

# 確率的ビジネス分析

この10年間の機械学習とAIのブームは、コンピュータビジョン、自然言語処理、予測モデルなどの多くの分野で驚くべき成果をもたらし、ビジネスコンサルティングの重要な手法としてのデータサイエンスの到来を後押ししました。データサイエンスプロジェクトは、通常、大規模なデータセットの中で関連性のパターンを発見することによって予測の問題を解決する従来の機械学習法に焦点を当てています。例えば、過去に記録されたデータに基づいて製品の販売をどのように予測することができるか、といった質問に答えます。

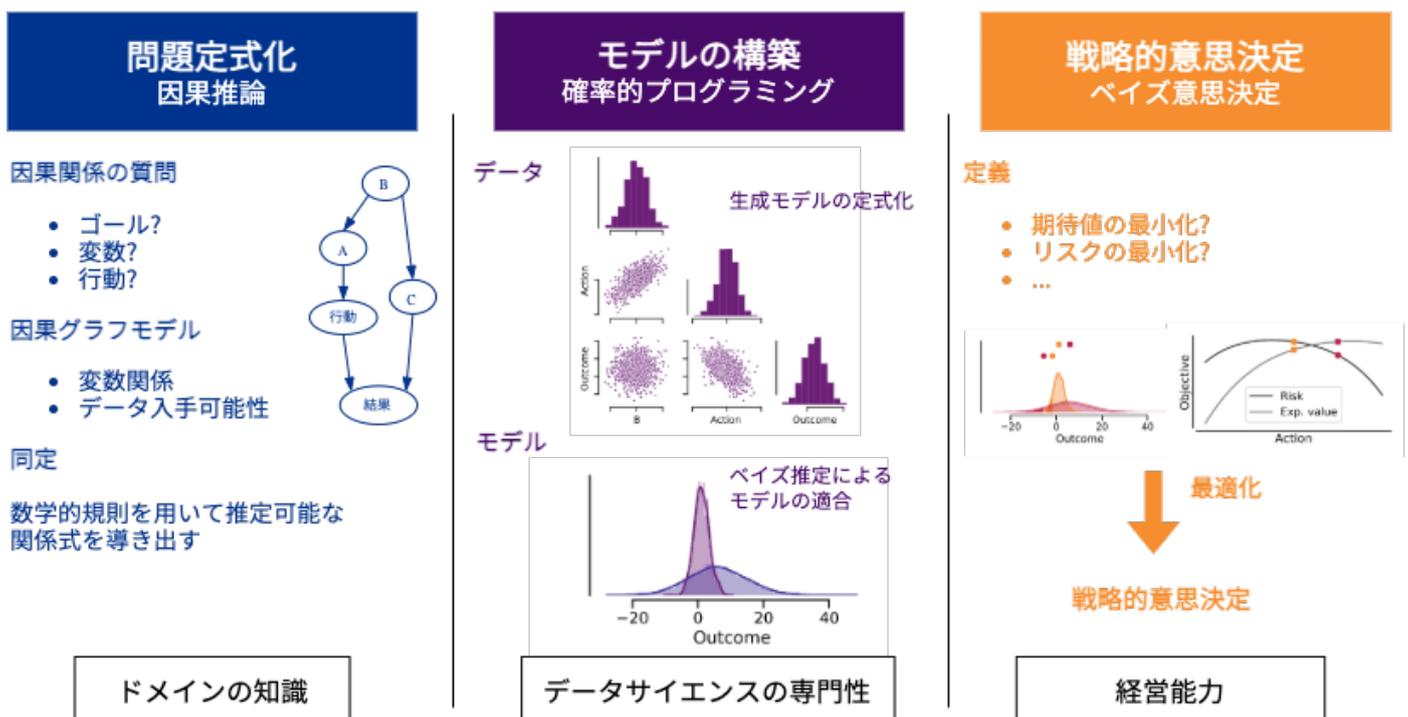
しかし、典型的なビジネスは、様々な課題を抱えています。例えば、価格設定（売り上げを伸ばすには製品の価格をどのように設定すればいいのか?）、プロセス最適化（どうすればサプライチェーンを最適化できるか?）、または顧客サービス（どうすれば顧客の減少を防ぐことができるか?）、といったことが少なくとも例として挙げられます。このような課題に答えるためには、行動（介入）と結果の間の因果効果を推定することによって従来の機械学習に勝る推測を行う、因果推論を用いる必要があります。

