# 2024 年卒業論文

教師あり学習を用いた放射光メスバウアー実験における実験継続性の評価

弘前大学理工学部数物科学科

学籍番号:20S1066 名前:坂西 和也

2024年2月20日

# 目次

第1章	はじめに	3
1.1	現状と問題点....................................	3
1.2	本研究における成果	3
1.3	プログラミングコードについて	3
第2章	実験概要	4
第3章	実験データの観察と作成	5
3.1	教師信号を持つデータ作成	5
3.2	期待値データの倍率変化に関する影響	6
3.3	非教師信号を持つデータ作成	8
第4章	モデルの説明	11
4.1	教師あり学習の概要	11
4.2	ロジスティック回帰	12
4.3	ニューラルネットワーク	13
4.4	畳み込みニューラルネットワーク	15
第5章	学習結果	16
第6章	今後の課題	18
付録 A	ポアソン分布の性質	19
A.1	ポアソン分布の確率質量関数	19
A.2	ポアソン分布の期待値	19
A.3	ポアソン分布の分散	20

付録 B	吸収が観測されないデータに対して行った統計的な方法に関する説明	22
B.1	Q-Q プロット	22
B.2	シャピロ・ウィルク検定	23
付録 C	ロジスティック回帰に関する計算	25
C.1	シグモイド関数	25
C.2	ベルヌーイ分布の最尤法	25
参考文献		28

## 第1章

# はじめに

### 1.1 現状と問題点

メスバウアー実験では、特定の吸収ピークの検出が重要な役割を果たすが、これには検出器を用いて長時間にわたり観測する必要がある複雑な過程だ。この吸収の発生は確率的な事象で、多くの時間を費やしても所望の吸収が得られない可能性がある。したがって、実験の成功は実験者の判断に大きく依存し、吸収が正常に観測されているか否かの判断を誤ると、実験時間が無駄になることがある。また、誤った判断により実験を早期に中断してしまうこともある。

## 1.2 本研究における成果

本研究では、上述の問題を解決するために、機械学習の教師あり学習モデルを用いて、 実験者の判断よりも迅速に正常な吸収の有無を判定する方法を検討した.具体的には、教 師データとして理論値に基づく吸収データに統計的な誤差を加えたデータを作成し、非教 師データとして、教師データ作成時と同様の統計誤差を持つ完全にランダムな正規乱数 を生成した.これらのデータを用いて、3種類の異なる機械学習モデルの性能を比較分析 した.

## 1.3 プログラミングコードについて

本研究に関するプログラミングコードは以下の GitHub ページに記載.

https://github.com/LABO-M/BANZAI.git

# 第2章

# 実験概要

メスバウアー実験の概要を参考文献 [1] を参考に説明する. 放射光メスバウアー分光法の主要な構成要素には放射光源,結晶分光器,測定試料,エネルギー基準物質,検出器,速度トランスデューサーが含まれる. それらの関係は図 2.1 に示す. 放射光は,シリコン結晶の結晶分光器を通過し,広い帯域幅のエネルギープロファイルを持つ. 本研究で想定するデータは,測定試料とエネルギー基準物質共にユウロピウムフッ化物 (EuF3) を使用したデータとなる. 図 2.1 に関して放射光は測定試料を透過した後,エネルギー基準物質に入射し,その散乱は大きな立体角をカバーする検出器で検出される. 速度トランスデューサーは光のドップラー効果を利用して核共鳴エネルギーを制御する. 今回想定するデータはこの検出器により測定されるデータとなる.

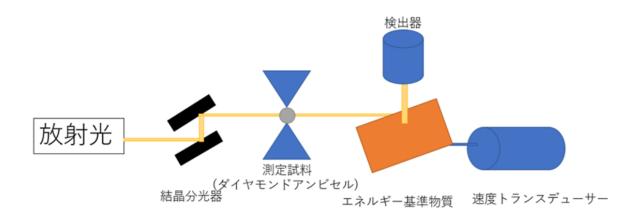


図 2.1 実験装置概略図 (2023 年卒業研究村上海斗氏の図を一部修正)

# 第3章

# 実験データの観察と作成

## 3.1 教師信号を持つデータ作成

教師信号として用いるデータの作成について説明する.

検出器から観測される期待データの形状は図 3.1 である.これは参考文献 [1] の  $\mathrm{Fig.}2.7$  にカウント数を付加したものとなっている.

