

SOA テキストに見る損保数理のアクチュアリー実務への応用 ＜ASTIN 関連研究会＞

あずさ監査法人 島本 大輔 君
共栄火災海上保険 佐野 誠一郎 君
大同火災海上保険 川上 良一 君

【司会】 時間となりましたので、セッションCの2、ASTIN 関連研究会による、「SOA テキストに見る損保数理のアクチュアリー実務への応用」を開始します。発表者は、あずさ監査法人の島本大輔さん、共栄火災の佐野誠一郎さん、大同火災の川上良一さんの3名です。なお、質疑応答に関しては、「Slido」に投稿された質問に対して、各発表者がそれぞれの発表直後に回答することとしています。コロナ対策により、最後にまとめる質疑応答の時間は取りませんので、質問のある方は、質問したい事項の発表者の発表中に質問を投稿するようにお願いします。また、時間の都合などで、回答する質問を選ばせていただく場合もありますので、あらかじめご了承ください。

それでは、島本さん、よろしく申し上げます。

2020年度 日本アクチュアリー会年次大会

SOAテキストに見る 損保数理のアクチュアリー実務への応用

2020年11月6日
ASTIN 関連研究会

あずさ監査法人 島本 大輔
共栄火災海上保険 佐野 誠一郎
大同火災海上保険 川上 良一

本日の発表は発表者個人の見解であり、
発表者の所属する企業とは関係ありません。

1

【島本】 それでは、ASTIN 関連研究会から発表させていただきます。あずさ監査法人の島本と申します。よろしく申し上げます。

本セッションの概要

- アクチュアリー試験の一次試験では数学や損保数理の科目において、様々な確率分布、統計学的手法、損害額や損害保険料率の推定方法等について学習する。しかし、**試験のための学習を実際にアクチュアリー実務へ応用するにあたっては、データの取扱いやモデル選択等の様々な課題に直面するのではないだろうか。**
- 本セッションでは、米国のアクチュアリー会SOA（Society of Actuary）試験において、日本の損保数理に相当するExam STAM：Short-Term Actuarial Mathematicsの指定テキストとされている「Loss Models」の第5版をベースに、**本書のトピックの1つである信頼性理論について改めて整理する。**さらに米国においては、信頼性理論の利用領域が損保だけでなく生命保険、健康保険、年金制度にも広がっていることなどに触れ、アクチュアリー実務への活用の可能性を改めて考える。その他にも、本書で触れられている**特殊なデータの取扱いやモデル選択の手法など、アクチュアリー実務担当者として理解しておくべきポイント**について紹介する。

2

はじめに、本セッションの概要になります。本セッションの課題意識としまして、本日参加されている皆様の中には、アクチュアリー試験を勉強されていらっしゃる方もいれば、既に合格されて、長く時間がたっている方、いろいろな方がいらっしゃると思います。皆様、数学や損保数理の科目において確率分布や統計に関して学習されてきた方がほとんどかと思っております。一方で、スライド1点目の太字に書かせていただいておりますけれども、実際に試験で学習した内容を実務に応用するにあたっては、データの取り扱いやモデル選択などの面で、さまざまな課題に直面するのではないかと考えております。

われわれASTIN関連研究会では米国アクチュアリー会のSOAのテキスト「Loss Models」の輪読を行っておりましたが、その輪読の中で、トピックの1つである信頼性理論が本書の中で非常に詳しく説明されているという点と、下の方にありますけれども、データの取り扱いやモデル選択といった、実際にアクチュアリー試験で勉強する内容を実務に落とし込むうえでの課題に対してヒントになる点が本書の中で取り上げられておりました。先ほど申し上げた背景を踏まえまして、今回、この「Loss Models」が実務への活用において役立つと思い、年次大会でご紹介させていただきたいと考えました。

本セッションの概要

- 構成
 - 第1部 Loss Modelsの紹介
担当：あずさ監査法人 島本大輔
内容：SOA指定テキスト「Loss Models」の概要を紹介する。
 - 第2部 Loss Modelsによる保険実務に即した統計手法の学習
担当：共栄火災海上保険 佐野 誠一郎
内容：Loss Modelsで紹介されている様々な手法の中から、「実務上の観点に基づく統計手法の学習」と「保険データ特有の課題に対する取扱い」の2つのテーマでその幾つかを紹介する。
 - 第3部 信頼性理論 (CREDIBILITY)
担当：大同火災海上保険 川上 良一
内容：有限変動信頼性理論、ベイズ的手法、最大精密信頼性理論といった様々な信頼性理論について、各手法の違いや応用例について紹介する。

3

本セッションの構成ですけれども、まず第1部は、私から、「Loss Models」というテキストについてご紹介させていただきたいと思います。続きまして第2部では、共栄火災の佐野さんから、「Loss Modelsによる保険実務に即した統計手法の学習」と題して、先ほどご紹介したような実務における課題への対応として紹介されている手法について、ご説明させていただきたいと思います。続いて第3部では、「信頼性理論」と題しまして、大同火災海上の川上さんから、有限変動信頼性理論やベイズ的手法といった信頼性理論について、ご説明させていただきたいと思います。

視聴者へのアンケート①

- 損保数理のテキストで学ぶ（学んだ）手法を実務で活用していますか？
 - ① していない
 - ② 時々使うことがある
 - ③ よく使う

4

本セッションの最初のパートでは、幾つか、視聴者の皆様へアンケートをさせていただきたいと思っております。3つアンケートをご用意しておりますので、「Slido」でご回答いただけ

ればと思います。

最初の質問は、「損保数理のテキストで学んだ手法を実務で活用していますか」ということで、「①していない」から「③よく使う」まで、3つの回答をご用意させていただきました。順次投稿可能ですので、しばらく発表の方を進めさせていただいたうえで、投票結果をご紹介しますと思います。

第1部 Loss Modelsの紹介

5

続きまして、第1部、「Loss Models」の紹介になります。

ASTIN関連研究会

- 目的・ミッション
 - ASTIN(Actuarial Studies In Non-life insurance) の名称のとおり、主に損害保険関連のアクチュアリー学の研究を目的
 - 実際の活動は損害保険に限らず、生保数理、データサイエンス等、幅広く研究しアクチュアリー会の会員へ還元
- メンバー
 - 44名(2020年11月6日時点)
 - 損害保険会社、生命保険会社、共済団体、再保険会社・ブローカー、信託銀行、監査法人、コンサルティング会社、その他事業会社等と多様なバックグラウンドを持つメンバーから構成

6

その前に、ASTIN 関連研究会について、少しご紹介させていただきたいと思います。ASTINですけれども、その名のとおり、損害保険のアクチュアリー学の研究を目的としていますが、実際の活動は損害保険に限らず、生保数理、データサイエンスなど、幅広く活動を行っております。

す。

また、メンバーも、現在 44 名おり、損害保険会社に限らず、生保、共済、保険会社以外の事業会社の方も含め、幅広いメンバーで構成されております。

ASTIN関連研究会の活動

- 現在の主な活動
 - **Loss Models 5th Editionの輪読**
 - ☞ **本日も紹介**
 - 損保数理のLecture notes (ETHのMario Wüthrich教授) 輪読
 - ☞ 12月から輪読開始
 - ASTIN Bulletin 掲載論文紹介
 - ☞ ASTIN Bulletin Vol.49掲載論文のいくつかについて深掘り研究中
 - 予測モデリングに関する応用事例調査・研究
 - ☞ 保険分野における予測モデリング手法について海外の事例を調査・研究中
 - IAA ASTIN Local Chapterとの連携
 - ☞ ASTIN Japan Chapterの立ち上げに向けて準備中

7

ASTIN 関連研究会の活動ですけれども、現在の活動としましては、本日も紹介する「Loss Models」の輪読以外にも、さまざまな活動を行っております。2点目の、「損保数理のLecture notes」は、12月から開始を予定しております。また、「ASTIN Bulletin」という IAA が発表している論文集のアブストラクトの日本語サマリーをジャーナルに投稿していたり、予測モデルに関する研究を行ってセミナーなどで会員の皆様に還元していたりもします。また、一番下になりますけれども、このたび IAA の ASTIN の日本支部を立ち上げるに当たって、ASTIN 関連研究会と ASTIN Local Chapter での連携も予定しております。

ASTIN関連研究会の活動

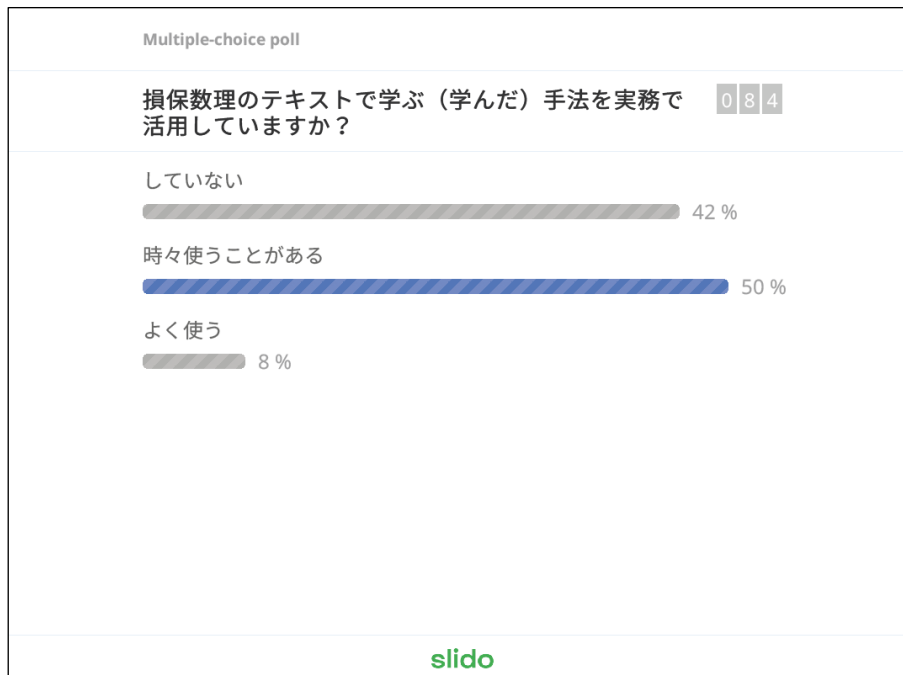
■ 近年の活動実績

年度	活動内容	テーマ・トピックなど
2020	ASTIN Bulletin翻訳	Volume. 49 / No.1-3の論文要旨
	ムーンライトセミナー講師	Accurate GLM (AGLM)の研究
2019	年次大会発表	保険における予測モデリングの応用（実践編）
	文献翻訳	Predictive Modeling Applications in Actuarial Science: Volume II (日本語タイトル：保険数理における予測モデリングの応用 第II巻)
	ASTIN Bulletin翻訳	Volume. 48 / No.1-3の論文要旨
2018	年次大会発表	アクチュアリーとモデル選択 (Accurate GLM)
	年次大会発表	支払備金の見積もりにおける機械学習・深層学習的手法の適用と従来手法との比較 (2018ASTIN WP 参加報告)
	ムーンライトセミナー講師	フレディティブ・モデリングの実践
	ASTIN Bulletin翻訳	Volume. 47 / No.1-3の論文要旨
2017	年次大会発表	フレディティブモデリングの保険データへの応用 ("ASTIN COLLOQUIUM 2017"参加報告)
	文献翻訳	Predictive Modeling Applications in Actuarial Science: Volume I (日本語タイトル：保険数理における予測モデリングの応用 第I巻)
	ASTIN Bulletin翻訳	Volume. 46 / No.1-3の論文要旨

上記の他、他委員会・他団体と共同でのセミナー開催、メンバーの海外論文発表のサポート等、幅広く活動

8

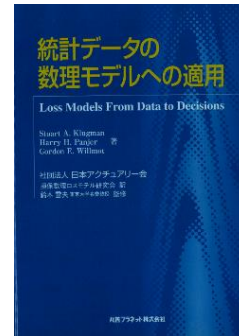
近年の活動実績としまして、表のように、年次大会の発表やムーンライトセミナーなど、さまざまなところで活動をさせていただいております。



では、一旦、皆様に投稿していただいた結果を参照させていただきたいと思っております。今、結果を画面に表示させていただきましたけれども、「①していない」という方と「②時々使う」という方がおおむね半々程度で、「③よく使う」という方が8%という結果になりました。損保数理のテキストで学んだ内容が、実際のところ、実務の上でほとんど使われていないということが、皆様のご回答で分かりました。私自身もそうですけれども、テキストで学んだ内容が、うまく実務に生かせていない。もしくは、テキストと実務の中に、まだギャップがあったりというところがあるのかと思います。

Loss Models

- 米国のアクチュアリー会SOA（Society of Actuary）試験において、日本の損保数理に相当するExam STAM：Short-Term Actuarial Mathematicsの指定テキストである。
- アクチュアリー会の損保数理ロスモデル研究会が翻訳を行い、2004年に「統計データの数理モデルへの適用」、2005年に「統計データの数理モデルへの適用－演習編－」としてアクチュアリー会から発行されている。
※画像は「統計データの数理モデルへの適用」



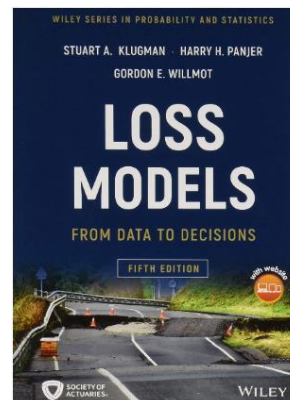
9

では、スライドの方に戻らせていただきます。続きまして、「Loss Models」のテキストの紹介になります。先ほどご紹介したとおり、SOAのテキストとなっております。

ご存じの方もいらっしゃるかもしれませんが、アクチュアリー会でも2004年と2005年に翻訳が行われておりまして、こちらの画像に載っている本が、アクチュアリー会から購入可能となっております。

Loss Models 5th Edition

- 本日のセッションで紹介する5th Editionは2019年に発行された。
- 初版が発行された1998年以降、SOAの試験カリキュラム変更に合わせて改訂が行われている。5th Editionにおいても試験制度の変更に合わせた改訂に加え、例題・演習の追加などが行われている。
- 確率統計の手法や損害額の推定方法などの基礎的な知識に加え、これらの手法のアクチュアリー実務への応用例として豊富な例題と演習問題が提供されており、日本のアクチュアリーにとっても有用なテキストであると考えられる。



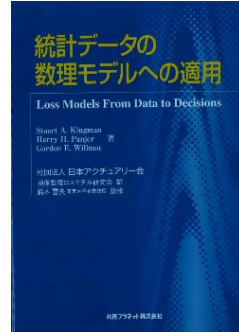
10

今回のセッションでご紹介するのは、この「Loss Models」の5th Editionになっておりまして、2019年、昨年発行されたものになります。初版は1998年の発行ですけれども、試験のカリキュラムの変更に合わせて、改訂が行われております。今回の第5版においても、演習問題の追加などの改訂が行われております。

視聴者へのアンケート②

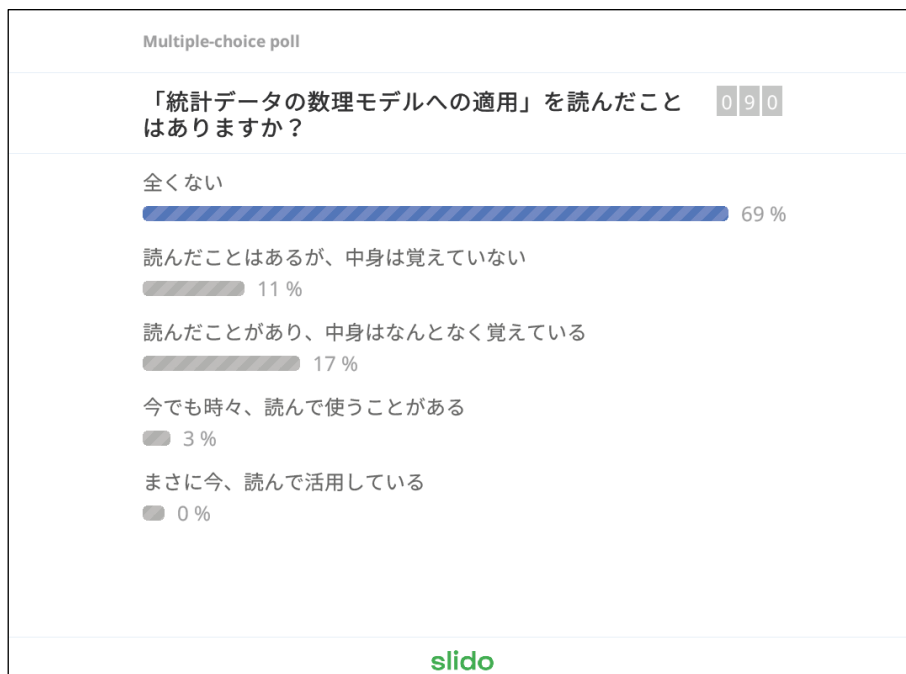
■ 「統計データの数理モデルへの適用」を読んだことはありますか？

- ① 全くない
- ② 読んだことはあるが、中身は覚えていない
- ③ 読んだことがあり、中身はなんとなく覚えている
- ④ 今でも時々、読んで使うことがある
- ⑤ まさに今、読んで活用している



11

では、続いて、2つ目のご質問をさせていただきたいと思います。今、ご紹介しました「統計データの数理モデルへの適用」、「Loss Models」の日本語訳ですけれども、読んだことがある方はどのくらいいらっしゃるかということで、アンケートを取らせていただきたいと思います。1番が「①全くない」、5番が「⑤まさに今、読んで活用している」ということで、手元に置いているような方は、5番を選んでいただければと思います。



結果がおおむね出たようですので、ご紹介したいと思います。全体の7割の方が、「①全くない」ということになっております。「②読んだことはあるが、中身は覚えていない」という方、「③読んだことがあり、中身はなんとなく覚えている」という方が、合計で30%弱でした。「①全くない」がほとんどかなと事前に予想していたのですが、思ったよりはいるかなと思っ

ています。私自身も、今回の発表や輪読にあたって初めて購入して、読み始めたという状況ですが、本セッションで興味を持った方も、ぜひ購入して読んでいただければと思います。

Loss Models 5th Edition目次

Part	Chapter
Part I 序論 (Introduction)	1. モデリング (Modeling)
	2. 確率変数 (Random Variables)
	3. 基本統計量 (Basic Distributional Model)

- Section1ではモデリングプロセスとして以下の6段階が示されている。

```

graph TD
    A[経験および事前知識] --> B[第1段階  
モデル候補の列挙]
    C[データ] --> D[第2段階  
モデルの形成]
    B --> D
    D --> E[第3段階  
モデルの有効性評価]
    E --> F{第4段階  
その他のモデル  
があるか?}
    F -- はい --> B
    F -- いいえ --> G[第5段階  
モデルの選定]
    G --> H[第6段階  
将来に向けた修正]
    
```

- Chapter2では確率変数、確率密度関数などを定義し、Chapter3ではモーメント、VaR、T-VaRといった統計量について説明されている。

ここからは、「Loss Models」のチャプターのご紹介です。Part I は、確率変数、基本統計量といった内容のご紹介になっております。

Loss Models 5th Edition目次

Part	Chapter
Part II 保険数理モデル (Actuarial Models)	4. 保険数理モデルの特徴 (Characteristics of Actuarial Models)
	5. 連続モデル (Continuous Models)
	6. 離散分布 (Discrete Distributions)
	7. 発展的な離散分布 (Advanced Discrete Distributions)
	8. 補償内容の変更を伴う場合の頻度と損害規模 (Frequency and Severity with Coverage Modifications)
	9. 総ロスモデル (Aggregate Loss Models)

- 本パートでは頻度分布と損害規模分布のそれぞれについて紹介されるとともに、日本の損保数理のテキスト第2章「クレームの分析」で取り上げられているクレーム総額の分布について説明されている。
- また、本書のChapter6では以下のような点についても取り上げられている。
 - 分布の特性についての実務上の解釈
 - 免責金額を変更したときの影響について、規模分布だけでなく頻度についても言及

➔第2部で紹介

Part II は、前半が確率分布、後半が損保数理でも取り上げられるような内容になっております。この中の一部については、後ほど第2部で、佐野さんからご紹介していただきたいと思っております。

Loss Models 5th Edition目次

Part	Chapter
Part III 数理統計学 (Mathematical Statistics)	10. 数理統計学入門 (Introduction to Mathematical Statistics)
	11. 最尤推定 (Maximum Likelihood Estimation)
	12. 離散分布の頻度推定 (Frequentist Estimation for Discrete Distributions)
	13. ベイズ推定 (Bayesian Estimation)

