-2} + 0.2937c_{x-1} + 0.5594c_x + 0.2937c_{x+1} - 0.0734c_{x+2}$$

(例) 5点移動平均法とGrevilleの方法（3次5項方式）の比較（43歳）

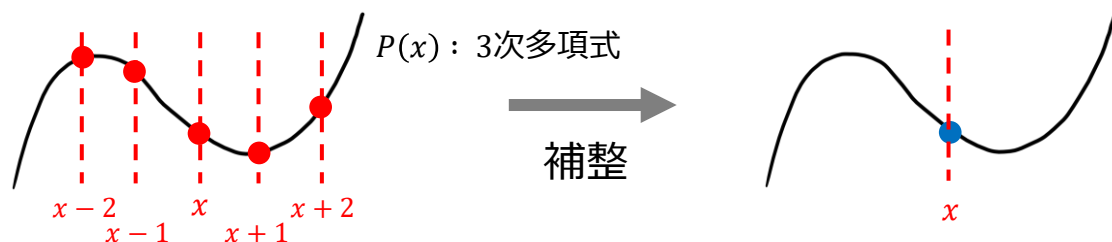


Grevilleの方法（推定値0.0435）は移動平均法（推定値0.0418）よりも中央の値を重視

1-4. 粗製脱退率の補整

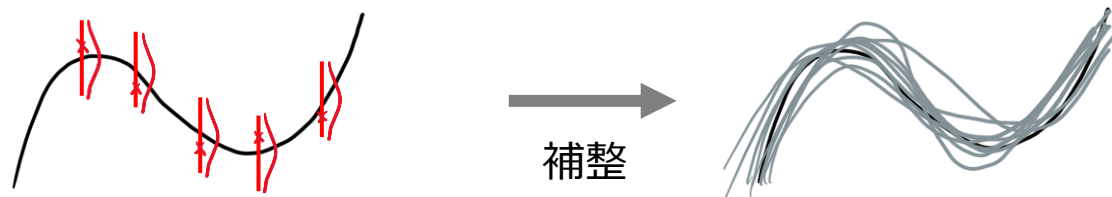
Grevilleの補整法の重み $\{L_t\}$ の算出方法 : 次の①②を満たすように決定

① 3次多項式再現条件



- ✓ 真の脱退率が3次関数であり、データがその上に乗っていると、**補整後も乗る**
- ✓ 条件に合致すれば、補整結果が実際の値と一致するものであり、**乖離度を意識したもの**

② 補整後の系列の**3回差分** $\Delta^3 q_x := q_{x+3} - 3q_{x+2} + 3q_{x+1} - q_x$ の**2乗和** $\sum_x (\Delta^3 q_x)^2$ を平均的に最小化



- ✓ 3回差分を全年齢で0に近づけることで、**鋭い変化（差分の4次以降の成分）を抑制**
- ✓ 補整後の系列が**滑らか**になる

(注) ここでは平滑度が小さいほど滑らかとする

- ✓ Grevilleの補整法は、①のように**乖離度（補整前後の系列の近さ）**を小さくすることと、②のように**平滑度（補整後の系列の滑らかさ）**を小さくすることを両立させた方法と言える
- ✓ この**乖離度**と**平滑度**という観点はデータサイエンスの観点では重要（後述）

1-4. 粗製脱退率の補整

(ケース2) もし31歳の加入者があと98人いたら、50人くらい脱退していたのでは？

課題 4（両補整方法について共通）

✓ 粗製脱退率の価値を、その算出分母の大きさに関わらず同一としている



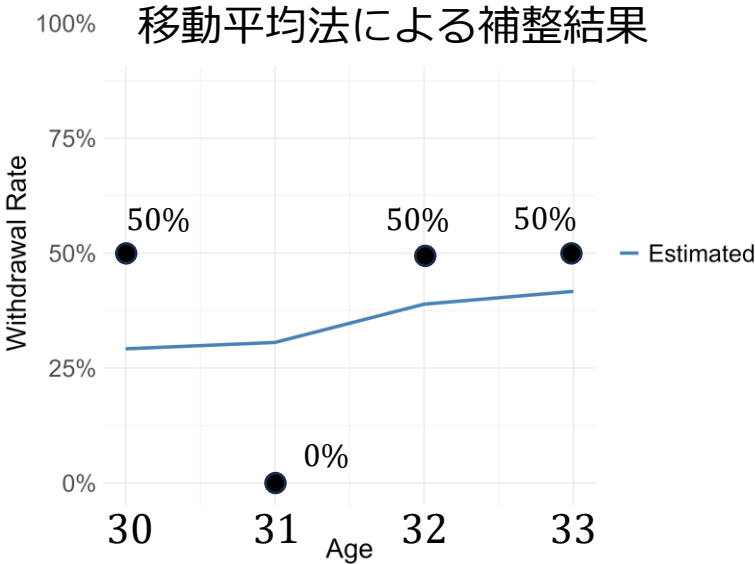
(例) 次の2つの脱退実績に対して、移動平均法による補整結果は**同じもの**になる

年齢	ケース 1	ケース 2	粗製脱退率
30	1/2 人	50/100 人	50%
31	0/2 人	0/2 人	0%
32	1/2 人	50/100 人	50%
33	1/2 人	50/100 人	50%

補整 →

ケース 1：全年齢で実績数同じ
→ 右図でも適切

ケース 2：31歳の実績だけサンプル小
→ 31歳の実績をほぼ無視したもの
(50%一定率) になってほしい



脱退データでなく、粗製脱退率に対して補整を行うことに起因している

1-5. 実務上の課題

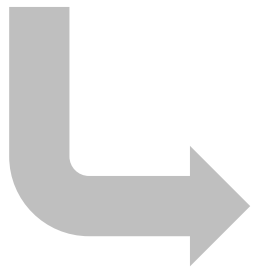
以上、実務上の課題をまとめると・・・

1. 【粗製脱退率】 **分母が0**のとき、そのまま計算できない
2. 【粗製脱退率】 「算定年齢幅の拡大」によって**結果が大きく変わり得る**
3. 【Greville】 補整結果が**マイナス**になり得る
4. 【補整】 移動平均やGrevilleなどの補整方法は、**粗製脱退率の価値を全て同一**と扱う



これらは特に**小規模集団**において顕著に現れる

(注) 現行の方法は、**集団の規模が年齢別に十分に大きければ有効な方法**
(各歳ごとの粗製脱退率の持つ価値が十分に大きいため)



- データサイエンスを活用し、脱退率算定に**一般化加法モデル (Generalized Additive Model ; GAM)** ^{*1} という確率モデルを導入することで、上記の課題に対して一定の解決が図れる
- GAMの方法は、**乖離度、平滑度の両方をバランスよく考慮でき**、これは**Grevilleの方法の発展形**になっているとも言える

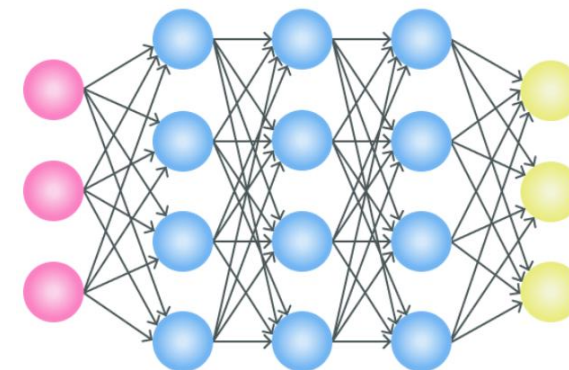
* 1 Hastie and Tibshirani [1986]

2. データサイエンスの活用

2-1. データサイエンスの活用

近年、深層学習といった機械学習手法がトレンド

- ✓ 大規模言語モデル（LLM）
- ✓ 画像認識
- ・ ・



脱退率推定に深層学習などの機械学習手法を用いると・・・

推定精度は上がるかもしれないが、なぜこの結果が得られたのか？を説明することは非常に困難（ブラックボックス化されている）

→ いくら精度が良かったとしても、顧客説明の観点からは望ましくない

（例）次のように脱退率算定について説明できることが望ましい

- ・ 「〇〇が最小となるように定めている」
- ・ 「〇〇という考え方で脱退率を推定している」

脱退率推定の問題を数学的に扱うため、ここでは「補整とは何か？」を数学的に定式化し、それを確率統計の問題と捉えて、その解を検討する

2-2. “補整”の定式化

“補整する”ということを捉え直す

補整前後で満たすべき性質

1. 決定脱退率は粗製脱退率からかけ離れない
2. 決定脱退率は全体に滑らかな形態をしている
3. 各年齢の粗製脱退率の価値には違いがある

↓
数学的に扱うため、これらを定量化する

1を乖離度として定量化

乖離度：粗製脱退率 c_x と補整後 q_x との乖離度合

(例)

- 2乗誤差 = $\sum_x (q_x - c_x)^2$ / 年齢ごとの価値の違いを反映した乖離度
- 重み付き2乗誤差 = $\sum_x w_x (q_x - c_x)^2$

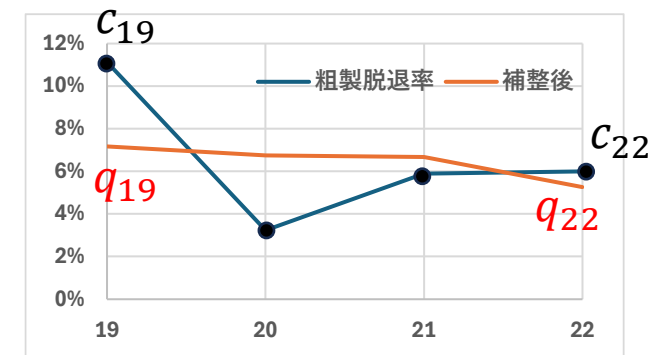
2を平滑度として定量化

平滑度：補整脱退率 q_x の滑らか度合

(例)

- m 次差分平滑度 = $\sum_x (\Delta^m q_x)^2$; $\Delta q_x := q_{x+1} - q_x$
- m 次微分平滑度 = $\int_x \{q_x^{(m)}\}^2 dx$

“小さい方が滑らか”



- ✓ “補整する”ということを、**乖離度、平滑度をバランス良く小さくなるように q_x を決める**という問題に置き換える
- ✓ 性質3は**乖離度を工夫して定義することで考慮する**

2-2. “補整”の定式化

乖離度と平滑度どちらか一方でなく、**両方をバランス良く小さくすることが良い**補整に繋がるのか



乖離度と平滑度：トレードオフの関係

✓ 乖離度のみ重視すると・・・

$$\text{乖離度} = 0 \Leftrightarrow \forall x, q_x = c_x$$

“過剰適合”という

となり、**平滑度合いが失われる（補整が行われず、粗製脱退率そのものになってしまう）**

✓ 平滑度のみ重視すると・・・

$$\text{平滑度} = 0 \Leftrightarrow q_x : \text{高々 } m-1 \text{ 次多項式 (} m \text{ 次微分平滑度のとき)}$$

となり、**曲線の形状が限定されてしまい、データへの当てはまり度合（乖離度）が失われる**



両者のバランスを取るパラメータ λ を **平滑化パラメータ** という

“補整する” ($c_x \mapsto q_x$) ということを、次の数学的な問題として**2ステップ**で定式化する

ステップ1. $\lambda > 0$ を任意に固定して、“**乖離度** + $\lambda \times$ **平滑度**” が**最小**となる関数 q_x を求める

ステップ2. この解 q_x は λ を変えるごとに異なるが、このうち“**最も良い**”と考えられる q_x を選ぶ

2-2. “補整”の定式化

今回提案する手法もこの形

(ステップ1) 補整の問題は適当な関数空間内での最小化問題として定式化できる

$$\min_f (\text{乖離度} + \lambda \times \text{平滑度}); \quad \lambda: \text{正の定数}$$

ステップ1の最小化問題 (λ : 固定) の解は一般に知られている!

命題

(自然スプラインについては後述)

平滑度: m 次微分 (乖離度は任意) の場合、解は自然 $2m - 1$ 次スプライン関数となる! ^{*1}

- ✓ 乖離度が同一のものの中で平滑度を最小にするのが自然 $2m - 1$ 次スプライン関数であることから、この命題が示される
- ✓ この命題を用いると、自然 $2m - 1$ 次スプライン関数に限定して最小化問題を解けば良いことになるが、この最小化問題の解は数値的に求めることができる!

今回提案する一般化加法モデルはこの統一的な解決策を与える

残された問題:

1. 人数規模の違いを補整に反映するために、乖離度をどのように設定すれば良いか?
2. 平滑化パラメータ λ をどのように決めれば良いか? (ステップ2)

*1 最小化問題の解となるスプライン関数は、全ての観測値をknotとしたもの。

歴史的には、Schoenberg[1964]により乖離度 = $\sum_x w_x (q_x - c_x)^2$ の場合にまず示された。一般の場合はカーネル法の表現定理などを用いて示される。

2-2. “補整”の定式化

自然スプライン関数とは

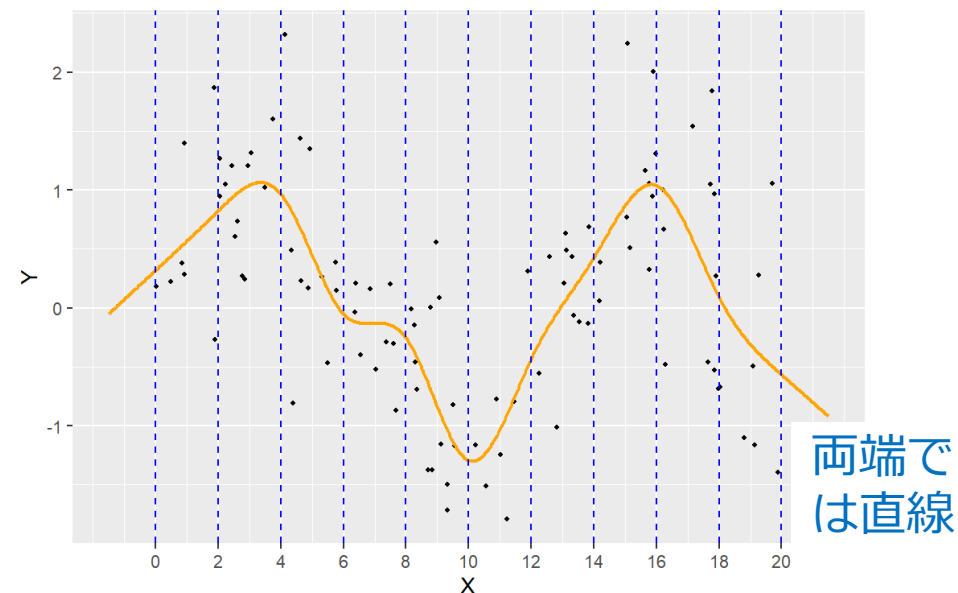
自然 $2m - 1$ 次スプライン関数 $s(x)$ とは、次を満たすもの

- 結び目(knot)ごとに異なる $2m - 1$ 次多項式
- これらは各knotにおいて $2m - 2$ 回微分まで連続
- 両端では高々 $m - 1$ 次式

→ パラメータ数は K (=knotの数) に等しい
(特に全体の集合は有限次元ベクトル空間)

$m = 2$ 、knotの個数 $K = 11$ の場合

natural cubic spline ($m = 2$, total knots = 11)



(注) 最小化問題の解となるスプライン関数は、全ての観測値をknotとしたもの

2-2. “補整”の定式化

“補整”の定式化の結果・・

- ✓ 適切な λ を設定した上で、最小化問題

$$\min_f (\text{乖離度} + \lambda \times \text{平滑度})$$

の解を求めれば良いことが分かった

- ✓ これはGrevilleの方法との連続性からも自然な考えである
- ✓ しかし次の2つの問題が残る：

1. 「人数規模の違いを反映した乖離度はどのように定めればよいか？」
2. 「平滑化パラメータ λ はどう決めればよいか？」

- ✓ これに対するアプローチの1つとして、「決定論的視点」から「確率論的視点」で現行手法を更に捉え直すことを考える
- ✓ 「データへの当てはまりが良い」（乖離度が小さい）という事象を統計の言葉で見つめ直す

2-3. 決定論的から確率論的な手法へ

“脱退するかどうか”を確率的な事象（二項事象）と捉える

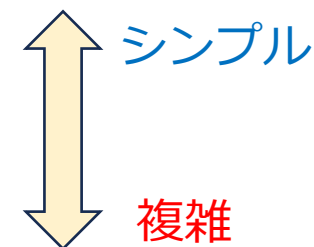
- ✓ 各年齢での脱退率は、確率モデルでいうと二項分布（脱退するかしないか）に従うと考えられる
- ✓ 具体的には、次の確率変数 Y は真の脱退率を q_x^{true} とした二項分布 $\text{Bin}(1, q_x^{\text{true}})$ に従うとみなせる

$$Y = \begin{cases} 1, & \text{if. } x \text{ 歳の加入者が次の1年間に脱退する} \\ 0, & \text{if. } x \text{ 歳の加入者が次の1年間に脱退しない} \end{cases}$$

- ✓ 一方、実務では、確率分布の状態のまま脱退率を取り扱うのは複雑であるし、算定にあたり様々な調整を行うこともあるため、
確率の要素を無くして二項分布の期待値をその年齢での脱退率として決定論的に設定している
- ✓ 乖離度や平滑度を定めるだけであれば、確率分布を考える必要は必ずしも無いが、確率分布を持つモデルであれば、尤度（確からしさ）の考えを取り入れられる等のメリットがあり、より複雑で精度が高いものを作ることができる