さらに、データサイエンスの応用においては、不確実性の影響はしばしば見落とされたり過小評価されたりします。データやビジネスプロセスの不確実性を考慮することは、分析の信頼性を向上させるだけでなく、意思決定に伴うリスクを調整する微妙な決断を可能にするためにも不可欠であると考えます。

本稿では、因果推論（因果効果）、確率的プログラミング（統計的モデリング）、およびベイズ意思決定（最適意思決定）を結びつける、原理に基づいたワークフローについて説明し、データサイエンス手法（図1）を用いたビジネス課題への取組みをご紹介します。ますます勢いを増すこれらの技術について説明していきます [1-5]。

## 原則に基づくワークフロー

ワークフローは3つのフェーズで構成されます。まず、因果推論を使用して問題を定義し、ドメインの専門知識を活用して問題内の変数間の因果関係について仮説を立てます。



これにより、機械学習を用いて推定可能な関係式が得られます。第2段階では、モデルに不確実性を組み込むことを可能にする手法である確率的プログラミングを用いて、この関係式を推定するモデルを構築します。最後のステップでは、ベイズ意思決定の中でこのモデルを用いて、クライアントと定義した目標に基づいた最適な戦略的意思決定に到達します。

## 因果推論による問題定式化

まず、解決すべき問題を定義します。システム内の変数は何か？、どのようなアクションをとることができるか？、つまりどの変数がクライアントによりコントロール可能か？、何をゴールとすべきか？ このゴールを、*interventional distribution* と呼ばれる特定の行動が選択された場合の結果の確率とします。次に、クライアントのドメイン専門知識を活用して、システム内の変数間の因果関係に関する仮説を立てます。どの変数がどの変数に影響するか？ こうした仮説は、因果グラフモデル (CGM) というグラフ表現に要約されます。

*interventional distribution* を直接推定するには、実験を行い、異なる行動の影響をテストする必要があります。これは通常、ビジネスのコンテキストでは不可能なので、最後のステップでは、*interventional distribution* を、観測された（過去の）データのみから推定できる観測分布に変換します。このステップは同定と呼ばれ、*do-calculus* を使って自動化されています。*do-calculus* は、CGM を使って変換が可能かどうかを調べ、可能であれば観測分布を導出する数学的規則の集合です。