- Chapter10では統計学の復習として、推定量の望ましい性質（不偏性、一致性等）について説明されるとともに、本書の数値例に使用するデータセットについて説明されている。
- Chapter11および12ではモーメント法や最尤法によるパラメータの推定手法が説明されている。特に最尤法については、以下のような場合についても説明されている。
 - データがグループ化されている場合の最尤法 →第2部で紹介
 - 切断や打ち切りがある場合の最尤法
 - 尤度比に基づく信頼区間の構成 →第2部で紹介
- Chapter13ではベイズ推定の基本的事項として、事前確率、事後確率、ベイズの定理などについて説明されている。ベイズ推定は信頼性理論のうちベイズ的手法の基礎でもある。

14

続いて Part III ですけれども、こちらは、統計に関する内容になります。最尤推定や尤度比といった日本の数学の試験で学ぶ内容に加えて、日本の数学、損保数理のテキストでは、あまり大きく取り扱われていないベイズ推定に比較的ページを割いて紹介されているところが、特徴的かと思います。

Loss Models 5th Edition目次

Part	Chapter
Part IV モデル構築 (Construction of Models)	14. 経験モデルの構築 (Construction of Empirical Models)
	15. モデル選択 (Model Selection)

- Chapter14ではChapter10で紹介したデータセットを用いて、経験分布の構築方法について説明されている。免責金額や支払限度額により、ロスの実態が分からないデータが混在しているときに対する規模分布の推定方法として以下が紹介されている。
 - Kaplan-Meier法 →第2部で紹介
 - Nelson-Åalen法
- また、Chapter14では、経験分布を連続分布の密度関数を近似する方法として、カーネル密度関数が説明されている。 →第2部で紹介
- Chapter15ではモデル選択のアプローチについて体系的に整理されている。 →第2部で紹介
- 例えばその1つである仮説検定については、コルモゴロフ-スミノフ検定（K-S検定）、アンダーソン-ダーリング検定（A-D検定）、カイ二乗検定、尤度比検定のそれぞれについて、どのような場合に適用可能か例示とともに説明されている。

15

Part IV は実務寄りの内容になっておりまして、先ほどの日本語版から、今回の第5版の英語版で、大きく改訂されている箇所になります。

経験モデルの構築について取り上げられています。また、モデル選択では、データを使って経験モデルを構築して、それをどのように連続モデルで近似するかといった内容や、どのモデルを選べばよいかといった、実際に実務で使えるようなアプローチが紹介されております。こ

ちらも、第2部でご紹介したいと思います。

Loss Models 5th Edition目次

Part	Chapter
Part V 信頼性理論 (Credibility)	16. 有限変動信頼性理論入門 (Introduction to Limited Fluctuation Credibility)
	17. 最大精密信頼性理論 (Greatest Accuracy Credibility)
	18. 経験ベイズパラメータ推定 (Empirical Bayes Parameter Estimation)

- 本パートでは日本の損保数理のテキスト第3章「経験料率」でも取り上げられている信頼性理論について説明されている。 →本パート全般について第3部で紹介
- Chapter16では有限変動信頼性理論について説明されている。
- Chapter17ではベイズ的手法、信頼性保険料、Bühlmannモデル、Bühlmann-Straubモデルについて説明されている。また、完全信頼性 (Exact Credibility) という概念についても説明されている。
- Chapter18ではBühlmannモデルとBühlmann-Straubモデルにおけるパラメータの推定方法について、ノンパラメトリックな場合とセミパラメトリックな場合のそれぞれについて説明されている。

16

Part V が信頼性理論になります。有限変動信頼性理論や、ビュールマンといった内容が紹介されております。こちらについても第3部でご紹介したいと思います。

視聴者へのアンケート③

- 信頼性理論を実務で使ったことがありますか？
 - ① 使ったことはない
 - ② 有限変動信頼性理論のみ使ったことがある
 - ③ ベイズ的手法・ビュールマンモデル等のみ使ったことがある
 - ④ 上記の手法を使い分けている

17

3番目のアンケートになります。有限変動信頼性理論、またはビュールマンといった信頼性理論について、実務でどの程度使われているか、アンケートでご回答いただければと思います。

Loss Models 5th Edition目次

Part	Chapter
Part VI シミュレーション (Simulation)	19. シミュレーション (Simulation)

- 本パートではシミュレーションに関する以下のような点について説明されている。
- 確率変数のシミュレーションにおける一般的なステップ
- 以下のような分布のシミュレーション法
 - 離散混合分布
 - 生命表からの死亡時期や年齢のシミュレーション
 - $(a, b, 0)$ クラスの分布 ※ポアソン分布、二項分布、負の二項分布
 - 正規分布、対数正規分布
- シミュレーションに必要なサンプルサイズ
- 保険数理モデリングの以下のようなケースへの適用例
 - 総ロスモデル
 - 変数間の独立が仮定できない場合
 - リスク指標の計算

18

Part VI が最後のパートになります。本パートでは、シミュレーションに関して、実際にサンプルデータを用いて手法の紹介が行われております。日本のアクチュアリー試験でも、モデリングのテキストの中でシミュレーションについて記載がありますが、それに加えて、生命表のシミュレーションや、一番下にあるように総ロスモデルや VaR、TVaR といったリスク指標について、シミュレーションによって計算する手法も、この中で取り上げられております。

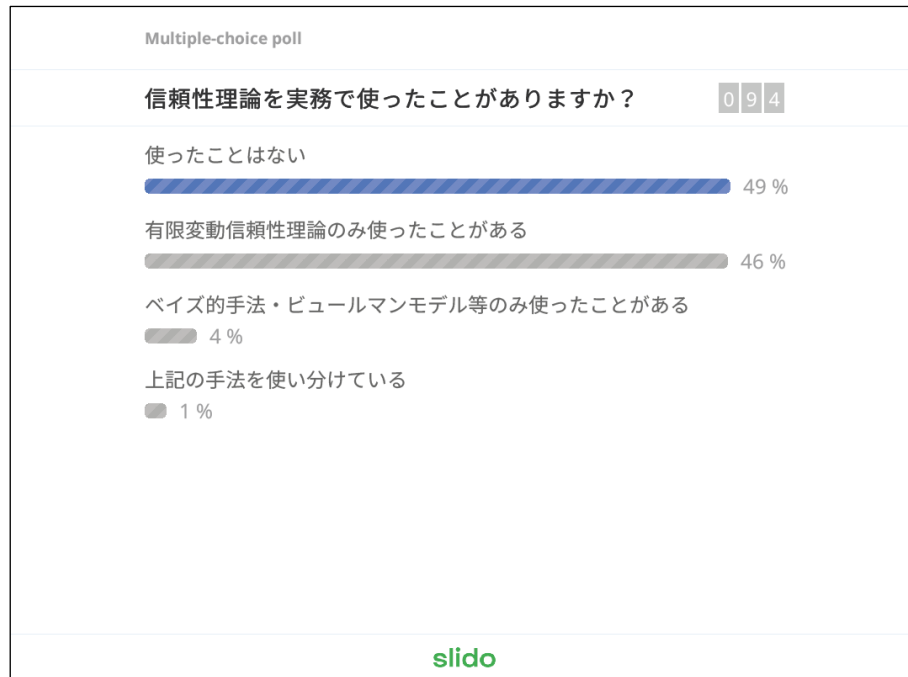
Loss Models 5th Edition目次

Part	Chapter
付録 (Appendix)	A. 連続分布の一覧 (An Inventory of Continuous Distributions)
	B. 離散分布の一覧 (An Inventory of Discrete Distributions)
	C. 頻度と損害規模の関係 (Frequency and Severity Relationships)
	D. 再帰公式 (The Recursive Formula)
	E. 損害規模分布の離散化 (Discretization of the Severity Distribution)

- 付録A、Bでは様々な分布の密度関数（確率関数）、分布関数、統計量がまとめられている。様々な分布が類似の分布族や型に類型化されている点が特徴である。
- 付録Cでは免責金額が設定された場合の総ロス件数の分布についてまとめられている。
- 付録DではChapter9で紹介されているロス総額分布の再帰式における、初期値（ロス総額 $S=0$ の確率）についてまとめられている。
- 付録Eでは連続な損害規模分布を離散分布に変換する手法として、ラウンド法と平均保持法が紹介されている。

19

最後に付録についても紹介します。A から E までありますが、大変幅広い種類の分布について公式が整理されているというところが特徴的だと思います。例えば、確率分布について、密度関数や分布関数だけではなくて、モーメントといったところについても一覧の形で整理されておりまして、非常に有用な内容かと思えます。



私のパートの最後に先ほど取らせていただいた、信頼性理論についてのアンケートの結果をご紹介します。①「使ったことがない」という方が49%、②「有限変動信頼性理論のみ使ったことがある」という方が46%、ベイズ的手法やビュールマンモデルなどを使用されたことがある方は僅か5%という結果になりました。信頼性理論を使ったことがないという方が多いのではないかと思いましたが、有限変動信頼性理論だけであれば使ったことがあるという方が、半分近くいることが分かりました。

本セッションの最後のパートで、川上さんの方から、信頼性理論について、全体の理論などについても丁寧に時間を取ってご説明させていただきますので、ぜひ本日のセッションを参考にして、有限変動信頼性理論以外の信頼性理論も、活用の機会が得られればと考えております。

アンケートなどで時間を取ってしまいましたので、頂いたご質問に回答する時間がなくなりましたけれども、第1部の私のセッションは、以上とさせていただきます。続きまして、第2部の佐野さんに交代させていただきたいと思っております。

【司会】 島本さん、ありがとうございます。発表者交代準備のため、少々お待ちください。それでは、佐野さん、よろしく申し上げます。

第2部 Loss Modelsによる 保険実務に即した統計手法の学習

20

【佐野】 はい。共栄火災の佐野と申します。第2部からは、本の内容について説明していきます。

Loss Modelsによる学習の利点

Part	Chapter
Part I 序論	1. モデリング
	2. 確率変数
	3. 基本統計量
Part II 保険数理 モデル	4. 保険数理モデルの特徴
	5. 連続モデル
	6. 離散分布
	7. 発展的な離散分布
	8. 補償内容の変更を伴う場合の頻度と損害規模
	9. 総ロスモデル
Part III 数理統計学	10. 数理統計学入門
	11. 最尤推定
	12. 離散分布の頻度推定
Part IV モデル構築	13. ベイズ推定
	14. 経験モデルの構築
Part V 信頼性理論	15. モデル選択
	16. 有限変動信頼性理論入門
	17. 最大精密信頼性理論
Part VI	18. 経験ベイズパラメータ推定
	19. シミュレーション
付録	

- 実務上の観点に基づく統計手法の解説
- 保険データ特有の課題に対する取扱い

➡ **重要である（と勝手に解釈した）
部分をピックアップして説明**

1. 保険実務の例による統計手法の学習
2. 分布の解釈
3. 免責金額の変更による頻度の影響
4. グループ化されたデータの最尤推定
5. 最尤推定量の（尤度比による）信頼区間
6. 免責・限度額が混在するデータのロス分析
7. カーネル密度関数による経験分布の平滑化
8. モデル選択

21

まず損保数理の手法を実務で活用するに当たって、「Loss Models」による学習は2つの利点があります。1つは、実務上の観点から必要な手法を、必要な形で学べること。

もう1つは、保険データ特有の課題に対する取扱いを学べることです。ここでは本の内容を全て説明する時間はありませんので、この2つの観点から重要と思われる部分、そして、教科書や試験勉強では学ぶことが難しい部分について、各章からピックアップして紹介していきます。ラインナップは、記載のとおりです。

Loss Modelsによる統計手法の学習

Part	Chapter
Part I 序論	1. モデリング
	2. 確率変数
	3. 基本統計量
Part II 保険数理 モデル	4. 保険数理モデルの特徴
	5. 連続モデル
	6. 離散分布
	7. 発展的な離散分布
	8. 補償内容の変更を伴う場合の頻度と損害規模
	9. 総ロスモデル
	10. 数理統計学入門
Part III 数理統計学	11. 最尤推定
	12. 離散分布の頻度推定
	13. ベイズ推定
Part IV モデル構築	14. 経験モデルの構築
	15. モデル選択
Part V 信頼性理論	17. 最大積密信頼性理論
	18. シミュレーション

1. 保険実務の例による統計手法の学習
2. 分布の解釈
3. 免責金額の変更による頻度の影響
4. グループ化されたデータの最尤推定
5. 最尤推定量の（尤度比による）信頼区間
6. 免責・限度額が混在するデータのロス分析
7. カーネル密度関数による経験分布の平滑化
8. モデル選択

22

まずは「保険実務の例による統計手法の学習」です。

「Loss Models」では基本統計から始まって、たくさんの実務に即した例や演習が用意されているために、それを見ることによって、その手法の使い方を具体的に学ぶことができます。

1. 保険実務の例による統計手法の学習

- 基本統計から、保険実務の例によってその使用方法を具体的に学ぶことができる
 - (例) 分布の収束（中心極限定理）

医療保険契約1,000件に対する1年間のロス実績を集計すると、
（契約1件あたりの支払は）標本平均1,300かつ標準偏差400だった。
次年度は2,500件の契約を見込んでいる。
次年度の支払総額が期待支払額の101%を超える確率を求めよ。



$$S_k = X_1 + \dots + X_k \text{ に対して}$$

$$\begin{cases} E(S_k) = E(X_1) + \dots + E(X_k) \\ V(S_k) = V(X_1) + \dots + V(X_k) \end{cases}$$

$$\lim_{k \rightarrow \infty} [S_k - E(S_k)] / \sqrt{V(S_k)} \sim N(0,1)$$

次年度の契約集団に対する支払総額の見込みは、
平均 $2,500 \times 1,300$ ・ 標準偏差 $\sqrt{2,500} \times 400$ の正規分布で近似できる

23

例えば中心極限定理は非常にベーシックな定理で、教科書では定理だけが載っているようなものですが、「Loss Models」ではこのような演習が用意されていて、いわゆるロスの統計量が分かれば、支払総額はその統計量をパラメーターとする正規分布で近似できるという説明があります。このような例や演習があれば、「定理は分かっているが、使えるかどうかという意味では理解していない」という場合であっても、自分のデータを例や演習に当てはめることで、自分が何をしているのか、何をしたいのかを理解できますので非常に便利です。

1. 保険実務の例による統計手法の学習

- 基本統計から、保険実務の例によってその使用方法を具体的に学ぶことができる

- (例) 仮説検定

平均クレーム額が1,200であるとして保険料を算出している商品について、保険料を引き上げたい。
 ※過去の実績に基づくクレーム額の平均は1,200を超えている。
 ただし監督当局は引上げの根拠を求めている。

- ・母集団の平均 μ に対する
 帰無仮説・対立仮説... $\begin{cases} H_0: \mu \leq 1,200 \\ H_1: \mu > 1,200 \end{cases}$
- ・棄却域、第1種・第2種の誤り
- ・有意水準、p値に基づく判断

次の点についても言及している
 ・仮説検定はデータ数が少ない場合に何らかの判断を下すための手法であり、データ数が大きければほとんどのモデルは棄却される
 ・有意水準は任意であり、その水準次第で結果は何とでもなり得る

⇒モデル選択における尤度比検定で再登場する

他にも仮説検定、これもベーシックなものですが、まず課題、しかも、監督当局から保険料引き上げの根拠を求められるという生々しい課題から始まって、その対応という形で話は進みます。ですから実務でそのようなものを扱っている人は、いやでも心で覚えてしまいますので、理解が深まるのではないかと思います。また、この課題をそのとおりに読むことも勉強になりますが、別の見方として、過去の実績による平均は1,200を超えているが、それだけを理由に保険料を引き上げなければならないのかという見方も、また一つ勉強になるのではないかと思います。

Loss Modelsによる統計手法の学習

Part	Chapter
Part I 序論	1. モチベーション
	2. 基礎統計学
	3. 基本統計学
Part II 保険数理モデル	4. 保険数理モデルの特徴
	5. 連続モデル
	6. 離散分布
	7. 発展的な離散分布
Part III 数理統計学	8. 線形モデル
	9. 線形モデル
	10. 線形モデル
	11. 最尤推定
Part IV 信頼性理論	12. ベイズ推定
	13. ベイズ推定
	14. ベイズ推定
	15. モデル選択
Part V 信頼性理論	16. 最大積乗性信頼性理論
	17. 最大積乗性信頼性理論
	18. シミュレーション
	19. シミュレーション

1. 保険実務の例による統計手法の学習
2. 分布の解釈
3. 免責金額の変更による頻度の影響
4. グループ化されたデータの最尤推定
5. 最尤推定量の(尤度比による)信頼区間
6. 免責・限度額が混在するデータのロス分析
7. カーネル密度関数による経験分布の平滑化
8. モデル選択

次が「分布の解釈」です。「Loss Models」では、頻度や損害規模のモデル化に使う幾つかの分布が紹介されていますが、ここでは、そのうち頻度モデルに使う離散型分布の解釈について

ピックアップします。

2. 分布の解釈

- 分布の特性について実務上の解釈を説明している
 - ポアソン分布の分解可能性に関する定理

定理

ある事象の件数 N が平均 λ のポアソン分布に従うとする。
さらに事象は独立した m 種類に分類され、
それぞれが $p_1 \sim p_m$ の確率で発生すると仮定する。
このとき $N_1 \sim N_m$ を各事象の件数とすると、
それぞれは独立で平均 $\lambda p_1 \sim \lambda p_m$ のポアソン分布に従う

解釈

例えば医療保険のクレーム件数がポアソン分布に従うとしたとき、
特定の疾病を分類または除去したとしても、
そのクレーム件数はポアソン分布に従う

26

まずは「ポアソン分布の分解可能性に関する定理」です。これは、ポアソン分布に従う確率変数が独立に分類されたとしても、それぞれがポアソン分布に従うというものです。ポアソン分布にはこの他にも便利な性質が幾つかありますが、実務で使う場合には、このような背景にある性質を理解せずに、例えば手法だけ引き継いで、何も考えずに計算してしまうような場合があるかもしれません。また、これ自体は教科書にも載っていませんので、便利な手法というよりは知っておくべき前提として紹介しておきます。

2. 分布の解釈

- 分布の特性について実務上の解釈を説明している
 - 負の二項分布
 - パラメータ λ がガンマ分布に従う（混合）ポアソン分布

適用するリスク

- ・自動車保険のある料率区分に属する契約集団のような、保険料水準から見るとリスクが同質な集団を考える
- ・その区分の保険料は同額であったとしても、リスク実態は様々な要因（例えば車庫の場所）により異なり得る
- ・この場合、個々の契約者の事故件数はポアソン分布に従うとしてもそのパラメータ λ は観測できない不確実性を持つ

ただし、負の二項分布は複数の概念から導き出されることを踏まえて、逆の解釈が成り立つわけではないとも説明している
(例えば、あるクレーム件数に負の二項分布がよく当てはまるからといって、それがパラメータがガンマ分布に従うポアソン分布だと必ずしも言えるわけではない)

27

次が、パラメータ λ がガンマ分布に従うポアソン分布であるということについて、負の二項分布について適用すべきリスクの解釈です。ここでは、保険料が同額という意味でリスクを

同質と判断している集団であっても、そのリスク実態がさまざまな要因により異なりうる場合は、保険料からは同質であって個々の事故件数がポアソン分布に従うとしても、保険料側からは観測できない不確実性を持っているため、このような場合には負の二項分布を適用することが妥当だという説明があります。教科書にも負の二項分布が混合ポアソン分布であるということ自体の記載はありますが、このような具体的なものではありませんし、例えば GLM でポアソン分布を当ててみたら過分散なので負の二項分布にするというようなこともよく見る形ですが、これを専門ではない人に説明すると「過分散？」のようなことになると思うので、そのようなときの説明のストックとして読んでおくと便利だと思います。

Part	Chapter
Part I 序論	1. はじめに
	3. 基本統計量
	5. 連続モデル
Part II 保険数理 モデル	7. 保険的な損益分布
	8. 補償内容の変更を伴う場合の頻度と損害規模
	9. 損失モデル
Part III 数理統計学	11. 最尤推定
	13. ベイズ推定
	15. モデル選択
Part V 信頼性理論	17. 最大積率信頼性理論
	19. シミュレーション

1. 保険実務の例による統計手法の学習
2. 分布の解釈
3. 免責金額の変更による頻度の影響
4. グループ化されたデータの最尤推定
5. 最尤推定量の（尤度比による）信頼区間
6. 免責・限度額が混在するデータのロス分析
7. カーネル密度関数による経験分布の平滑化
8. モデル選択