この期待データの元になるのは参考文献 [1] より以下の式に関するもので

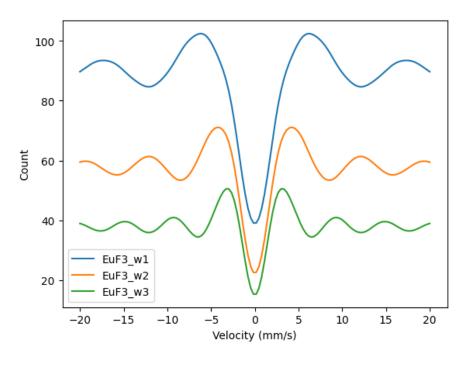


図 3.1 期待値データに関する図

$$I(\omega_s) = I_A(\omega_s) + I_C(\omega_s) + I_B$$

ただし,

$$I_A(\omega_s) = C_A \int_{\tau_1}^{\tau_2} d\tau \int_0^{z_s} dz \left| \int \frac{d\omega}{2\pi} \exp(-i\omega\tau) E_A(\omega, \omega_s, z) \right|^2$$
$$I_C(\omega_s) = C_C \int_{\tau_s}^{\tau_2} d\tau \int_0^{z_s} dz \left| \int \frac{d\omega}{2\pi} \exp(-i\omega\tau) E_C(\omega, \omega_s, z) \right|^2$$

であり、 $I_B$  はその他の  $\omega_s$  に独立な過程である.図 3.1 の 3 本の曲線の違いは  $\tau_1$  の違いである.

実験により観測される値は統計誤差を含んでいる。その誤差は期待値カウント数の平均と分散にもつポアソン分布であることが知られている。シミュレーションでは、その統計誤差の形状を平均と分散が等しい正規分布と近似してシミュレーションを行うこととする。期待データに統計誤差を付加した様子を図 3.2 に示した。

### 3.2 期待値データの倍率変化に関する影響

期待値データを c 倍した影響について説明する. 重要な点は、データを c 倍に圧縮してもノイズの標準偏差が c 倍にならないということである.

具体的に、Y を元データに乱数を加えたデータ X を元データとし、次のように仮定する:

$$Y|X \sim \mathcal{N}(X,X)$$

ここで、 $\sigma = \sqrt{Var[Y|X]} = \sqrt{X}$  となる.

次に X を cX に置き換えると,次のようになる:

$$\sqrt{Var[Y|cX]} = \sqrt{cX} \approx \sqrt{c}\sigma$$

この結果から,元データを c 倍に圧縮すると,ノイズの標準偏差は約  $\sqrt{c}$  倍になり,乱数の影響が大きくなることが分かる.期待データを定数倍してから統計誤差を付加した様子を図 3.3 に示した.吸収率が倍率変化させる定数項が小さくなるにつれて小さくなることが確認される.

