

# リスク管理のための予測モデリング

2021年11月13日

早稲田大学大学院会計研究科 客員教授

岩沢宏和

# 話す項目

- 自己紹介
- リスクを扱うための予測モデリングとは何か
- リスクを扱うための予測モデリングを学ぶ
- リスクを扱うための予測モデリングを創る
- おわりに

# 話す項目

- 自己紹介
- リスクを扱うための予測モデリングとは何か
- リスクを扱うための予測モデリングを学ぶ
- リスクを扱うための予測モデリングを創る
- おわりに

# 自己紹介

- 早稲田大学大学院会計研究科 客員教授。アクチュアリー専門コースの学生を主に対象とした数科目，特に「予測モデリング論」「リスクデータサイエンスワークショップ」。
- データサイエンス（以下「DS」）関係は，本学のほかに
  - 日本アクチュアリー会（以下「ア会」）で教育，普及，研究活動
  - 東北大学でも今年度12月に集中講義（特任教授）
- 法人に属してアクチュアリー実務（年金アクチュアリー）をしていたのは20世紀。

→実務家でも研究者でもなく教育者。また，今日はふだんの授業と違ってテクニカルな話はあまりせず，「考え」を伝える「講話」であり，学生のみなさんにメッセージを送ることを主眼とする。

# 話す項目

- 自己紹介
- リスクを扱うための予測モデリングとは何か
- リスクを扱うための予測モデリングを学ぶ
- リスクを扱うための予測モデリングを創る
- おわりに

# リスクを扱うための予測モデリングとは何か アクチュアリーの視点

- 会のスローガン：

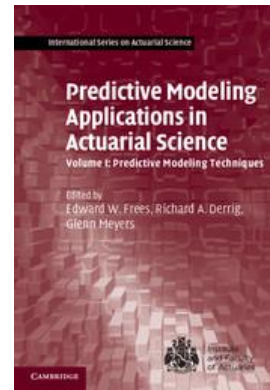
***Think the Future, Manage the Risk***

- 会の創設は1898年.
- アクチュアリー\*の扱うリスク：
  - 第1種：生保数理（17世紀～）
  - 第2種：損保数理，確率論的なリスク理論（20世紀初頭～）
  - 第3種：資産運用，ALM（1980年代～）
  - 第4種：統合的リスク管理（ERM）（21世紀初頭～）
  - 第5種：ビックデータ（2010年代～）

\* 職名としての「アクチュアリー」は19世紀から。アクチュアリーにあたる職業的専門家は18世紀から。基本となる保険数理の研究がはじまったのは17世紀から。損害保険の歴史は古いが、アクチュアリーが関わるようになったのは歴史上は比較的新しい。

# リスクを扱うための予測モデリングとは何か アクチュアリーへの予測モデリングへの注目

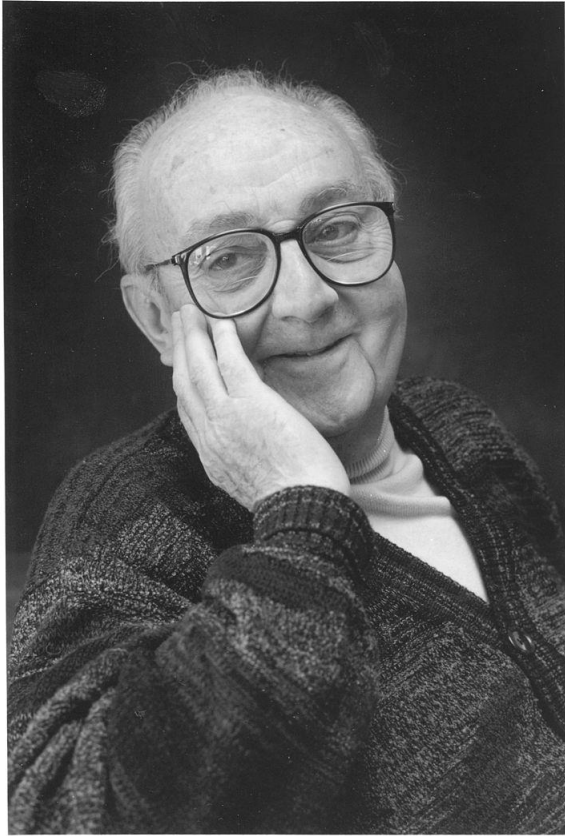
- 少なくとも北米のアクチュアリーたちはすでに2000年代から predictive modeling や predictive analytics というキーワードを使って、（「ビッグデータ」という名前が出る前から）ビッグデータ時代の統計モデリングを模索していた。
- 2014年には、Predictive Modeling Applications in Actuarial Science という教科書が発行された。