## 確率的プログラミングによるモデル構築

この段階では、最初の段階で同定した表現を推定するモデルを構築します。

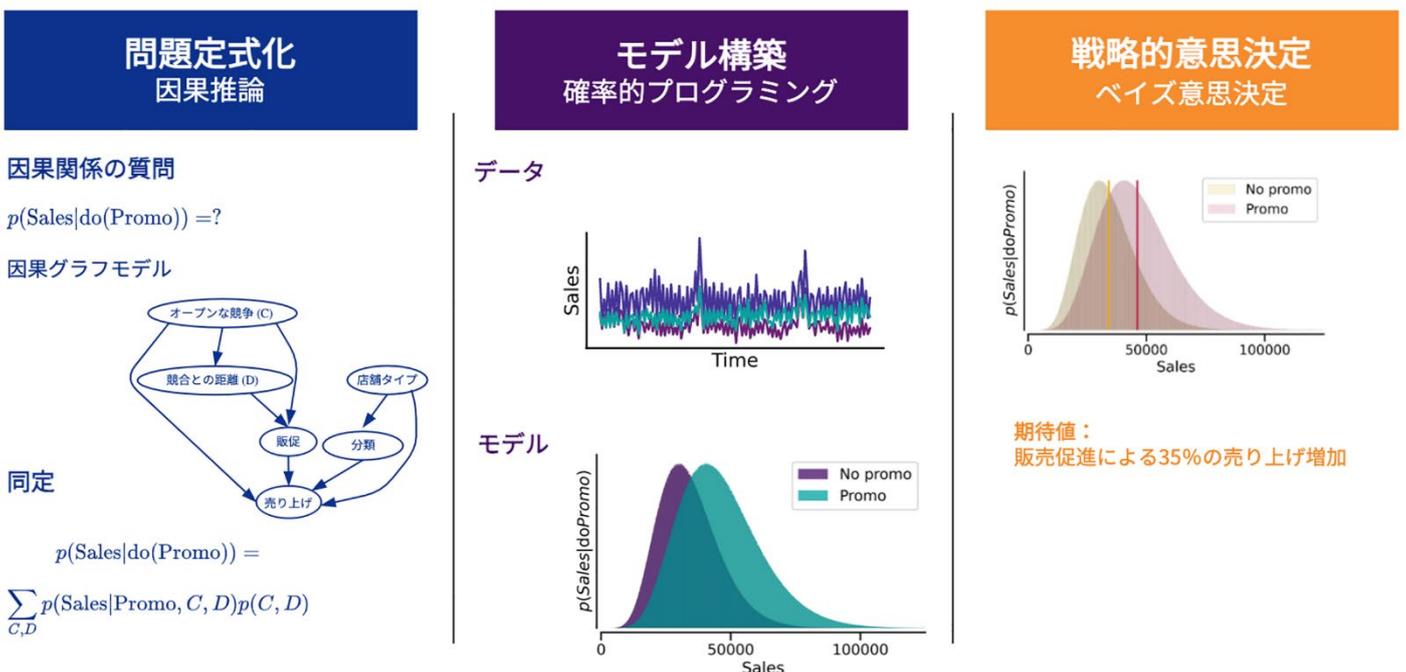
データがどのように生成されたかを想像し、これを確率的プログラムとして定式化します。（決定論的）コンピュータプログラムと同様に、アルゴリズムは出力値を計算します。しかし、変数は 確率論的な、実際の過程に関する不確実性とデータの不確実性を捕捉する確率分布から導かれます。結果のモデルは、最初の段階で定義した問題（特定の行動を選んだ場合の結果の確率はどれくらいか？）の推定値を導き出します。

## ベイズ意思決定による戦略的意思決定の導出

ビジネスの課題では、出力変数の目的関数の最適化のように、行動に対する最適な意思決定を求めます。目的関数は経営上の意思決定を反映しています。クライアントは、例えば期待される結果を最大化することにより、よりリスク許容度を高くするか、または不良な結果のリスクを最小化することにより、より保守的になるかを選択できます。ベイズ意思決定は、問題の不確実性を取り込み、選択された行動によって引き起こされる可能性のあるすべての結果を考慮するので、このように様々な選択肢を得ることができます。

## ビジネス分析には因果推論が欠かせない

因果推論は、変数間の真の因果効果を同定し定量化することを可能にします。従来の統計学や機械学習が連想関係に焦点を当て、「Yの値がわかっている時、Xについて何が言えるか？」のような疑問を呈するのに対し、因果推論は、「YはどのようにしてXを引き起こすのだろうか？」や、「Yを値に設定するとXはどうなるか？」といった質問を投げかけます。つまり、因果推論は行動の効果を推定するものであり、ビジネスの結果に影響を及ぼす意思決定を求めるビジネス分析において重要です。



観測データ中の誤った相関が全体像を歪めることがあるので、相関は因果関係ではないというのは統計学でよく言われる言葉です。これが真実である1つの状況は、行動と結果の間の相互作用が、両方に影響を及ぼす追加的な要因によって重ね合わされる場合です。ある企業が過去にその店舗で販促キャンペーンを行ったことがあり、その販促キャンペーンによって売上が増加したかどうかを知りたい場合を考えます（図2）。

ワークフローの最初の段階では、店舗に近い競合他社の存在が、過去にプロモーションを行うという決定に影響を与えていたことがわかります。近隣の店舗から競争圧力がある場合、販売促進を使用して売り上げ増加を試みる傾向があります。それと同時に、競争の激しい店舗は、他の追従を許さない店舗よりも売り上げが低いのかもかもしれません。