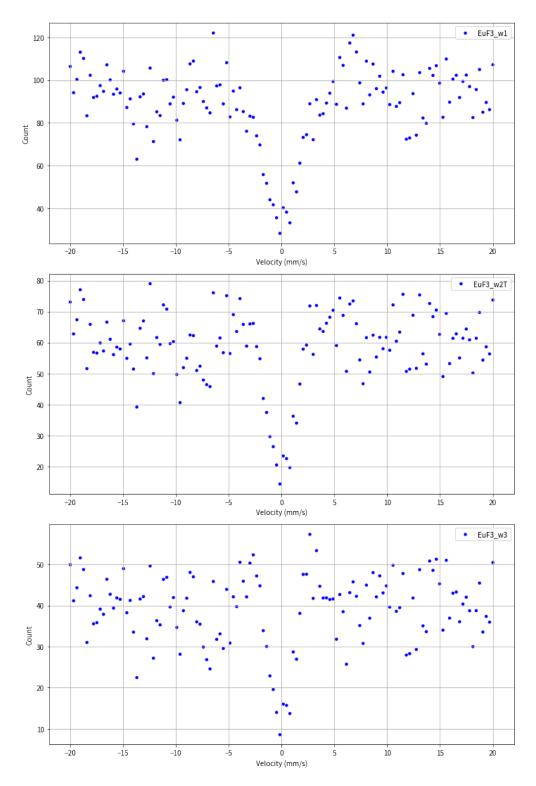


図 3.2 期待値データにノイズを付加した様子

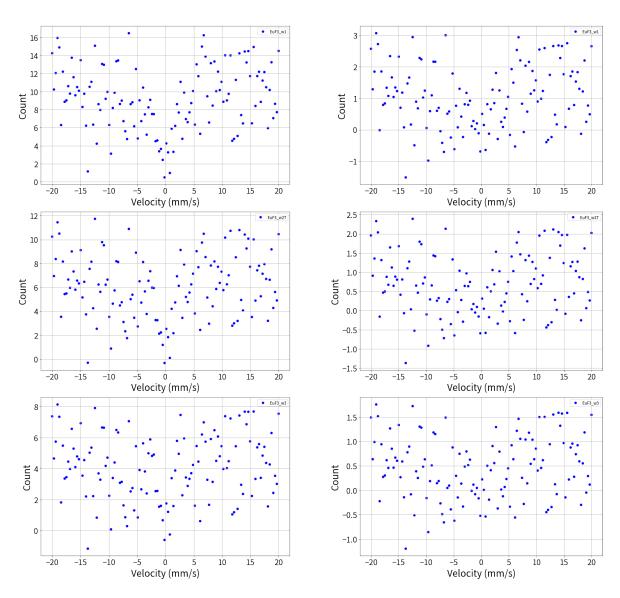


図 3.3 左:0.1 倍した様子、右:0.01 倍した様子

## 3.3 非教師信号を持つデータ作成

実験を継続しても吸収が観測できないデータは完全な正規乱数と仮定する.実験から観測される吸収が観測されないデータの様子を図 3.4 に示した. その確認のために実験から得られたデータに対してそれぞれ正規乱数か否かの統計的仮設検定を行う.

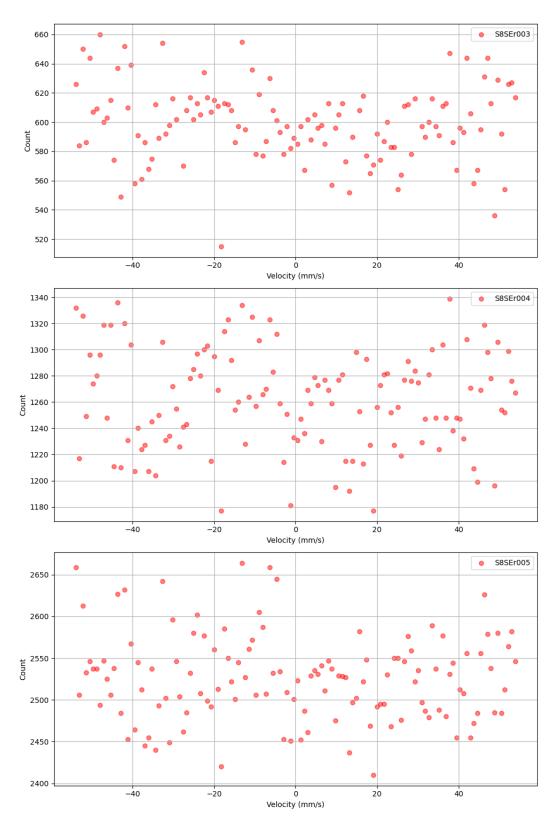


図 3.4 実際に吸収が観測されない実験データ

まず視覚的な検証として,ヒストグラム及び Q-Q プロットを行う.その様子を図 3.5 に示す.

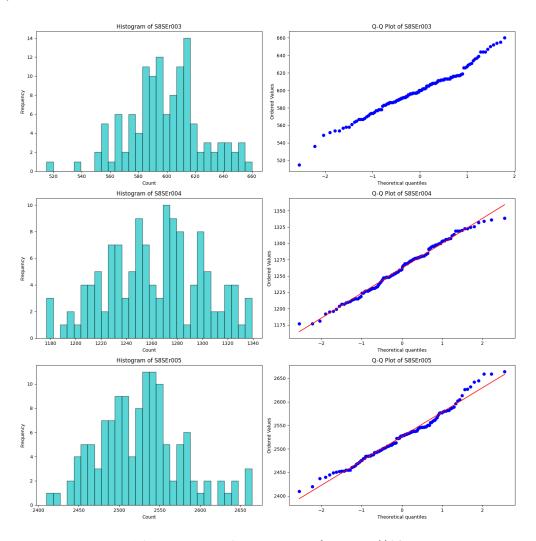


図 3.5 ヒストグラムと Q-Q プロットの結果

次にシャピロ・ウィルク検定を行う. 結果は 3.6 となり、それぞれ正規乱数であることが統計的に有意であることが示された.

```
{'S8SEr003': {'p-value': 1.0},
    'S8SEr004': {'p-value': 0.27501988410949707},
    'S8SEr005': {'p-value': 0.06507931649684906}}
```