# リスクを扱うための予測モデリングとは何か 「予測モデリング文化」

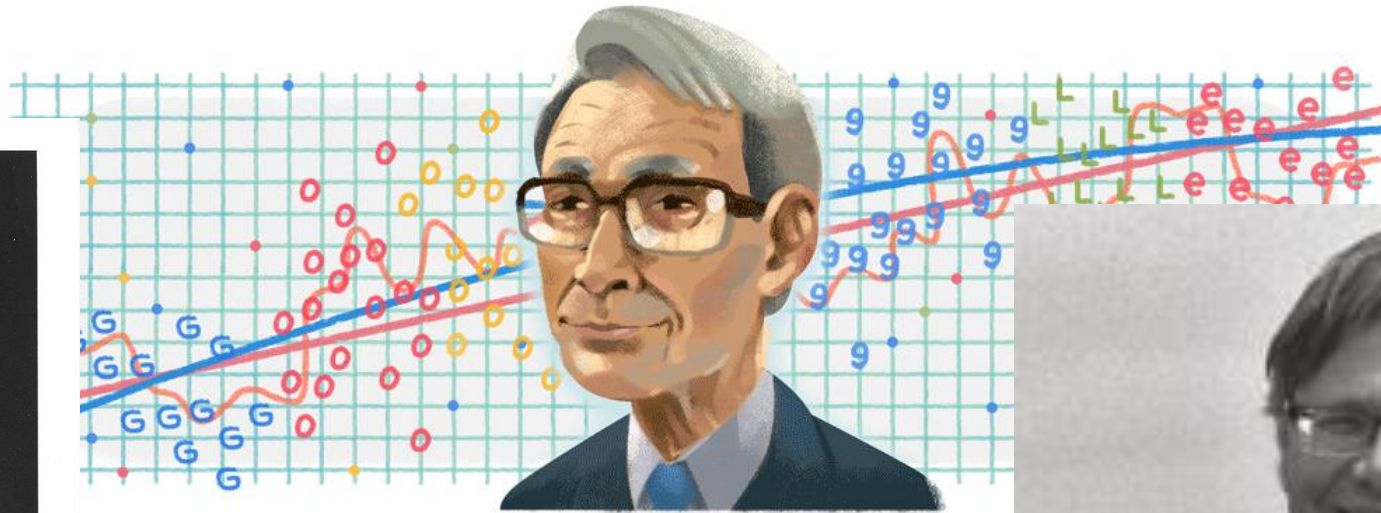
- 統計学者David Donoho (1957-)が2015年に行った有名な講演（→[2017年の論文](#)）で、近年の機械学習分野の大成功を「予測モデリング文化」というキーワードで説明。対立概念は「生成モデリング文化」。
- 鍵となるのは「予測の視点」。真のモデルは誰もわからないが、予測が当たったか外れたかならわかる。
- “All models are wrong, but some are useful.”---George Box (1919-2013)
- 「予測の視点」自体は、真新しいものでなく、赤池広次 (1927-2009) のAICの根本思想。





## George Box (1919-2013)

By DavidMCEddy at en.wikipedia, CC BY-SA 3.0,  
<https://commons.wikimedia.org/w/index.php?curid=14941622>



赤池広次 (1927-2009)  
Google's Doodle, Nov. 15, 2017



David Donoho (1957-)  
Stanford University

# リスクを扱うための予測モデリングとは何か

## 予測モデリングの特徴

- 予測モデリングの特徴をいくつか挙げれば、次のとおり。
  - A. 説明変数の候補（特徴量）が多くてもかまわない。
  - B. 複雑なモデルも採用できる。
  - C. 多くのモデル候補を同時に検討することもできる。
- 上記の特徴はもちろんどれもコンピュータの性能の向上によって支えられているが、理解にとってより重要なのは、従来の統計モデリングではあまり重視されていなかった発想（正則化、最適化、自動化）が大きく寄与している点。
- そしてそれらの発想を共通して支えているのが「予測の視点」。