次が「補償内容の変更を伴う場合の頻度と損害規模」から、「免責金額の変更による頻度の影響」です。ここでは、いわゆる支払限度額や免責金額、コ・インシュアランスやインフレなどを設定することによって、ロスの分布から支払金額の分布がどのように変化するかについての説明があります。それ自体は試験勉強でももちろん勉強することですが、「Loss Models」では、そのうち、免責金額を変更したときに補償の対象が増減することについての頻度の影響についても言及しています。

3. 免責金額の変更による頻度の影響

- 免責金額を変更したときの影響について、規模分布だけでなく頻度についても言及している

支払金額 X 、免責 d に対して $v = \Pr(X > d)$ としたときの頻度分布パラメータの調整値が、一覧として載っている

ロスの頻度分布	支払の頻度分布のパラメータ
ポアソン分布	$\lambda^* = v\lambda$
ゼロ調整ポアソン分布	$p_0^{M*} = \frac{p_0^M - e^{-\lambda} + e^{-v\lambda} - p_0^M e^{-v\lambda}}{1 - e^{-\lambda}}, \lambda^* = v\lambda$
二項分布	$q^* = vq$
ゼロ調整二項分布	$p_0^{M*} = \frac{p_0^M - (1-q)^m + (1-vq)^m - p_0^M(1-vq)^m}{1 - (1-q)^m}, q^* = vq$
負の二項分布	$\beta^* = v\beta, r^* = r$
ゼロ調整負の二項分布	$p_0^{M*} = \frac{p_0^M - (1+\beta)^{-r} + (1+v\beta)^{-r} - p_0^M(1+v\beta)^{-r}}{1 - (1+\beta)^{-r}}, \beta^* = v\beta, r^* = r$

29

具体的には、頻度モデルに使う離散型分布のパラメーターが、免責金額を変更することでどのように変化するかについての一覧が載っています。概念自体は納得感のあるものなので、あとは計算だけが面倒だというときであっても、このような表があれば、その計算も省くことができ非常に便利だということで、ここで紹介しております。

Loss Modelsによる統計手法の学習

Part	Chapter
Part I 序論	1. モデル링
	2. 基本統計量
	3. 確率モデル
	4. 連続的な確率分布
Part III 数理統計学	11. 最尤推定
	12. 離散分布の頻度推定
	13. ベイズ推定
	14. モデル選択
	15. モデル選択
Part V 信頼性理論	17. 最大積密度信頼性理論
	18. カルネン密度関数による経験分布の平滑化
	19. シミュレーション
	20. 補遺

1. 保険実務の例による統計手法の学習
2. 分布の解釈
3. 免責金額の変更による頻度の影響
4. グループ化されたデータの最尤推定
5. 最尤推定量の（尤度比による）信頼区間
6. 免責・限度額が混在するデータのロス分析
7. カーネル密度関数による経験分布の平滑化
8. モデル選択

30

次が、最尤法から2つです。

4. グループ化されたデータの最尤推定

■ 分布関数を用いた最尤推定

- (連続) 支払金額がグループ化されている場合

支払金額の範囲	観測値の数
0-7,500	99
7,500-17,500	42
17,500-30,000	29
Over 30,000	17

$$\Rightarrow L = \prod_{j=1}^k [F(c_j) - F(c_{j-1})]^{n_j}$$

- (離散) 一定数以上がグループ化されている場合

クレーム件数	観測値の数
0	47
1	97
2	109
3+	62

$$\Rightarrow L = p_0^{n_0} p_1^{n_1} p_2^{n_2} (1 - p_0 - p_1 - p_2)^{n_3}$$

31

1つ目は「グループ化されたデータの最尤推定」。これは、データがグループ化されている場合や、一般統計にあるような一定の数より上がまとまってしまっている場合、このような場合は尤度関数に分布関数を使うというものです。これ自体はそれほど珍しいものではありませんし、日本語版を読んでいる人にとっては当たり前ですが、たまにこれを知らない人がいて、この手法を使ったときに「何をもって」と聞かれることがあるので、そのようなときに「これです」と言うために、ここで紹介しております。

5. 最尤推定量の(尤度比による)信頼区間

■ 漸近正規性ではなく、尤度比に基づくもの

- パラメータが複数
- 標本数が少ない
- 尤度関数が非対称

こういうときに有効

- 尤度関数を最大化するパラメータ値の周辺 (ある c に対する $\{\theta; l(\theta) \geq c\}$) を考える

$$\Rightarrow 100(1 - \alpha)\% \text{信頼区間 } \{\theta; l(\theta) \geq l(\hat{\theta}) - 0.5\chi_{\alpha}^2\}$$

- θ : パラメータベクトル = $(\theta_1, \theta_2, \dots, \theta_n)^T$
- $l(\theta)$: 対数尤度関数
- $\hat{\theta}$: 最尤推定量

32

次が「最尤推定量の(尤度比による)信頼区間」です。これは、いわゆる教科書にあるような漸近正規性を仮定するのではなく、尤度比に基づくものです。パラメーターが複数で相関が見られる場合や、標本数が少ない場合、尤度関数が非対称である場合に有効であるといわれて

います。考え方としては、最尤法が尤度関数を最大化するパラメータを探すものなので、その値の周辺、ここの式にある c に当たる部分を探すというものです。

これに対して、尤度比検定などで使う、尤度比がカイ二乗分布で近似できることを利用して信頼区間を作ります。

5. 最尤推定量の（尤度比による）信頼区間

- ガンマ分布の場合（95%信頼区間）

下記ロスデータから推定

27	82	115	126	155	161	243	294	340	384
457	680	855	877	974	1,193	1,340	1,884	2,558	15,743

$$\Rightarrow (l(\alpha, \theta) =) (\alpha - 1) \sum_j^n \ln x_j - \frac{1}{\theta} \sum_j^n x_j - n \ln \Gamma(\alpha) - n\alpha \ln \alpha$$

$$\geq l(\hat{\alpha}, \hat{\theta}) - 0.5\chi_{\alpha}^2$$

$$= -165.289$$

を満たす α, θ

33

これがロスデータの例と、実際にそれをガンマ分布で当ててみたものです。この図にある白い部分が信頼区間になります。ガンマ分布はパラメータが二つで、尤度関数をこのロスデータに当ててみるとこのようないびつな形になりますので、信頼区間も図にあるいびつな形となります。

信頼区間そのものでこのような形のものを見ても、あまり興味が湧かない。それよりは平均の信頼区間などが興味の対象だと思いますので、それについての説明もあります。

5. 最尤推定量の（尤度比による）信頼区間

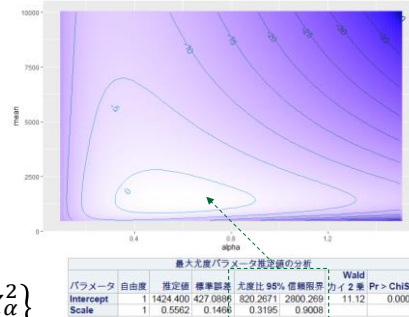
- ガンマ分布の場合（平均の95%信頼区間）
 - 平均パラメータ（ $\mu = \alpha\theta$ ）を用いて密度関数を変換

$$f(x) = \frac{x^{\alpha-1} e^{-x\alpha/\mu}}{\Gamma(\alpha)(\mu/\alpha)^\alpha}$$

- パラメータベクトルのうちのひとつ θ_j に対する尤度比による信頼区間

$$\{\theta_j; \max_{\theta_{-j}} l(\theta) \geq l(\hat{\theta}) - 0.5\chi^2_{\alpha}\}$$

をプロフィール尤度信頼区間という



34

ガンマ分布の場合は、平均パラメータを作って、それによって密度関数を変換してから出すというやり方です。この図で言うと、縦軸が平均パラメータになりますので、この等高線の山を縦軸に投影してできる山の頂上付近が信頼区間になります。

このような、尤度比による信頼区間のうちパラメータ1つに対するものを、このように縦軸や横軸に投影するという意味で、「プロフィール尤度信頼区間」といいます。この信頼区間は、例えば平均パラメータといえば指数型分布族、指数型分布族といえば GLM なので、GLM を使うときなどで見る場合があります。

Loss Modelsによる統計手法の学習

Part	Chapter
Part I 序論	1. はじめに
	2. 基本統計量
	3. 連続モデル
Part II 経験分布	4. 経験分布
	5. 経験分布関数
	6. 経験分布関数の推定
	7. 経験分布関数の推定
Part III 数理統計学	8. 最尤推定
	9. ベイズ推定
	10. ベイズ推定
Part IV モデル構築	14. 経験モデルの構築
	15. モデル選択
Part V 信頼性理論	16. 最大積乗信頼性理論
	17. 最大積乗信頼性理論
	18. シミュレーション

1. 保険実務の例による統計手法の学習
2. 分布の解釈
3. 免責金額の変更による頻度の影響
4. グループ化されたデータの最尤推定
5. 最尤推定量の（尤度比による）信頼区間
6. 免責・限度額が混在するデータのロス分析
7. カーネル密度関数による経験分布の平滑化
8. モデル選択

35

次が「経験モデルの構築」から「免責・限度額が混在するデータのロス分析」です。ここでは、データが十分にあって、それをうまく表現できればロスの分析ができるような場合の、経

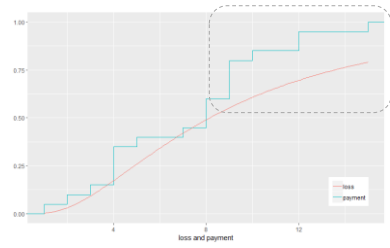
験分布についての手法が説明されています。

6. 免責・限度額が混在するデータのロス分析

- 免責金額や支払限度額により、ロスの実額が分からないデータが混在しているときに対する損害規模モデルの経験分布

(例) 支払金額の一覧
*のついた観測値は支払限度額が適用されている

1	2	3*	4	4	4*	4*	5	7*	8
8	8	9	9	9	9	10*	12	12	15*



限度額の影響を考慮しないとロスの推定が上手くいかない

➡ Kaplan-Meier法を使う

36

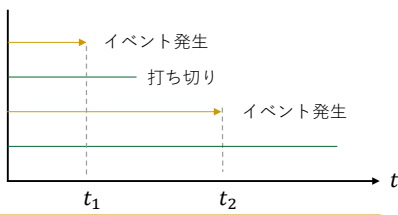
これがロスデータの例です。このうち、*のついているものには支払限度額が適用されているために、実際の損害額はこの額以上だということしか分かりません。このようなデータは物保険や入院保険金などで見られるものですが、限度額が適用されているデータをそのまま使うと分布の裾の方の推定がおかしくなりますし、かといって捨ててしまってもったいない。そのような場合に対応する手法として、Kaplan-Meier法を使います。

6. 免責・限度額が混在するデータのロス分析

- Kaplan-Meier法
 - 打ち切りがあるデータに対して、イベント（死亡など）発生までの時間を解析するノンパラメトリックな手法
 - 打ち切りデータはその時点まで対象としてカウントする
 - 生存関数 $S(t) = Pr(T > t)$ の推定量が

$$\hat{S}(t) = \prod_{t_i < t} \left(1 - \frac{d_i}{r_i}\right)$$

- ・ t_i : 時点
- ・ r_i : 時点 t_i での対象数
- ・ d_i : 時点 t_i でのイベント発生数



37

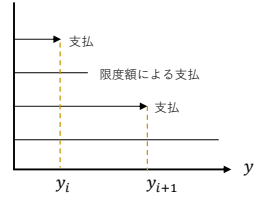
Kaplan-Meier法とは、生存時間や要介護期間のような時間解析に用いるノンパラメトリックな手法です。ポイントとして、観察期間中にデータが打ち切られて実際の時間が分からないようなものについて、その打ち切りの時点までは発生率の母数としてカウントします。それから

イベント発生ごとの発生率を算出し、1-発生率を累積することで生存関数を推定します。

6. 免責・限度額が混在するデータのロス分析

- ロスデータに対するKaplan-Meier法
 - イベント発生の時点を支払金額、打ち切り時点を限度額に置き換える
 - ロスの確率変数 Y に対して
 - ・ $y_1 < \dots < y_k$: 観測値が取る支払金額 (限度額による支払は含まない)
 - ・ y_0 : Y の取り得る最小値 ($y_0 < y_1$)
 - ・ y_{max} : 観測値が取る最大値 ($y_k \leq y_{max}$)
 - ・ s_i : 支払金額が y_i の観測数
 - ・ b_i : 限度額による支払で、その金額が $[y_i, y_{i+1})$ にある観測数
 - ⇒ y_i に対するnumber at risk (r_i)

$$r_i = \begin{cases} n, & i = 1 \\ r_{i-1} - s_{i-1} - b_{i-1}, & i = 2, 3, \dots, k + 1 \end{cases}$$



i	y_i	s_i	b_i	r_i
0	0	-	-	-
1	1	1	0	20
2	2	1	1	20-1-0=19
3	4	2	2	19-1-1=17
4	5	1	1	17-2-2=13
5	8	3	0	13-1-1=11
6	9	4	1	11-3-0=8
7	12	2	1	8-4-1=3
max	15	-	-	3-2-1=0

38

ロスデータに対しては、イベント発生までの時間を支払金額、打ち切り時点を限度額に置き換えてこの手法を使います。

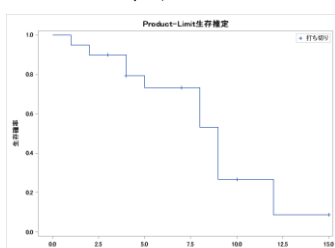
まず、各支払金額の発生率の母数である number at risk を、その金額以下のデータをデータ全体から差し引いていくことで算出します。右にあるものが、先ほどのロスデータの例に対して実際に計算したものです。

6. 免責・限度額が混在するデータのロス分析

- ロスデータに対するKaplan-Meier法
 - $\lambda_j = Pr(Y = y_j | Y \geq y_j)$ の推定量が最尤法により

$$\hat{\lambda}_j = \frac{s_j}{r_j}$$
 - 生存関数 $S(x)$ の推定量が

$$\hat{S}(y) = \begin{cases} 1, & y < y_1 \\ \prod_{i=1}^j \left(1 - \frac{s_i}{r_i}\right), & y_j \leq y < y_{j+1} \\ \prod_{i=1}^k \left(1 - \frac{s_i}{r_i}\right), & y_k \leq y < y_{max} \end{cases}$$



i	y_i	s_i	r_i	$\hat{S}(y_i)$
1	1	1	20	1-1/20=0.950
2	2	1	19	0.950(1-1/19)=0.900
3	4	2	17	0.900(1-2/17)=0.794
4	5	1	13	0.794(1-1/13)=0.733
5	8	3	11	0.733(1-3/11)=0.533
6	9	4	8	0.533(1-4/8)=0.267
7	12	2	3	0.267(1-2/3)=0.089

39

時間を支払金額に置き換えること以外は、Kaplan-Meier法そのものと手法は同じです。

支払金額ごとの発生率を、そのデータ数から先ほどの number at risk で割ることにより算出して、それから 1-発生率を累積することで、生存関数の推定値を出します。右にあるもの

が、先ほどのデータで実際に生存関数を計算したもの、それに対するプロットが上にあります。

Kaplan-Meier は離散的な手法ですので、生存関数のプロットは、このような階段状のものとなります。

6. 免責・限度額が混在するデータのロス分析

■ Kaplan-Meier法（平均の推定）

□ $\mu = E(Y) = \int_0^{\infty} S(y)dy$ を使う

□ $\tau(x) = \int_x^{\infty} S(y)dy$ とすると

$$\hat{\mu} = y_1 + \sum_{j=2}^k (y_j - y_{j-1}) \prod_{i=1}^{j-1} \left(1 - \frac{s_i}{r_i}\right) \\ + (y_{max} - y_k) \prod_{i=1}^k \left(1 - \frac{s_i}{r_i}\right) + \tau(y_{max})$$

⇒ $\hat{\mu}$ は $\tau(y_{max})$ によって決まる

・・・本書では3つの案が提示されている

40

ロスの分析で生存関数だけ見て終わりということは、あまり興味としては無いと思いますので、平均の推定についても説明しています。ここでは、生存関数を積分すれば平均になることを利用します。ですので先ほどの階段状のプロットを積分するということです。式にするとここに書いてあるとおりですが、最後にある $\tau(y_{max})$ は、このデータ上の最大値より先の生存関数です。

これは平均の推定にも必要ですが、データ上はそこより先のデータはありませんので、推定のしようがありません。ですから、みなしの推定が必要になりますが、「Loss Models」ではみなしの推定について3つの案が提示されています。

6. 免責・限度額が混在するデータのロス分析

■ Kaplan-Meier法（裾の推定）

□ $\tau(y_{max})$ の推定に対する3つの案

種類	$\hat{S}(y)$ ($y \geq y_{max}$ のとき)	$\tau(y_{max})$
Efron's tail correction	0	0
Klein and Moeschberger	γ ($\geq y_{max}$) (最終年齢など) に対して $\hat{S}(y) = \begin{cases} \hat{S}(y_k), & y < \gamma \\ 0, & \gamma \leq y \end{cases}$	$(\gamma - y_{max})\hat{S}(y_k)$
Brown, Hollander, and Korwar's exponential tail correction	$e^{-\frac{y[\ln \hat{S}(y_k)]}{y_{max}}} = [\hat{S}(y_k)]^{\frac{y}{y_{max}}}$	$\frac{y_{max}\hat{S}(y_k)}{-\ln \hat{S}(y_k)}$

41

1つ目が0にしてしまうという案。もう1つが、例えば年齢に対する最終年齢のような客観的な最大値を設定して、そこまでは生存関数を伸ばし、そこから先は0とする案。もう1つは、データ上の最大値より先は指数的に生存関数を減少させる案です。これらを使うことで平均の推定ができます。

Loss Modelsによる統計手法の学習

Part	Chapter
Part I 序論	1. モデルリング
	2. 基礎統計量
	3. 連続モデル
	4. 発展的な連続分布
Part II 経験統計学	5. 経験分布
	6. 経験モデル
	7. 経験分布の推定
Part III 数理統計学	8. 最大推定
	9. ベイズ推定
Part IV モデル構築	10. 経験モデルの構築
	11. モデル選択
Part V 信頼性理論	12. 最大推定信頼性理論
	13. シミュレーション
	14. 補足

1. 保険実務の例による統計手法の学習
2. 分布の解釈
3. 免責金額の変更による頻度の影響
4. グループ化されたデータの最尤推定
5. 最尤推定量の（尤度比による）信頼区間
6. 免責・限度額が混在するデータのロス分析
7. カーネル密度関数による経験分布の平滑化
8. モデル選択

42

次が「カーネル密度関数による経験分布の平滑化」。これは同じ章からの紹介です。

7. カーネル密度関数による経験分布の平滑化

- 母集団は連続型であると考えられるデータに対する経験分布を、連続分布のモデルで近似する

- ロスの確率変数に対して

$$\hat{f}(x) = \sum_{j=1}^k p(y_j) k_{y_j}(x)$$

- ・ $y_1 < \dots < y_k$: 観測値が取る支払金額
- ・ $p(y_j)$: 経験分布に基づく y_j の確率

- $k_y(x)$ をカーネルという
- カーネルの種類によって分布の形が異なる
 - ・・・本書では3つのカーネルが紹介されている

43

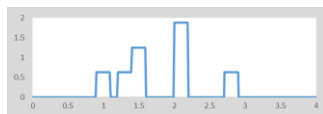
カーネル密度関数とは、経験分布は各データ点に確率を与えるものなので離散型になりますが、元々のロスの分布が連続型であると考えられる場合に、離散型の分布を連続分布で近似するための関数です。式にすると、ここにあるとおりになります。各データ点に与えられた確率 p に対して、カーネルと呼ばれる関数を掛けることで密度関数を作ります。

カーネルの種類によって分布の形が異なり、「Loss Models」では3つのカーネルが紹介されています。

7. カーネル密度関数による経験分布の平滑化

- 3つのカーネル（バンド幅により形が異なる）

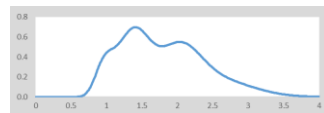
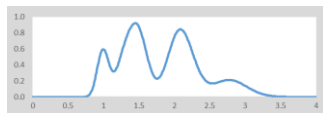
- 一様カーネル



- 三角カーネル



- ガンマカーネル



44

1つ目が、データ点の周辺に一様分布を当てる一様カーネル。ヒストグラムは、まさに一様カーネルに当たります。一様分布の幅を「バンド幅」と言いますが、そのバンド幅を調整することで、左のプロットから右のプロットのように調整することができます。2つ目は、三角カ

カーネル。これはデータ点を中心とした三角形を作るものです。これも同じように幅を変えることで、右のように調整できます。3つ目が、ガンマカーネル。これはデータ点に対してガンマ分布の密度関数を当ててるものです。これは、上の2つよりもより分布のような形になって、同じように幅を調整することで右のようになります。