図 3.6 検定結果

## 第4章

# モデルの説明

### 4.1 教師あり学習の概要

教師あり学習は、機械学習の一分野であり、モデルがラベル付きトレーニングデータから学習する方法である。この学習プロセスでは、アルゴリズムは入力データとそれに対応する正解ラベルのセットを使用して、データの特徴とラベル間の関係を学習する。学習が完了すると、アルゴリズムは新しい、未知のデータに対して正確な予測や分類を行うことが期待される。

教師あり学習におけるプロセスは以下のようなものが主流である.

- 1. データの収集と前処理: データを収集し, クリーニング, 正規化, 特徴選択などの前処理を行う.
- 2. モデルの選択:分類や回帰などのタスクに適したモデルを選択する.
- 3. トレーニング: ラベル付きトレーニングデータを使用してモデルをトレーニング する.
- 4. 評価:テストデータセットを使用してモデルの性能を評価する.
- 5. チューニング:必要に応じてモデルを調整し、パフォーマンスを最適化する.

本研究は、分類タスクに当たる.以下では、本研究で使用するモデルの説明を行う.これらモデルは、分類タスクにおける教師あり学習の基本的なモデルでもある. GitHub サイトには参考文献 [2] 及び [3] を参考にサンプルコードを載せてある. また理論的側面については参考文献 [4] 及び [5] を参考にしている.

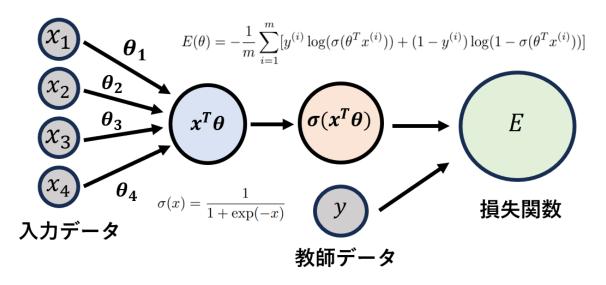


図 4.1 ロジスティック回帰のイメージ

#### 4.2 ロジスティック回帰

ロジスティック回帰は、分類問題に適用される教師あり学習アルゴリズムの一つである。特に、二項分類(目的変数が2つのクラスのいずれかに属する場合)に使用される。このアルゴリズムは、入力特徴とロジスティック関数(シグモイド関数)を使用して、出力が特定のクラスに属する確率をモデル化する。計算に関する基本的な知識は付録に記載してある。このロジスティック回帰が最も基本的なモデルとなる。

### ロジスティック回帰の数学的モデル

ロジスティック回帰モデルは,線形回帰モデルとシグモイド関数を組み合わせて構成される.線形回帰モデル  $\theta^T x$  の出力をシグモイド関数に通すことで,クラス 1 に属する確率を表現する.数式は以下のようになる:

$$p(y=1|x;\theta) = \sigma(\theta^T x) = \frac{1}{1 + e^{-\theta^T x}}$$

ここで、x は入力特徴、 $\theta$  はモデルのパラメータ、 $\sigma$  はシグモイド関数を表す.

#### コスト関数と最適化

ロジスティック回帰モデルの学習には、コスト関数(損失関数)が使用される. このコスト関数は、モデルの予測が実際のラベルとどれだけ異なるかを測定する. 一般的なコスト関数は以下のようになる:

$$E(\theta) = -\frac{1}{m} \sum_{i=1}^{m} [y^{(i)} \log(\sigma(\theta^T x^{(i)})) + (1 - y^{(i)}) \log(1 - \sigma(\theta^T x^{(i)}))]$$

ここで,m はトレーニングセットのサンプル数, $y^{(i)}$  は i 番目のサンプルのラベルを表す.