# リスクを扱うための予測モデリングとは何か 生成モデリング vs. 予測モデリング

- 生成モデリングの典型的な内容
  - 注目すべき説明変数をあらかじめ厳選.
  - 真の母数の値さえわかれば（近似的には）真となると考えられるモデルを用意.
  - 確率的モデルが有効となるように的確に無作為化して実験実施.
  - 実験結果から母数を推測.

「実験計画的」

「真か偽か」

- 予測モデリングの典型的な内容
  - 有用かはわからない特徴量（説明変数の候補）がたくさんある.
  - 使えそうなモデルをいくつも試す.
  - モデルどうしが比較可能なように無作為化してモデル構築.
  - 選択したモデルで将来の観測値を予測.

「データ駆動的」

「当たるか外れるか」

# リスクを扱うための予測モデリングとは何か

## リスクを扱うための予測モデリング

- 予測モデリングは大成功しているのだから採用して然るべきだとしても、リスクを扱うには、これまで成功してきた課題と異なった観点が重要なため、予測モデリングを適用するためには、新たな分野を切り開く必要がある。

# リスクを扱うための予測モデリングとは何か リスクは「不確定」

- まず、リスクは「不確定」である点に注目。
- 機械学習分野で研究されてきた文字認識、画像認識、音声認識、機械翻訳等は、典型的には、人間には答えがわかるものを機械に学習させようというものであり、その予測の対象は（機械には最初はわからないものの）答え自体は確定しているという意味で、不確定ではない。
- 予測の対象がこの意味で不確定でない場合の課題を、以下では  
**パターン認識の課題**  
とよぶ。

# リスクを扱うための予測モデリングとは何か

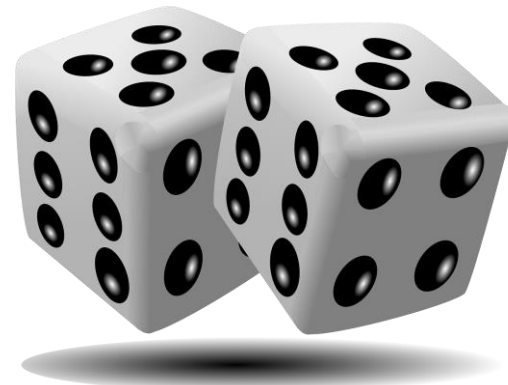
## リスクは「不確定」

- パターン認識の課題においては,
  - 異なった観測対象のデータどうしが完全に一致していれば, 正解が何であるかも一致する.
  - 与えられたデータが (何らかのうまい測定方法のもとで) 似通っていれば, 答えも似通っていると期待できる.
- これに対して, リスクに関する課題は, 異なった観測対象のデータどうしが完全に一致していても, 正解が何であるかが一致するとは限らない.

たとえば, 生命保険の加入者の属性データをもとに, ある契約者の翌年の保険金の期待値を的確に予測したとしても, 一般にはそのとおりの保険金になることはまずない. 実際には, ゼロか非常に大きな額かのいずれかとなり, 期待値とは一致しない.

1014007353  
8913312075  
8620236997  
8949213114  
9144263774  
7519022391  
2150634810  
3962647141

「8」なら「8」.  
予測モデリングの  
得意分野



まったく同質でも同じ目が出るわけではない。  
予測モデリングの対象外

では、  
保険金の予測は？

# リスクを扱うための予測モデリングとは何か

## リスクは「不確定」

- リスクのような不確定な事象に対しては，何らかの確率モデルを想定してモデリングするのが基本.
- ただし，確率モデルに基づかないモデルで，パターン認識の課題に対して強力なモデルが，リスクを扱うための予測モデリングにおいて無効とは限らない.
- 単純にいえば，たとえば次のようなことが起こる.
  - パターン認識に近い課題→ランダムフォレスト（後述）を単純にあてはめるだけで，GLM（後述）やその発展モデルよりもずっと高い予測力を示す.
  - 不確定の度合いが高い課題→ランダムフォレストを単純にあてはめるだけではうまくいかない（チューニングが難しい）.