確率分布を持つモデルのうち、回帰モデルの例（他も多数あり）

- 線形回帰モデル（Linear Model ; LM）
- 一般化線形モデル（Generalized Linear Model ; GLM）
- 一般化加法モデル（Generalized Additive Model ; GAM）



2-3. 決定論的から確率論的な手法へ

LM、GLM、GAMの比較（1変数の場合）

- ✓ LMは正規分布しか扱えないが、GLMやGAMは二項分布等も扱える
- ✓ 目的変数 Y と説明変数 x との関係：LMは線形だが、GAMは非線形

モデル	期待値	確率分布	回帰式	回帰式の線形/非線形
線形回帰モデル (LM)	$E[Y x] = \beta_0 + \beta_1 x$	正規分布	$Y = \beta_0 + \beta_1 x + \varepsilon$	線形
一般化線形モデル (GLM)	$E[Y x] = \mu = g^{-1}(\eta),$ $\eta = \beta_0 + \beta_1 x$	指数型分布族 (二項分布を含む)	$g(\mu) = \beta_0 + \beta_1 x$	線形
一般化加法モデル (GAM)	$E[Y x] = \mu = g^{-1}(\eta),$ $\eta = f(x) : \text{非線形関数}$	指数型分布族 (二項分布を含む)	$g(\mu) = f(x)$ $f(x) : \text{非線形関数}$	非線形

一般化加法モデル（GAM）を用いることで、

✓ 目的変数と説明変数の間の非線形な関係も捉えることができる

✓ 二項事象に対してもモデリングができる

脱退率に適用できそう！

2-3. 決定論的から確率論的な手法へ

人数規模の違いを
考慮した乖離度として
尤度を使えば良いのか



回帰式はどのように決定するか？

- LM、GLMなど、確率モデルの一般的な回帰式の決定方法は**最尤法**
 - これは、**乖離度** $= -\log(\text{尤度})$ （ここでは**負の対数尤度**と呼ぶ）が最小となるようにパラメータを決定するもの
 - 尤度を使用することで、**データへの当てはまりを確率論的に表現**できる
 - しかしGAMにおいては、非線形関数をどう選ぶかが必要となり、最尤法では決定できない
 - そこで乖離度だけでなく、**平滑度**の考えも取り入れ、
“乖離度 $+\lambda \times$ 平滑度”の最小解である自然スプライン関数で回帰する
- 尤度を用いると、データへの当てはまり（乖離度）を確率論的に表現できる
 - 対数尤度は観測の寄与を総和するため、**サンプルサイズの違いが値に内在的に反映**される
 - LM、GLM、GAMの中では、**GAMのみが乖離度と平滑度の両方を考慮**している
 - Grevilleの方法も同様であり、**GAMはGrevilleの方法の発展形**と考えることもできる

2-3. 決定論的から確率論的な手法へ

詳細は後述

提案手法：一般化加法モデル（GAM）

- 乖離度：負の対数尤度 ($= -\log(\text{尤度})$)
- 平滑度： m 次微分平滑度

として、 λ を固定するごとに、“乖離度 $+\lambda \times$ 平滑度”の最小解を求め、 λ ごとの解のうち、最も予測精度が高いものとして、統計的規準を用いて λ を定める

残された問題に対する解

1. 「人数規模の違いを反映した乖離度はどのように定めればよいか？」
→ 尤度を使用することで対応
(1人1人に対して尤度を算定するため、人数規模が大きいところの価値が高まる)
2. 「平滑化パラメータ λ はどう決めればよいか？」
→ 統計的規準を用いることで、“最も予測精度が高くなる”ように選択できる（後述）

一般化加法モデルを用いて得られる脱退率は、標語的には次のように言える

“GAMという確率モデルを使って、
乖離度と平滑度を（予測の観点で）ベストな割合で最小化を図ったもの”

3. 一般化加法モデルを用いた脱退率算定

3-1. GAMの概要

【前提】 次のデータが得られているとする

- ✓ 加入者数 l_x (年齢 x)
- ✓ 脱退者数 d_x

加工

加入者1人ごとのレコードに分解

- ✓ x : 年齢
- ✓ y : 脱退→1、非脱退→0

加入者数0の年齢
があっても良い

年齢 x	加入者数 l	脱退者数 d
19	5	1
20	2	1
21	3	1
22	4	1
23	0	0
24	1	1
25	1	1
26	2	0
27	10	3
28	30	2
29	20	10
30	18	10
31	40	21
32	20	10
33	8	4
34	20	0
35	2	1
36	18	0
37	17	0
38	16	1
39	17	1

年齢 x	加入者数 l	脱退者数 d
40	1	1
41	2	1
42	18	2
43	16	3
44	15	1
45	14	0
46	13	1
47	12	1
48	11	0
49	12	0
50	13	0
51	12	1
52	13	0
53	14	0
54	15	0
55	12	0
56	14	1
57	6	0
58	8	0
59	4	1

脱退率算定の実務
で使用している
年齢別脱退データ
と同じもの

```
> head(df, 20)
   x y
1  19 1
2  19 0
3  19 0
4  19 0
5  19 0
6  20 1
7  20 0
8  21 1
9  21 0
10 21 0
11 22 1
12 22 0
13 22 0
14 22 0
15 24 1
16 25 1
17 26 0
18 26 0
19 27 1
20 27 1

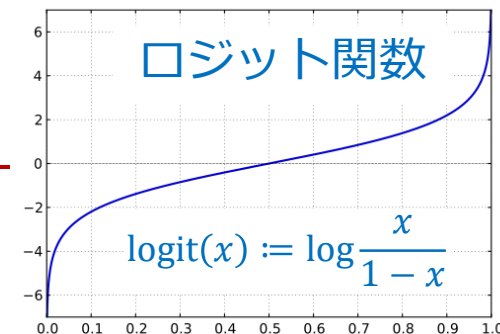
> tail(df, 20)
   x y
460 56 0
461 56 0
462 57 0
463 57 0
464 57 0
465 57 0
466 57 0
467 57 0
468 58 0
469 58 0
470 58 0
471 58 0
472 58 0
473 58 0
474 58 0
475 58 0
476 59 1
477 59 0
478 59 0
479 59 0
```

y は年齢 x を固定するごとに、
脱退率 q_x に従う二項標本で
あるとみなせる

3-1. GAMの概要

一般化加法モデルを用いた脱退率推定

- 観測データ $\{(x_1, y_1), \dots, (x_n, y_n)\}$ について、 y は二項標本 $\text{Bin}(1, q_x)$ とみなす
- 自然 $2m - 1$ 次スプライン関数 $s(x)$ を用いて、次のようにモデル化
$$\text{logit } q_x = s(x)$$



【GAM】 $E[Y|x] = q_x = \text{logit}^{-1}(\eta)$
 $\eta = s(x)$: 非線形

- 観測データを用いて $s(x)$ を次の**2ステップ**で求め、 $\text{logit}^{-1}(\hat{s}(x))$ を**予定脱退率**とする

ステップ1 λ を任意に固定し、 $s(x)$ を次の最小化問題の解として定める

$$\min_{s(x)} L(s(x_1), y_1, \dots, s(x_n), y_n) + \lambda \int \left(s^{(m)}(x) \right)^2 dx$$

乖離度 (二項分布の負の対数尤度)

小さいほどデータへの当てはまりが良い

平滑度 (m 次微分)

小さいほど曲線が滑らか

$$\text{乖離度} = -\log \left\{ \prod_{i=1}^n q_{x_i}^{y_i} (1 - q_{x_i})^{1-y_i} \right\}$$

観測データ (x_i, y_i) の
二項分布に基づく発生確率(=尤度)

ステップ2 あらゆる λ の中で、**最も予測精度が高いもの**を与える λ を、**統計的規準**により決定 (次頁)

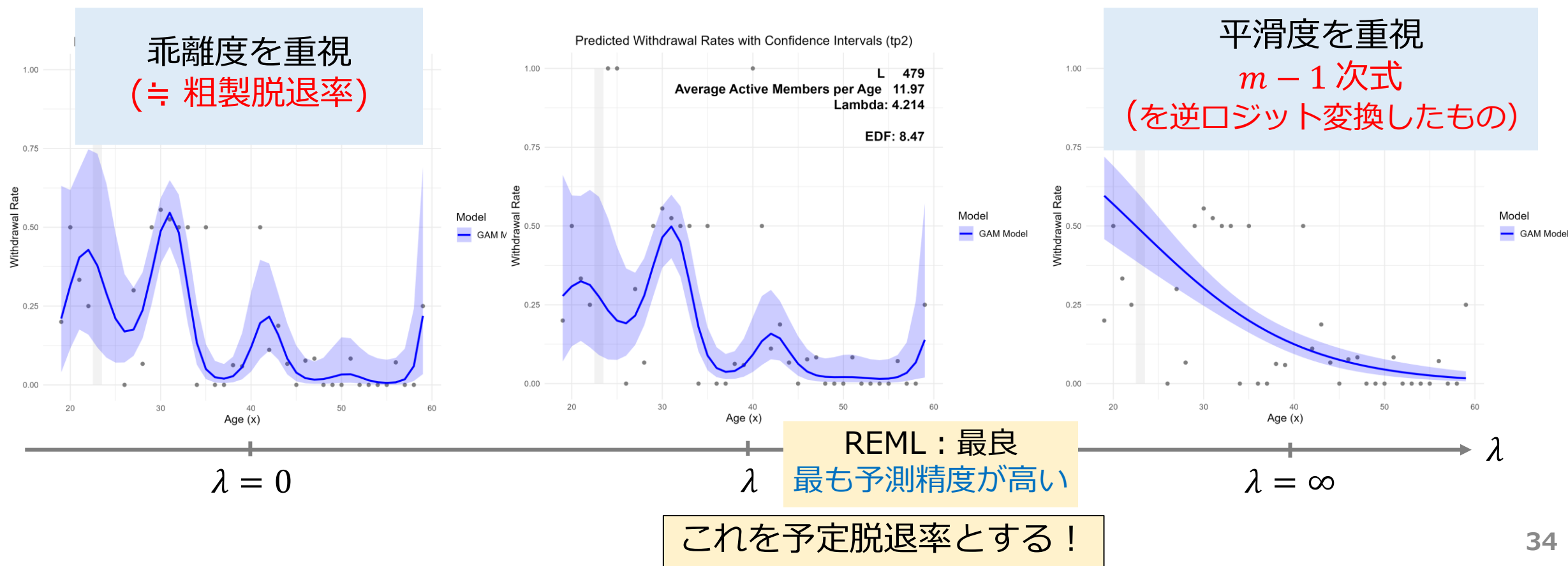
- ✓ ロジット関数で変換するのは、推定結果が**0~1の値を取ることを保証**するため
(統計モデリングにおいては標準的な取扱い)
- ✓ 標語的に言えば、得られる脱退率は**"GAMという確率モデルを使って、乖離度と平滑度を (予測の観点で) ベストな割合で最小化を図ったもの"** (を逆ロジット変換したもの) "

3-1. GAMの概要

平滑化パラメータ λ の選び方（ステップ2）

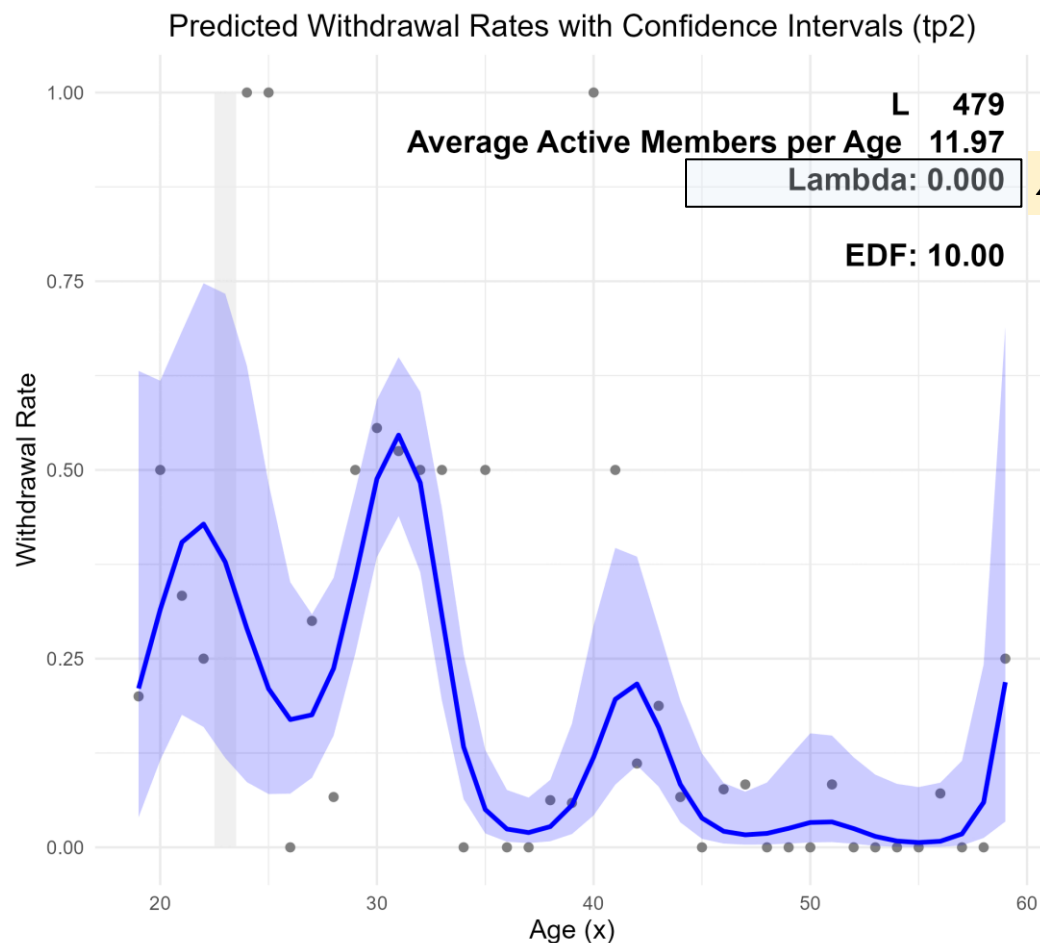
ステップ1により、あらゆる λ について推定結果が計算されているものとする

→ この中で“最も予測精度が高い”ものを統計的手法（REMLなど）により判定

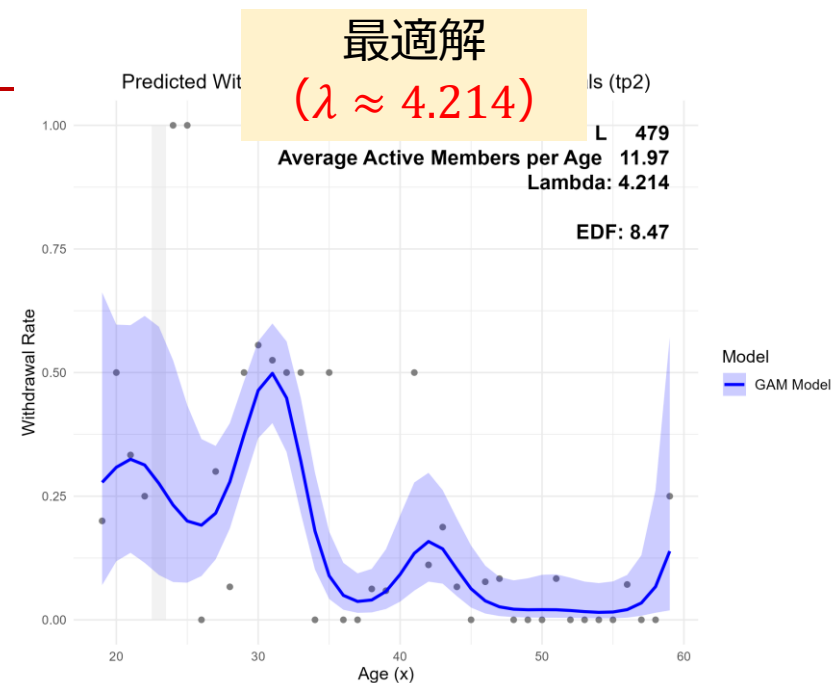


3-1. GAMの概要

【アニメーション】 平滑化パラメータ λ を動かす



λ を大きくする



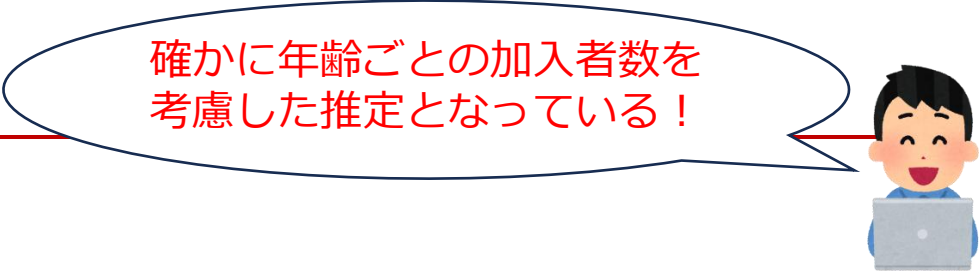
乖離度を重視
($\equiv$ 粗製脱退率)

平滑度を重視
 $m-1$ 次式
(を逆ロジット変換したもの)