Loss Modelsによる統計手法の学習

Part	Chapter
Part I 序論	1. モデルシナリオ
	3. 基本統計量
	5. 連続モデル
Part II 基礎的な経験分布	7. 基礎的な経験分布
	9. ロロスモデル
	11. 最尤推定
	13. ベイズ推定
Part III 数理統計学	
Part IV モデル構築	15. モデル選択
Part V 信頼性理論	17. 最大積密信頼性理論
	19. シミュレーション

1. 保険実務の例による統計手法の学習
2. 分布の解釈
3. 免責金額の変更による頻度の影響
4. グループ化されたデータの最尤推定
5. 最尤推定量の（尤度比による）信頼区間
6. 免責・限度額が混在するデータのロス分析
7. カーネル密度関数による経験分布の平滑化
8. モデル選択

45

最後が「モデル選択」。ここでは章の全体的な話をします。

8. モデル選択

- 目的に応じた手法を体系的に網羅している

項目	手法
データの可視化 プロットにより経験分布と比較して当てはまりを確認する	<ul style="list-style-type: none"> ・ 経験分布とモデルの比較 (cdf, pdf) ・ p-pプロット
仮説検定 主張するモデルが受け入れられるものであることを確認する	<ul style="list-style-type: none"> ・ K-S検定 ・ A-D検定 (K-S検定よりも裾を重視する) ・ カイ2乗検定 (グループ化したデータにも対応) ・ 尤度比検定
スコアに基づく選択 複数のモデルを統計量により比較する	<ul style="list-style-type: none"> ・ 仮説検定による統計量 ・ SBC (またはBIC) ・ AIC

正しい問いに対する近似的な解答は、
 誤った問いに対する正確な解答よりはるかに良い
 —John Turkey

46

「モデル選択」ではいわゆる手法について、目的に応じて体系的に、大まかなものから細かいものまで網羅していますので、その流れに沿って読むことも勉強になると思いますが、手法そのものよりも、モデルというものが絶対的ではないということを踏まえて、課題を理解し整理することが、モデル選択の手法を使うより前に重要であると書いてあります。それを踏まえ

て、曖昧な課題を、いかにモデル選択の手法に落とし込んで処理して判断するかということについての例が幾つかありますので、手法そのものは知っているという人であっても、いかに曖昧な課題を具体的に落とし込むかという部分について読むことが勉強になるのではないかと思います。

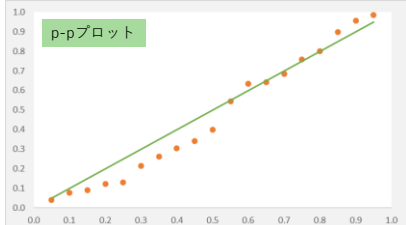
例を2つ紹介します。

8. モデル選択

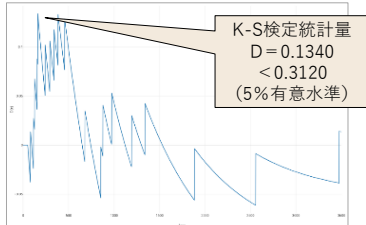
- (例) ロスデータに指数分布を当てはめる
 - 下記ロスデータから推定
(フランチャイズ50が適用されているため、50で切断した分布を使う)

	82	115	126	155	161	243	294	340	384	
	457	680	855	877	974	1,193	1,340	1,884	2,558	3,476

- データの可視化と仮説検定



p-pプロット



K-S検定統計量
D=0.1340
<0.3120
(5%有意水準)

47

1つ目が「ロスデータに指数分布を当てはめる」ということで、このデータに指数分布を当てはめた例です。下にある p-p プロットは、モデルと経験分布のデータ点ごとの分布関数のずれを表したもので、右にあるコルモゴロフ-スミルノフ検定は、そのずれの最大値をもってモデルの当てはまりを判定する手法です。

この例自体は曖昧な課題に対するというものではありませんが、ここではロスデータにフランチャイズが適用されているため切断分布を使う必要があって、これを再現しようとする、すんなり R などに落とし込めるような作りにはなっていません。「Loss Models」では、時々すんなりと再現できないようなデータを扱った例があって、再現しようとする非常に手間がかかるのですが、別の見方をすれば、もし実務できれいではないデータに出くわしたとしても、最悪でも手作業で対応できるような手法が学べるという意味で、ここで紹介しています。

8. モデル選択

■ (例) アクチュアリーの数に変化したのか

会員数	会社の数 (1949年)	会社の数 (1957年)
1	17	23
2	7	7
3-4	3	3
5-9	2	3
10+	0	1

- 各社最低一人はアクチュアリーが在籍
・・・ゼロ切断頻度分布を使う
- 2データ (1949年と1957年を区分したもの・合算したもの) をモデル化して、スコアを比較する
・・・NLL (負の対数尤度) は合算モデルの74.35に対して区分モデルの合計が74.15であり、0.20しか改善しない

⇒平均は1.83人(1949年)から2.10人(1957年)に増加しているが、データが十分ではなく変化したとはいえない

48

次が「アクチュアリーの数に変化したのか」という例です。このような見出しで書くと興味を持つ人がいるのではないかと思って、ここでピックアップしています。アクチュアリーの数別の会社の数が2年分集計されています。1949年から1957年にわたって、アクチュアリーの数に変化したと言えるのかという問いです。このような曖昧な問いや、変則的な、最低1人はアクチュアリーが在籍するというモデルの条件、また、興味を引いてしまうデータのようなものを全て合わせて、「Loss Models」における総合テストのような例になっています。

手順としては、2つモデルを計算します。1つはこの2年分のデータを区分したもの、もう1つはこの2年分のデータを合算したものです。それから、その2つのモデルの対数尤度の改善やAICを見て、どちらのモデルを採用するか考えます。そのときに区分したモデルの方が採用されるのであれば、この2つのデータは異なると言えるので、アクチュアリーの数に変化したと言える、そうでない場合は、アクチュアリーの数に変化したとは言えないと結論づけます。この例の場合ですと、区分したモデルは合算モデルに対して対数尤度やAICの改善が十分ではないため、平均そのものを計算すると増加はしているものの、数に変化したとは言えないと結論づけています。

この例だけをぱっと見ても「まあそうだろうな」と思う人が多いとは思いますが、恐らくそれは、モデル選択の章で説明するというバイアスがかかっているからだと思うので、そのような場合は一度ここから離れて、この例だけがぼつんと独立して落ちてきたとしたら、そのときにどのように対応するかという観点で見ただけだと、勉強になるのではないかと思います。

Loss Modelsによる統計手法の学習

Part	Chapter
Part I 序論	1. モデリング
	2. 基本統計量
	3. 連続モデル
	4. 発展的な連続分布
Part II 数理統計学	7. 発展的な連続分布
	8. ロボモデル
	9. 最尤推定
Part III 数理統計学	11. 最尤推定
	13. ベイズ推定
Part IV 信頼性理論	15. モデル選択
	17. 最大積密信頼性理論
Part V 信頼性理論	18. シミュレーション
	付録

1. 保険実務の例による統計手法の学習
2. 分布の解釈
3. 免責金額の変更による頻度の影響
4. グループ化されたデータの最尤推定
5. 最尤推定量の（尤度比による）信頼区間
6. 免責・限度額が混在するデータのロス分析
7. カーネル密度関数による経験分布の平滑化
8. モデル選択

49

最後は「付録」です。

9. 付録

- 付録には頻度・規模分布の一覧が載っている

- 限度付モーメント

$$\dots E[(X \wedge u)^k] = \int_{-\infty}^u x^k f(x) dx + u^k [1 - F(u)]$$

- (例) 対数正規分布

$$\exp\left(k\mu + \frac{1}{2}k^2\sigma^2\right) \Phi\left(\frac{\ln x - \mu - k\sigma^2}{\sigma}\right) + x^k \left[1 - \Phi\left(\frac{\ln x - \mu}{\sigma}\right)\right]$$

- 対数t分布（自由度 r のt分布 Y に対する $\exp(\sigma Y + \mu)$ ）

- ・ ・ ・ t分布が正規分布より裾が重いと同様に、
対数正規分布より裾が重い分布

$$f(x) = \frac{\Gamma\left(\frac{r+1}{2}\right)}{x\sigma\sqrt{\pi}\Gamma\left(\frac{r}{2}\right) \left[1 + \frac{1}{r}\left(\frac{\ln x - \mu}{\sigma}\right)^2\right]^{(r+1)/2}}$$

50

「Loss Models」の最後には「付録」があって、そこにはたくさんの頻度や規模の分布、パラメーターなどが載っていて、そのうち例えばここで言うと、限度額を適用したときの平均の計算ができるようになっているとか、他にも下にある対数t分布のような、いつ使うのかよく分からないような分布なども紹介されているので、それらを見ることも一つ興味深いのではないかと思います。

Loss modelsによる統計手法の学習（まとめ）

- 実務上の観点に基づく統計手法の解説
 - 保険実務の例による統計手法の学習
 - 分布の解釈
 - 最尤推定量の（尤度比による）信頼区間
 - モデル選択

- 保険データ特有の課題に対する取扱い
 - 免責金額の変更による頻度の影響
 - グループ化されたデータの最尤推定
 - 免責・限度額が混在したデータのロス分析
 - カーネル密度関数による経験分布の平滑化

51

第2部では、2つの観点に基づいて、本の中から重要そうなものをピックアップして説明しました。それらを大まかに分類すると、このような感じになります。全てに興味を持つことはあまりないかもしれませんが、このうち1つでも興味を持ってみようと思うものがあれば、「Loss Models」を読んでみる価値があるのではないかと思います。

以上で、第2部の説明を終わります。

少し時間があるので、質問に答えます。

【質問】 アメリカの実務例がありましたら、ご紹介いただきたいです。

【回答】 「Loss Models」にあるものはまさにその例だと思いますので、先ほど紹介したような実務に関する例そのものが、この質問に対する回答になるかと思います。

【質問】 生保での活用例

【回答】 例えば先ほどのKaplan-Meier法などは、どちらかという生保で使っているようなものだと思います。それを損保のロス分析に使ったらどうなるかという例ですから、それをもう一度逆輸入して、入院期間などと同じように戻してみるということも、いいかなと思います。

【司会】 佐野さん、ありがとうございます。

引き続き沖縄から、川上さんに発表していただきます。川上さん、よろしく申し上げます。

第3部 信頼性理論(CREDIBILITY)

52

【川上】 こんにちは。第3部を担当いたします、大同火災の川上と申します。ASTIN 関連研究会に所属しています。どうぞよろしくお願いいたします。

第3部の信頼性理論につきましては、「Loss Models」のPart Vとなりまして、ページ数でいきますと全体で77ページとなっております。また、「Loss Models」のテキスト自体がA4サイズとなります。日本の損保数理のテキストについてはB5サイズで31ページということで、単純比較はできませんがかなりの量を割いているのかなと思っています。また、先ほど説明がありましたとおり、ベイズ推定については別途CHAPTER13で説明していますので、かなり詳しく信頼性理論について触れられています。初学者の方ですと、なかなか条件付き確率等についてはとっつきにくいかと思われるのですが、テキスト自体は、非常にシンプルな離散分布を用いて、分かりやすい説明等がされています。本日は、信頼性理論の全般について説明させていただきます。

はじめに

- 信頼性理論は、基礎科目「損保数理」の出題範囲に含まれており、アクチュアリーにとってはなじみの深いものであるはず
- にもかかわらず実務での活用は、特定の用途（損害保険の料率算定）における特定の手法（有限変動信頼性理論）を除き、それほど広がっていないように見受けられる
- Loss Models（第5版）に記載されている内容は、初版（2004年翻訳発行の「統計データの数理モデルへの適用」）からそれほど変わっておらず、日本のアクチュアリーにとっても目新しいものではないかもしれない
- それでも、ここで改めてLoss Modelsの内容に沿って信頼性理論の概要を振り返り、実務での活用可能性を考えるきっかけとしたい

53

まずはじめに、信頼性理論は、基礎科目「損保数理」の出題範囲に含まれており、アクチュアリーにとってはなじみ深いものだと思います。事前に ASTIN 関連研究会のメンバーを経由して、信頼性理論の活用状況について感触を伺ってみたところ、損保での実務の活用については、特定の用途として損害保険の料率算定や、特定の手法として有限変動信頼性理論を除き、それほど使われていないのではないかと考えていました。また、日本の生保分野でも、ほとんど使われていないのかなという感触を得ています。先ほどのアンケートの結果ですけれども、最終的な結果を見ますと、使われていない方が 48%、有限変動信頼性理論を使われている方が 46%、ベイズ的手法あるいはビュールマンモデル等まで使われている方が 4%、使い分けている方が 1%でしたので、使われていない方を除きますと、有限変動信頼性理論は、ベイズ的手法等の 10 倍程度使われているということになるかと思えます。

「Loss Models」第 5 版に記載されている内容については、先ほどありました初版の「統計データの数理モデルへの適用」からそれほど変わっていないもので、日本のアクチュアリーにとっても、もしかすると目新しいものではないかもしれません。それでも、ここで改めて「Loss Models」の内容に沿って信頼性理論の概要を振り返りまして、実務における活用の可能性を考えるきっかけとしたいと考えています。

タリフ料率における問題意識と信頼性理論

タリフ料率では、料率は料率区分ごとに各料率区分の支払見込みを反映するように設定されており、料率区分内ではリスクが同質であることを仮定している

しかし、いかなる料率体系も完璧ではなく、常に不均一性が残る

信頼性理論を使うと

経験料率

同一料率区分内のリスク(またはリスクのグループ)に対し、支払実績に基づき将来の保険料を調整することができる

料率区分の細分化

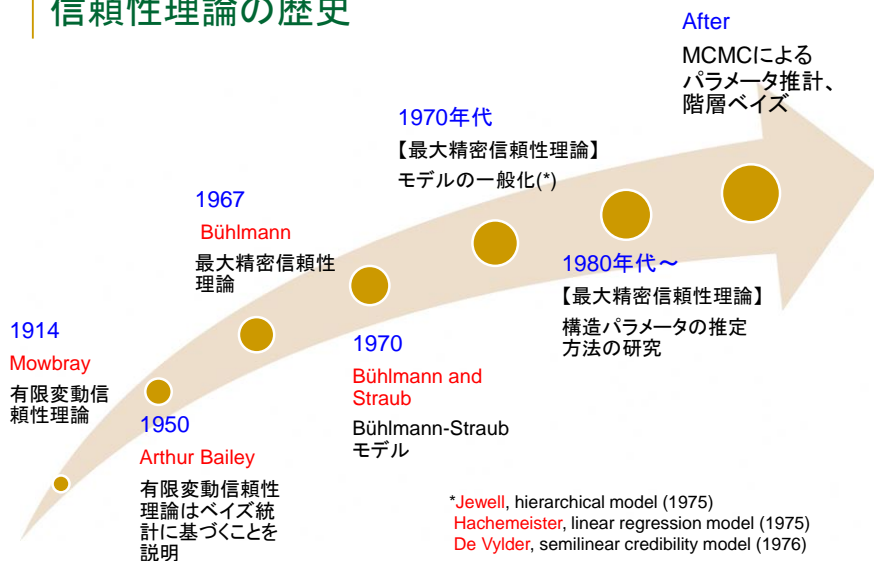
料率区分を細分化し、データがほとんどないような区分に対しても支払実績を加味した料率を設定することができる

54

続きまして、「タリフ料率における問題意識と信頼性理論」というトピックです。まず、信頼性理論の問題意識ですが、タリフ料率では、料率区分内では料率が同質であるということを仮定しています。しかし、いかなる料率体系も完璧ではなく、常に不均一性が残る形となっています。信頼性理論につきましては、次の2点の観点で主に使われています。まず1点目が、「経験料率」ですが、これは、将来の料率調整で用いており、料率リスク区分に対して、支払実績に基づいて将来の料率を調整しています。

2点目が、料率細分ですね。こちらは、料率の区分を細分化し、データがほとんどないような区分に対しても、支払実績を加味した料率設定という使い方をしてしています。

信頼性理論の歴史



55

ここで、簡単に信頼性理論の歴史について触れさせていただきますと、1914年に有限変動信

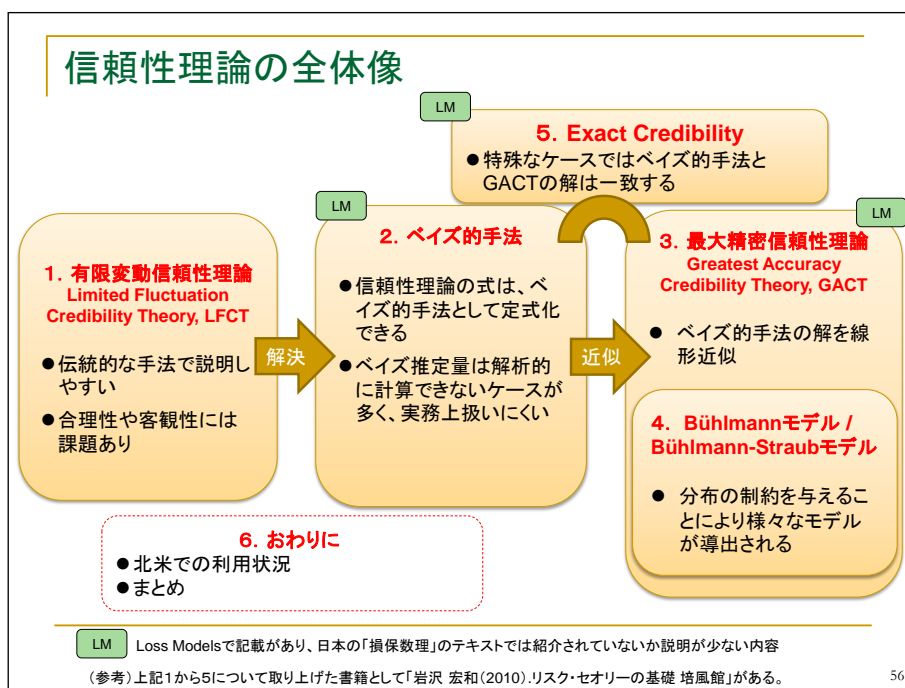
信頼性理論が開発されています。その後、1950年に、有限変動信頼性理論は実はベイズ統計に基づくことであることが、Arthur Baileyによって説明されています。1967年に、日本でも有名な Bühlmann によって、最大精密信頼性理論が発表されています。こちらは少しややこしいのですが、Bühlmann は、最大精密信頼性理論という理論を提唱されていて、その中の一つが、Bühlmann モデルという形になっています。ですので、最大精密信頼性理論については、もう少し広い概念になります。

続きまして、1970年に、Bühlmann-Straub モデルが発表されています。

その後も70年代には、最大精密信頼性理論のモデルの一般化ということで、下の方に記載していますが、Jewell や Hachemeister という方々によってモデル化されています。

80年代以降は、後ほど説明する構造パラメーターというものがあるのですが、パラメーター推定の方法について研究がなされています。その後、MCMC (マルコフチェーンモンテカルロ) によるパラメーター推定や、階層ベイズなどが考案されています。

信頼性理論について、保険料の話ですので期待値を求める話と、信頼度 Z 、すなわち 0 から 1 の値を推定する話なのですが、100年以上にわたってこのような理論が発展しているという経緯が、非常に興味深く感じています。



本発表における信頼性理論の全体像について、ここで紹介させていただきます。一部追加もございますが、「Loss Models」のテキストの分類に従ったものになります。まず1番目に有限変動信頼性理論を取り上げます。先ほどのアンケートからも分かる通り、日本でもなじみやすいものなのかなと考えています。こちらは伝統的手法で、説明しやすいという特徴がありますが、合理性や客観性には課題があることが知られています。

続きまして、その課題を解決するためのものとして、ベイズ的手法が取り上げられています。信頼性理論はベイズ的手法として定式化できることが知られていますが、ベイズ推定量は解析的に計算できないケースが多く、実務上は取り扱いにくいということも指摘されています。なお、ベイズ的手法を線形近似したものが最大精密信頼性理論になっており、GACTと略されるこ

ともあります。先ほどの説明のとおり、最大精密信頼性理論の1つのモデルとして、Bühlmannモデル、あるいは、Bühlmann-Straubモデルがあります。例えばBühlmannモデルですと、分布に制約を与えることによってモデル化しているようなこととなります。線形近似と書いてありますが、実は「Exact Credibility」と呼ばれる、特殊なケースにおいては、ベイズ的手法と最大精密信頼性理論が一致するということが知られています。