# リスクを扱うための予測モデリングとは何か

## リスクは「避けたい」

- もう1つの観点は、リスクは「避けたい」ものだということ。
- 「予測が当たればうれしい」という面よりも、「予測が外れたら困る」という面が重要。
- マーケティングやレコメンデーションに予測モデリングを使う場合は…。
- リスクを扱う場合（たとえば保険会社は、各契約に関して保険金がどれくらい発生するかを予測する）、予測がぴたりと当たっても「大儲け」ということにはならないが、予測が外れた場合には、会社にとっては大打撃になるかもしれない。

# リスクを扱うための予測モデリングとは何か

## リスクを扱うための予測モデリングの要件

1. 確率モデルに基づいたモデリング手法が典型.
  2. 原理がわかりやすい.
  3. もとのデータと予測値との関係がわかりやすい.
  4. 頑健で安定的で高い再現性をもつ手法を用いる.
- このうちでも特に2と3に含意される「説明力の高さ」は最重要.
  - その意味では、リスクを扱うための予測モデリングとは、  
**予測の視点の下での説明力の高い統計モデリング**  
のこと.

# 話す項目

- 自己紹介
- リスクを扱うための予測モデリングとは何か
- リスクを扱うための予測モデリングを学ぶ
- リスクを扱うための予測モデリングを創る
- おわりに

# リスクを扱うための予測モデリングを学ぶ IAAシラバスと予測モデリング

[IAA Syllabus](#) (2016年に内容が実質的に確定, 2017年に採択)

## A. Supporting Learning Areas

1. Statistics
2. Economics
3. Finance
4. Financial Systems
5. Assets

## B. Core Learning Areas

6. **Data and Systems** ← ココ
7. Actuarial Models
8. Actuarial Risk Management
9. Personal and Actuarial Professional Practice

# リスクを扱うための予測モデリングを学ぶ IAAシラバス「6. Data and Systems」のねらい

To enable students to apply methods from statistics and computer science to real-world data sets in order to answer business and other questions, in particular with application to questions in long and short term insurance, social security, retirement benefits, healthcare and investment.

学習者が、統計学とコンピュータ科学に由来する手法を、現実のデータセットに適用することで、業務その他の問いに答えられるようにすること。その適用先には、特に、長短さまざまな契約期間の保険、社会保障、退職給付、ヘルスケア、投資といった分野における問いが含まれる。

# リスクを扱うための予測モデリングを学ぶ IAAシラバス「6. Data and Systems」の5つの節

6.1 DATA AS A RESOURCE FOR PROBLEM SOLVING

6.2 DATA ANALYSIS

6.3 STATISTICAL LEARNING

6.4 PROFESSIONAL AND RISK MANAGEMENT ISSUES

6.5 VISUALIZING DATA AND REPORTING

# リスクを扱うための予測モデリングを学ぶ ア会では

- 関連する講演（とりわけ2017年度ころから多数）を会員向けに不定期に実施.
- 定期的には、2019年度よりアクチュアリー専門講座「データサイエンス」を会員向けに3時間×12回で実施.
- 今後ますます充実していくもよう.

# リスクを扱うための予測モデリングを学ぶ おすすめの教科書類

- [Predictive Modeling Applications in Actuarial Science](#) 2巻ある。协会会员向けには無料で読める邦訳もある。
- [The Elements of Statistical Learning](#) 無料でも読める。邦訳『統計的学習の基礎』（もちろん？有料）もある。
- [Computer Age Statistical Inference](#) 無料でも読める。邦訳『大規模計算時代の統計推論』（同上）もある。
- [Computational Actuarial Science with R](#) 無料では読めない。
- [Interpretable Machine Learning](#) 無料で読める。邦訳もあって、それも無料で読める。
- [『入門 Rによる予測モデリング——機械学習を用いてリスク管理のために』](#) 手前味噌。有料。