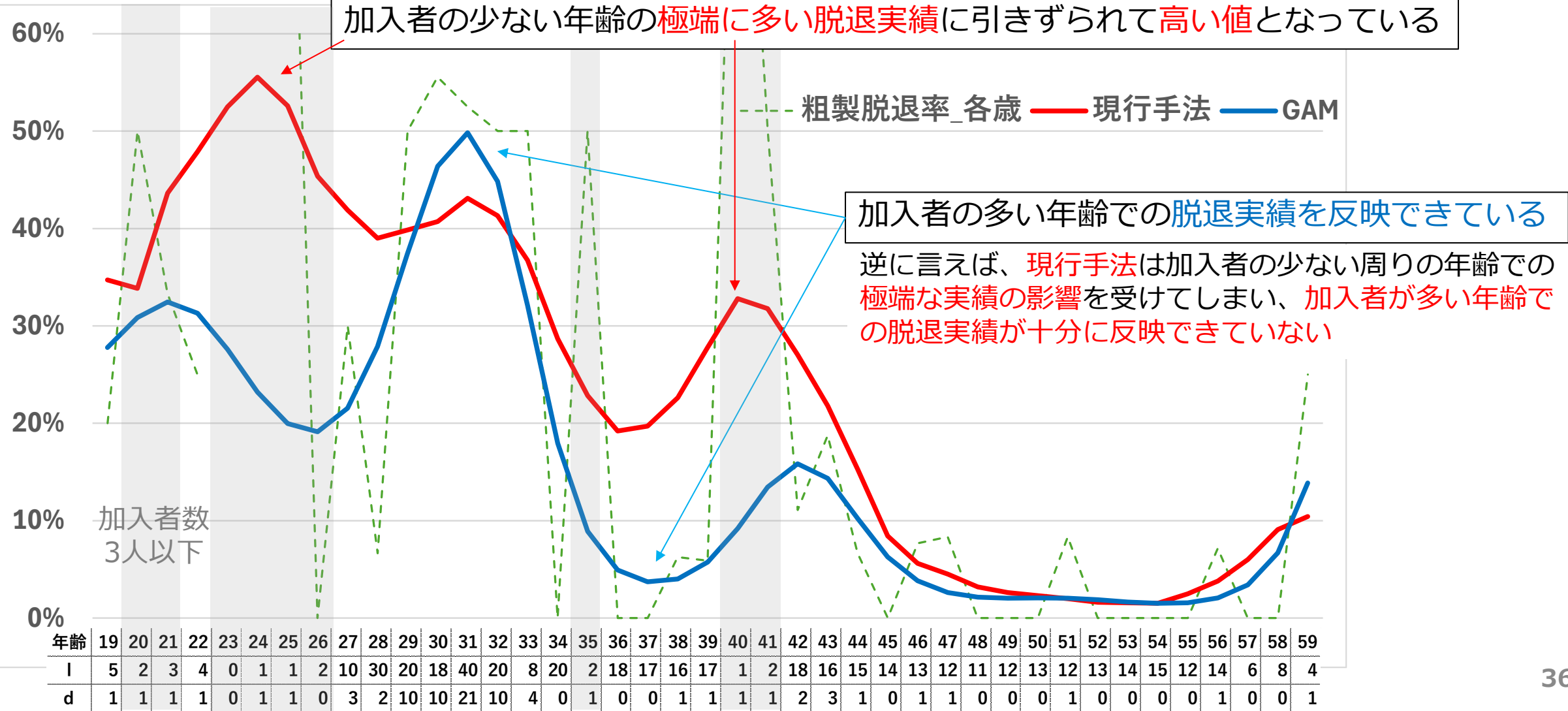


あらゆる脱退率曲線の中で、最も予測精度に優れた脱退率を統計的手法(REMLなど)を使用して自動選択

3-2. GAMの実行例



現行手法とGAMを比較



3-3. GAMの実装

パッケージがあるので
実装は非常に簡単！



Rコード例 (Rパッケージ"mgcv"のgam関数を使用)

```
> head(df, 20)
  x y
1 19 1
2 19 0
3 19 0
4 19 0
5 19 0
6 20 1
7 20 0
8 21 1
9 21 0
10 21 0
11 22 1
12 22 0
13 22 0
14 22 0
15 24 1
16 25 1
17 26 0
18 26 0
19 27 1
20 27 1
```

```
library(mgcv)
# data_original: 加工前データ (年齢別脱退データ)
# df: 加工後データ (1レコードずつに分解)

# モデルあてはめ (基底関数としてtp(m=2)を使用する場合)
model_base <- gam(
  y ~ s(x), data = df, family = binomial
)

# 年齢範囲を作成
all_ages <- seq(min(data_original$x, na.rm = TRUE),
  max(data_original$x, na.rm = TRUE))
newdata <- data.frame(x = all_ages)

# 予測値の計算
pred_base <- predict(
  model_base, newdata = newdata, type = "response"
)
```

脱退率の推定結果

```
> pred_base
      1      2      3      4      5      6
0.27790218 0.30846742 0.32471876 0.31304181 0.27589761 0.23220001
      7      8      9     10     11     12
0.19976263 0.19135149 0.21560168 0.27896396 0.37424367 0.46395363
     13     14     15     16     17     18
0.49846272 0.44853964 0.32118958 0.17956948 0.08903722 0.04947107
     19     20     21     22     23     24
0.03717332 0.04010188 0.05748380 0.09184837 0.13455065 0.15840630
     25     26     27     28     29     30
0.14340949 0.10222043 0.06280655 0.03827989 0.02626722 0.02146370
     31     32     33     34     35     36
0.02033320 0.02057560 0.02037114 0.01883637 0.01659785 0.01510923
     37     38     39     40     41
0.01580000 0.02060332 0.03406248 0.06700238 0.13885323
```

- ✓ Rパッケージ"mgcv"では一般化加法モデルの計算がgam関数で全て実装されている！
→ 最小化問題などを手で実装する必要なし！
- ✓ gam関数にはいくつか引数があり、結果に大きく影響を与えるものもある
→ 引数をどのように設定するかの問題はある
- ✓ 引数の例 (詳細は次頁)
 - bs: 自然スプラインの基底関数
 - m: 微分平滑度の次数
 - method: 平滑化パラメータ λ の選び方

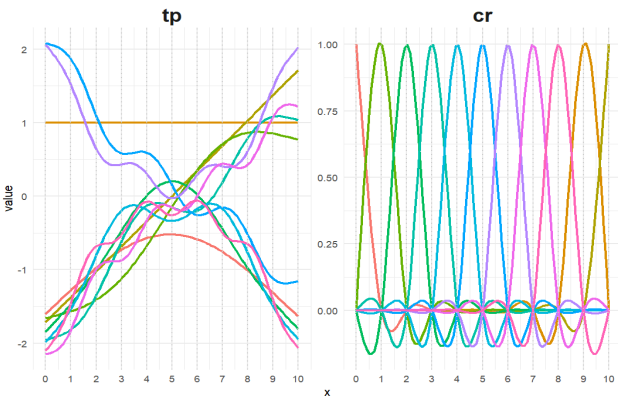
本発表では、最適な引数の設定方法として、
bs="tp"、m=2 or 3、method="REML"
を提案する

3-3. GAMの実装

【参考】mgcvパッケージの関数gamの引数

(1) 自然 $2m - 1$ 次スプラインの基底関数(bs)

tpの使用を提案する



	【既定値】薄板スプライン (tp)	3次回帰スプライン (cr)
微分平滑度の次数 m	任意 (既定値 : 2)	2
適用できる説明変数の数	任意	1次元
knot	指定する必要はなく、観測データの全ての観測値をベースに自動で決定される (過剰適合が生じないよう、重要な基底から上位10個選ばれる)	指定する必要がある (既定値 : 10個 (説明変数を10等分))

(2) 平滑化パラメータ λ の選び方(method)

既定値はGCV.Cpだが、ここではREMLを提案する

	【既定値】 GCV.Cp	REML
λ の選び方	最も予測精度が高いもの (予測誤差基準)	ベイズ的な観点から周辺尤度を最大化するものを選ぶ (周辺尤度基準)
計算時間	短い	長い

3-3. GAMの実装

【参考】mgcvパッケージの関数gamの引数

本発表では、**bs="tp"、 $m = 2$ or 3、method="REML"**を推奨の設定として提案する

根拠

- ✓ 次の**数値実験の結果**から、この場合が**最も良い結果**となった^{*1}
- ✓ 引数bsは"cr"よりも"tp"の方が、また引数methodは"GCV.Cp"よりも"REML"の方が**予測精度の観点で優れている**と、パッケージ"mgcv"の作成者であるWood氏の文献^{*2}にも言及されている
- ✓ 引数methodについて、懸念点であった実行時間について、今回の脱退率の推定においては、取り得る年齢の値が高々50個程度と少なく、実行時間は全く問題にならない

微分平滑度 m の選び方

- ✓ パッケージの既定値である $m = 2$ とするのは**標準的な設定**
- ✓ 3次差分平滑度を使用する**Grevilleとの連続性から $m = 3$ とすることも想定される**
- ✓ 後述の通り、人数規模が小さく、無情報に近い場合の推定結果は $m - 1$ 次式（を逆ロジット変換したもの）に近づくが、これを**直線としたい場合は、 $m = 2$ とすれば良い**

* 1 車谷&岩沢 [2025]

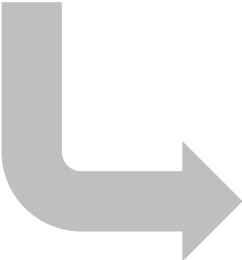
* 2 Wood [2025]

3-4. GAM：実用上の課題の解決策

GAMによる推定が、実務上の課題を解決していること

実務上の課題（再掲）

1. 【粗製脱退率】 **分母が0**のとき、そのまま計算できない
2. 【粗製脱退率】 「算定年齢幅の拡大」によって **結果が大きく変わり得る**
3. 【Greville】 補整結果が **マイナス** になり得る
4. 【補整】 移動平均やGrevilleなどの補整方法は、 **粗製脱退率の価値を全て同一** と扱う

- 
1. 特定の年齢で加入者数（分母）が0であっても、一般化加法モデルは適用可能
（実績無し → 尤度には何も反映されない）
 2. そもそも **粗製脱退率を作成しない** のでこの問題は生じない
 3. 一般化加法モデルではロジット関数で変換したものを推定対象としているため、**推定結果は必ず0から1までの範囲** となる
 4. 乖離度として **負の対数尤度** を用いているため、加入者数が多い年齢では **その脱退実績分だけ尤度に反映され、各年齢における実績数を自動的に考慮** した方法となる
（例：次頁）



3-4. GAM : 実用上の課題の解決策

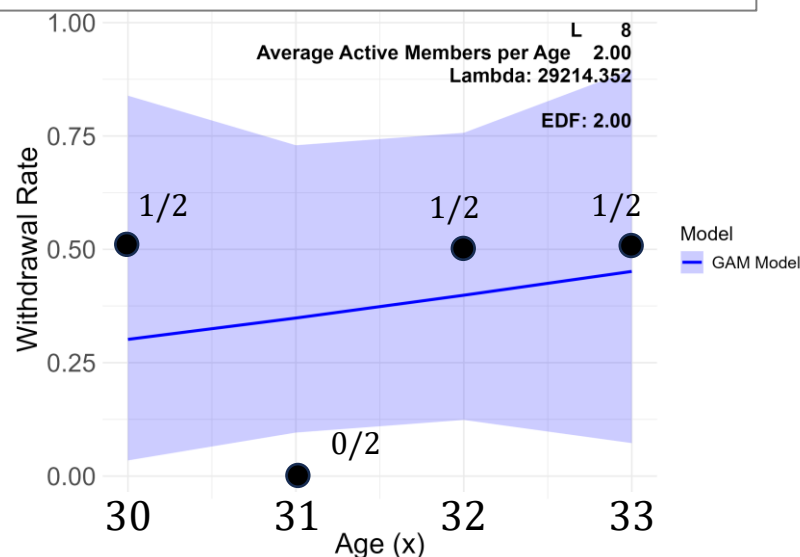
確かに年齢ごとの加入者数の違いを考慮した推定となっている！



(課題4の例)

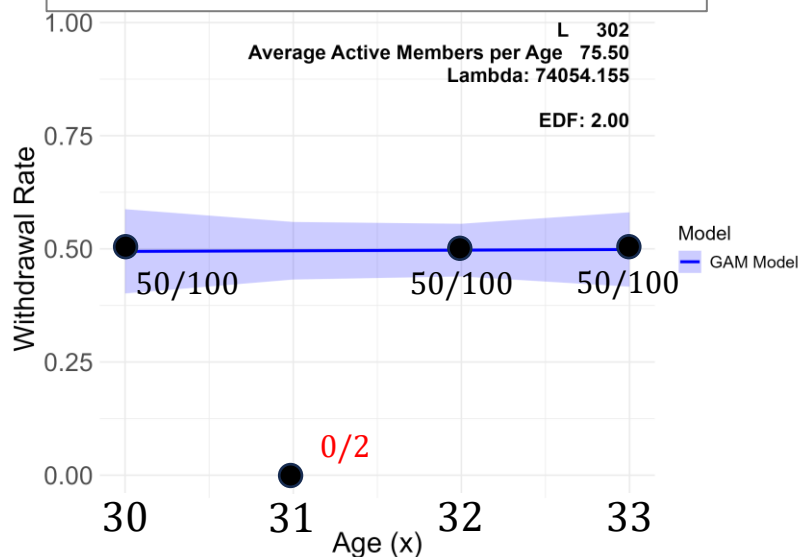
年齢	ケース1	ケース2	粗製脱退率
30	1/2 人	50/100 人	50%
31	0/2 人	0/2 人	0%
32	1/2 人	50/100 人	50%
33	1/2 人	50/100 人	50%