終わりに、北米での利用状況などについて紹介させていただきまして、まとめとしています。信頼性理論全般なのですが、地域等によって違う用語や枠組みで語られることもありますので、ここでは、あくまで「Loss Models」のテキストに沿った分類となっています。また、先ほどのアンケートのとおり、ベイズなどについては、日本では活用があまりされていないという状況が伺えるのですが、1から5について取り上げられた書籍として、アクチュアリー会の各種講師等でもご活躍されている岩沢先生の「リスク・セオリーの基礎」という書籍を、参考として記載しています。このテキストについては、なかなか入手困難という状況が続いているようなのですが、最近重版され、実際の書店等に行けば入手できる可能性があるかと伺っています。



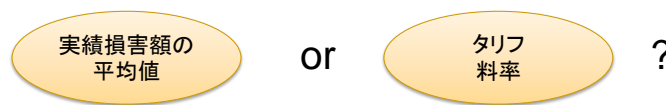
有限変動信頼性理論 (LIMITED FLUCTUATION CREDIBILITY)

57

続いて、「有限変動信頼性理論」に移らせていただきます。

基本的な問題意識

- ある契約に、タリフ料率 M を適用してこれを引き受け、 n 年間の（または、 n 件の契約における1年間の）実績損害額として、 X_1, \dots, X_n が得られた
- 翌年保険料の算定において、過去の実績損害額の平均値 \bar{X} を採用すべきか、タリフ料率 M を採用すべきか
- いくつかの実績データがあれば、 \bar{X} は料率算出の基礎データとして信頼に足るといえるのか
- 十分なデータがない場合にはどうすればよいのか



58

基本的な問題意識としましては、タリフ料率を適用して、 n 年間の実績があつて、実績の損害額として X_1 から X_n までが得られるような状況です。翌年度の保険料について、平均値の \bar{X} を用いるべきか、あるいはタリフ料率 M を用いるべきかということが、まず問題意識となります。また、幾つのデータがあれば、 \bar{X} は基礎データとして信頼に足るのか。あるいは、十分なデータがない場合にはどうすればいいのかということが、基本的な信頼性理論の問題意識となっています。

有限変動信頼性理論(全信頼)

- X (期待値 ξ , 標準偏差 σ) の標本平均 \bar{X} が $100p\%$ の確率で真のクレームコスト ξ の $\pm 100r\%$ の範囲に収まるときに「実績損害額の平均」が使えると考える (全信頼)

$$\Pr(|\bar{X} - \xi| \leq r\xi) \geq p$$

- $\frac{\bar{X} - \xi}{\sigma/\sqrt{n}}$ が $N(0,1)$ で近似できれば、 $y_p = \Phi^{-1}((1+p)/2) \leq \frac{r\xi}{\sigma/\sqrt{n}}$ であればよい

$$\Rightarrow n \geq \lambda_0 (\sigma/\xi)^2, \text{ ただし } \lambda_0 = (y_p/r)^2$$

λ_0 の具体的な値は右表のとおり

- ξ, σ は通常未知なのでデータから推定

$p \setminus r$	0.30	0.20	0.10	0.05	0.01
0.900	30	68	271	1,082	27,060
0.950	43	96	384	1,537	38,416
0.990	74	166	663	2,654	66,358
0.999	120	271	1,083	4,331	108,274

- X の分布の仮定により様々な式が得られる
 - 例1: X がポアソン分布 (頻度 λ) のとき $n\lambda \geq \lambda_0$ (λ は推定値を用いる)
 - 例2: X が複合ポアソン分布 (頻度 λ , 損害規模の平均 θ_Y ・分散 σ_Y^2) のとき

$$n\lambda \geq \lambda_0 \left[1 + \left(\frac{\sigma_Y}{\theta_Y} \right)^2 \right]$$

59

ここで、まずはじめに、有限変動信頼性理論における全信頼に必要な件数について、考えていきます。 X の標本平均 \bar{X} について、 $100p\%$ の確率で、真のクレームコスト、 ξ の $\pm 100r\%$ の範囲に収まるときに、全信頼を与えることを考えます。式で表すと、スライド式のとおりになりま

す。細かい説明は省きますけれども、この λ_0 について r と p を与えることによって、件数が与えられます。これは損保数理のテキストでも載っているようなものなので、例えば r が 0.05 で、 p が 0.9 の場合は 1,082 件ということが分かり、そこから更に、 X の分布によってパラメータが変わり、実際に全信頼に必要な件数が分かります。ここでは例 1、2 として挙げていますが、 X の分布の仮定により、さまざまな式が得られます。

有限変動信頼性理論(部分信頼)

- データが足りない場合(部分信頼)、両者の加重平均を用いる(信頼性保険料)

$$Z \times \text{実績損害額の平均} + (1 - Z) \times \text{タリフ料率}$$
- Z を信頼度といい、さまざまな手法が提案されているが理論的な正解はない
 - 例1

$$Z = \frac{n}{n+k}$$

k は何らかの方法で決める必要がある(理論的には後述の最大精密信頼性理論で正当化される)
 - 例2

信頼性保険料の分散 = データ数が全信頼の下限の場合の「実績損害額の平均」の分散となるように定める

$$Z = \sqrt{\frac{\text{実績クレーム件数}}{\text{全信頼に必要なクレーム件数}}}$$

続きまして、「有限変動信頼性理論(部分信頼)」ですが、データが足りない場合については部分信頼と呼ばれ、両者の加重平均を用いて信頼性保険料を計算します。ご承知のとおり、 Z は「信頼度」といいまして、さまざまな手法が提案されていますが、理論的な正解はないものと考えられます。例 1 として、 $Z = \frac{n}{n+k}$ となっていますが、 k については何らかの方法で求める必要がありますが、理論的には最大精密信頼性理論で正当化される内容になっています。また、例 2 では、損保数理の有限変動信頼性理論の信頼度のような形で表われるものになっています。さまざまなパターンが信頼度にもあるということについて、紹介させていただきます。

例題1 (有限変動信頼性理論①)

- ある保険契約について、契約者(id)ごとの1年目のクレーム件数 N_{id} が下表のように与えられている
- 一方、類似商品の統計によれば合計年間クレーム件数の期待値 M は200である
- 保険金はクレーム1件につき定額1とする
- N_{id} がポアソン分布に従うとき、有限変動信頼性理論により全体の信頼性保険料を求め、これを用いて(id=1)の個別保険料を求めよ($p = 0.9, r = 0.05$ とする)

id	N_{id}
1	19
2	23
3	26
4	20
5	34
6	4
7	32
8	11
合計	169

年間クレーム額はポアソン分布に従うので、
全信頼に必要なクレーム件数は $n\lambda \geq \lambda_0 = 1,082$ 件

$$\text{信頼度 } Z = \sqrt{\frac{169}{1082}} = 0.3952$$

$$\begin{aligned} \text{全体保険料} &= Z \times 169 + (1 - Z) \times 200 \\ &= 0.3952 \times 169 + 0.6048 \times 200 \\ &= 187.75 \end{aligned}$$

$$\text{個別保険料}_{id=1} = \frac{187.75}{8} = 23.47$$

61

続いて、例題1として、有限変動信頼性理論の簡単な例題を考えてみたいと思います。ある契約者について、契約者(id)ごとの1年目のクレーム件数、 N_{id} が、スライドの表のとおり与えられています。idとして、1から8までの契約者がいるという状況です。それに対して、1年目のクレーム件数の実績値が与えられています。一方、類似の商品統計によれば、合計の年間クレーム件数の期待値 M は、200であるとされています。また、保険金は、クレーム1件につき定額1とします。 N_{id} がポアソン分布に従うとき、有限変動信頼性理論により全体の信頼性保険料を求め、これを用いてid=1の個別保険料を求めます。ここで、 p は0.9、 r は0.05として、この例では、 p と r によってのみ全信頼に必要な件数が分かるという形になっています。先ほどのとおり、全信頼に必要な件数としては、1,082件となります。

これに対して信頼度の計算なのですが、先ほどの合計の169件と、この1,082件により、信頼度については0.3952となります。全体保険料としては、簡単ではありますが、169と $M(200)$ の加重平均値となり、最終的には187.75となります。個別保険料は、それを契約者数8で割って、23.74となります。

例題2(有限変動信頼性理論②)

- 例題1の集団で10年分の実績が蓄積され、集団全体では全信頼(p=0.9,r=0.05)
- 契約者ごとに事故件数に差異が認められるため契約者ごとに保険料を細分する
- 有限変動信頼性理論によりid=1に対する保険料を計算せよ

id	$N_{id,1}$	$N_{id,2}$	$N_{id,3}$	$N_{id,4}$	$N_{id,5}$	$N_{id,6}$	$N_{id,7}$	$N_{id,8}$	$N_{id,9}$	$N_{id,10}$	計	平均
1	19	12	25	19	23	15	15	20	8	11	167	16.70
2	23	15	21	22	17	23	15	17	26	21	200	20.00
3	26	35	34	36	25	22	23	27	18	22	268	26.80
4	20	27	17	19	14	22	21	16	17	24	197	19.70
5	34	26	31	29	28	28	33	32	22	29	292	29.20
6	4	5	7	5	10	11	13	9	10	14	88	8.80
7	32	34	29	37	36	31	35	35	32	28	329	32.90
8	11	13	17	12	14	19	12	13	10	22	143	14.30
合計	169	167	181	179	167	171	167	169	143	171	1,684	168.40
平均	21.13	20.88	22.63	22.38	20.88	21.38	20.88	21.13	17.88	21.38	210.50	21.05

$$\text{全体保険料} = \frac{1684}{80} = 21.05 \text{ (これを } M \text{ として、id=1の個別保険料を求める)}$$

$$Z_{id=1} = \sqrt{\frac{167}{1082}} = 0.3929$$

$$\begin{aligned} \text{個別保険料}_{id=1} &= Z_{id=1} \times 16.70 + (1 - Z_{id=1}) \times 21.05 \\ &= 0.3929 \times 16.70 + 0.6071 \times 21.05 \\ &= 19.34 \end{aligned}$$

62

続きまして例題2です。先ほどの例題1の集団の10年分の実績が蓄積されたような状況を考えます。全体の事故件数としては、10年間で1,684件となっていますので、集団全体では、全信頼の1,082件を超えていることとなります。契約者ごとに事故件数に差異が認められそうだとということが分かり、例えばid=6ですと事故件数が少なそうだとことが分かりますので、料率を細分することを考えます。

ここで、有限変動信頼性理論を用いて、契約者id=1の保険料を計算したいと思います。

まず、全体保険料としては21.05となり、これをMとして、id=1の個別保険料を算出することとなります。id=1の事故件数は167件になっていますので、全信頼に必要な件数で割ってあげて、0.3929となります。この16.70と、全体の21.05を加重平均して、19.34がid=1の個別保険料となります。この例題としては、Mとして、外部データではなくて、全体のクレーム平均を用いています。有限変動信頼性理論においては、外部データを用いた料率調整、あるいは、支払実績データを用いた料率細分で使用することができます。

pとrが保険料に与える影響

- 有限変動信頼性理論ではpとrの選択次第で信頼度が変わるため、恣意性が働く可能性がある
- 例題2についてpとrを変えた場合の影響は下表のとおり

全信頼に必要なクレーム件数

p \ r	0.30	0.20	0.10	0.05	0.01
0.900	30	68	271	1,082	27,060
0.950	43	96	384	1,537	38,416
0.990	74	166	663	2,654	66,358
0.999	120	271	1,083	4,331	108,274

例題2 id=1 の信頼度

p \ r	0.30	0.20	0.10	0.05	0.01
0.900	1.0000	1.0000	0.7850	0.3929	0.0786
0.950	1.0000	1.0000	0.6595	0.3296	0.0659
0.990	1.0000	1.0000	0.5019	0.2508	0.0502
0.999	1.0000	0.7850	0.3927	0.1964	0.0393

例題2 id=1 の個別保険料

p \ r	0.30	0.20	0.10	0.05	0.01
0.900	16.70	16.70	17.64	19.34	20.71
0.950	16.70	16.70	18.18	19.62	20.76
0.990	16.70	16.70	18.87	19.96	20.83
0.999	16.70	17.64	19.34	20.20	20.88

63

ここで簡単に、pとrが保険料に与える影響を見てみたいと思いますが、有限変動信頼性理論では、pとrの選択次第で全信頼に必要な件数が変わってきてまして、最終的には信頼度にも、個別保険料にも影響が出てくるという形になります。こちらの例ですと、最小値では16.70、最大値では20.88となり、保険料に差が出てくることが分かります。

有限変動信頼性理論の問題点

有限変動信頼性理論はシンプルな解を与えるが、理論的根拠に欠ける部分がある

- 関心のある確率変数 X の分布に理論的なモデル(M との違い等)を仮定しておらず、 $Z\bar{X} + (1-Z)M$ の形の保険料が適切で M より望ましい理由がない
- p, r の選択に関する指針がない(恣意的である)
- M は推定値であり、それ自身が信頼できるものであるとは限らない

Loss modelsより抜粋

このアプローチ(引用注:有限変動信頼理論のこと)の難しさは、これらの方法の使用を正当化するための健全な基礎となる数学的理論が欠如していることである。とはいえ、このアプローチはこの問題を独自に扱ったものであり、今日でも使われている(p.389)

私たちがまた、初期のアクチュアリーたちに脱帽しなくてはならない。彼らは、単純な数学の道具を自由に使って、機能するだけでなく、この部分で私たちが注意深く開発したものと非常によく似た計算式を考え出すことができたのである(p.390)

使い方を工夫すれば、有限変動信頼性理論でも十分に機能する!

64

ここで有限変動信頼性理論の問題点なのですけれども、ただいま見ていただいたことも含めて、理論的根拠に欠ける部分があるということが、「Loss Models」では指摘されています。1点目ですが、「関心ある確率変数 X の分布に理論的モデルを仮定しておらず、 $Z\bar{X} + (1-Z)M$ の形の保険料が適切で、 M より望ましい理由がない」となっています。2点目として、pとrに選択に関する指針がなく、恣意性が働く可能性があるということが指摘されています。また、3点目

としまして、外部データ M については、M 自体が推定値であり、また、それ自体が信頼できるものとは限らないということが指摘されています。

しかしながら、ここはテキストからの抜粋なのですが、歴史のところでも触れましたが、「これらは 100 年前に作られた理論となり、1950 年には、数学的道具立てが洗練されていない時代に、後に説明するベイズ的手法に似た、そして、何より実際に機能する数式を考えた初期のアクチュアリーたちには、脱帽すべきであろうと思われる」と記載されています。また、次のスライドでも説明しますが、使い方を工夫すれば、有限変動信頼性理論についても、十分に使えるものと考えられています。

(参考)ASOP No. 25 Credibility Procedures

- 米国の ASB (Actuarial Standards Board) は、ASOP (Actuarial Standard of Practice) により、アクチュアリーが業務遂行にあたり遵守すべき事項を定めている
- ASOP no.25 Credibility Procedures は、信頼性理論の手法の選択・開発やデータへの適用に関する指針の提供を目的とし、具体的には以下を求めている
 - 対象の経験 (\bar{X}) が全信頼を有するかどうかを判断する際、また対象の経験を関連する経験 (M) とブレンドする際に、適切な信頼性理論の手法を用いるべきである
 - 信頼性理論の手法を選択・開発する際には、以下の基準を考慮すべきである
 - ・ 合理的な結果をもたらすことが期待できるか
 - ・ 意図された用途および目的に適しているか
 - ・ コストと便益を考慮した場合に、その手法を採用することは現実的か
 - ✓ 関連する経験の選択・利用にあたり、専門的判断力を発揮し注意を払うべきである
関連する経験は、対象の経験に類似した特徴を持つべきである
 - ・ 考慮すべき特徴... 性・年齢・補償範囲・頻度・規模、その他特定可能なリスク特性等
 - 手法の選択・開発・使用にあたっては、専門的な判断を用いるべきである
信頼性理論の手法の利用は、必ずしも厳密な数学のプロセスではない
 - 対象の経験と関連する経験の両方の同質性を考慮すべきである
また、データの同質性とデータセットのサイズとのバランスを考慮すべきである

<http://www.actuarialstandardsboard.org/standards-of-practice/>

65

ここでは、米国のアクチュアリーが実務遂行によって遵守すべき基準である ASOP について、簡単に説明させていただきます。ASOP には幾つかの基準がありまして、現時点で確認したところ 56 の基準があり、この中で、ASOP No.25 の「Credibility Procedures」というものがあります。ここでは、信頼性理論の手法の選択・開発や、データの適用に関する指針の提供を目的として、具体的に考慮すべき事項を定めています。

まず、対象の経験 (\bar{X}) なのですが、これが全信頼に有するかどうかの判断をする際、また、対象の経験と、関連する経験 (先ほどの外部データ M) を連動する際に、適切な信頼性理論の手法を用いるべきとしてしています。手法選択の場合には、考慮点として、合理的な結果をもたらすことが期待されるか。意図された用途および目的に適合しているか。コストと便益を考慮したうえで、その手法を採用することが現実的かということが掲げられています。

また、関連する経験 (M) の選択・利用に当たり、専門的判断力を発揮し注意を払うべきであるとされています。関連する経験は、対象の経験 (\bar{X}) に類似した特徴を持つべきであり、性、年齢、補償範囲、頻度等によって、考慮すべき特徴が挙げられています。また、手法の選択・開発に当たっては、「専門的な判断を用いるべきである」としまして、信頼性保険料の手法の利用は、必ずしも厳密的な数学というプロセスではないということも記載されています。

最後ですが、対象の経験と関連する経験の両方の同質性を考慮すべきであるとされています。

て、データの同質性と、データセットの最適バランスを考慮すべきであるとされています。

このように、有限変動信頼性理論の枠組みの中だけで考えると妥当性が保証されていないような部分について、実務基準を定めることによって、一定の妥当性を確保できるようになっています。

ベイズ的手法 (THE BAYESIAN METHODOLOGY)

66

続いて「ベイズ的手法」に移ります。ベイズ的手法については、先ほどのスライドで説明させていただいたとおり、有限変動信頼性理論の弱点を解決するようなものとなっています。

ベイズ的手法の導入

条件なしの平均 $E[E(X_{n+1}|\Theta)] = E(X_{n+1})$ を μ で表し、**集合的保険料** という

構造の仮定

母集団のリスク特性は θ (事前分布 $\pi(\theta)$) で表され、 $\theta = \theta$ を持つ保険契約者のクレーム (件数・金額) は、 θ が与えられた場合の条件付分布 $f_{X|\theta}(x|\theta)$ に従う

$\pi(\theta)$

理想

θ が既知であれば、特定の契約者の保険料は $f_{X_{n+1}|\theta}(x_{n+1}|\theta)$ から $E(X_{n+1}|\Theta = \theta)$ を計算することにより求まる

代替案

しかしながら θ は未知であることが多いため、 $X = x$ が観測されたときの**事後分布** $\pi_{\theta|X}(\theta|x)$ を用い**予測分布の期待値** $E(X_{n+1}|X = x)$ を計算

$E(X_{n+1}|X = x) = \int E(X_{n+1}|\Theta = \theta) \pi_{\theta|X}(\theta|x) d\theta$ を**ベイズ保険料** という

67

まず、一般的なベイズ的手法の導入部分ですが、概念としましては、「構造の仮定」を行います。

クレームの従う確率変数 X なのですが、それらも何らかのパラメーターに従うというものに

なっており、母集団のリスク特性 Θ は事前分布 $\pi(\theta)$ で表し、 $\Theta = \theta$ を持つ契約者のクレームは、 θ が与えられた場合の条件付き分布に従うこととなります。ちなみに、条件なしの期待値のことを、「集合的保険料」と言います。

続いて「理想」と書いてあるところですが、個々の契約者ごとのリスク特性が分かっている状態、すなわち θ が既知の場合については、特定の契約者の保険料については、こちらのとおり、 $\Theta = \theta$ の下での X の期待値、ここでは翌年度保険料という意味で添え字が $n+1$ としていますが、このような条件付き期待値から計算することができます。これを $\mu(\theta)$ で表して、「仮説平均」と呼びます。これは、個々の契約者のパラメーターが分かったときの期待値となっています。

理想に対して「代替案」と書いてありますが、実際は θ は未知であることが多いため、 $X=x$ については、標本値ベクトルが観測されたときの事後分布を用いて、予測分布の期待値を計算することとなります。先ほどの式と比べてみると、実績データが与えられたもとの条件付き期待値となっており、これを計算することになっています。この算式については、スライド下の式で表すことができ、先ほどの仮説平均 $\mu(\theta)$ を事後分布の上に期待値を取ったものとして表わされておりまして、これをベイズ保険料と言います。

例題3(ポアソンガンマモデル- θ が既知の場合)①

- 例題2と同じ状況において、事故件数はポアソン分布に従い、契約者ごとのパラメータ θ はガンマ分布に従うとする
- (現実的にはあり得ないが) θ は既知とする

θ の分布: $\Gamma(8, 0.4)$

↓

事故件数 N_i の分布: $\theta = \theta$ の条件の元で $Po(\theta)$

id	θ	$N_{id,1}$	$N_{id,2}$	$N_{id,3}$	$N_{id,4}$	$N_{id,5}$	$N_{id,6}$	$N_{id,7}$	$N_{id,8}$	$N_{id,9}$	$N_{id,10}$
1	15.8	19	12	25	19	23	15	15	20	8	11
2	19.9	23	15	21	22	17	23	15	17	26	21
3	24.6	26	35	34	36	25	22	23	27	18	22
4	18.5	20	27	17	19	14	22	21	16	17	24
5	31.3	34	26	31	29	28	28	33	32	22	29
6	8.1	4	5	7	5	10	11	13	9	10	14
7	33.0	32	34	29	37	36	31	35	35	32	28
8	14.2	11	13	17	12	14	19	12	13	10	22

id=1の仮説平均保険料 $E(X|\theta = \theta) = \theta = 15.8$

例題2と同じ

θ が分かっている場合はその値を個別保険料とすればよい!