このように、競合他社の存在は、販売促進キャンペーンを実施するかどうかの決定と販売の両方に影響を与えます。したがって、販売促進キャンペーンの販売に対する真の因果効果を見つけるには、販売に対する販促の因果効果の偏りのない視点を得るために、分析において競合の存在を調整する必要があります。問題を体系的に定義することにより、競争が分析に与える潜在的なバイアス効果が明確で扱いやすくなります。

因果推論では、闇雲に実験を行うのではなく、観測されたデータを使って推定できる値を特定します。このワークフローでは、変数とその関係を明示的に定義し、分析とクライアントとのコミュニケーションの両方を容易にします。

### 確率論的モデリングはより多くの情報に基づいた意思決定を可能にする

ビジネスは不確実な世界に置かれています。顧客の需要は時間の経過とともに変化し、取引はリスクを伴い（商品に欠陥がある、サプライチェーンが不安定など）、世界で起きている様々な要因（世界全体の経済状況、自然災害やパンデミックのよう

な壊滅的な出来事）が大きな影響を及ぼす可能性があります。したがって、ビジネス問題をモデル化するには、このような不確実性を考慮する必要があります。

確率的プログラミングのフレームワークは、単純かつ体系的な方法でこれを行うことを可能にします。それは、問題における行動のための異なる選択に基づいた、起こり得る結果のための確率的予測を叩き出します。これにより、期待される結果を最適化するのか、それともより保守的になり、リスクを抑えて望ましくない結果が生じる可能性を最小限にしたいのか、といったように、クライアントとの間で目標を柔軟に定義できるようになります。

サプライチェーンを運営するために最適な台数の車両を選択するという問題について考えてみましょう（図3）。片方では、クライアントは将来の顧客の需要を満たすために十分な台数を確保したいと考えています。その一方で、メンテナンスコストの最小化にも努めています。これら2つの相反する利害のバランスをとるには、顧客の需要を十分に予測して、一定数の車両が存在する場合のコストの可能性を予測する必要があります。モデルを構築してコストの見積もりを割り出した後、クライアントは、コストの期待値を最小化することによってリスク耐性を高めるか、または高いコストに直面するリスクを最小化して保守的になるか、のいずれかを選択できます。この例では、3台の車を選択すると、メンテナンスコストを節約することによってコストの期待値が低くなり、4台の車を使用すると、台数の増加によって顧客の需要を満たす可能性が高くなるため、リスクが低くなります。

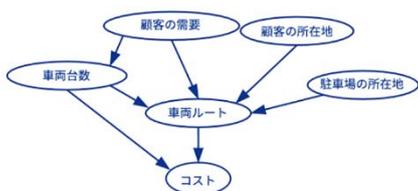
これらは、私たちのアプローチが可能にする戦略的な選択です。将来のためにたった1つの起こり得るシナリオに焦点を当てるのではなく、すべての可能な未来（と関連する確率）を考慮することによって、私たちは、これらの選択を明確にし、クライアントにより多くの情報に基づいた見解を提供することができます。

## 問題定式化 因果推論

### 因果関係の質問

$$p(\text{Costs}|\text{do}(\text{VehicleNumber}))=?$$

### 因果グラフモデル

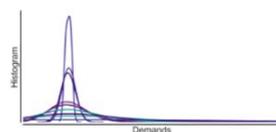


### 同定

$$p(\text{Costs}|\text{do}(\text{VehicleNumber})) = \int_D p(\text{Costs}|\text{VehicleNumber}, D)p(D)$$