ケース1：どの年齢も実績数同じ



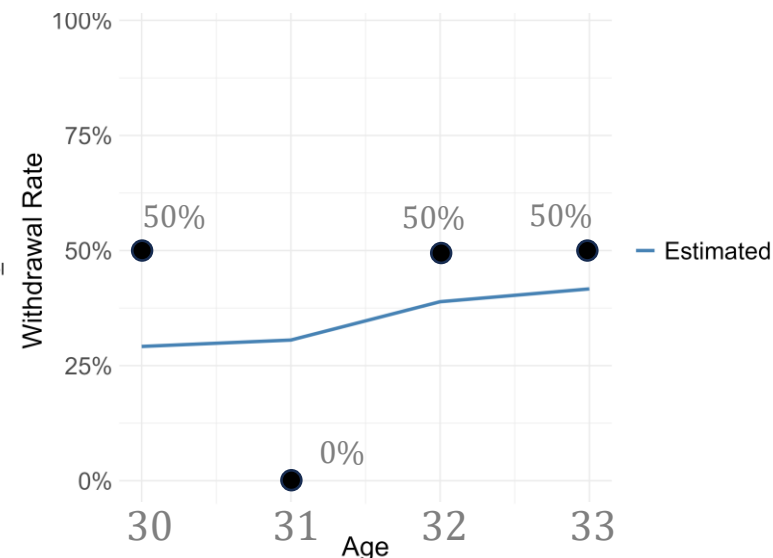
全ての年齢の実績を同列に扱っている

ケース2：31歳だけ実績小



31歳の実績はほぼ無視されている

(参考) 移動平均法による補整結果



3-5. GAM : 人数規模を踏まえた推定

今まで粗製脱退率に近づけるかは
アクチュアリーが判断していたが、
自動で考慮してくれるとは！



特徴：人数規模に応じて推定結果は異なる！

1. 実績が少ない場合 → 推定結果は $m - 1$ 次式(の逆ロジット変換)に近づく
2. 実績が多い場合 → 推定結果は 粗製脱退率に近づく

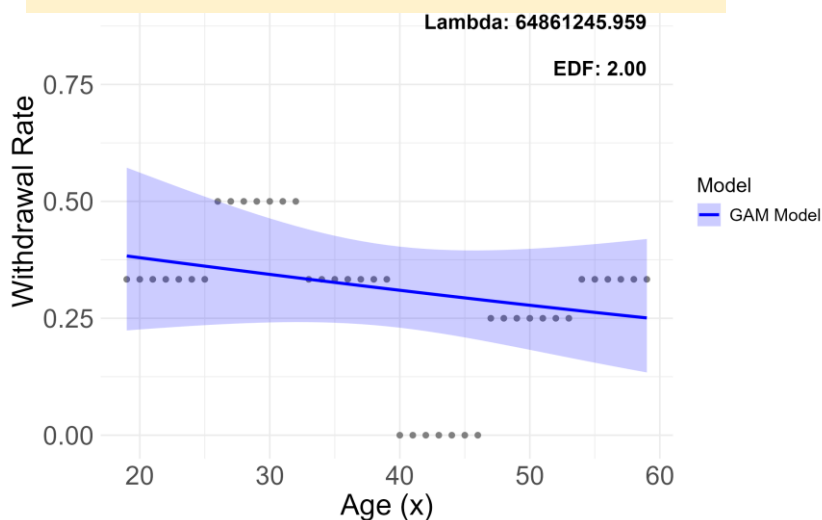
人数規模

小

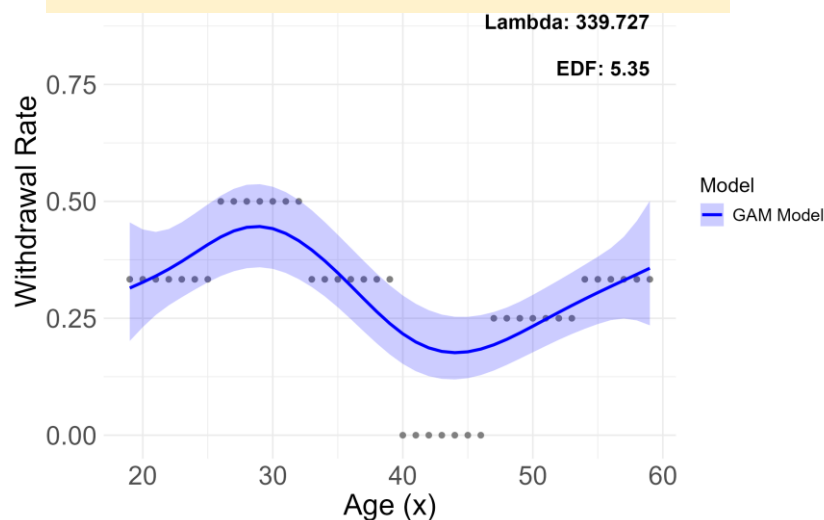
中

大

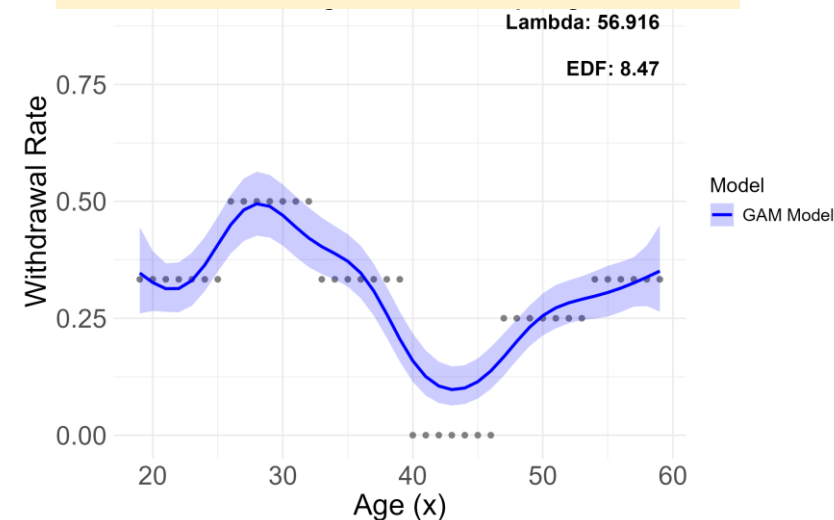
各年齢3人 ($\lambda \approx 6,500$ 万)
→ ほぼ直線($m = 2$ の場合)



各年齢15人 ($\lambda \approx 340$)
→ 直線と粗製脱退率の中間



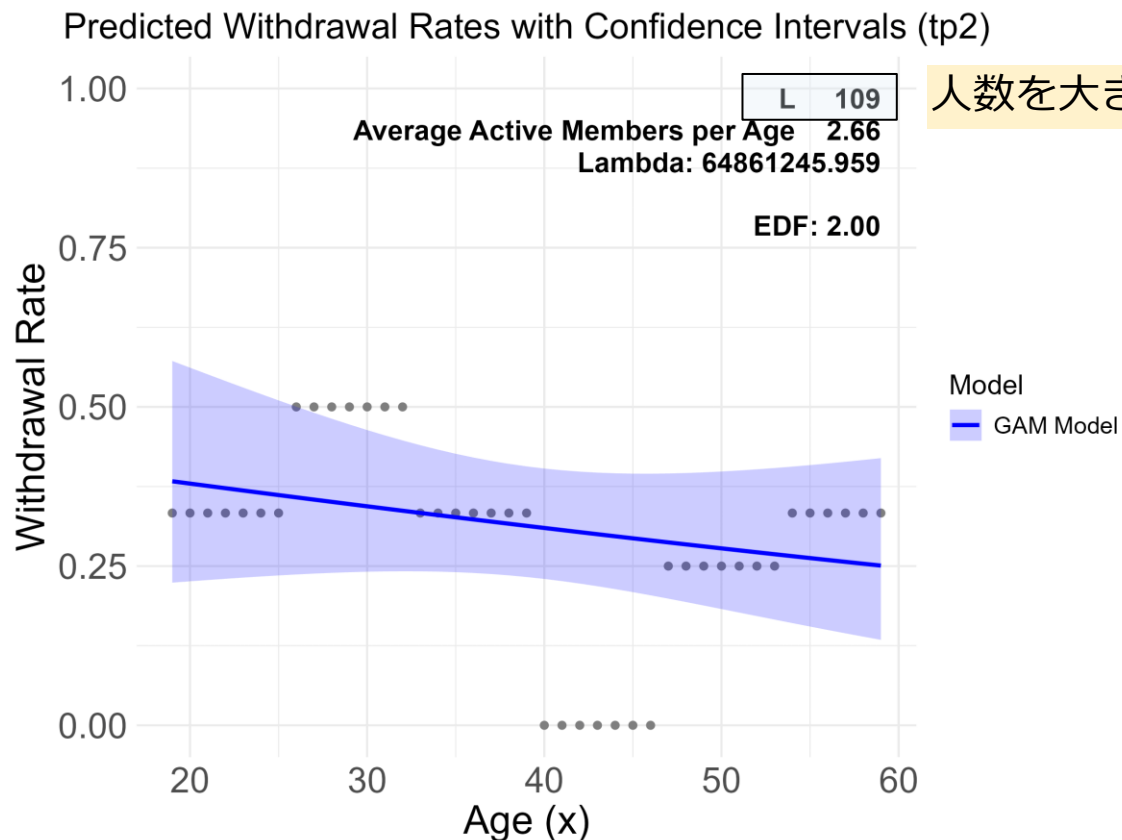
各年齢50人 ($\lambda \approx 57$)
→ ほぼ粗製脱退率



(注)全て粗製脱退率は同じ

3-5. GAM : 人数規模を踏まえた推定

【アニメーション】粗製脱退率を固定したまま、人数規模を大きくする



人数を大きくする

小規模集団：直線(の逆ロジット変換)

粗製脱退率を保ったまま
人数規模を大きくしていく

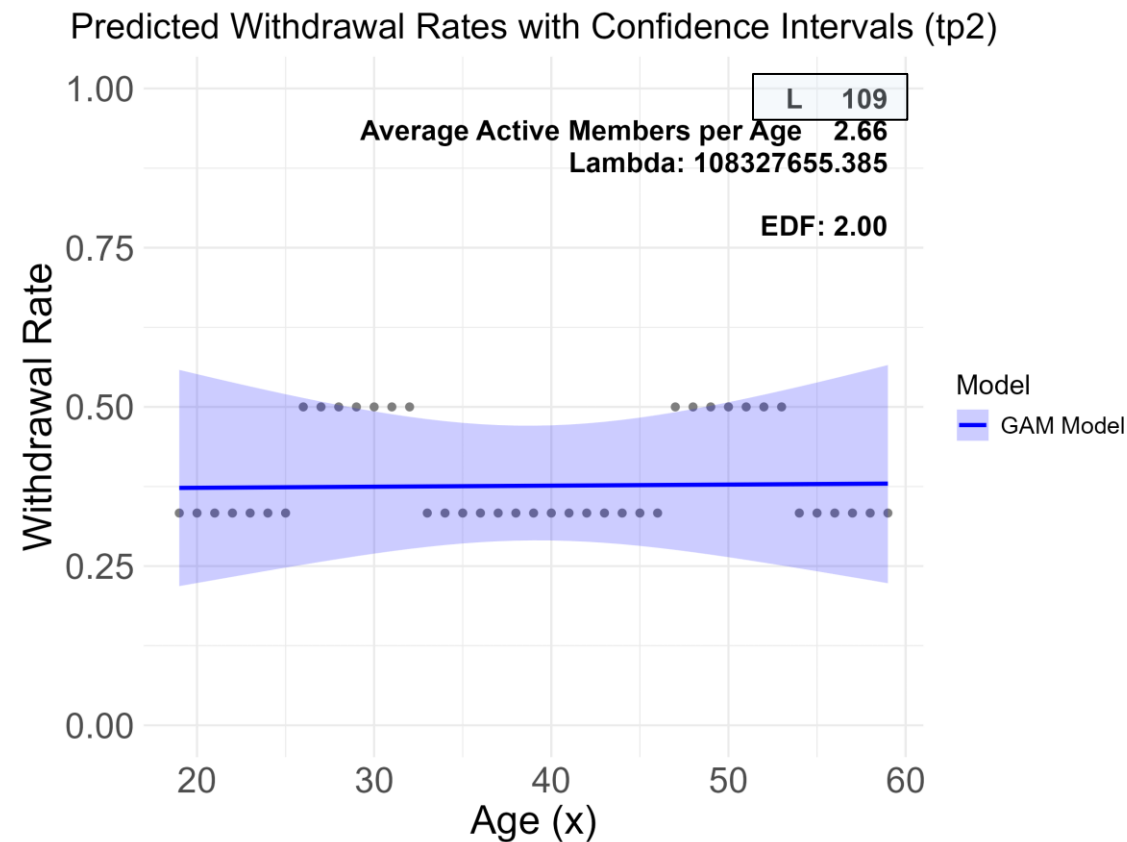
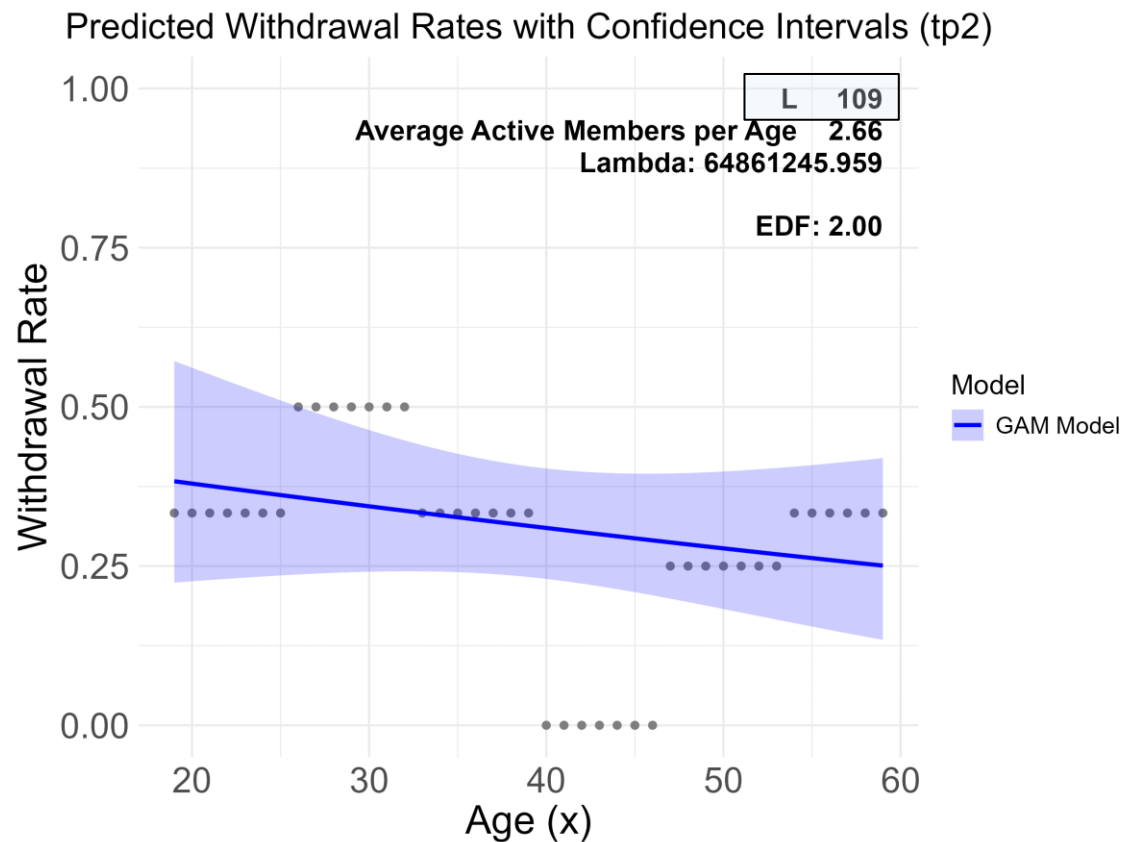
大規模集団：粗製脱退率

「情報が乏しい→推定結果を直線的にする」
というのは1つの考えとしてあり得る

(注) 全て粗製脱退率は同じ

3-5. GAM : 人数規模を踏まえた推定

【参考・アニメーション】 “直線的”でなくなる年齢は脱退データによって変わる



こちらの方がある程度
人数規模を大きくしても直線のまま

3-5. GAM : 人数規模を踏まえた推定

【参考】 数学的な理由

method="GCV.Cp"の場合 : λ の選択方法はAICの最小化と同値

$$\text{AIC}(\lambda) = L(\lambda) + 2 \cdot \text{EDF}(\lambda)$$

・ $L(\lambda)$: 負の対数尤度 (乖離度)

・ $\text{EDF}(\lambda)$: 有効自由度 $\left(= m + \sum_{j=1}^{10-m} \frac{1}{1+\lambda\delta_j} ; \delta_j > 0 \right)$