なお、集合的保険料は $E[E(X|\theta)] = E(\theta) = \frac{8}{0.4} = 20$

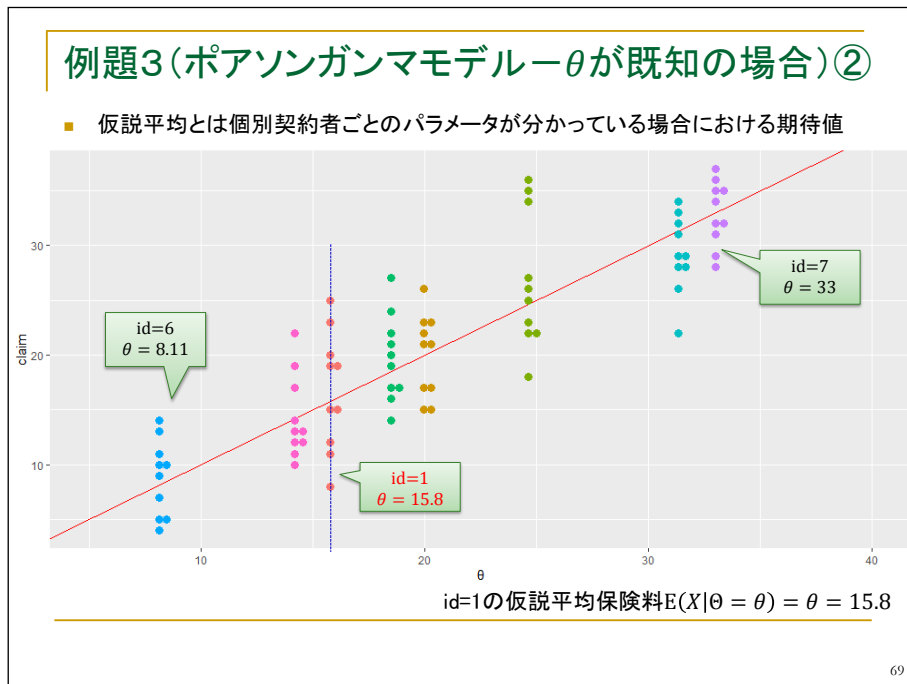
68

先ほどの例題の続きですが、ポアソンガンマモデルで、 θ が既知の場合を考えます。

例題 2 と同じ状況で、事故件数はポアソン分布に従い、契約者ごとのパラメーター θ はガンマ分布に従うとします。現実的にはありえないのですが、 θ は既知といたします。リスク特性について、ガンマ分布を仮定しており、今の例では、 $\Gamma(\alpha = 8, \beta = 0.4)$ の分布を考えています。

先ほどの例題 1 と 2 は、実は事故件数の数値は全く一緒なのですがけれども、先にガンマ分布からサンプルを 8 件選んできておりまして、その θ に従うポアソン分布という形で事故件数を発生させております。ですので、id=1 ですと、 θ は 15.8 と選ばれて、この値を平均とするポアソン乱数によって、10 年分のクレーム件数が発生しています。このような形で、契約者 8 まで発生させています。id=1 の仮説平均保険料というものを考えてみますと、リスクの構造とし

て、パラメータについては、15.8 と分かっています、平均 15.8 のポアソン分布なので、結局のところ、仮説平均保険料としては、15.8 が期待されます。θ が分かっている場合については、その値を個別保険料とすればよいことが分かります。



ここで、プロット図なのですが、横軸はθ、縦軸はクレーム件数を表しています。45度線の赤い線については、θ とクレーム件数が等しい所ですので、ポアソン分布の期待値になります。id=1 のときはθ が 15.8 なのですが、この 15.8 に近い所にクレーム件数が散らばっていることが分かります。id=7 については、θ が 33 ですので、33 の近くにクレーム件数が散らばっていることが伺えます。

例題4 (ポアソンガンマモデル - θ が未知の場合) ①

- 例題3の状況において、θは未知とする
- id=1の10年分のクレーム実績からベイズ保険料を求める
- ベイズ保険料は予測分布の期待値であり、この例題では解析的に解くことが可能

X|θのモデル

θ = θ の条件付で X の確率関数は

$$f_{X|\theta}(x|\theta) = \frac{\theta^x e^{-\theta}}{x!}$$

θの事前分布

θ ~ Γ(α, β)

$$\pi(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta}$$

Γ(8, 0.4) 期待値 $\frac{8}{0.4} = 20$

Xの分布

X ~ NB(k, p)

$$E(X) = \frac{k(1-p)}{p} \quad k = \alpha, p = \frac{\beta}{1+\beta}$$

ベイズ更新

id	N _{id,1}	N _{id,2}	N _{id,3}	N _{id,4}	N _{id,5}	N _{id,6}	N _{id,7}	N _{id,8}	N _{id,9}	N _{id,10}	計	平均
1	19	12	25	19	23	15	15	20	8	11	167	16.70

θの事後分布

$$\pi_{\theta|X}(\theta|\mathbf{x}) = \frac{\beta_*^{\alpha_*}}{\Gamma(\alpha_*)} \theta^{\alpha_*-1} e^{-\beta_*\theta}$$

α_{*} = α + ∑ x_j, β_{*} = β + n

Γ(8 + 167, 0.4 + 10) = Γ(175, 10.4)

期待値 $\frac{175}{10.4} = 16.83$

Xの予測分布

X ~ NB(k_{*}, p_{*})

$$k_* = \alpha_* = 175, p_* = \frac{\beta_*}{1+\beta_*} = \frac{10.4}{1+10.4}$$

$E(X|\mathbf{x}) = \frac{k_*(1-p_*)}{p_*} = 16.83$

続いて例題 4 となりますが、ポアソンガンマモデルで、θ が未知の場合を考えます。すなわ

ち、例題3の状況において、 θ は未知とします。実際には θ が分かることはほとんどありえないと思いますので、実際にはこのような仮定となります。id=1の10年分のクレーム実績から、ベイズ保険料を求めることを考えます。ベイズ保険料は、予測分布の期待値となりますが、この例題では、解析的に解くことが可能です。

まず、 Θ の条件付きの X のモデルですけれども、条件付きで X の確率関数は、ポアソン分布の形で与えられる形となります。先ほどの例のとおり、 Θ の事前分布としては、ガンマ分布を考えるということになってはいますが、個別の θ については、分からないような状況ですね。ここに対して、事前分布 $\Gamma(8, 0.4)$ があり、期待値は20となります。先ほどの佐野さんのスライドの方でも説明がありましたが、ポアソンガンマモデルの場合は、 X の分布としては、負の二項分布になります。

今は、この事前分布、ガンマ分布を考えるのですが、これを、観測値を得るごとに事後分布が更新されていくというところを表わしておきまして、これをベイズ更新と言いますが、例えば10年分の実績値が得られます。ここで、元のガンマ分布に対して、ベイズ更新されたあとの分布については、 $\alpha_* = \alpha + \sum x_j$ となり、つまり観測値の合計数となります。 β_* につきましては、 $\beta + n$ （観測年数を足したもの）となっています。ですので、10年後のガンマ分布の形はどのように推定されるかということ、サンプルが得られるごとにパラメーターが更新されて、 $\Gamma(175, 10.4)$ という形で表わされ、期待値としては16.83となります。

今の場合、ポアソン分布を仮定していましたので、事後分布の期待値がベイズ保険料になるので、16.83と表わされます。また、負の二項分布の期待値として求めてみても、同じ16.83という数値が得られます。ここも、パラメーターは、 k_* が α_* になって、 p_* が $\frac{\beta_*}{1+\beta_*}$ になるという形で計算されます。

例題4(ポアソンガンマモデル- θ が未知の場合)②

- ポアソン分布の共役事前分布はガンマ分布であり、事後分布もガンマ分布となる
- したがって予測分布は負の二項分布であり、パラメータは解析的に得られる
- ベイズ保険料は予測分布の期待値であり、次のとおり表現できる

$$E(X|\mathbf{x}) = \frac{k_*(1-p_*)}{p_*} = \frac{\left(k_* = \alpha + \sum x_j, p_* = \frac{\beta + n}{1 + \beta + n}\right)}{\frac{\beta + n}{1 + \beta + n}}$$

$$= \frac{n}{\beta + n} \frac{\sum x_j}{n} + \left(1 - \frac{n}{\beta + n}\right) \frac{\alpha}{\beta}$$

$$= Z\bar{X} + (1 - Z)\mu$$

$$Z = \frac{n}{\beta + n} = \frac{10}{10 + 0.4} = 0.9615$$

ベイズ保険料 = $Z\bar{X} + (1 - Z)\mu$
 = $0.9615 \times 16.70 + 0.0385 \times 20$
 = 16.83

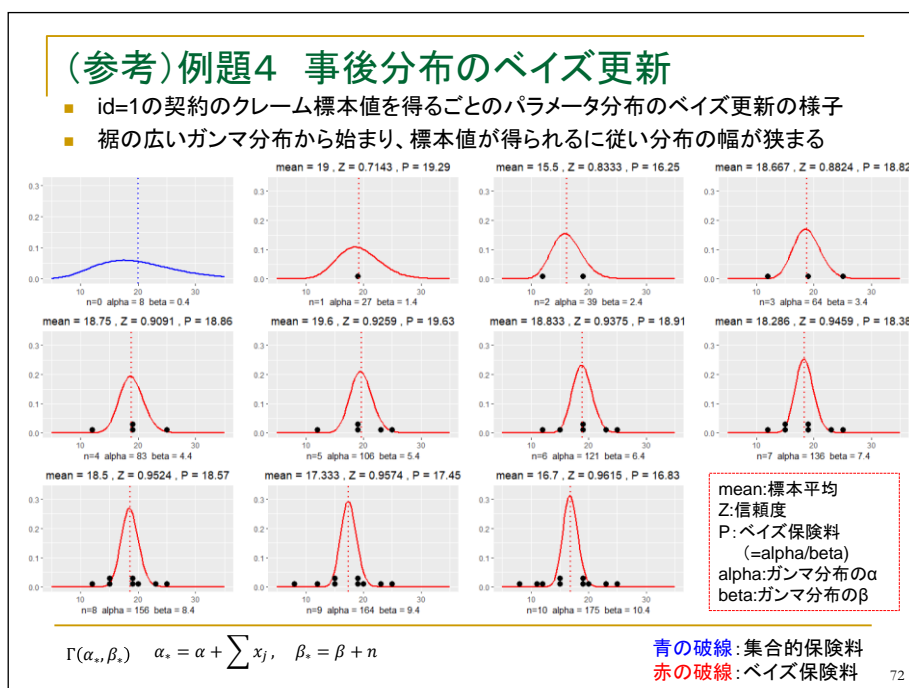
この例題のベイズ保険料は標本平均と集会的保険料の加重平均になっている！

71

続きまして、もう少しかみ砕いたものになりますが、ポアソン分布の共役事前分布はガンマ分布であり、事後分布もガンマ分布となったということは、先ほどの事例のとおり、分かりま

した。従って、負の二項分布ですので、パラメータを解析的に得られた結果となっています。ベイズ保険料は、このような形で観測値が得られたもとの期待値となっており、負の二項分布の形で期待値を書くと、このとおりになるのですが、これを展開すると、2行目の式のような形になっていまして、 $\frac{\sum x_j}{n}$ につきましては、 \bar{x} となります。また、 $\frac{\alpha}{\beta}$ なのですが、こちらは元の事前分布の期待値となりまして、 μ となります。

ここで、 $\frac{n}{\beta+n}$ をZとすると、4行目の形式で書けることが分かります。信頼度をZとして表わすのですが、 $\frac{n}{\beta+n}$ については0.9615となっており、ベイズ保険料の算式で書きますと、 $Z\bar{x} + (1-Z)\mu$ という形で記載でき、すなわちベイズ保険料についても、有限変動信頼性理論と同じような形で、信頼度による加重平均値として表わされるということが分かります。



続きまして、今の例題のベイズ更新の事例ですが、事前分布として、 $\Gamma(\alpha=8, \beta=0.04)$ という、割と広めのガンマ分布の形状から始まっていきまして、観測値が得られるごとにパラメータが更新されていく様子を表しています。1年目の標本値は19というクレーム件数でしたので、 α については27と記載されておりますが、これは8に19を足したものとなっており、 β は0.4に1を足したものとなっております。この α と β から、Pというベイズ保険料も計算されますし、信頼度も計算されます。このように10年分繰り返していくと、最終的には下段の一番右の図の形となっていて、事後分布について幅が狭くなっていくということで、観測値を入れるごとに、どんどん予測が更新されていくことを表わしております。

ベイズ的手法の問題点

問題点1

ベイズ保険料は解析的に計算できないケースが多く、実務上は扱いにくい

解決策

MCMC (Markov Chain Monte Carlo) を用いて数值的に計算することは可能

問題点2

事前分布の選択に恣意性が入る

解決策

経験ベイズ法や無情報事前分布を活用することで恣意性を排除できる

詳細については2014年年次大会のパネルディスカッション「アクチュアリーとベイズ統計学」参照

73

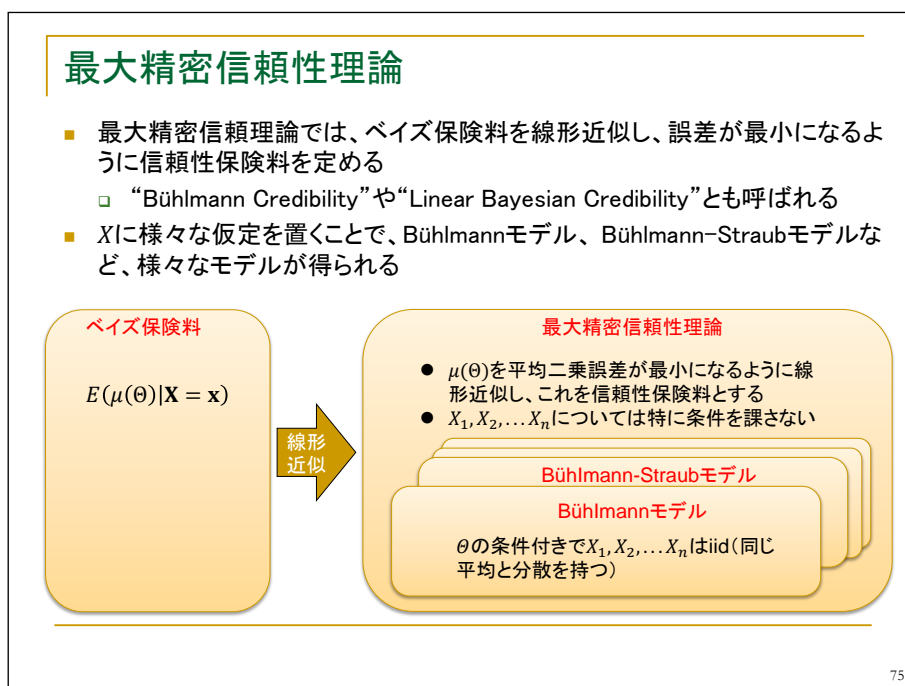
続きまして、ベイズ的手法の問題点ですけれども、ベイズ保険料は、解析的に計算できない例が多いということが挙げられます。今のケースは割と特殊な例で解析的に解けたのですが、一般的にはこのような問題点があります。解決策としては、マルコフチェーンモンテカルロ (MCMC) を用いて、数值的に計算することが可能となっています。

また、問題点2としまして、事前分布の選択に恣意性が入る可能性があります。解決策は、経験ベイズ法や無情報事前分布を活用することで、恣意性を排除することができます。詳細については、2014年度年次大会のパネルディスカッションで「アクチュアリーとベイズ統計学」の中で、先ほどの岩沢先生も講演されておりますが、そちらを参照いただければと思っております。

最大精密信頼性理論 (GREATEST ACCURACY CREDIBILITY)

74

続きまして、「最大精密信頼性理論」について説明させていただきます。



ベイズについては、計算しにくいというところがありますので、それを線形近似したものが、最大精密信頼性理論となります。ここで注意なのですが、最大精密信頼性理論については、ビュールマンが発表したことから、「Bühlmann Credibility」という名前や、あるいは「Linear Bayesian Credibility」とも呼ばれております。また、 X にさまざまな仮定を置くことで、異なるモデルができます。

まず、ベイズ保険料というものが考えられるのですが、最大精密信頼性理論はそれを線形近似するものとなっています。ここで留意事項ですが、 X_i については、特に条件を課さないという形になっておりますので、個々の X_i は別の分布でもいいというのが、最大精密信頼性理論の特徴となっています。この中に、Bühlmannモデルがありますが、 θ の条件付きで、 X_i については iid、すなわち同じ平均と分散を持つと、条件を強めたモデルとなっています。同様に、Bühlmann-Straubモデルなど、さまざまなモデルがあることが知られています。

最大精密信頼性理論の概要

- 個別契約者の翌年度保険料 $\mu_{n+1}(\theta) = E(X_{n+1}|\theta = \theta)$ の推定に当たり、 θ は未知のため θ ではなく過去の実績データ X_1, \dots, X_n を用いた次の線形式を考える

$$\alpha_0 + \sum_{j=1}^n \alpha_j X_j$$

- 上式係数の推定値 $\tilde{\alpha}_0, \tilde{\alpha}_1, \dots, \tilde{\alpha}_n$ は次の平均二乗誤差を最小化することで求まる

$$Q = E \left\{ \left[\mu_{n+1}(\theta) - \alpha_0 - \sum_{j=1}^n \alpha_j X_j \right]^2 \right\}$$

- $\partial Q / \partial \alpha_j = 0$ とすると、 $\tilde{\alpha}_0, \tilde{\alpha}_1, \dots, \tilde{\alpha}_n$ について以下の式が得られる(正規方程式)

$$E(X_{n+1}) = \tilde{\alpha}_0 + \sum_{j=1}^n \tilde{\alpha}_j E(X_j), \quad \text{Cov}(X_i, X_{n+1}) = \sum_{j=1}^n \tilde{\alpha}_j \text{Cov}(X_i, X_j), \quad i = 1, \dots, n$$

- 信頼性保険料はこれを解いて得られ、 $\tilde{\alpha}_0, \tilde{\alpha}_1, \dots, \tilde{\alpha}_n$ を用いて次のとおり表される

$$\tilde{\alpha}_0 + \sum_{j=1}^n \tilde{\alpha}_j X_j$$

- これは、 $E(X_{n+1}|\theta)$ のほか、 $E(X_{n+1}|\mathbf{X})$ や X_{n+1} の最良線形推定量でもある

76

最大精密信頼性理論の概要ですけれども、個別契約者の翌年度保険料 ($\mu_{n+1}(\theta)$) を求めたいという形になっており、本来ですと、先ほどのとおりリスクの θ が分かればいいのですが、 θ については未知なので、過去の実績データ X_i を用いた次の線形式を考えます。線形式を考えるので、 α についていろいろな値を取るのですが、この係数を最少とするような推定値を $\tilde{\alpha}$ と言っていますが、これについては、次の平均二乗誤差を最小化することで求められます。差を取って二乗して、期待値を取ったものなのですが、これを最小化するために各 α_j で偏微分して、得られた解を $\tilde{\alpha}_0$ など、 \sim (チルダ) として表わします。これを線形方程式と言います。

結局のところ、信頼性保険料については、これについて得られた $\tilde{\alpha}$ によって表わして、線形結合の形で表わされることとなっています。この線形結合で表わされた形式については、このような形で、幾つかの最良線形推定量ということが知られています。