# 話す項目

- 自己紹介
- リスクを扱うための予測モデリングとは何か
- リスクを扱うための予測モデリングを学ぶ
- リスクを扱うための予測モデリングを創る
- おわりに

# リスクを扱うための予測モデリングを創る 未整備の現状

- 今後、より本格的にリスクを扱うための予測モデリングを実践していくためには、現状はまだまだ未整備。
- 単純化していえば、現状は、次の2種類のものがある。
  1. 説明力はあるがビッグデータに対する予測力は十分でない手法
  2. ビッグデータに対する予測力は大いに期待できるが説明力が十分でない手法
- したがって、アプローチとしては、次の2つがある。
  - A) 説明力を維持しつつビッグデータに対する予測力も十分である手法を自ら開発する。
  - B) 予測力に期待できる手法に十分な説明力を加える方法論を確立する。

# リスクを扱うための予測モデリングを創る 新手法開発の例：AGLM

- ア会ASTIN関連研では，講演者が2017年に発案した手法AGLMを開発.
- Rパッケージaglm（いまではCRAN登録済）も構築.
- 同手法に関して2020年3月に執筆した次の論文が[2021年のHachemeister賞](#)を受賞.

[AGLM: A Hybrid Modeling Method of GLM and Data Science Techniques](#)

- 損保アクチュアリー分野での表彰だが，手法としては，生保などその他の分野でも十分に利用可能.

# リスクを扱うための予測モデリングを創る

## AGLMの受賞理由

- この論文で主張するハイブリッドモデリングアプローチは、データサイエンス技術による正確性の向上とGLMの優れた説明力の両方をバランスよく得たいという実際的な要求に応えようとしている。
- 自動車保険の例を用いて、既存のモデリング手法に対するAGLMの利点を定性的かつ定量的に評価している。
- 論文とともに統計ソフトウェアRのパッケージも用意しており、AGLMを簡単に使用できるように配慮している。
- 論文では自動車保険の価格設定の例を示しているが、提示される概念は他の損害保険分野、更にはその他の応用にも発展することが期待される。
- 全体として、委員会は、この論文はすべてのアクチュアリーにとって興味深いものであり、実用的な使用と独創的な思考を含み、大変優れた論文であると評価した。

# リスクを扱うための予測モデリングを創る 説明力向上のための方法論：説明力とは何か

- Interpretable Machine Learning, Explainable AI (XAI)
- 「わかりやすい」といってもいろいろな観点がある。
  1. 原理がわかりやすい。
  2. 特徴量（入力）と予測値（出力値）との関係がわかりやすい。
  3. リスクの特性がわかりやすい。
- 上記1は，方法論で解決するものではない。
- 上記2に資する方法論は，機械学習分野で広く開発されている。  
先に挙げた次の文献参照。 [Interpretable Machine Learning](#)
- 上記3に資する方法論は，アクチュアリー自らが開発すべき。

# リスクを扱うための予測モデリングを創る 予測誤差分解ツールの開発

- アクチュアリーは伝統的に、予測誤差をプロセス誤差\*，パラメータ誤差\*\*，その他の誤差\*\*\*の3つに分解してリスク管理に役立ててきたが，機械学習手法の場合に予測誤差を分解する方法は確立されていない。
- そこで，ア会DS基礎調査WGでは，機械学習手法にも適用可能な予測誤差分解ツールを開発中。
- 2020年3月にア会のジャーナルで報告し，2021年5月の国際大会でも発表。その後も引き続き内容を発展させており，今後も国際大会で発表していく予定。

\* 予測モデルが真だとしても存在する本質的な確率変動に起因する誤差

\*\* 学習データに限られるためにパラメータ等が真の値と乖離することによる誤差

\*\*\* 予測誤差のうちプロセス誤差とパラメータ誤差以外の誤差。モデル誤差ともいう

# 話す項目

- 自己紹介
- リスクを扱うための予測モデリングとは何か
- リスクを扱うための予測モデリングを学ぶ
- リスクを扱うための予測モデリングを創る
- おわりに

# おわりに

- ア会のスローガン（再掲）：

*Think the Future, Manage the Risk*

- リスク管理のための予測モデリングは、これからどんどん成長していく分野。

**ぜひ一緒に学びましょう**

**ぜひ一緒に活用しましょう**

さらには、

**ぜひ一緒に新たなものを創っていきましょう**