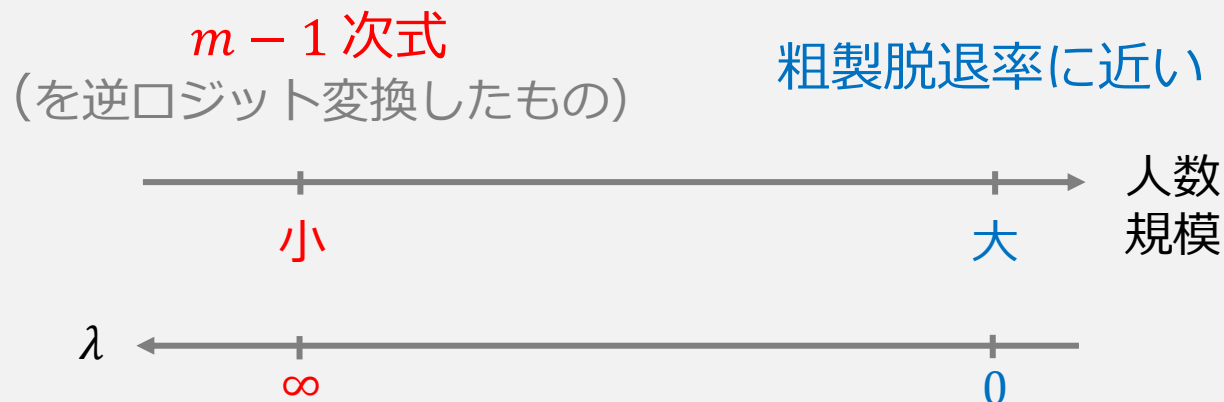
分かりやすさのためGCV.Cpを考えるが、REMLでも同じ議論

人数規模 : 小の場合、乖離度は相対的に小さく、AICの最小化はEDF最小化が重視される

→ 平滑度重視 (大きな λ) となる

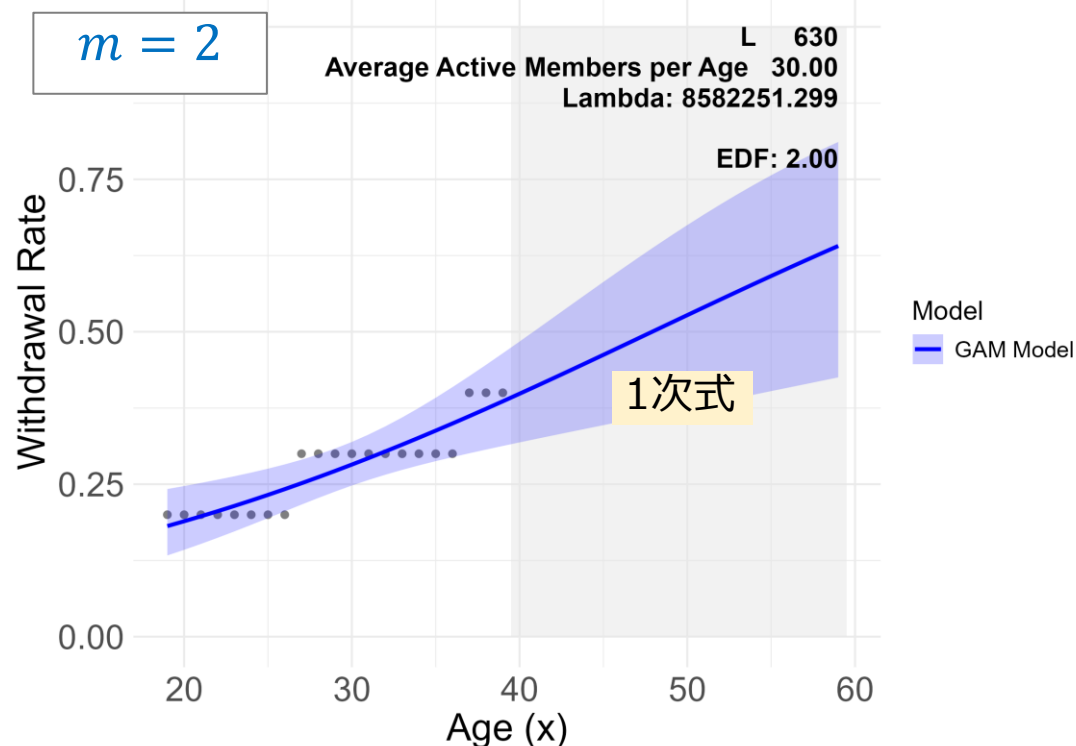
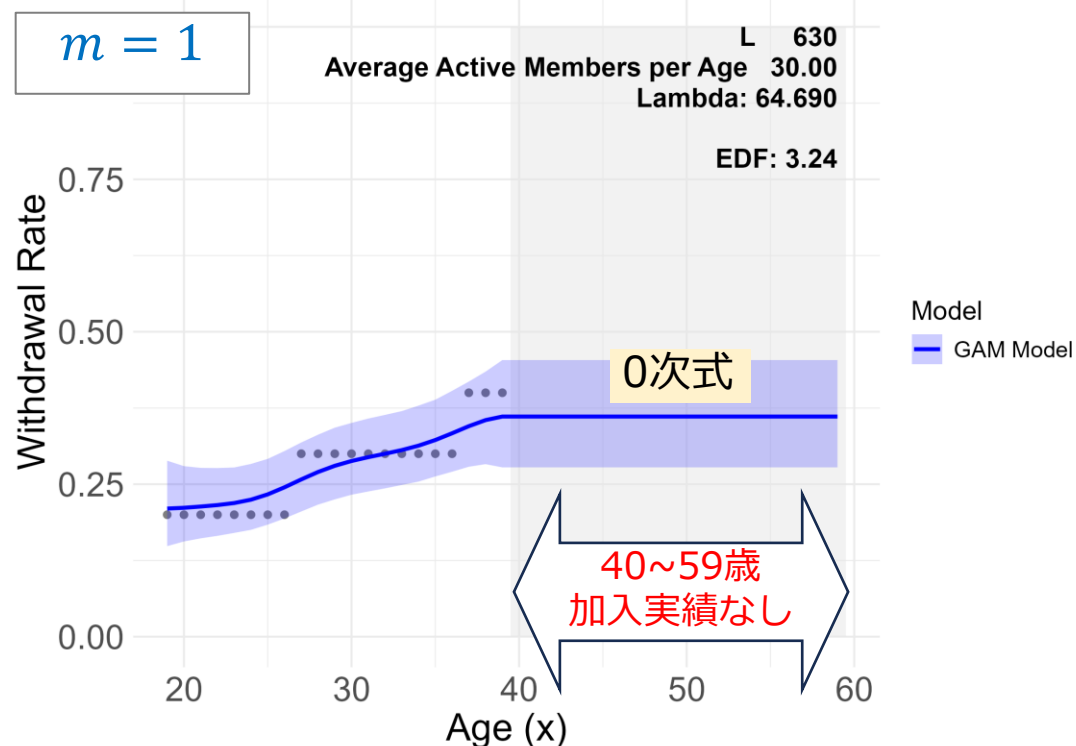
人数規模 : 大の場合、乖離度は相対的に大きく、AICの最小化は乖離度最小化が重視される

→ 乖離度重視 (小さな λ) となる



3-6. GAM : 補外

【参考】補外部分では、自然スプライン関数は $m - 1$ 次関数となる（自然スプラインの定義）
つまり、 $\text{logit } q_x = \beta_0 + \beta_1 x + \cdots + \beta_{m-1} x^{m-1}$ の形になる



- ✓ $m = 1$: 30歳台後半での脱退率がそのまま使用される
- ✓ $m = 2$: 30歳台後半での脱退傾向のまま右上がりの直線が引かれる
(厳密には逆ロジット関数で変換するので直線ではない)

3-7. GAM : まとめ

	現行方法	GAM
【課題 1】 ある年齢で加入者数が0人の場合であっても、そのまま計算できる	×	○
【課題 2】 算定年齢幅の拡大など、粗製脱退率作成に関わる問題が生じない	×	○
【課題 3】 補整結果が必ず0から1の範囲に入る	移動平均法:○ Greville :×	○
【課題 4】 各年齢の粗製脱退率の価値の違いを考慮している	×	○
決定脱退率は粗製脱退率からかけ離れない	○	○
決定脱退率は全体に滑らかな形態をしている	○	○
集団の規模が大きくなるにつれ、決定脱退率は粗製脱退率に近くなる	×	○
システムを用いて容易に算出され、計算速度は速く、乱数を使用しない	○	○
推定結果はノイズに引きずられず、その予測精度は高い	×	○
高度な理論を使用している（その分、説明が難しい）	移動平均法:×	○
	Greville :△	
加入者期間など、年齢以外の説明変数を推定に取り入れられる	×	○
外部データ脱退率 q_x^{ex} を推定に取り入れられる	△	○
脱退率に関する信頼区間が得られる（リスク評価に応用できる）	×	○

後述

4. 数值実験

4-1. 数値実験の概要

GAMを用いた提案手法の優位性は理論的に確かめられたが、**実際に理論通りになっているか？**

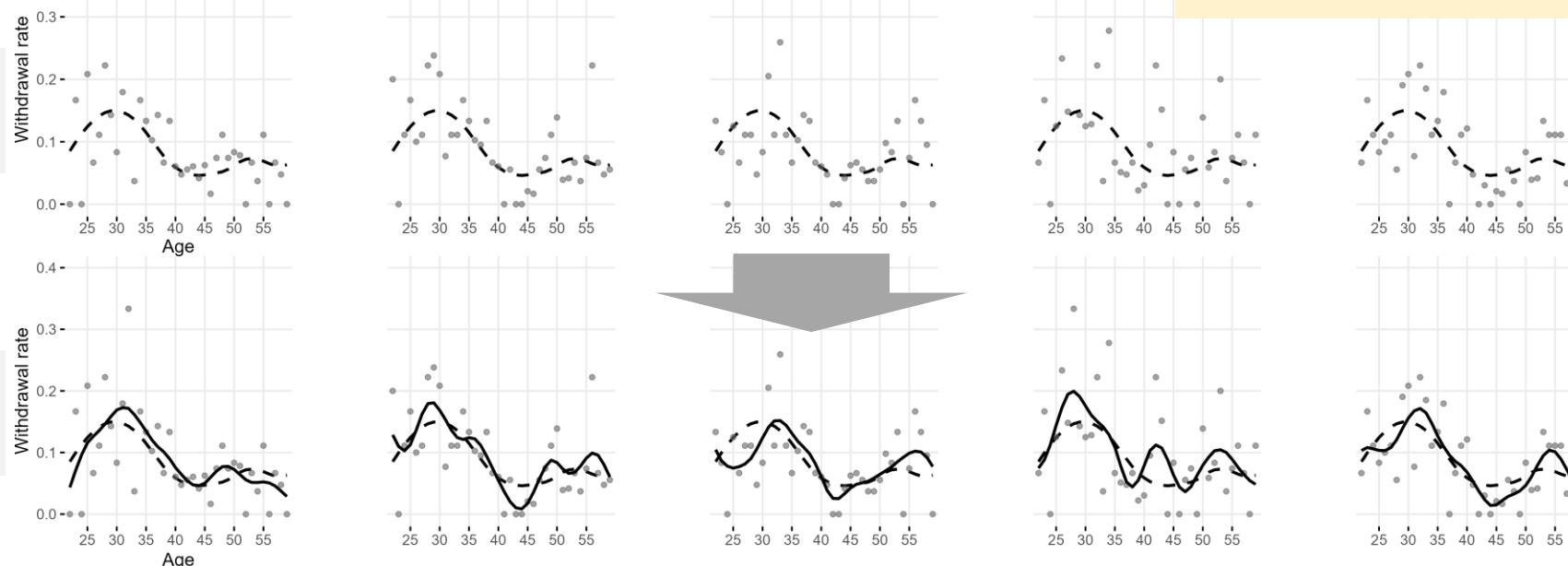
→ **人工データを用いた数値実験を行う**

手順

1. 人員構成 l_x と真の脱退率 q_x を1つ固定する
 2. 真の脱退率に基づき発生させた二項乱数をもとに、脱退データを300個作成
 3. このうち、**3個合算**で脱退率を1つ推定（つまり100個分の脱退率を推定）
- 真の脱退率を仮定
- 3年実績で推定する
実務の取扱いを踏襲

脱退データを乱数で
多数発生

各々、脱退率を推定
(計100個計算)



4-1. 数値実験の概要

どちらの方がより
真の脱退率に近い結果
となっているのか？



検証内容：「真の脱退率」を用いて、現行手法と提案手法を比較

現行手法

- ・ Greville 3次13項
- ・ 5点移動平均法

算定年齢幅の拡大(5歳幅)

行う、行わない、の2パターンを実施

提案手法

- ・ 一般化加法モデル (微分平滑度 m として、2,3の2通りを想定、bs="tp"、method="REML")

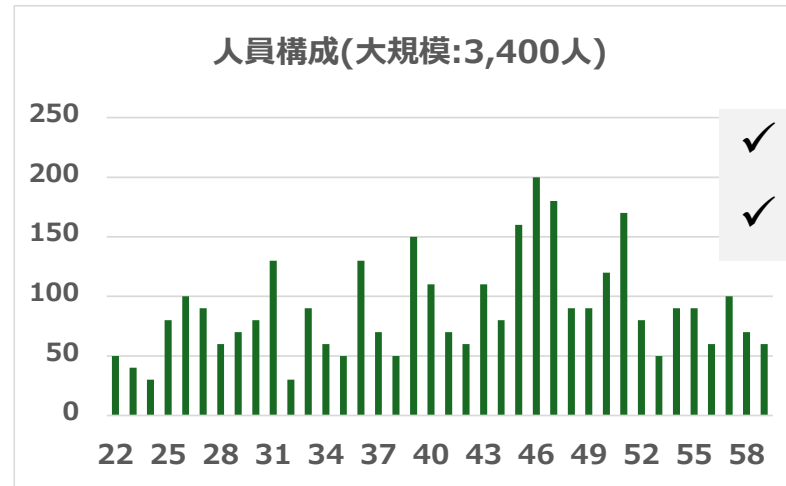
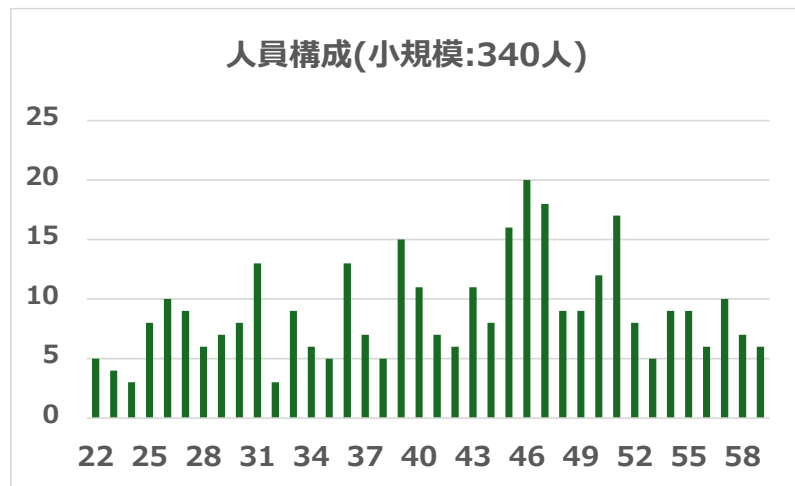
真の脱退率 q_x^{true} と推定脱退率 q_x の近さを表す指標として、ここでは
RMSE (二乗平均平方根誤差) を使用 (RMSEが小さいほど両者は近い)

$$RMSE := \sqrt{\frac{1}{x_r - x_e} \sum_{x=x_e}^{x_r-1} (q_x^{\text{true}} - q_x)^2} \quad ; \quad (x_e, x_r) = (22, 60)$$

この数値実験は真の脱退率を仮定した検証であり、**実証分析ではない**ことに注意

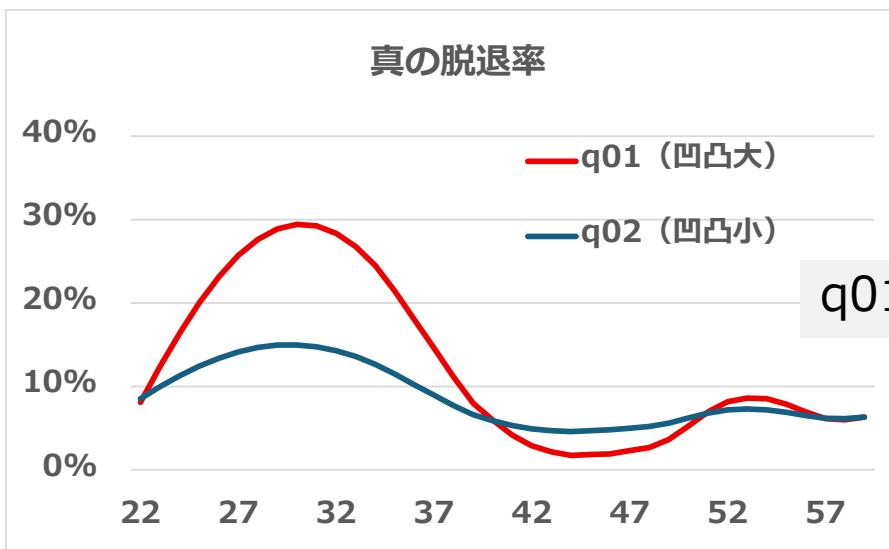
4-1. 数値実験の概要

- ✓ 人員構成 l_x : 小(340人)、大(3,400人) の2種類



- ✓ 両者の構成割合は同じ
- ✓ 各年齢集団の規模はまばら

- ✓ 真の脱退率 q_x : 次の2種類



q01の方が凹凸が大きい

4-2. 結果：100本の軌跡図（真の脱退率 q_{01} ）

小規模
集団

Greville
算定年齢幅
拡大：無

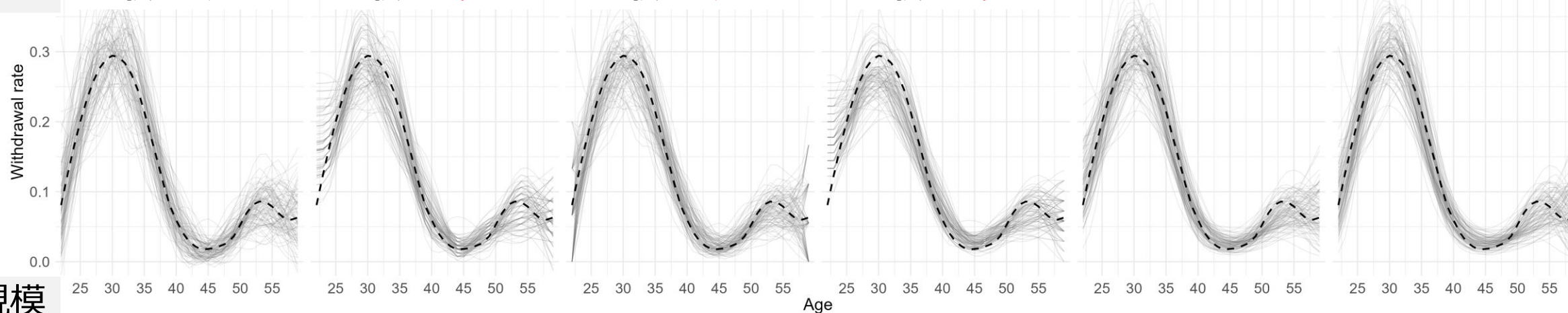
Greville
算定年齢幅
拡大：有

5点移動平均法
算定年齢幅
拡大：無

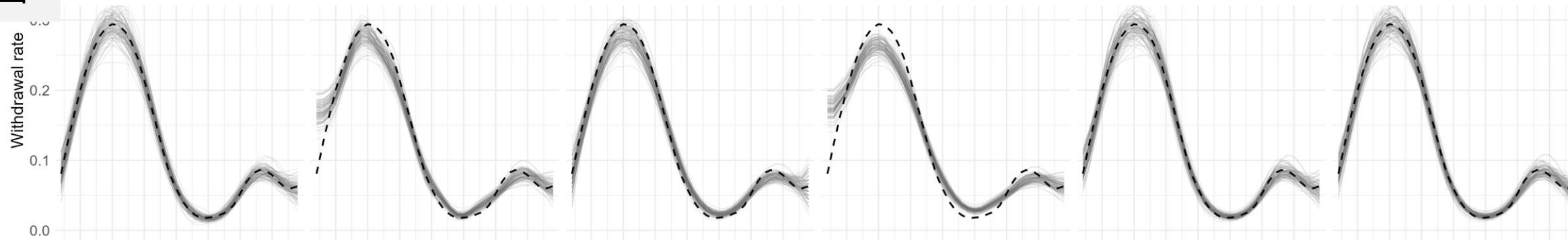
5点移動平均法
算定年齢幅
拡大：有

GAM
($m = 2$)

GAM
($m = 3$)