最適化は、コスト関数を最小化するパラメータ  $\theta$  を見つけるプロセスである。これは通常、勾配降下法やその他の最適化アルゴリズムを使用して行われる。

### 4.3 ニューラルネットワーク

#### 4.3.1 フォワードプロパゲーション

入力と重みの総和:

$$z^{(l)} = W^{(l)}a^{(l-1)} + b^{(l)}$$

ここで

- l は層の番号
- W<sup>(l)</sup> は l 層の重み行列
- $a^{(l-1)}$  は l-1 層の活性化
- $b^{(l)}$  は l 層のバイアスベクトル

活性化関数を適用:

$$a^{(l)} = f(z^{(l)})$$

ここで、f は活性化関数 (例: Sigmoid, ReLU など)

## 4.3.2 バックプロパゲーション

出力層の誤差:

$$\delta^{(L)} = (a^{(L)} - y) \odot f'(z^{(L)})$$

#### ここで

- L は最後の層
- y は正解ラベル
- f' は活性化関数の導関数
- ⊙ は要素ごとの積

隠れ層の誤差:

$$\delta^{(l)} = (W^{(l+1)T}\delta^{(l+1)}) \odot f'(z^{(l)})$$

重みとバイアスの勾配:

$$\frac{\partial E}{\partial W^{(l)}} = \delta^{(l)} a^{(l-1)T}$$
 
$$\frac{\partial E}{\partial b^{(l)}} = \delta^{(l)}$$

ここで、E は損失関数 (例:二乗誤差, 交差エントロピーなど)

回帰タスクには損失関数を二乗誤差とし、分類タスクには交差エントロピーを用いることが多い. 理論的な側面では、ロジスティック回帰と同様に最尤法における最適化と同様である.

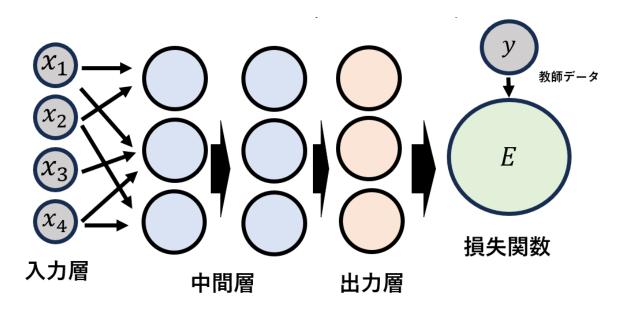


図 4.2 ニューラルネットワークのイメージ

### 4.4 畳み込みニューラルネットワーク

#### フォワードプロパゲーション

#### 畳み込み層

畳み込み層では、フィルタ(カーネル)を入力特徴マップに適用して特徴を抽出する. フィルタは入力データ上を滑らかに移動し、各位置での入力とフィルタの要素間の積の合計を計算する.これにより、出力特徴マップが生成される.数式は以下のように表される:

$$O_{xy} = \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} I_{(x+i)(y+j)} K_{ij}$$

ここで、 $O_{xy}$  は出力特徴マップの x,y 位置の値、I は入力特徴マップ、K はフィルタ、m,n はフィルタのサイズを表す.

#### プーリング層

プーリング層では、入力特徴マップのサブリージョンを単一の値(通常は最大値または 平均値)に縮小し、特徴マップの次元を減少させる.これは、モデルの計算量を減らし、 過学習を防ぐのに役立つ.

#### バックプロパゲーション

バックプロパゲーションでは、ネットワークの出力と期待される出力との間の誤差を計算し、この誤差をネットワークを通じて逆方向に伝播させ、各層の重みを更新する. 畳み込み層のフィルタの重み更新は以下のようになる:

$$\frac{\partial E}{\partial K_{ij}} = \sum_{x=0}^{X-1} \sum_{y=0}^{Y-1} I_{(x+i)(y+j)} \frac{\partial E}{\partial O_{xy}}$$

ここで,E は誤差関数, $\frac{\partial E}{\partial O_{xy}}$  は出力特徴マップの x,y 位置の誤差に対する偏微分,I は入力特徴マップ, $K_{ij}$  はフィルタの重みを表す.

# 第5章

# 学習結果

作成したデータに対して、ロジスティック回帰及びニューラルネットワーク、畳み込みニューラルネットワークを各定数倍に対して学習を行い、テストデータに対する正解率の推移を評価したグラフを図 5.1, 図 5.2, 図 5.3 にそれぞれ示す.