Bühlmannモデル

- 最大精密信頼性理論において、 θ の条件付きで X_1, X_2, \dots, X_n はiidを仮定(従って同じ平均 $\mu(\theta) = E(X|\theta = \theta)$ と分散 $v(\theta) = V(X|\theta = \theta)$ を持つ)
- このとき、正規方程式を解くと、信頼性保険料は以下のとおりとなる

$$\tilde{\alpha}_0 + \sum_{j=1}^n \tilde{\alpha}_j X_j = Z\bar{X} + (1-Z)\mu$$

$$\mu = E[\mu(\theta)]$$

ここで、 Z (Bühlmann信頼度)は以下のとおり

$$Z = \frac{n}{n + v/a}$$

$$v = E[v(\theta)], a = V[\mu(\theta)]$$

- μ, v, a は構造パラメータといい、一旦既知とする(推定方法は後述)

77

続いて「Bühlmann モデル」ですけれども、最大精密信頼性理論において、 θ の条件付きで、 X_i はiidを仮定しています。すなわち、同じ平均値と分散を仮定しています。このとき、正規方程式を解くと、信頼性保険料はスライド上段の形で表わされまして、先ほどのとおり、 $Z\bar{X} + (1-Z)\mu$ という形で表わされます。信頼度 Z はスライド下段の形で表わせまして、 $n+v/a$ が分母、分子は n となっていますが、損保数理のテキストでも記載されているものと同様となっています。ここで n, v, a は、構造パラメーターと言いまして、既知とします。

Bühlmann-Straubモデル

- 期間ごとのエクスポージャが異なる場合、各 X_j の分散が異なるためBühlmannモデルでは扱えない
- 各年のエクスポージャを m_1, \dots, m_n ($m = \sum m_i$)、平均クレームを X_1, \dots, X_n とする
- X_i は θ の条件付で独立で、次の平均・分散を持つ

$$E(X_i|\theta = \theta) = \mu(\theta)$$

$$V(X_i|\theta = \theta) = \frac{v(\theta)}{m_i} \left(X_j|\theta = \theta \text{は} m_j \text{個の独立な確率変数} \right)$$

(平均 $\mu(\theta)$ ・分散 $v(\theta)$)の平均

- このとき、正規方程式を解くと、信頼性保険料は以下のとおりとなる

$$\tilde{\alpha}_0 + \sum_{j=1}^n \tilde{\alpha}_j X_j = Z\bar{X} + (1-Z)\mu$$

$$\text{ただし、} \bar{X} = \frac{1}{m} \sum_{j=1}^n m_j X_j$$

ここで、 Z (Bühlmann-Straub信頼度)は以下のとおり

$$Z = \frac{m}{m + v/a}$$

78

続きまして、「Bühlmann-Straub モデル」を簡単に紹介します。こちらについては、期間ごとのエクスポージャーが異なる場合となり、そのままでは扱えないということで、各年のエクスポージャーを m_i としまして、 m としてはその合計値となり、平均クレームを X_i とします。です

ので、総ロスをエクスポージャーで割ったようなものが X_i となります。

X_i は、 θ の条件付きでスライドの形式で平均、分散が表せまして、正規方程式を解くと、先ほどと同じような形式で表わせます。Bühlmann-Straub 信頼度の Bühlmann 信頼度との違いについては、 n であったところが m で表わされるという違いがあります。

構造パラメータの推定

- 最大精密信頼性理論の信頼性保険料には、密度 $\pi(\theta)$ に関連する未知のパラメータ μ, v, a が含まれる(構造パラメータ)
- 構造パラメータは推定しなければならないことが多い
- パラメータの推定方法は以下の3種類に分類できる

ノンパラメトリック

クレーム X 、パラメータ θ のいずれも特定の確率分布を仮定せず構造パラメータの推定を行う

フルパラメトリック

クレーム X 、パラメータ θ のいずれも分布の仮定を置く

セミパラメトリック

クレーム X は何等かの確率分布を仮定し、パラメータ θ は分布の仮定を置かない

79

続いて「構造パラメータの推定」ですけれども、最大精密信頼性理論の信頼性保険料については、先程の未知のパラメータ、 μ 、 v 、 a が含まれていまして、これを構造パラメータと言います。構造パラメータは、実際は推定が必要となっており、推定方法について3種類紹介されています。1点目がノンパラメトリックな手法ですが、クレーム X 、パラメータ θ 、いずれも特定の確率分布を仮定せずに、構造パラメータの推定を行う方法となっています。続いてセミパラメトリックと呼ばれるもので、クレーム X は何らかの確率分布を仮定し、パラメータ θ は、分布の仮定を置かないようなものとなっています。最後のフルパラメトリックについては、いずれも分布の仮定を置くものとなっています。

ノンパラメトリック推定

- クレーム分布 X 、パラメータ分布 θ のいずれも特定の確率分布を仮定しない
- Bühlmann モデルにおいて、契約者 i の j 年のクレームを X_{ij} ($i = 1, \dots, r$ $j = 1, \dots, n$)とすると、以下の $\hat{\mu}, \hat{v}, \hat{a}$ は μ, v, a の不偏推定量となる

$$\hat{\mu} = \bar{X}$$

$$\hat{v} = \frac{1}{r(n-1)} \sum_{i=1}^r \sum_{j=1}^n (X_{ij} - \bar{X}_i)^2$$

$$\hat{a} = \frac{1}{r(n-1)} \sum_{i=1}^r (\bar{X}_i - \bar{X})^2 - \frac{\hat{v}}{n}$$

ここで、 $\bar{X}_i = \frac{1}{n} \sum_{j=1}^n X_{ij}$, $\bar{X} = \frac{1}{r} \sum_{i=1}^r \bar{X}_i = \frac{1}{rn} \sum_{i=1}^r \sum_{j=1}^n X_{ij}$

80

「ノンパラメトリック推定」ですけれども、クレーム分布 X 、パラメーター分布 θ のいずれも仮定しないものとなっていますので、Bühlmann モデルの場合は、損保数理のテキストでも記載がありますが、スライドの式で不偏推定量が求められるということになっており、例えば μ の推定値ですと、 \bar{X} で表わされるなどということが知られています。

セミパラメトリック推定

- クレーム X は何らかの分布を仮定し、 θ については分布を仮定しない
- μ, v および a に関数的な関係が生じ、パラメータの推定が容易になる
- 例: Bühlmann モデルにおいて $X|\theta$ が $Poisson(\theta)$ に従う場合
 - ポアソン分布の性質より、 $E(X|\theta) = V(X|\theta) = \theta$
 - これを利用すると、構造パラメータは以下のように推定できる

$$\hat{\mu} = \hat{v} = \bar{X}$$

$$\hat{a} = \widehat{V(X)} - \hat{\mu} \quad \widehat{V(X)} \text{は } V(X) \text{の推定量(不偏標本分散等)}$$

$$\left(\begin{array}{l} v = E(V(X|\theta)) = E(\theta) = \mu \\ V(X) = E(E(X|\theta)) + E(V(X|\theta)) \\ \quad = v + a \\ \quad = \mu + a \end{array} \right)$$

81

続いて「セミパラメトリック推定」ですが、クレーム X は何らかの分布を仮定し、 θ については分布を仮定しないものとなっています。 μ, v および a に関数的な関係性が生じるため、パラメーター推定が容易になるという特徴があります。例として、Bühlmann モデルにおいて、この θ の条件付きの X が $Poisson(\theta)$ に従う場合を考えると、期待値と分散値が θ で表わされますので、構造パラメーターの推定が容易であることが分かります。

例題5 (Bühlmannモデル) ノンパラメトリック①

- 例題2~例題4と同じ状況でid=1のBühlmann保険料を求める。

id	$N_{id,1}$	$N_{id,2}$	$N_{id,3}$	$N_{id,4}$	$N_{id,5}$	$N_{id,6}$	$N_{id,7}$	$N_{id,8}$	$N_{id,9}$	$N_{id,10}$	計	平均
1	19	12	25	19	23	15	15	20	8	11	167	16.70
2	23	15	21	22	17	23	15	17	26	21	200	20.00
3	26	35	34	36	25	22	23	27	18	22	268	26.80
4	20	27	17	19	14	22	21	16	17	24	197	19.70
5	34	26	31	29	28	28	33	32	22	29	292	29.20
6	4	5	7	5	10	11	13	9	10	14	88	8.80
7	32	34	29	37	36	31	35	35	32	28	329	32.90
8	11	13	17	12	14	19	12	13	10	22	143	14.30
合計	169	167	181	179	167	171	167	169	143	171	1,684	168.40
平均	21.13	20.88	22.63	22.38	20.88	21.38	20.88	21.13	17.88	21.38	210.50	21.05

$$\hat{\mu} = \bar{X} = 16.70$$

$$\hat{v} = \frac{1}{r(n-1)} \sum_{i=1}^r \sum_{j=1}^n (X_{ij} - \bar{X}_i)^2 = 18.25$$

$$\text{信頼度 } Z = \frac{n}{n+v/a} = \frac{10}{10 + \frac{18.25}{63.52}} = 0.9721$$

$$\hat{a} = \frac{1}{r(n-1)} \sum_{i=1}^r (\bar{X}_i - \bar{X})^2 - \frac{\hat{v}}{n} = 63.52$$

$$\begin{aligned} \text{Bühlmann保険料} &= Z \times 16.70 + (1-Z) \times 21.05 \\ &= 0.9721 \times 16.70 + 0.0279 \times 21.05 \\ &= 16.82 \end{aligned}$$

82

例題 5 としまして、Bühlmann モデルのノンパラメトリック推定について紹介させていただきます。例題 2 から 4 までと同じような状況で、id=1 の Bühlmann 保険料を、ノンパラメトリック推定で求めたものになっています。データから $\hat{\mu}$ については \bar{X} で表わされて、16.70、 \hat{v} については 18.25、 \hat{a} については 63.52 となっていて、これから、信頼度は 0.9721 となります。Bühlmann 保険料につきましても、先ほどと同様の考えで個別と全体の信頼度の平均値となっており、16.82 が答えになっています。

$E(V(X|\Theta))$ と $V(E(X|\Theta))$

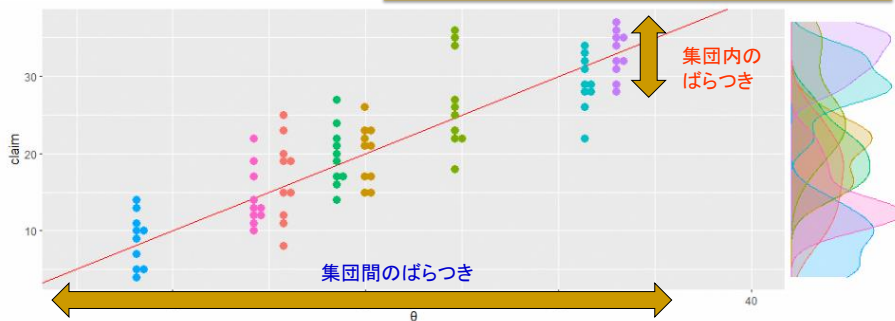
$$Z = \frac{n}{n + \frac{v}{a}} = \frac{n}{n + \frac{E(V(X|\Theta))}{V(E(X|\Theta))}}$$

$E(V(X|\Theta))$: 条件付分散の期待値 \Rightarrow その集団内のばらつき

集団内のばらつきが小さければ信頼度が上がり、その集団についての標本平均の重みが大きくなる

$V(E(X|\Theta))$: 条件付期待値の分散 \Rightarrow 集団間のばらつき

集団間のばらつきが小さければ信頼度が下がり、全体平均の重みが大きくなる(集団間の違いが重要ではない)

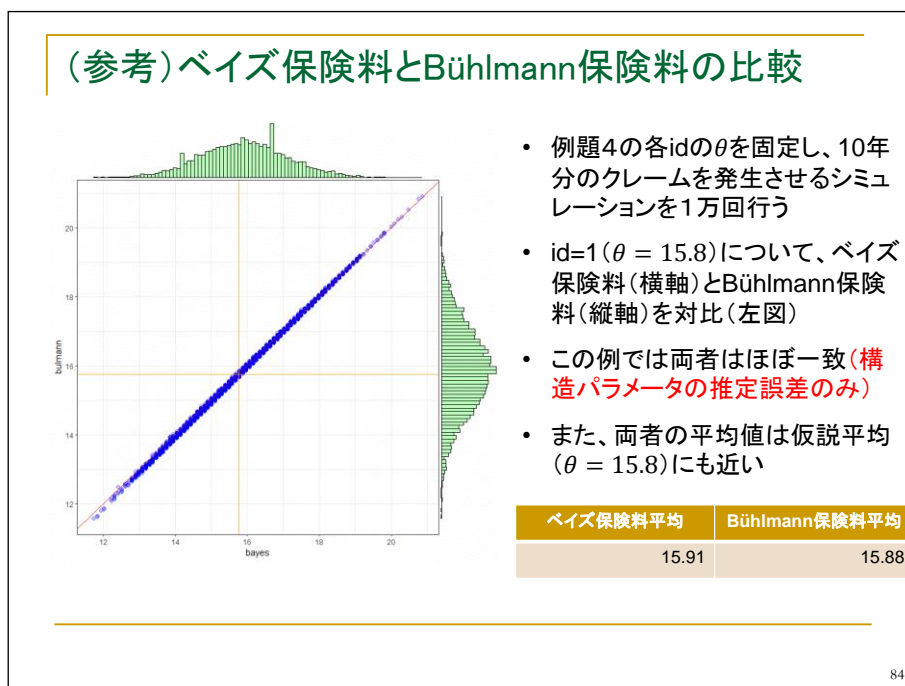


83

続いて、信頼度について、もう少しみ砕いて説明したスライドになりますが、信頼度の計算においては、分母が $n + \frac{v}{a}$ となります。これは、条件付きで書きますと右辺のとおりになって

いまして、このうちの分子が条件付き分散の期待値、分母が条件付き期待値の分散となっています。この赤字の所ですが、条件付き分散の期待値につきましては、この図でいくと、個々の集団のばらつきの期待値となっており、集団内のばらつきを表わしたものとなっています。横の方に図が記載されていますが、これは、各 id ごとの分布図となっています。これが広いほど、ばらついているという状況になります。条件付き分散の期待値は、これが小さいとリスクパラメーターから計算される期待値に近くなり、標本平均からそれほど離れていないことが期待され、信頼度が上がるという形になっています。

続いて分母の方ですが、条件付き期待値の分散なのですが、これは集団間のばらつきを表しています。集団間のばらつきが小さければ信頼度が下がり、全体の重みが大きくなっておりまして、集団が近しい場合については、差異が分かりにくくなるということかと考えています。



参考スライドですが、ベイズ保険料と Bühlmann 保険料の比較をしたものです。例題 4 の各 id の θ を固定しまして、クレーム件数について、改めてランダムでシミュレーションするというのを 1 万回行った結果を図示しています。図は id=1 について、横軸のベイズ保険料と、縦軸の Bühlmann 保険料を対比した図となっていますが、45 度線に乗っていると両保険料が等しいということですので、この結果、今回の場合はほぼ一致しています。また、両者の分布が記載されていますが、平均値は、先ほどの仮説平均 15.8 に近い結果となっています。

Bühlmann-Straubモデルの μ の推定(TP=TL)

- 構造パラメータの不偏推定量は一意には定まらず、複数の手法が存在し得る
- 例えば、Bühlmann-Straubモデルの構造パラメータ μ の推定について、損保数理のテキストでは、クレームの加重平均 $\bar{X} = \frac{1}{m} \sum_{i=1}^r m_i \bar{X}_i$ を用いている
- このとき、過去の観測期間におけるクレーム総額(TL)と、同じ期間に信頼性保険料を適用していた場合に得られる収入保険料(TP)は一致しない(TP \neq TL)

$$TL = \sum_{i=1}^r m_i \bar{X}_i$$

$$TP = \sum_{i=1}^r m_i [\hat{Z}_i \bar{X}_i + (1 - \hat{Z}_i) \hat{\mu}] = \sum_{i=1}^r m_i \frac{\hat{k}}{m_i + \hat{k}} (\hat{\mu} - \bar{X}_i) + \sum_{i=1}^r m_i \bar{X}_i \neq TL \quad (\hat{k} = \hat{v}/\hat{\mu})$$

$\neq 0$

- 実務上、料率算定においては、TP=TLである方が説明しやすい
この場合は μ の推定量 $\hat{\mu}$ について次が期待される

$$0 = \sum_{i=1}^r m_i \frac{\hat{k}}{m_i + \hat{k}} (\hat{\mu} - \bar{X}_i) \Leftrightarrow \hat{\mu} = \frac{\sum_{i=1}^r \hat{Z}_i \bar{X}_i}{\sum_{i=1}^r \hat{Z}_i}$$

- どちらの方法も不偏推定量を与えるが、後者の方法にはTP=TLという利点がある
- また後者の方法は、 \bar{X}_i の加重平均のなかで $\hat{\mu}$ の分散を最小にする

85

続いて Bühlmann-Straub モデルの構造パラメータについての補足的な説明です。構造パラメータの推定については、不偏推定量は一意に定まらず、複数の手法が存在していることが知られています。例えば、Bühlmann-Straub モデルの構造パラメータ μ の推定について、損保数理のテキストにおいては、 \bar{X} として、スライドに記載のとおり形で加重平均を用いています。しかしながら、「Loss Models」のテキストにおいては課題が指摘されていまして、過去の観測値におけるクレーム総額（トータルロス：TL）と、信頼性保険料を用いた場合の収入保険料（トータルプレミアム：TP）が一致しない。いわゆる収入と支出が一致しないという課題が、指摘されています。

具体的には、TLはスライド式の赤字の形式で表わせまして、一方TPは、計算していくとその中にTLが入ってきますので、下線部がゼロではないので、両者が一致しないという結果になります。実務上は、料率算定においてはTP=TLという方が説明しやすいと言われておりまして、その場合、どのようにしたら良いかが「Loss Models」では示されています。 μ の推定量 $\hat{\mu}$ について、スライドの枠内の式で計算すると記載されていまして、具体的に言うと、Zでの加重平均値となっています。2つの推定方法について、どちらについても不偏推定量を与えますが、後者にはTP=TLという利点があり、更に推定量 $\hat{\mu}$ の分散を最小にする、最小不偏分散推定量ということが知られています。

(参考)Rの「actuar」パッケージ

- Rの「actuar」パッケージ(*1)の「cm」関数では各種Credibility Modelが実装されており、BühlmannモデルやBühlmann-Straubモデル等の計算が簡単に実行できる
- Bühlmann-Straubモデルでは標準で「TP=TL」のパラメータ推定方法(信頼度による加重平均値)が実装されている

```
> p_bul<-cm(~id,data=data_pois_gamma,ratios = c(1:year_n),method = "Buhlmann-Gisler")
> summary(p_bul)
Call:
cm(formula = ~id, data = data_pois_gamma, ratios = c(1:year_n),
    method = "Buhlmann-Gisler")

Structure Parameters Estimators

Collective premium: 21.05

Between id variance: 63.515
within id variance: 18.25

Detailed premiums
```

id	Indiv. mean	Weight	Cred. factor	Cred. premium
1	16.7	10	0.9720692	16.821499
2	20.0	10	0.9720692	20.029327
3	26.8	10	0.9720692	26.639398
4	19.7	10	0.9720692	19.737707
5	29.2	10	0.9720692	28.972364
6	8.8	10	0.9720692	9.142153
7	32.9	10	0.9720692	32.569020
8	14.3	10	0.9720692	14.488533

例題5 id=1の
ビュールマン保険料

*1 C. Dutang, V. Goulet and M. Pigeon (2008). actuar: An R Package for Actuarial Science. Journal of Statistical Software, vol. 25, no. 7, 1-37. URL <http://www.jstatsoft.org/v25/i07>

参考情報として、Rの「actuar」パッケージを紹介させていただきます。この中で「cm」関数があり、Credibility Modelの略だと思われそうですが、信頼性理論が実装されており、Bühlmannモデル、Bühlmann-Straubモデル等の計算が、簡単に実行できるようになっています。また、Bühlmann-Straubモデルの場合は、前のスライドで説明させていただいたTP=TLのパラメータ推定が、標準で実装されています。パッケージは非常に簡単に使えるものとなっており、例題5についても改めて計算したところ、全く同じ数値が出る結果となりました。なお、このパッケージではBühlmann-Straubモデルより発展的なモデルや、ベイズの階層モデルについても算出でき、例えばBühlmannモデルについては、エクスポージャー変動を考慮したBühlmann-Straubがあるのですが、それに加えて、パラメーターが時間的に変化するようなモデルの推定もできるようになっています。