大規模
集団



- ✓ 小規模集団の場合、特に「算定年齢幅の拡大：無」の場合にバラツキは大きい
- ✓ 大規模集団の場合、「算定年齢幅の拡大：有」は形状をうまく捉えられていない

4-2. 結果：100本の軌跡図（真の脱退率 q_{02} ）

小規模
集団

Greville
算定年齢幅
拡大：無

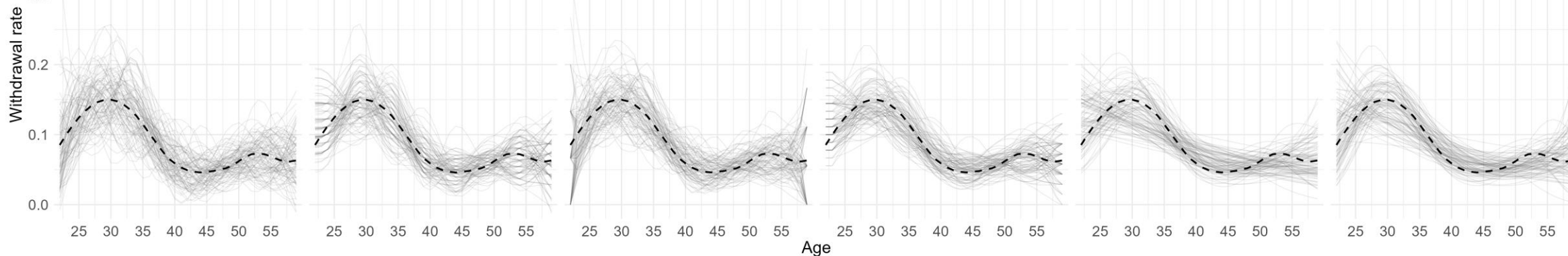
Greville
算定年齢幅
拡大：有

5点移動平均法
算定年齢幅
拡大：無

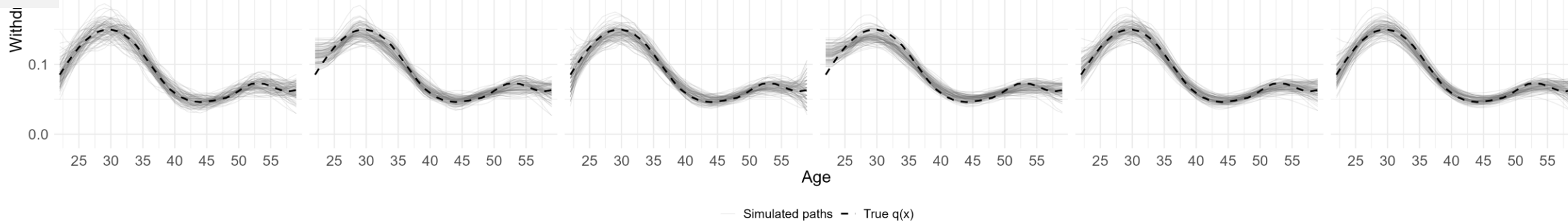
5点移動平均法
算定年齢幅
拡大：有

GAM
($m = 2$)

GAM
($m = 3$)

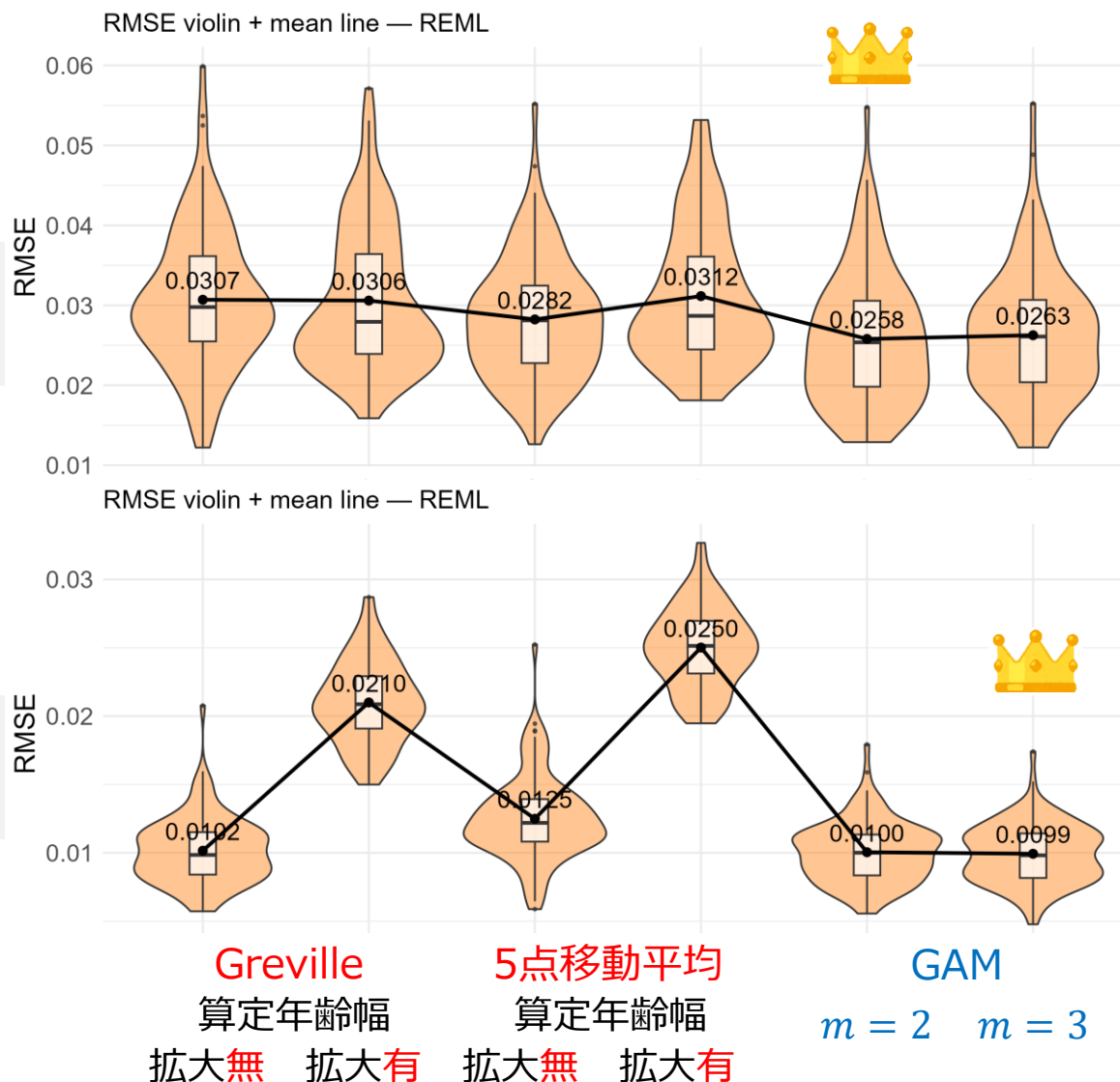


大規模
集団



- ✓ 小規模集団で提案手法の一部の結果に曲がり具合が少ない($m = 2$ の場合、直線的)ものが確認される
- ✓ 他の傾向は q_{01} と概ね同じ

4-2. 結果 : RMSEのviolin plot (真の脱退率q01)

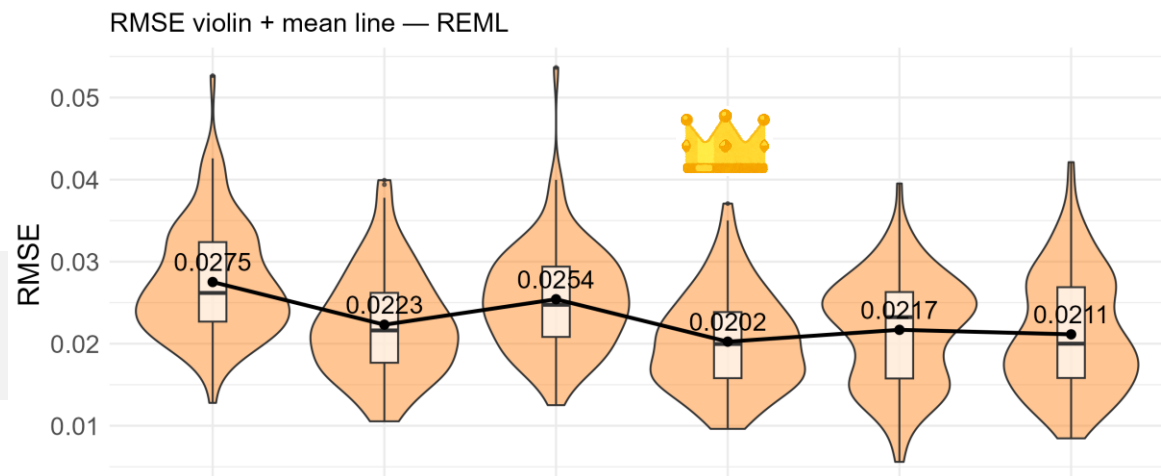


縦軸：真の脱退率と推定脱退率との
RMSE (二乗平均平方根誤差)

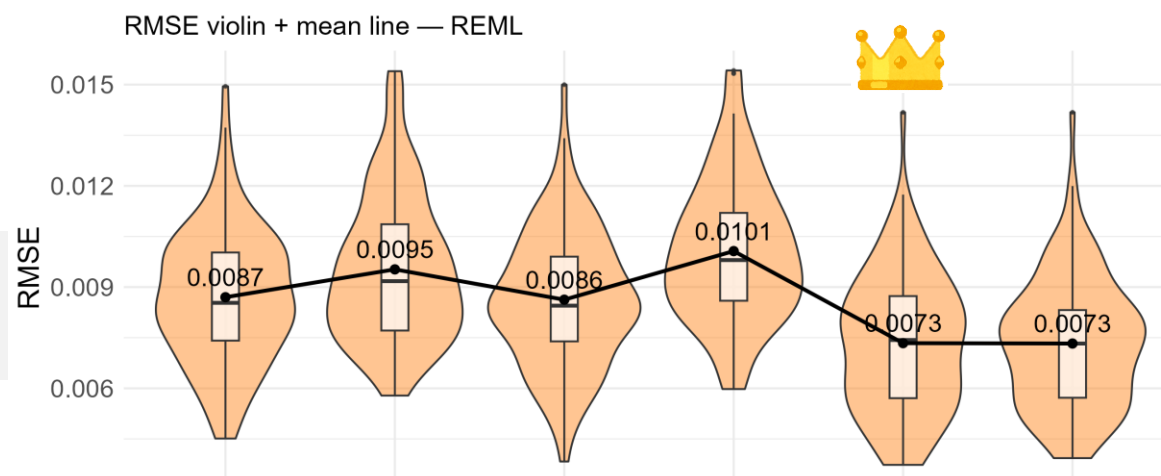
- ✓ 提案手法は、いずれの集団においても、
現行手法と比べて、**平均的かつ安定的に
真の脱退率に近い結果**を返す
- ✓ 「算定年齢幅の拡大」は小規模集団でも
あまり有効でなく、大規模集団では**精度
の大幅な悪化を引き起こしている**
→ 現行手法において、**拡大するかどう
かのアクチュアリー**の判断が必要
- ✓ 「算定年齢幅の拡大：無」の場合に
大きく外している結果が含まれる

4-2. 結果 : RMSEのviolin plot (真の脱退率q02)

小規模
集団



大規模
集団



Greville
算定年齢幅
拡大無 拡大有

5点移動平均
算定年齢幅
拡大無 拡大有

GAM
 $m = 2$ $m = 3$

- ✓ 小規模集団では、5点移動平均法（算定年齢幅の拡大：有）が最も良い結果となった（提案手法は2番目に良い）
→ 「算定年齢幅の拡大」を行うと、提案手法よりも良い結果となる場合がある
- ✓ 一方で大規模集団では、5点移動平均法（算定年齢幅の拡大：有）は最も悪い結果となった（提案手法は最も良い）
→ 現行手法において、拡大するかどうかのアクチュアリー判断が非常に重要
- ✓ 提案手法はどちらの場合でも良い結果を示した

4-3. 検証結果

提案手法の良いところ

いずれの人員構成、真の脱退率の形状についても、
(特に何も手を加えずとも) **安定して良い結果が得られるところに強みがある**と言える

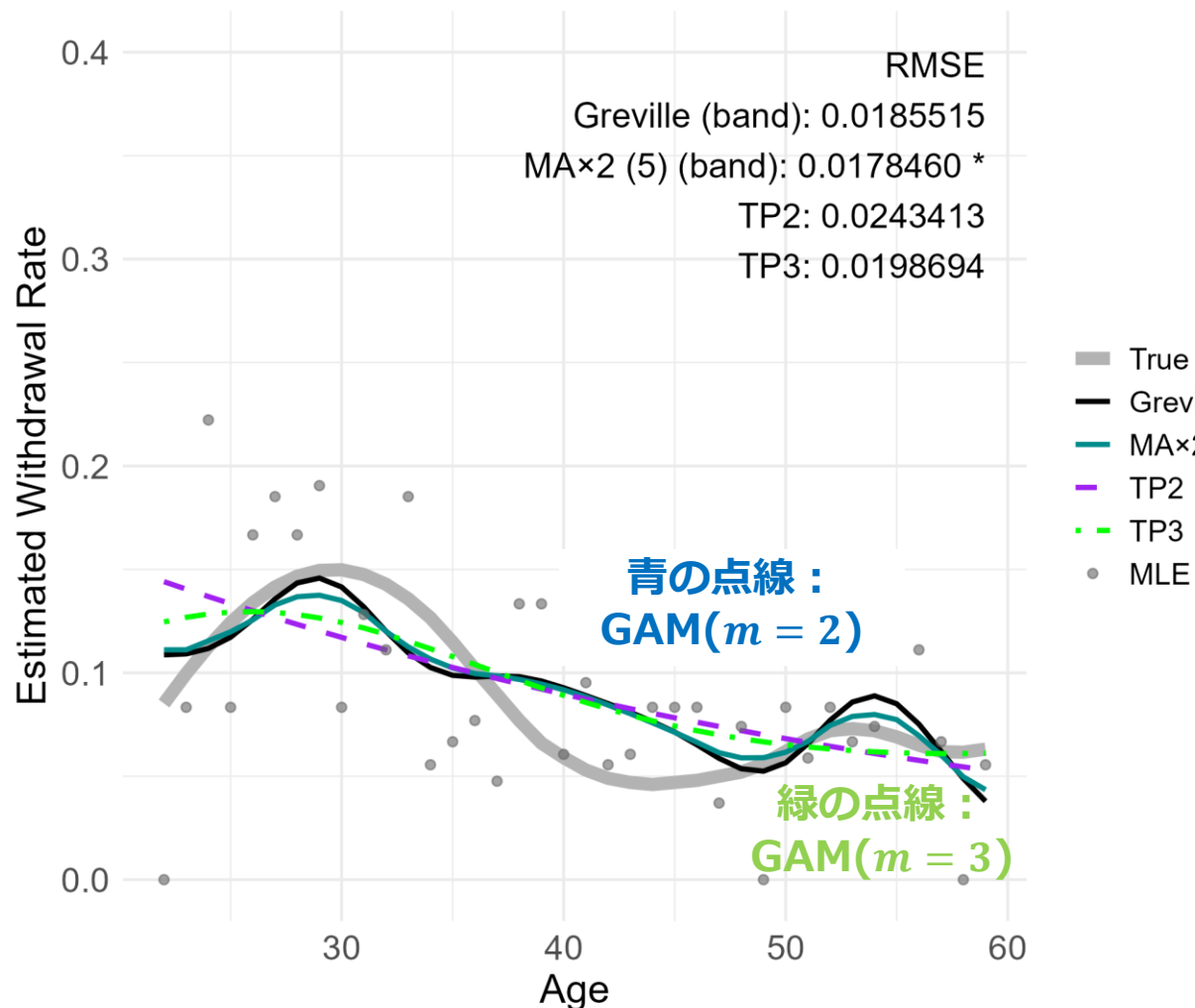
現行手法との比較

- ✓ 小規模集団において「算定年齢幅の拡大」の実施有無を適切に判断すると、提案手法を上回る精度を示す場合もある
- ✓ ただし、**人員構成（、真の脱退率の形状）**に応じて適切な判断をしなければ**精度悪化を招く可能性**もある

4-3. 検証結果

【参考】一部の推定結果で確認される曲がり具合が少ない（“直線的になる”）事象

Estimated Withdrawal Rates (Simulation 47) [method=REML]



- ✓ 小規模集団かつ真の脱退率 q_0 の場合のみに見られた（全体の1割～2割程度発生）
- ✓ 現行手法では手を加えない限り、ほとんど発生しない事象である
- ✓ “乖離度よりも平滑度を重視した方が予測精度が高い”と判断されたためと考えられる
- ✓ 推定結果として特に問題はないと考えられる

5. 応用

5-0. 応用

脱退率推定に更に期待したいこと

1. 加入者期間など、**現在年齢以外の説明変数**を導入したい
2. **外部情報**を脱退率推定に取り入れたい
3. 脱退率算定に関わる**リスク評価**を行いたい