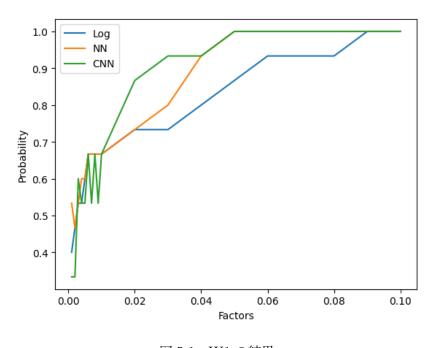


図 5.1 W1 の結果

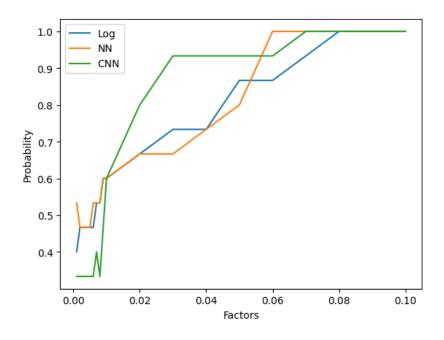


図 5.2 W2 の結果

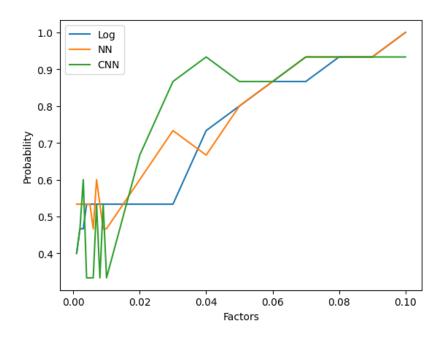


図 5.3 W3 の結果

# 第6章

# 今後の課題

本研究では、ロジスティック回帰及びニューラルネットワーク、畳み込みニューラルネットワークの3つのモデルについて、シミュレーションによりカウント数が少ない、つまり実験時間が短い場合でも、吸収が起こるか否かの判断が可能かについて評価を行った.

しかし、シミュレーションと実際の実験データとの間にはしばしば乖離が存在する. 実際の実験データでは、吸収位置や吸収の大きさが毎回異なるため、シミュレーションモデルをこれらの実データに適用し、その有効性を検証する必要がある.

今後の課題として、以下の点が挙げられる:

- 1. 実験データの統合と分析: 実際の実験データを収集し、それを既存のシミュレーションモデルに統合することで、モデルが実世界の複雑さをどの程度捉えているかを評価する.
- 2. データの多様性への対応: 実験データの吸収位置や大きさのばらつきを考慮に入れ、これらの変動が結果に与える影響を分析する. モデルのロバスト性を高めるための改善策を検討する.
- 3. **シミュレーションと実データの比較:** シミュレーション結果と実験データとを比較し、両者の一致度を評価する. 不一致が見られる場合は、モデルの改善点を特定する.

これらの課題に取り組むことで、モデルの現実への適用性を高め、より信頼性のある結果を得ることが期待される.

## 付録 A

# ポアソン分布の性質

### A.1 ポアソン分布の確率質量関数

以下参考文献 [6] を参考に統計的な計算に関する説明を付録として付け加えておく. ポアソン分布の確率質量関数は以下のように与えられる.

$$P(X = k) = \frac{e^{-\lambda} \lambda^k}{k!}$$

ここで、X はポアソン分布に従う確率変数、k は非負の整数、 $\lambda$  は平均および分散を表すパラメータ.

## A.2 ポアソン分布の期待値

期待値 E[X] は以下のように計算される

$$E[X] = \sum_{k=0}^{\infty} k \cdot P(X = k)$$
$$= \sum_{k=0}^{\infty} k \cdot \frac{e^{-\lambda} \lambda^k}{k!}$$

この式を展開し、 $k! = k \cdot (k-1)!$  という事実を利用して

$$E[X] = \sum_{k=1}^{\infty} \frac{e^{-\lambda} \lambda^k}{(k-1)!}$$

$$= \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!}$$

ここで, $\sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!}$  は  $e^{\lambda}$  のテイラー級数展開に他ならない. したがって

$$E[X] = \lambda e^{-\lambda} e^{\lambda}$$
$$= \lambda$$

以上より、ポアソン分布の期待値はパラメータ  $\lambda$  に等しい.