例題1～5の保険料比較

- 有限変動信頼性保険料は、 Z の計算方法や p, r の選択に依存
例題2ではベイズ的手法やBühlmannモデルと大きく異なる結果となった
($Z = n/(n+k)$ とし、 k をBühlmannモデルの方法で決めれば一致する)
- ベイズ保険料とBühlmann保険料(ノンパラメトリック推定)は近い値となる

例題	保険料区分	数式	説明	信頼度	保険料
例題1	有限変動信頼性保険料①	$Z\bar{X} + (1-Z)M$	関連データを M として用いた保険料で、契約者間のリスク特性の違いは考慮しない (関連データを用いているため以下の例題と単純比較はできない)	0.3952	23.47
例題2	有限変動信頼性保険料②	$Z\bar{X} + (1-Z)M$	全体平均を M として用いた契約者ごとの保険料	0.3929	19.34
例題3	仮説平均	$E(X_{n+1} \theta) = \mu(\theta)$	パラメータ θ が既知の場合の特定契約者の保険料	-	15.80
例題4	ベイズ保険料	$E(X_{n+1} X = x)$	予測分布の期待値	0.9615	16.83
例題5	Bühlmann保険料	$Z\bar{X} + (1-Z)\mu$	最大精密信頼性理論のうち一番単純なモデルでノンパラメトリック手法により推定	0.9721	16.82

87

ここで、例題1から5の保険料比較を行います。有限変動信頼性理論については、例題1と2で取り上げましたが、 Z の計算方法や、 p と r の選択に依存する形となっていました。例題1は関連データを用いたため、簡単に比較できない状況となっていますので、例題2を考えてみますと、下の例題3、4、5、特に例題4、5と、水準感が違うような形になっています。仮説平均が一番正しいだろうと思われているものなのですが、実際にパラメーターが見えない形ですので、推定したベイズ保険料とBühlmann保険料については掲載の数値となっていますが、両者の結果が近くなっているのは偶然ではなくて、ある条件のもとでは、例題5の保険料についてはベイズ保険料と一致することが知られています。

Exact Credibility(完全信頼性)①

- 信頼性保険料はベイズ保険料への最良の線形近似となるが、信頼性保険料がベイズ保険料に等しいとき**完全信頼性(exact credibility)**と呼ぶ
- BühlmannモデルやBühlmann-Straubモデルにおいて、 $X|\theta$ の分布が線形指数族で、 θ の分布がその共役事前分布である場合に典型的に見られる

- 線形指数族
パラメータ θ を持つ確率変数 $X|\theta$ の確率(密度)関数が以下で与えられるとき、 $X|\theta$ は線形指数族の確率分布に従うという

$$f_{X|\theta}(x|\theta) = \frac{p(x)e^{r(\theta)x}}{q(\theta)}$$

- 線形指数族の共役事前分布
線形指数族について、 θ の確率(密度)関数 $\pi(\theta)$ が以下で与えられるとき(k, μ はパラメータ)、事後確率(密度)関数 $\pi_{\theta|X}(\theta|x)$ は $\pi(\theta)$ と同じ型となる

$$\pi(\theta) = \frac{[q(\theta)]^{-k} e^{-\mu kr(\theta)} r'(\theta)}{c(\mu, k)}$$

- 事後分布のパラメータは、 $k^* = k + n, \mu^* = \frac{k}{k+n}\mu + \frac{n}{k+n}\bar{x}$ となる(n : 標本数、 \bar{x} : 標本平均)

88

次のスライドで、そのことについて説明させていただきます。それが、完全信頼性(Exact

Credibility) という概念となります。信頼性保険料は、ベイズ保険料の最良の線形近似となりますが、信頼性保険料がベイズ保険料に等しいとき、完全信頼性 (Exact Credibility) と呼びます。特に Bühlmann モデル等において、条件付きのクレーム分布が線形指数族に従い、パラメーター分布がその共役事前分布の場合に、典型的に見られることとなっています。

「線形指数族」について簡単に説明させていただきますと、従う分布としては、ポアソン分布やガンマ分布、正規分布などがありますが、 Θ の条件のもとでの X の密度関数が、スライド式の形で表わされるものとなっています。

「線形指数族の共役事前分布」というものがありますが、このような形で与えられる分布になりまして、この場合、線形指数族に共役事前分布を選びますと、事後密度関数が元の事前分布と同じような形になっておりまして、先ほどのベイズ更新の例で見ていただいたとおり、事前分布と観測値によって得られる事後分布が、同じガンマ分布に従うということとなっています。

加えまして、事後分布のパラメーターについては、 k^* は $k+n$ 、 μ^* は $\frac{k}{k+n}\mu + \frac{n}{k+n}\bar{x}$ となることが知られています。

このような仮定を置くことで、計算上、取り扱えるモデルになりますが、本当にクレーム自体が線形指数族に従うかどうか、あるいは事前分布が共役事前分布かどうかは分かりませんが、モデルの正しさよりも、扱いやすさを優先したということになるかと思われます。このように扱いやすさを優先する仮定をすることは、GLM でも見られることかと思われます。

Exact Credibility (完全信頼性) ②

- Bühlmannモデルの前提条件の場合 (Θ の条件付きで X_1, X_2, \dots, X_n は iid)
 - 事前分布 $\pi(\theta)$ のパラメータ μ, k について

$$E[\mu(\Theta)] = \mu + \frac{\pi(\theta_0)}{kr'(\theta_0)} - \frac{\pi(\theta_1)}{kr'(\theta_1)} \quad ((\theta_0, \theta_1) \text{ は } \pi(\theta) \text{ の台})$$

$$k = \frac{E[v(\Theta)] + \mu(\theta_0) \frac{\pi(\theta_0)}{r'(\theta_0)} - \mu(\theta_1) \frac{\pi(\theta_1)}{r'(\theta_1)}}{V[\mu(\Theta)]}$$

$\frac{\pi(\theta_0)}{r'(\theta_0)} = \frac{\pi(\theta_1)}{r'(\theta_1)}$ のとき $\mu = E[\mu(\Theta)]$ 、 $\mu(\theta_0) \frac{\pi(\theta_0)}{r'(\theta_0)} = \mu(\theta_1) \frac{\pi(\theta_1)}{r'(\theta_1)}$ のとき $k = \frac{E[v(\Theta)]}{V[\mu(\Theta)]}$ ① ②
 - ベイズ保険料 $E[X_{n+1} | \mathbf{X} = \mathbf{x}]$ は以下のとおり

$$E[X_{n+1} | \mathbf{X} = \mathbf{x}] = \mu_* + \frac{\pi_{\Theta|\mathbf{X}}(\theta_0|\mathbf{x})}{k_* r'(\theta_0)} - \frac{\pi_{\Theta|\mathbf{X}}(\theta_1|\mathbf{x})}{k_* r'(\theta_1)}$$

$\frac{\pi_{\Theta|\mathbf{X}}(\theta_0|\mathbf{x})}{r'(\theta_0)} = \frac{\pi_{\Theta|\mathbf{X}}(\theta_1|\mathbf{x})}{r'(\theta_1)}$ のとき ③

$$E[X_{n+1} | \mathbf{X} = \mathbf{x}] = \mu_* = \frac{n}{n+k} \bar{X} + \left(1 - \frac{k}{n+k}\right) \mu$$

$= Z\bar{X} + (1-Z)\mu$

さらに、①②のとき、
 $\mu = E[\mu(\Theta)]$, $k = \frac{E[v(\Theta)]}{V[\mu(\Theta)]}$
 であり、Bühlmannモデルと一致

次のスライドでは、理論モデルの前提では両者が一致するということが式で示してありまして、詳細は割愛させていただきますが、①、②、③という条件が成り立つときに、Bühlmann 保険料と一致することが示されています。

完全信頼性となる線形指数族と共役事前分布の例

- 代表的な線形指数族と共役事前分布は次のとおり(*)
- この場合、ベイズ保険料が信頼性保険料と等しくなり、ベイズ保険料を解析的に求めることができる

$X \theta = \theta$	θ	ベイズ保険料	信頼度
$Poisson(\theta)$	$Gamma(\alpha, \lambda)$	$\frac{\alpha + \sum_{i=1}^n X_i}{\lambda + n}$	$\frac{n}{n + \lambda}$
$Exponential(\theta)$	$Gamma(\alpha, \lambda)$	$\frac{\lambda + \sum_{i=1}^n X_i}{\alpha + n - 1}$	$\frac{n}{n + \alpha - 1}$
$Normal(\theta, \sigma_2^2)$	$Normal(\mu, \sigma_1^2)$	$\frac{\sigma_1^2 \sum_{i=1}^n X_i + \sigma_2^2 \mu}{n\sigma_1^2 + \sigma_2^2}$	$\frac{n}{n + \sigma_2^2 / \sigma_1^2}$
$Binomial(v, \theta)$	$Beta(a, b)$	$\frac{v(a + \sum_{i=1}^n X_i)}{a + b + nv}$	$\frac{n}{n + (a + b)/v}$
$Geometric(\theta)$	$Beta(a, b)$	$\frac{b + \sum_{i=1}^n X_i}{a + n - 1}$	$\frac{n}{n + a - 1}$

* C. Dutang, V. Goulet and M. Pigeon (2008). actuar: An R Package for Actuarial Science. Vignettes (credibility)
<https://cran.r-project.org/web/packages/actuar/vignettes/credibility.pdf>

90

次に、完全信頼性となる線形指数族、共役事前分布の事例です。先ほどのポアソンガンマモデルや、あるいは、よく知られたものとして、指数ガンマ、ノーマルノーマル、二項ベータ、幾何ベータというものが挙げられています。この場合は、ベイズ保険料や信頼度については、パラメータから簡単に求めることができます。

(参考)ポアソン・ガンマモデルの事例

- スイスの自動車保険(疑似データ)について、地域(26区分)別の事故件数から、ポアソン・ガンマモデルを仮定しBühlmann-Straubモデルで信頼度を算出(*)

N_i : 地域*i*の事故件数、 v_i : 地域*i*のエクスポージャー

事故件数: $N_i | \Lambda \sim Po(\Lambda v_i)$

リスク・パラメータ: $\Lambda \sim \Gamma(\gamma, c)$

条件付密度: $\pi(\theta | N_i) \propto e^{-\theta v_i} \frac{(\theta v_i)^{N_i}}{N_i!} \frac{c^\gamma}{\Gamma(\gamma)} \theta^{\gamma-1} e^{-c}$
 $\propto \theta^{\gamma+N_i-1} e^{-(c+v_i)\theta}$

事後分布: $\Lambda | N_i \sim \Gamma(\gamma + N_i, c + v_i)$



$N_1, N_2, \dots, N_n: N_i | \Lambda \sim Po(\Lambda v_i)$

$N = \sum_{i=1}^n N_i | \Lambda \sim Po\left(\Lambda \sum_{i=1}^n v_i\right)$

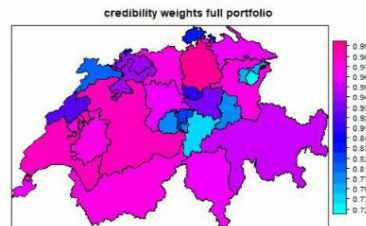
信頼度

$$\alpha_i = \frac{v_i}{c + v_i} \in (0, 1) \quad \text{and} \quad \hat{\lambda}_i = \frac{N_i}{v_i}$$

結果

地域	エクスポージャー	信頼度
全体	253,022	-
最大地域	42,644	99%
最小地域	1,194	72%

それぞれの地域の分散の違いをポアソンガンマモデルで表現



(Wuthrich (2020), Fig.4.1)

* Wuthrich, Mario V. and Buser, Christoph, Data Analytics for Non-Life Insurance Pricing (September 10, 2020). Swiss Finance Institute Research Paper No. 16-68, <https://ssrn.com/abstract=2870308>, Section 4.1.3 Example in motor insurance pricingより

91

ここでポアソンガンマモデルの事例です。これは、スイスの自動車保険。疑似データを用いた事例です。こちらについては、下の方にありますが、スイスの「Data Analytics for Non-Life Insurance Pricing」という文献からの紹介になります。元データはフランスの実績データだそうですが、これを用いて、Bühlmann-Straubモデルで地域別の信頼度を計算しています。こちらにありますとおり、全体のエクスポージャーとしては25万件程度ありまして、最少地

域では 1,194 件なのですが、信頼度は 72%になるという計算をしています。また、こちらでは割愛していますが、特定の年齢条件ごとの信頼度等についても計算されています。信頼度については、先ほどのスライドと同じような形式で表わされることが分かります。

終わりに

92

終わりに、米国の状況等を中心に説明させていただきます。

北米での活用状況①

- 損保分野での調査は見当たらず(使っていて当たり前?)
- 生保・年金分野
 - 米国
 - 2009年アンケート^{*1}(調査対象190社中有効回答19社)
 - 利用または利用予定 死亡率17社、解約率14社、経費4社
 - 最大精密信頼性理論の利用(予定)はゼロ
 - 用途はValuation, Actuarial Opinion, Pricingなど様々
 - NAIC Valuation Manual VM-20(生保契約の原則主義方式準備金)
 - 内国歳入庁 DB年金制度における現在価値決定に関する死亡率表
 - カナダ:2018年アンケート^{*2}(回答11社、保険料シェア約70%)
 - 死亡率では9社が利用、解約率での利用は少数、その他の利用4社
 - 利用しているすべての会社が有限変動信頼性理論を利用
 - CIAガイドライン、手法が単純、業界で利用、自社実績と整合

^{*1} Klugman, S., Rhodes, T., Purushotham, M., Gill, S., & Solutions, M. I. B. (2009). Credibility theory practices.
^{*2} CIA (2019). The Application of Credibility Theory in the Canadian Life Insurance Industry.

93

北米での信頼性理論の活用状況ですけれども、損保分野での調査は見当たらなかったのですが、使っていて当たり前なのかなというところがございます。特徴としては、生保・年金分野でも使われているという結果となります。こちらは 2009 年のアンケートで、有効回答数は少ないものの、回答のあった会社につきましては、死亡率や解約率の設定に信頼性理論を使われているということでした。

カナダでも同様に、死亡率や解約率等で使用がなされています。

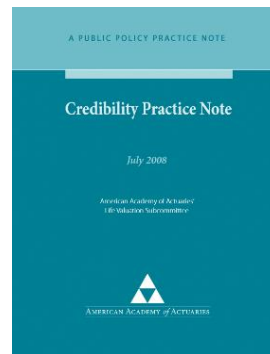
日本の生保分野では、ほとんど使っていないと伺っていますが、日本とアメリカなどの状況がだいぶ違うことが影響しているのではないかと、感触を聞いています。例えば定期保険については、米国では新規契約と更新契約の保険料格差がかなりあり、更新率の設定が重要であることなども影響していると考えています。

北米での活用状況②

『Credibility Practice Note』(American Academy of Actuaries)

2.2 Credibility Theory and Regulatory Practice – An Example: Credibility of Annual Claims Experience Data Reported by Health Insurance Companies to Regulatory Agencies

- 個人向けの健康保険を販売する保険会社
- 保険料の値上げにあたっては、州の規制当局から事前承認が必要
- 当局が「妥当」と考える料率水準⇒例：医療費用保険の損害率70%以上(カリフォルニア州の場合)
- 保険会社と規制当局が直面する重要課題は、将来のクレームコストのBest Estimateを算出するために、会社の過去データを外挿することが、どの程度正しく適切であるかということ
- この課題に対し、本プラクティスノートでは、有限変動信頼性理論と最大精密信頼性理論を用いる方法を提示している



94

続いて事例②ですが、こちらについては、AAA (American Academy of Actuaries) が出しているプラクティスノートです。ここでは、個人向けの健康保険を販売している保険会社向けにプラクティスノートを出しておりまして、具体的な計算手法などを掲載しています。有限変動信頼性理論と、最大精密信頼性理論を用いた事例を紹介しています。

3手法のメリットデメリット・活用状況

	有限変動信頼性理論	ベイズ的手法	最大精密信頼性理論
メリット	<ul style="list-style-type: none"> ✓ 伝統的に活用され、理解が得られやすい ✓ 全信頼に必要な件数を算出できる ✓ 他のデータで補完ができる 	<ul style="list-style-type: none"> ✓ 事後分布により、期待値だけではなくパラツキについての情報も得られる 	<ul style="list-style-type: none"> ✓ スプレッドシートやRで計算可能 ✓ パラメトリックだけでなく、ノンパラメトリックな場合も扱える
デメリット	<ul style="list-style-type: none"> ✓ p, rの選択に恣意性が入る ✓ Mの選択に制約がなく結果の妥当性の保証がない 	<ul style="list-style-type: none"> ✓ 特殊なケースを除いては解析解が得られない ✓ 事前分布に恣意性が入る懸念がある 	<ul style="list-style-type: none"> ✓ モデルの内容について理解が得られにくい

Frees教授からのコメント

- ✓ 有限変動信頼性理論は、1920年代に労働者災害補償保険分野で開発された古典的なアプローチ。料率算定に信頼性理論を適用すべきかという基本的な問いに答え、シンプルかつ直感的な公式であり、結果の解釈や説明が行い易い。
- ✓ ベイズ的手法は最も有用である。しかしながらMCMCなどの技法を用いるので、事前分布とクレームの分布が既知でなければならない(事前に決定する必要がある)。
- ✓ データに関する情報が少ない場合(分布が分からないなど)には、ノンパラメトリック手法が使える最大精密信頼性理論が望ましい。

(2020.9.8 ASTIN関連研からの照会メールに対する回答より)

95

続きまして、3手法のメリットとデメリットですが、有限変動信頼性理論は伝統的に活用され、理解されやすいのですが、恣意性が入る可能性がある。ベイズ的手法は、期待値だけではなく、ばらつきの情報も得られることが期待できますが、特殊なケースを除いて、解析的に解が得にくいということがあります。最大精密信頼性理論は、スプレッドシートやRで計算可能ということがあります。モデルの内容について、理解が得にくい可能性があります。

下の方にございますが、米国の Wisconsin School of Business のリスク保険学部の教授で、ASTIN 関連研究会で翻訳を行った「Predictive Modeling Applications in Actuarial Science」のエディターでもあります Frees 氏から、信頼性理論についてコメントをいただいています。有限変動信頼性理論は、古いものですが、古典的なアプローチで、「シンプルかつ直感的な方式で結果の解釈や説明がしやすい」となっています。ベイズ的手法は最も有用ですが、MCMC などの技法を用いるので、事前分布とクレーム分布が既知でなければならないということがあります。また、データが少ない場合には、ノンパラメトリックが使える最大精密信頼性理論が望ましいとなっています。

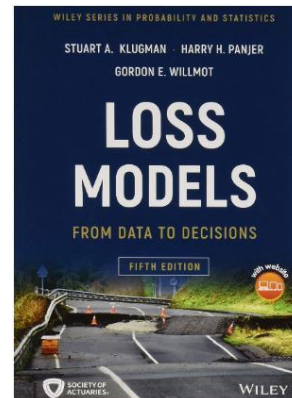
まとめ

- 与えられている状況(データや情報の有無/計算時間の制約等)によって各種手法を使い分けると良い
- MCMCによりベイズ的手法のハードルは下がってきたが、その線形近似としての最大精密信頼性理論にも十分価値がある
 - 最大精密信頼性理論をベイズ的手法の線形近似と考えれば、料率算定以外(リザービング、将来予測前提設定等)にも活用できるのではないか
- 新商品の料率算定等、利用できるデータが限られている場合には、有限変動信頼性理論も有用
 - p, r 選択の恣意性や M 選択の妥当性の懸念は健全な実務慣行や保険計理人の関与等により低減できる

96

まとめとして、与えられている状況、データの情報の有無、計算時間の制約等によって、各種手法を使い分けることがいいのかなと考えています。それぞれメリット、デメリットがありますので、使い分けが重要ということですね。MCMC により、ベイズ的手法のハードルは下がってきましたが、その線形近似としての最大精密信頼性理論についても十分価値があるのではないかと考えています。保険料以外にも、リザービング等で使えるのではないかと考えています。また、新商品の料率算定等、利用できるデータが限られている場合には、有限変動信頼性理論も有用となっています。先ほどの恣意性の課題については、健全な実務基準等があれば、軽減できるものと考えています。

ご清聴ありがとうございました



97

第3部の発表は、以上となります。時間の都合上、Q&Aについては割愛させていただきたいと思っております。これで、ASTIN 関連研究会の発表を終わらせていただきます。ありがとうございました。

【司会】 以上をもちまして、セッションCの2、「SOA テキストに見る損保数理のアクチュアリー実務への応用」を終了します。発表された3名の方、ありがとうございます。