これらについても、一般化加法モデル（GAM）による方法を応用し、**一定の解決策を図る**ことができる

5-1. 新たな説明変数の導入



現行の脱退率の説明変数は**年齢のみ**（現行の補整方法は**2変数以上に対応していない**）

一般化加法モデルは、**説明変数が2つ以上の場合にも対応**している！

（自然スプラインも多次元に拡張可能）

つまり、**“2次元脱退率”**が推定でき、より精緻な**債務評価**に応用できる！

（注：実務対応はこれから）

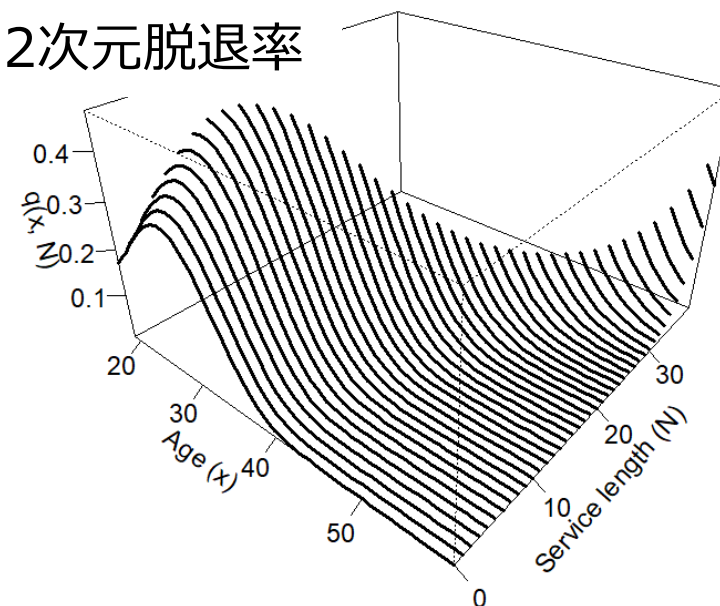
脱退データに
加入者期間 N の列を追加

```
> head(df, 20)
  x N y
1 19 0 0
2 19 1 0
3 20 0 0
4 20 1 1
5 21 0 0
6 21 1 0
7 21 2 1
8 22 0 0
9 22 1 0
10 22 2 0
11 22 3 1
12 24 3 1
```

Rコードを実行

```
model <- gam(
  y ~ te(x, N, bs = c("tp", "tp")),
  data = df,
  family = binomial,
  method = "REML"
)
```

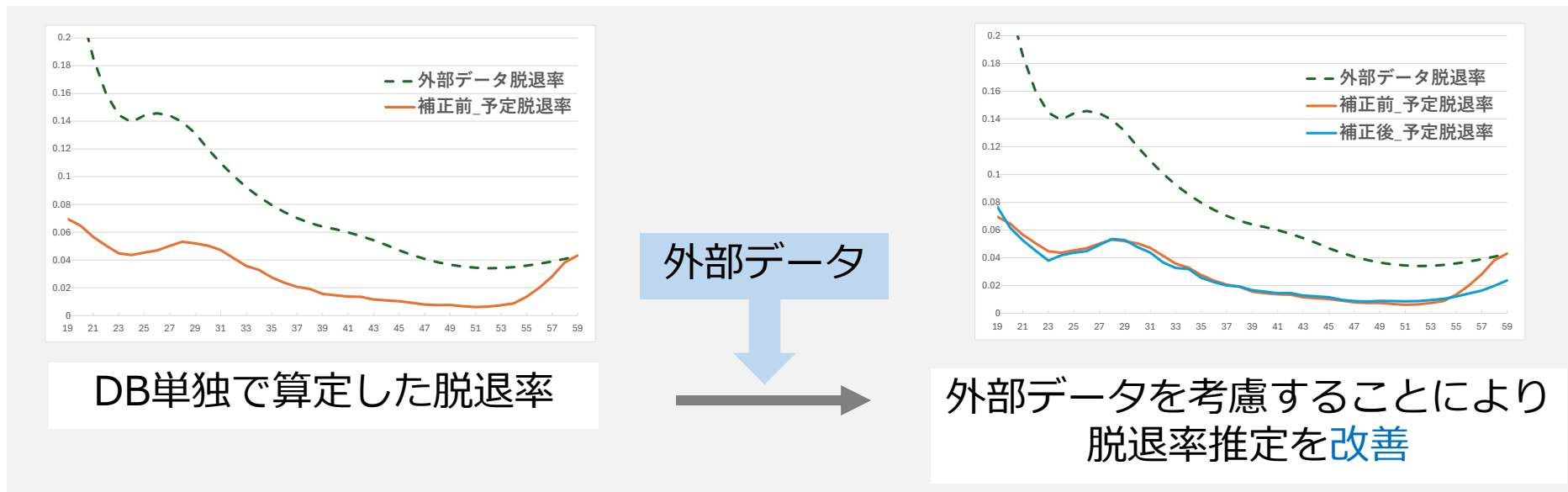
2次元脱退率



5-2. 外部データ脱退率の使用

小規模DBにおける脱退率として次を使用できる（DB数理実務ガイダンス）

1. 同種の企業で使用されている脱退率
 2. 業界毎の統計資料から推定した脱退率
 3. 1または2に対して合理的な補正を行った脱退率
- 本発表で 外部データ脱退率 と表記



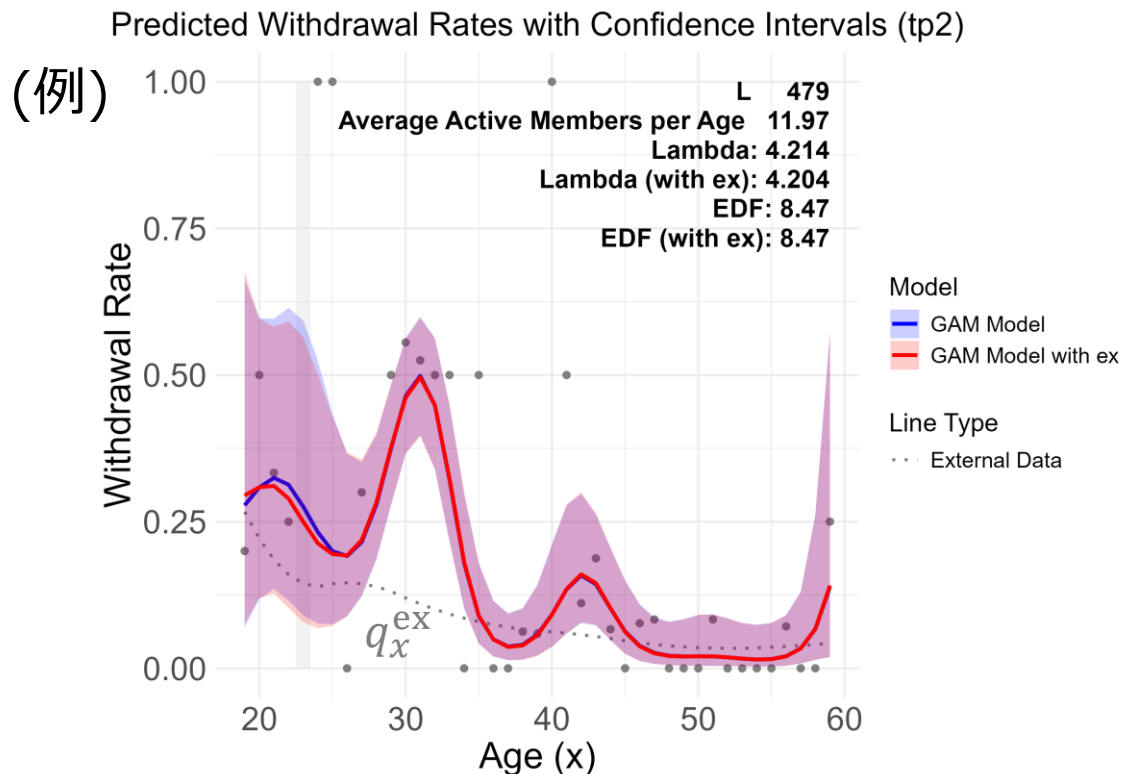
仮に外部データ脱退率を適切に取得できたとして、その情報をどのように推定に組み込むべきかについて、明確な基準は存在せず、實際上ほとんど使用されていない



5-2. 外部データ脱退率の使用

GAMを用いて外部データ脱退率 q_x^{ex} を推定に組み込む

(例) モデル式を $\text{logit } q_x = s(x)$ でなく、
 $\text{logit } q_x = s(x) + \text{logit } q_x^{\text{ex}}$ 説明変数：年齢 x 、オフセット項： q_x^{ex}
とし、一般化加法モデルを適用する



- ✓ オフセット項として外部データ脱退率を取り入れる方法は、対数オッズ比を推定対象とする考え方に基づくもの
- ✓ 他の取り入れ方として、説明変数として外部データ脱退率を追加する方法がある（例えば次の2通り）
 1. $\text{logit } q_x = s(x) + \alpha \cdot q_x^{\text{ex}}$
 2. $\text{logit } q_x = s(x) + \alpha \cdot \text{logit } q_x^{\text{ex}}$

5-2. 外部データ脱退率の使用

【参考】どのモデルを選んだら良いのか？

AIC最小のものを選択することで、**予測精度が最も高いモデル**を客観的に選択できる

(例)モデル選択 : 次のモデル1~4から最も予測精度が良いモデルを選びたい

1. $\text{logit } q_x = s(x)$
2. $\text{logit } q_x = s(x) + \text{logit } (q_x^{\text{ex}})$
3. $\text{logit } q_x = s(x) + \alpha \cdot q_x^{\text{ex}}$
4. $\text{logit } q_x = s(x) + \alpha \cdot \text{logit } (q_x^{\text{ex}})$

→
全てのモデル
のAICを計算

```
> # モデルの構築
> models <- list(
+   model1 = gam(y ~ s(x), data = df, family = binomial),
+   model2 = gam(y ~ offset(qlogis(ex)) + s(x), data = df, family = binomial),
+   model3 = gam(y ~ ex + s(x), data = df, family = binomial),
+   model4 = gam(y ~ I(qlogis(ex)) + s(x), data = df, family = binomial)
+ )
> # 予測値の計算
> all_ages <- seq(min(data_original$x, na.rm = TRUE),
+                 max(data_original$x, na.rm = TRUE))
> newdata <- data.frame(
+   x = all_ages,
+   ex = data_original$qex[match(all_ages, data_original$x)]
+ )
> predictions <- lapply(models, predict, newdata = newdata, type = "response")
> # AICの表示
> aic_values <- sapply(models, AIC)
> print(aic_values)
  model1  model2  model3  model4 
345.1002 345.4049 345.5506 344.9587
```

Model4が最小！

この例だと予測精度の観点からは**Model4** ($\text{logit } (q_x^{\text{ex}})$ を説明変数として追加) が最良
(一般に人員データによって最良となるモデルは変わり得る)

5-3. DBにおけるリスク評価に関する現行の取扱い

財政悪化リスク相当額^{*1}とは

- 財政悪化リスク相当額は、20年に1回の頻度で生じると想定されるリスクに耐えうる額として算定
- 次の2種類のリスクが想定されている
 - ✓ 価格変動リスク：年金資産の運用結果により積立金の額が変動することで積立不足が生じるリスク
 - ✓ 負債変動リスク：予定利率等の基礎率の変動に伴い債務が変動することで積立不足が生じるリスク

負債変動リスク：次の2種類に分類（DB数理実務ガイダンス）

1. 【財政再計算時に発生する差損】基礎率が変動することによる数理債務の変動
2. 【財政決算時に発生する差損】数理債務及び積立金の単年度の変動の実績と予定の乖離

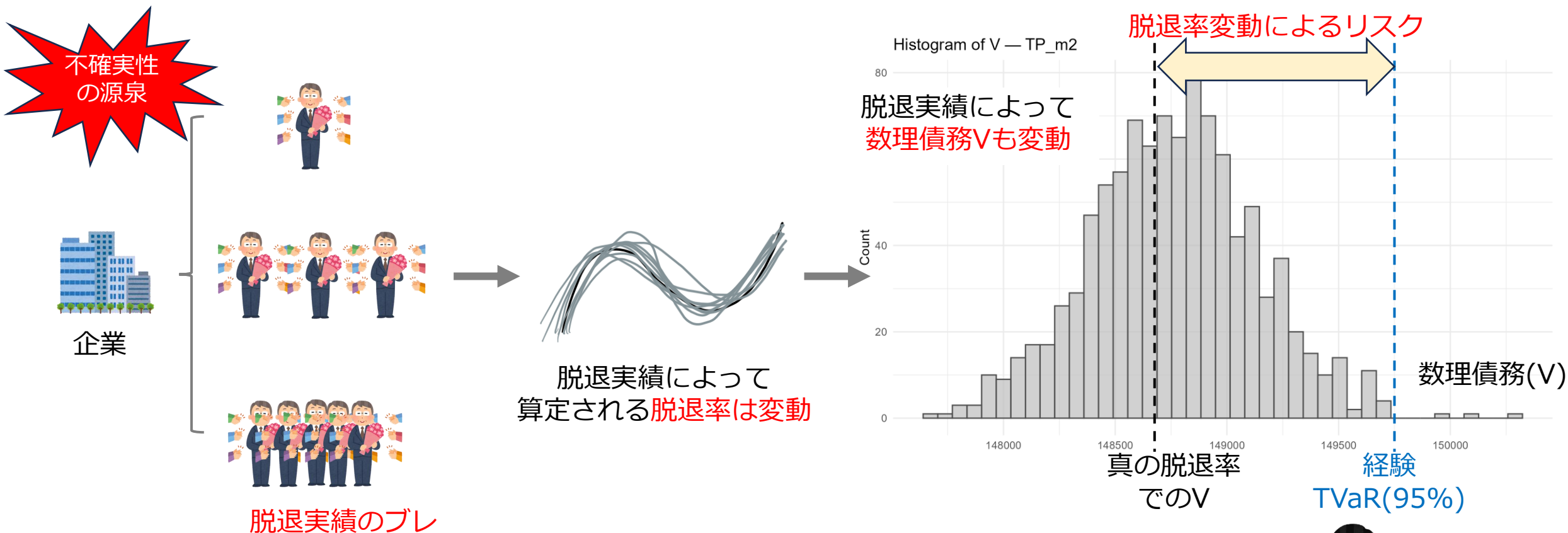
ここでは負債変動リスクの「財政再計算時に発生する差損」に関し、**脱退率算定時のリスク**を定量化する手法を例示する

*1 ここではリスク分担型企業年金以外のDBを想定

5-3. DBにおけるリスク評価に関する現行の取扱い

定量化するリスクのイメージ

脱退率の算定結果が変わると数理債務も変わるが、この変動のうち、**20年に1回の頻度で生じるものを算定したい**



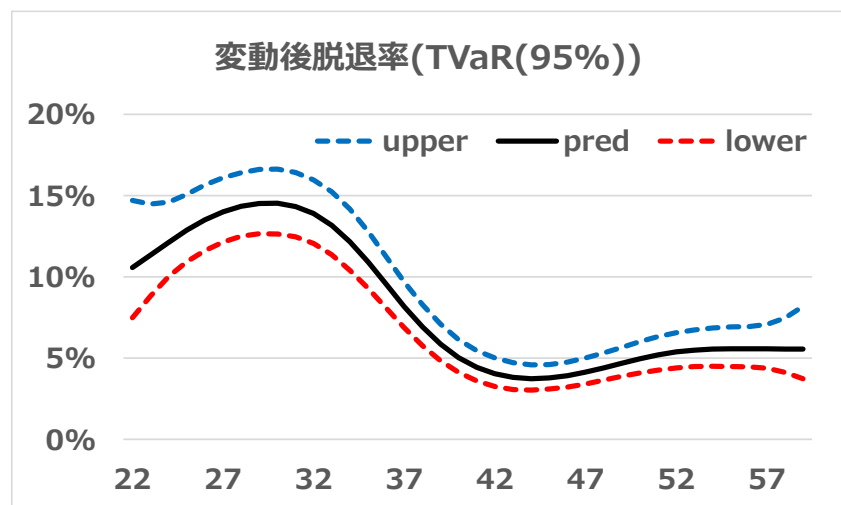
数理債務を多数計算するのは実務上現実的でなく、この方法では算定できない



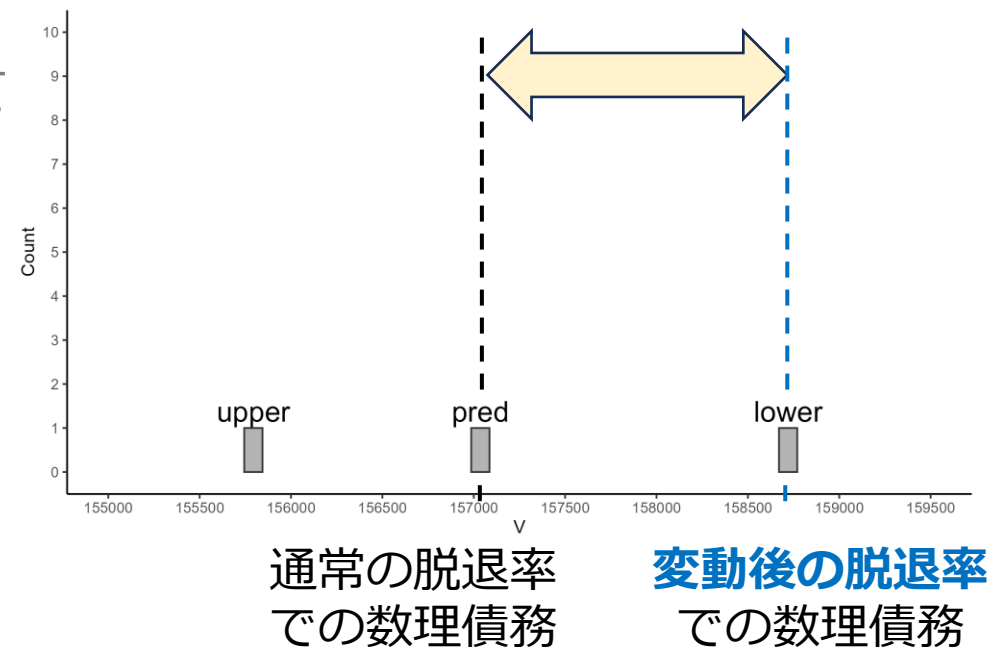
5-3. DBにおけるリスク評価に関する現行の取扱い

数理債務の計算回数が少ないリスク算定方法

変動後の脱退率を作成する方法（DB数理実務ガイダンス）



各脱退率に対して
数理債務を計算



- ✓ リスク = (変動後の脱退率での数理債務) - (通常脱退率での数理債務)
- ✓ つまり、リスク尺度 **TVaR(95%)** を用いて変動後の脱退率を作成する必要がある
- ✓ この方法は、脱退率の変動の動きと数理債務の変動の動きが**一方向であることを仮定**している