### A.3 ポアソン分布の分散

ポアソン分布の分散 Var(X) は期待値と同様に計算できる.

$$Var(X) = E[X^2] - (E[X])^2$$

まず、 $E[X^2]$ を計算する.

$$E[X^{2}] = \sum_{k=0}^{\infty} k^{2} \cdot P(X = k)$$
$$= \sum_{k=0}^{\infty} k^{2} \cdot \frac{e^{-\lambda} \lambda^{k}}{k!}$$

ここで、 $k^2 = k \cdot (k-1) + k$  と分解できることを利用する

$$E[X^{2}] = \sum_{k=1}^{\infty} (k \cdot (k-1) + k) \cdot \frac{e^{-\lambda} \lambda^{k}}{k!}$$

$$= \sum_{k=1}^{\infty} k \cdot (k-1) \cdot \frac{e^{-\lambda} \lambda^{k}}{k!} + \sum_{k=1}^{\infty} k \cdot \frac{e^{-\lambda} \lambda^{k}}{k!}$$

$$= \sum_{k=2}^{\infty} \frac{e^{-\lambda} \lambda^{k}}{(k-2)!} + \lambda$$

ここで、最初の項は $\lambda^2$ を因数として外に出すことができる

$$\begin{split} E[X^2] &= \lambda^2 \sum_{k=2}^{\infty} \frac{e^{-\lambda} \lambda^{k-2}}{(k-2)!} + \lambda \\ &= \lambda^2 e^{-\lambda} \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} + \lambda \\ &= \lambda^2 e^{-\lambda} e^{\lambda} + \lambda \\ &= \lambda^2 + \lambda \end{split}$$

既に知っている  $E[X] = \lambda$  を使って分散を求める

$$Var(X) = E[X^{2}] - (E[X])^{2}$$
$$= (\lambda^{2} + \lambda) - \lambda^{2}$$
$$= \lambda$$

以上より、ポアソン分布の分散もパラメータ λ に等しい.

## 付録 B

# 吸収が観測されないデータに対して 行った統計的な方法に関する説明

### B.1 Q-Q プロット

Q-Q (Quantile-Quantile) プロットは、二つの確率分布が同じであるかどうかを視覚的に評価するために使用されるグラフィカルツールだ.一般的には、サンプルデータが特定の理論分布 (例えば正規分布) に従っているかどうかを評価するのに用いられる.

### Q-Q プロットの作成方法

Q-Q プロットを作成するためには、以下の手順を行う:

- 1. データの準備: 分析するサンプルデータセットを用意する.
- 2. **分位数の計算:** サンプルデータセットの分位数(または順序統計量)を計算する. これは、データセットを昇順に並べた際の各値が占める位置だ.
- 3. 理論分布の分位数との比較: 各サンプルデータ点に対して, 理論分布の同じ分位数を計算する.
- 4. プロット: サンプルデータの分位数を横軸に, 理論分布の分位数を縦軸にプロットする.

サンプルのi番目のデータ点に対する分位数は以下のように計算される:

$$q_i = G^{-1} \left( \frac{i - 0.5}{n} \right)$$

ここで, $G^{-1}$  は理論分布の逆累積分布関数(quantile function),n はサンプルサイズだ.i はデータ点の順位(1 から n まで)だ.

#### Q-Q プロットの解釈

- **直線上にデータ点が配置されている場合**: サンプルデータは理論分布に従っていると考えられる.
- **直線から逸脱しているデータ点がある場合**: サンプルデータが理論分布から逸脱している可能性がある. 逸脱のパターンは, データが理論分布とどのように異なるかについての手がかりを提供する.

Q-Q プロットは, データが特定の分布に従っているかどうかを定量的に判断するものではないが, 視覚的な評価として非常に有用だ. 特に, データの正規性の検証に頻繁に使用される.

### B.2 シャピロ・ウィルク検定

シャピロ・ウィルク検定は、サンプルデータが正規分布に従っているかどうかを検定するための統計的手法だ.この検定は特に小さなサンプルサイズに対して有効であり、データが正規分布に従うかどうかを評価するために広く使用されている.