## モデル構築 確率的プログラミング

### データ

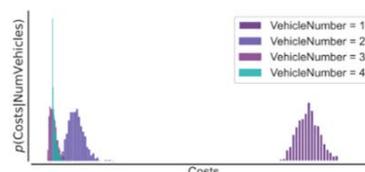


コスト = 車両の固定数 + ペナルティ料金

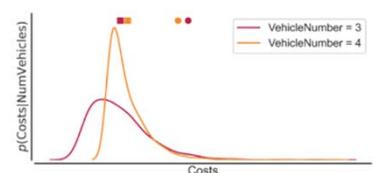
より多くの車両 = より高い固定費

より少ない車両 = より高いペナルティ料金

### モデル



## 戦略的意思決定 ベイズ意思決定



最適な車両台数は？

3台：より低いコスト期待値

4台：高コストになるリスクがより低い

## 結語

複雑なビジネス問題に取り組むための、因果推論、確率的プログラミング、およびベイズ意思決定の方法を結びつける原理に基づいたワークフローを紹介しました。このワークフローにより、アクションが結果に及ぼす因果関係の影響の特定、将来の展開に関する柔軟で正確な予測の作成、それらの予測に基づいた戦略的な意思決定、といった最も重要な課題に対処します。

## 用語集

**ベイズ意思決定：** 事前の知識とデータ駆動型モデルの両方の情報によって与えられる好い結果が起こる可能性に基づいて意思決定を行うプロセス。

**確率的プログラミング：** データをコンピュータプログラムとして定式化することにより、データの生成モデルを構築する方法。

**推定式：** 観測データのみから推定できる、すなわち、能動的な実験を行う必要のない数式。

**目的：** 「売上高の期待値を最大化したい」などの大まかな目的を数学的に定式化したもの。

**観測分布：** 他の変数の値が観測された場合の事象の確率を定量化する確率分布。

**interventional distribution：** 他の変数に介入した場合の事象の確率を定量化する確率分布で、積極的に望ましい値に設定。

**因果グラフモデル：** 変数間の因果関係をグラフ化したもの。変数はノードであり、因果関係は変数間の矢印で表される。

## 参考文献

[1] Silvia Chiappa and William Isaac. Causal bayesian networks: A flexible tool to enable fairer machine learning, 2019.

URL: [https://deepmind.com/blog/article/Causal\\_Bayesian\\_Networks](https://deepmind.com/blog/article/Causal_Bayesian_Networks)

[2] Totte Harinen and Li Bonnie. Using causal inference to improve the uber user experience.

URL: <https://Eng.Uber.Com/Causal-Inference-At-Uber/>

[3] Paul Hünermund, Jermain Kaminski, and Carla Schmitt. Causal machine learning and business decision making, 2021. URL: <https://causalscience.org/blog/causal-data-science-in-practice>

[4] Cloudera Fast Forward Labs. Causality for machine learning, 2020.

URL: <https://ff13.fastforwardlabs.com/>

[5] Jeffrey Wong and Colin McFarland. Computational causal inference at Netflix, 2020.

URL: <https://netflixtechblog.com/computational-causal-inference-at-netflix-293591691c62>

---

## 連絡先

KPMG Ignition Tokyo

〒100-0004

東京都千代田区大手町 1-6-1

大手町ビル7F

マクシミリアン・シュミット

シニアエンジニア

DXソリューション2

+81 70 1043 7493

maximilian.schmidt@jp.kpmg.com

宮部俊吾

シニア・サイエンティスト

スマートトランザクション

[home.kpmg/jp/KIT](https://home.kpmg/jp/KIT)

---

ここに記載されている情報はあくまで一般的なものであり、特定の個人や組織が置かれている状況に対応するものではありません。私たちは、的確な情報をタイムリーに提供できるよう努めておりますが、情報を受け取られた時点及びそれ以降においての正確さは保証の限りではありません。何らかの行動を取られる場合は、ここにある情報のみを根拠とせず、プロフェッショナルが特定の状況を綿密に調査した上で提案する適切なアドバイスをもとにご判断ください。

© 2021 KPMG Ignition Tokyo, Inc., a company established under the Japan Companies Act and a member firm of the KPMG global organization of independent member firms affiliated with KPMG International Limited, a private English company limited by guarantee. All rights reserved. 21-1071

The KPMG name and logo are trademarks used under license by the independent member firms of the KPMG global organization.