5-4. 信頼区間を用いたリスク評価

一般化加法モデルにおける信頼区間

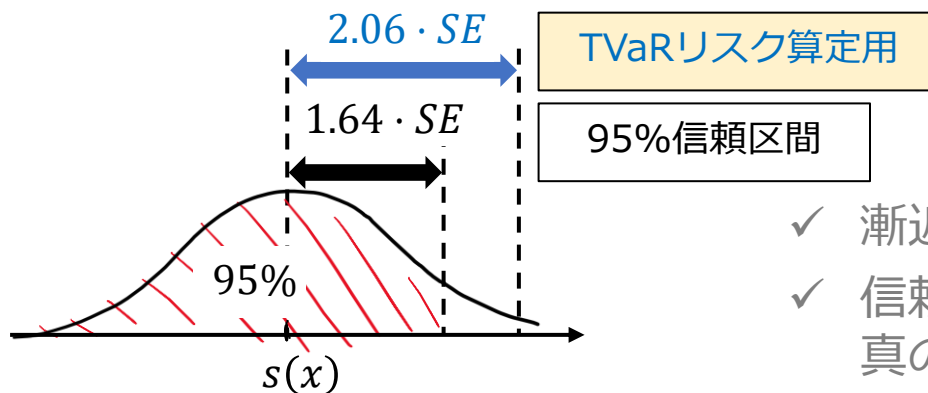
真の脱退率 $\text{logit } q_x = s(x)$ の一般化加法モデルによる推定結果

$$\text{logit } \hat{q}_x = \hat{s}(x)$$

に対し、この標準誤差を $SE(x)$ としたとき、**95%信頼区間**（片側）は次式で与えられる

$$\text{logit}^{-1}(\hat{s}(x) \pm z \cdot SE(x)) \quad ; z \approx 1.6449(\text{片側})$$

これを**TVaR(95%)**になるようにするには、 $z \approx 2.0627$ とすればよい



- ✓ 漸近的に $\hat{s}(x) \sim N(s(x), SE(x)^2)$
- ✓ 信頼区間の性質より、各年齢で、真の脱退率が95%信頼区間内に含まれる確率は概ね95%

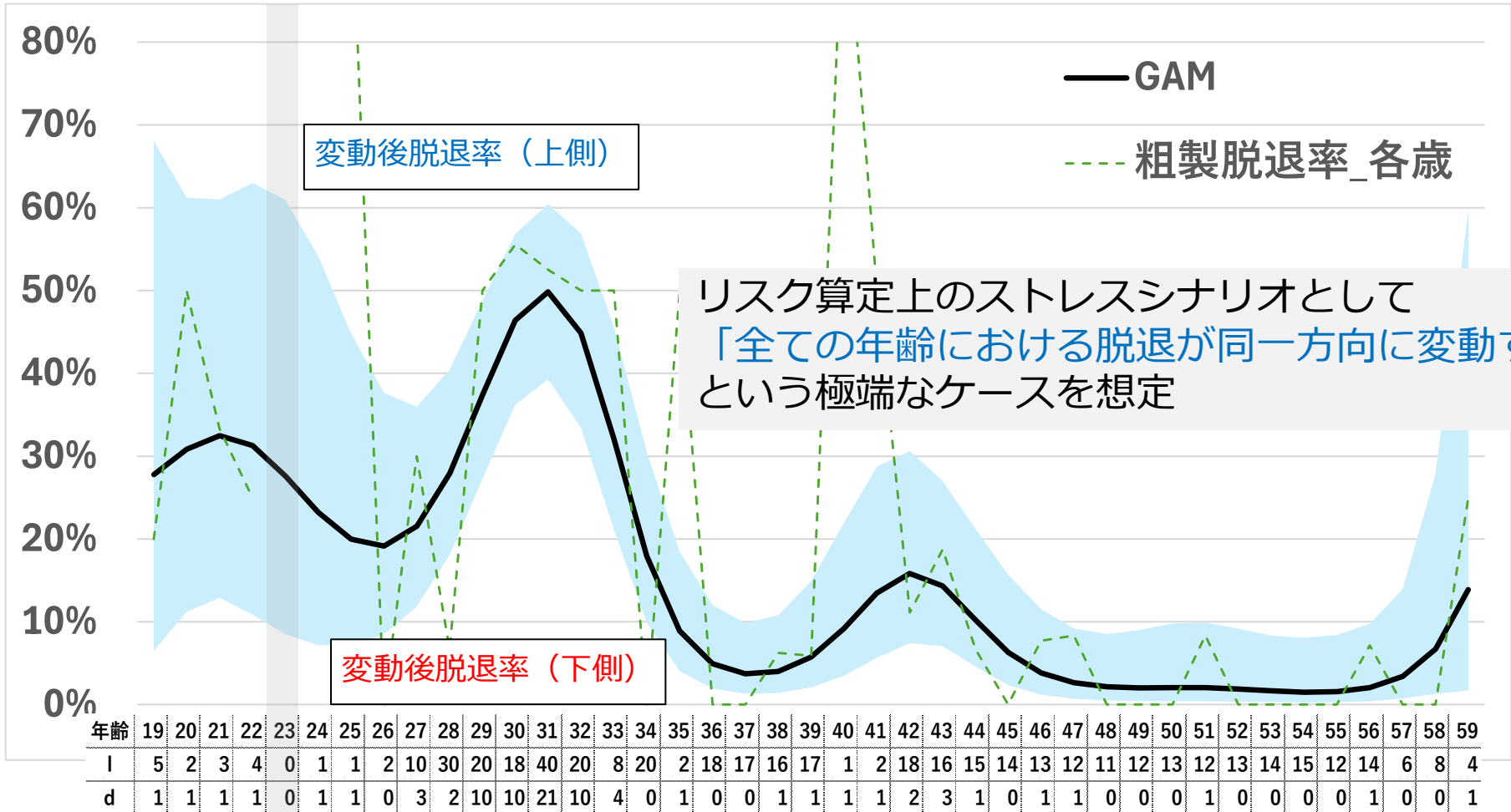
$\text{logit}^{-1}(\hat{s}(x) \pm z \cdot SE(x)) \quad ; z \approx 2.0627(\text{片側})$ で計算される脱退率を“変動後脱退率”とする！



5-4. 信頼区間を用いたリスク評価

(例) “変動後脱退率”

不確実性が大きいほど
リスク大になるはずだが、
確かに実績数が多いところは幅が狭く、
少ないところは幅が広がっている！

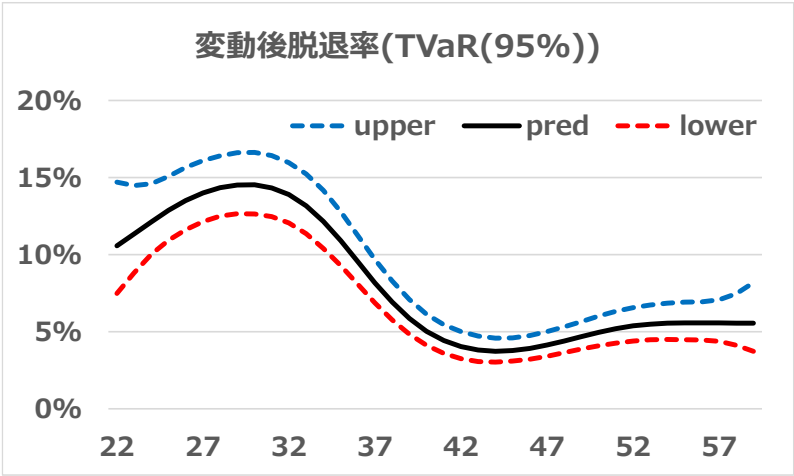
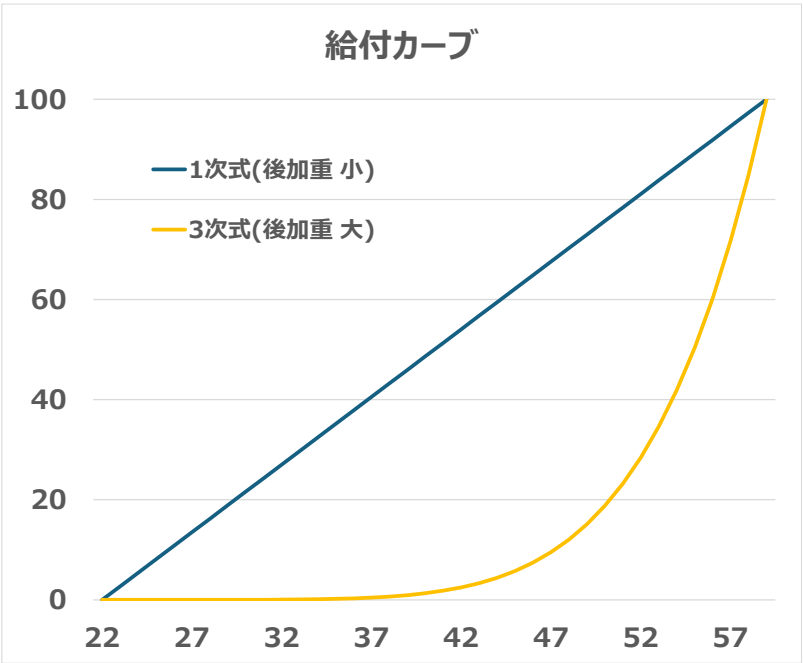


* 1 このように年齢ごとに完全な正の相関を仮定する考え方は、資産クラスごとに完全な正の相関を仮定する価格変動リスクの考え方と整合的

5-5. リスク評価の具体例

具体例

- ✓ 加入者数3,400人のDB
- ✓ 給付カーブは次の2通りを想定



各脱退率に対して、数理債務を計算

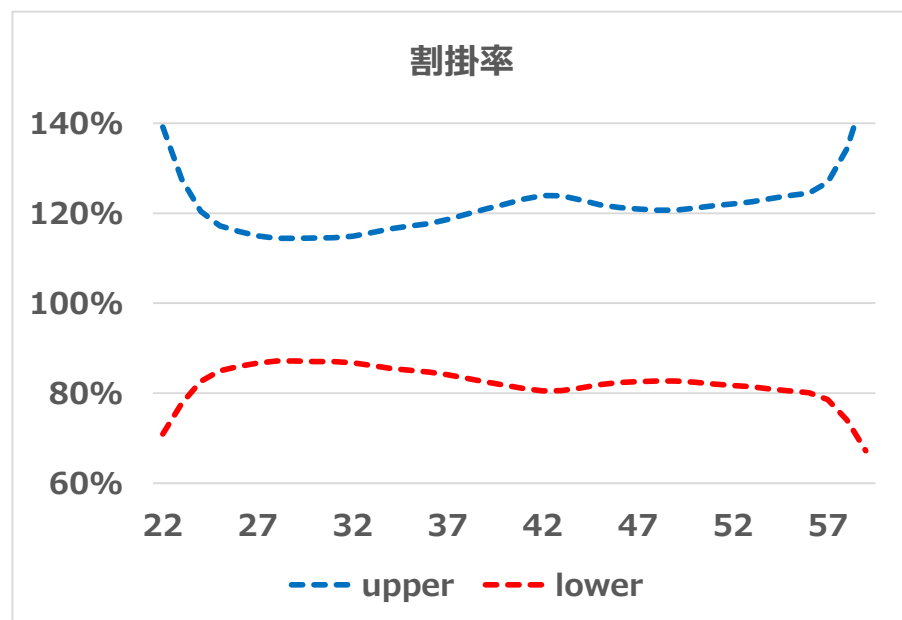
脱退率	1次式 (後加重性：小)		3次式 (後加重性：大)	
上側	158,720	+1% 1,684	132,072	-5%
予測	157,036	—	138,454	—
下側	155,794	-0.8%	143,604	+4% 5,150
脱退率に関する 負債変動リスク	1,684		5,150	

5-5. リスク評価の具体例

【参考】割掛率との関係

保守的な財政運営を行う観点から、基礎率（死亡率や脱退率）に割掛率を乗じることで、債務や掛金を保守的に計算する実務が行われている

（例）変動後脱退率と通常脱退率との比



- ✓ 概ね15%～20%程度の割掛に相当
- ✓ 両端の年齢で少し乖離する

GAMによって割掛率に対する統計的な根拠を与えることができる

6. まとめ

6-1. まとめと今後の課題

現行の脱退率算定方法にはいくつか課題がある



データサイエンスの知見から、GAMを適用すると課題が解決される

数値実験の結果、様々な状況で安定して良い結果が得られることが確かめられた

GAMによる方法はリスク評価等にも応用できるなど、応用可能性を秘めている



今後の課題：実証分析（実データを用いた分析など、実務導入に向けた研究）

6-2. まとめ（再掲）

	現行方法	GAM
【課題 1】 ある年齢で加入者数が0人の場合であっても、そのまま計算できる	×	○
【課題 2】 算定年齢幅の拡大など、粗製脱退率作成に関わる問題が生じない	×	○
【課題 3】 補整結果が必ず0から1の範囲に入る	移動平均法:○ Greville :×	○
【課題 4】 各年齢の粗製脱退率の価値の違いを考慮している	×	○
決定脱退率は粗製脱退率からかけ離れない	○	○
決定脱退率は全体に滑らかな形態をしている	○	○
集団の規模が大きくなるにつれ、決定脱退率は粗製脱退率に近くなる	×	○
システムを用いて容易に算出され、計算速度は速く、乱数を使用しない	○	○
推定結果はノイズに引きずられず、その予測精度は高い	×	○
高度な理論を使用している（その分、説明が難しい）	移動平均法:×	○
	Greville :△	
加入者期間など、年齢以外の説明変数を推定に取り入れられる	×	○
外部データ脱退率 q_x^{ex} を推定に取り入れられる	△	○
脱退率に関する信頼区間が得られる（リスク評価に応用できる）	×	○

Appendix

Appendix A. 脱退率に関する先行研究

脱退率に関する先行研究

- 岩沢[1991]
- ✓ 乖離度 $\sum_x w_x (q_x - c_x)^2$ に対して、 $w_x := \frac{l_x}{q_x(1-q_x)}$ として年齢ごとの集団の大きさを考慮
 - ✓ 乖離度を尤もらしい値(確率変数としての期待値)に固定し、3次差分平滑度の最小化を図る
-
- ✓ 年齢ごとの集団の規模を考慮した脱退率推定が可能
 - ✓ 3次差分平滑度を使用することでGrevilleの補整法との連続性を確保
 - ✓ ただし、解が一意に定まらない場合がある、満足のいくような計算アルゴリズムが出来ていないなどの問題が残る
-
- 小川[1992]
- ✓ 年齢ごとの集団の規模を考慮した別の脱退率推定方法
 - ✓ ベイズ統計を使用
 - ✓ 粗製脱退率を作成する際に経験則を加味することで、この課題の解決を図ろうとするもの
-
- ✓ 年齢ごとの集団の規模を考慮した脱退率推定が可能
 - ✓ ただし、事前情報として「少数集団の脱退率を多数集めた集団」が必要となり、導入の負荷が大きいものと言える

参考文献

- 岩沢宏和[1991],「小規模集団における脱退率の作成方法についての考察」,『日本アクチュアリー会会報』,第44号,第2分冊,105-130頁.
- 小川裕之[1992],「企業年金制度における脱退率のモデル化」,『日本アクチュアリー会会報』,第45号,第2分冊,261-291頁.
- 車谷優樹 & 岩沢宏和[2025],「一般化加法モデルを用いた脱退率推定の改善」(査読中)
- 日本年金数理人会[2014],『厚生年金基金実務基準第1号(平成26年11月改訂)代行保険料率の計算に関する細目』.
- 日本年金数理人会[2025],『確定給付企業年金に関する数理実務ガイダンス(最終改定2025年2月19日)』.
- Greville, T. N. E. [1979], "Moving-weighted-average smoothing extended to the extremities of the data", *MRC Technical Summary Report #2025*.
- Hastie, T., & R. Tibshirani [1986], "Generalized additive models", *Statistical Science*, 1(3), 297-310.
- Schoenberg, I. J. [1964], "Spline functions and the problem of graduation", *Proceedings of the National Academy of Sciences of the U.S.A.*, 52, 947-950.
- Wood, S. N. [2017], "Generalized Additive Models: An Introduction with R", 2nd Edition, Chapman & Hall/CRC.
- Wood, S. N. [2025], "mgcv: Mixed GAM Computation Vehicle with Automatic Smoothness Estimation", <https://cran.r-project.org/web/packages/mgcv/index.html> (2025/8/1).