#### シャピロ・ウィルク検定の統計量

シャピロ・ウィルク検定の統計量 W は次のように計算される:

$$W = \frac{\left(\sum_{i=1}^{n} a_i x_{(i)}\right)^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

ここで、 $x_{(i)}$  はサンプルの i 番目の順序統計量(昇順に並べた際の i 番目の値)、 $\bar{x}$  は サンプルの平均、 $a_i$  は正規分布から得られる特定の定数である。n はサンプルサイズを表す。

#### 統計量の分布と検定の実施

シャピロ・ウィルク検定では、計算された統計量 W を用いて、データが正規分布に従っているかどうかを判断する。W 値が大きいほど、データは正規分布に近いと考えられる。

具体的には、あらかじめ定められた有意水準(例えば 0.05)に対して W 値を比較し、 W 値が小さい場合(有意水準よりも小さい場合),帰無仮説(データが正規分布に従っている)を棄却する.

この検定は特にサンプルサイズが小さい場合に有効であるが, サンプルサイズが大きい場合は他の検定(例えばコルモゴロフ・スミルノフ検定)を使用することが推奨されることもある.

# 付録 C

# ロジスティック回帰に関する計算

### C.1 シグモイド関数

シグモイド関数の定義

$$\sigma(x) = \frac{1}{1 + \exp(-x)}$$

導関数

$$\frac{d}{dx}\sigma(x) = \sigma(x)(1 - \sigma(x))$$

ロジット関数の定義

$$logit(x) = log \frac{x}{1-x}$$

ロジット関数にシグモイド関数を代入

$$logit(\sigma(x)) = x$$

## C.2 ベルヌーイ分布の最尤法

ベルヌーイ試行の結果を表す確率変数を X とすると、その確率質量関数 (pmf) は次のように表される:

$$P(X = x) = p^{x}(1 - p)^{1 - x}, \quad x \in \{0, 1\}$$

ここで、x は試行の結果を表し、x=1 は成功、x=0 は失敗を意味する. サンプルデータ  $x_1, x_2, \ldots, x_n$  が与えられた場合、尤度関数は次のようになる:

$$L(p) = \prod_{i=1}^{n} p^{x_i} (1-p)^{1-x_i}$$

尤度関数を最大にする p の値を見つけるために、対数尤度関数を用いることが一般的である.対数尤度関数は次のように表される:

$$\log L(p) = \sum_{i=1}^{n} \left[ x_i \log(p) + (1 - x_i) \log(1 - p) \right]$$

これを p で微分し、0 と等しいと置くことで、最尤推定量は次のように求められる:

$$\hat{p} = \frac{\sum_{i=1}^{n} x_i}{n}$$

ここで  $\hat{p}$  はデータセットにおける成功の平均値であり、最尤推定量となる

# 謝辞

本研究を進めるにあたり、絶えず的確な助言と温かい支援を賜った増田亮先生に深く感謝する. 先生の広範な知識と熱意に満ちた指導は、本研究の質を高めるだけでなく、私の学問的な成長にも大きな影響を与えた.

また、機械学習に関連する諸問題について、貴重な意見と洞察を提供してくれた守真太郎先生にも感謝の意を表する。先生の深い専門知識と実践的なアドバイスは、研究の方向性を見出す上で不可欠だった。

さらに、研究過程でご支援を頂いたすべての方々に対し、心からの感謝を申し上げる. 皆様の助けがなければ、この論文は完成していなかった.

最後に、私の研究を支え、忍耐強く励ましてくれた家族に最大の感謝を捧げる. 皆様の 絶え間ない支援と愛情が、困難な時期を乗り越える力となった.

# 参考文献

- [1] Synchrotron-Radiation-Based Energy-Domain Mössbauer Spectroscopy, Nuclear Resonant Inelastic Scattering, and Quasielastic Scattering Using Mössbauer Gamma Rays(Springer2021,Makoto Sato,Ryo Masuda,Makina Saito)
- [2] ゼロから作る Deep Learning(オライリー・ジャパン 2016, 斎藤 康毅)
- [3] 最短でわかる PyTorch & 深層学習プログラミング (日経 BP, 赤石 雅典)
- [4] パターン認識と機械学習 上 (丸善出版 2012, クリストファー・ビショップ)
- [5] 深層学習 改訂第 2 版 (講談社 2022, 岡谷 貴之)
- [6] 現代数理統計学の基礎 (共立出版 2017, 久保川 達也)