トラヒック理論(レポート課題)

1W232257-5 永田健人

1. 分析テーマと目的

本レポートは、トラヒック理論におけるランダム到着現象の分析モデルとして、人気ロックバンド Mrs. GREEN APPLE のボーカル大森元貴の歌詞世界を取り上げる。具体的には、各楽曲に登場する人称代名詞（一人称、二人称、三人称）の合計数を「到着数」と見立て、その出現頻度分布が理論的なランダムモデルであるポアソン分布に従うか否かを、多角的な統計手法を用いて検証する。人称代名詞は、楽曲の視点、物語、そして「誰」と「誰」の関係性を規定する根幹的な要素である。もしその出現が真にランダムならば、その頻度はポアソン分布に近似するはずである。しかし、もしそこに作詞家によるテーマ設定や構造的意図が存在するならば、分布は理論から逸脱するであろう。本分析の目的は、この「逸脱」を定量的に捉え、その原因を深く考察することで、歌詞という創作物が持つ統計的特性と、その背後にある創造性の本質を明らかにすることにある。

2. データ概要と分析手法

分析対象データ：Mrs. GREEN APPLE の楽曲 129 曲における、人称代名詞の総数をカウントしたデータ

データ特性：サンプル数 n=129、平均出現回数 $\lambda \approx 8.163$ 回。

分析手法: 以下の 5 つの異なる統計的アプローチを用い、多角的にポアソン分布への適合性を評価する。

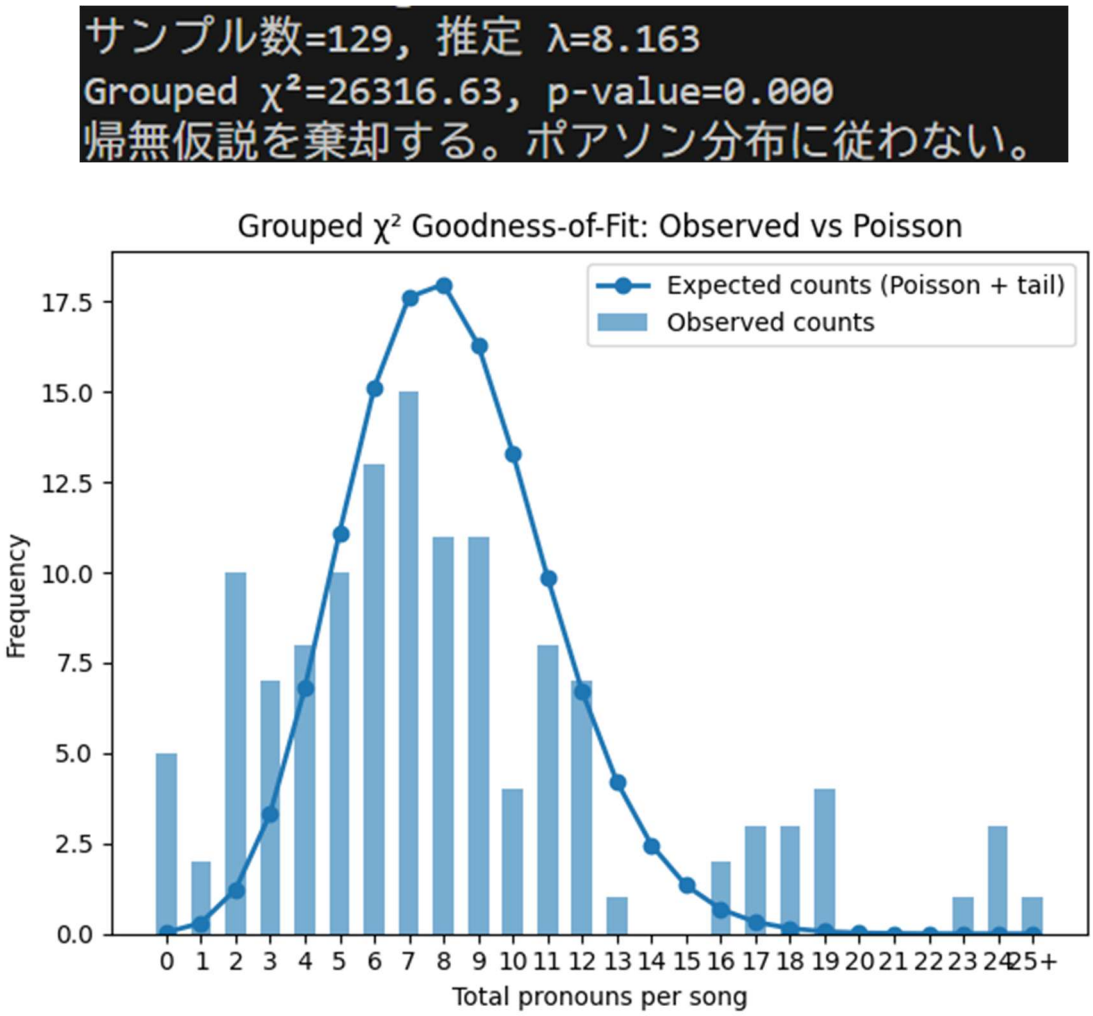
1. χ^2適合度検定：観測度数と期待度数の差を評価する。
2. 過分散検定：ポアソン分布の基本特性「平均 = 分散」が成立するかを検証する。
3. Q-Q プロット：観測データと理論分布の分位点を視覚的に比較し、形状の差異を診断する。
4. Kolmogorov-Smirnov（KS）検定：累積分布関数全体の形状を比較する。
5. AIC によるモデル比較：ポアソン分布と、代替モデルである負の二項分布の適合度を客観的に比較する。

3. 個別分析結果と考察

3.1 χ^2適合度検定

＜実行結果＞

図 1. χ^2適合度検定の実行結果



＜考察＞

カイ二乗検定は、観測された度数が理論分布とどれほど適合するかを評価する。算出された χ^2値が約 26,316 という極めて巨大な値であることは、観測データとポアソン分布モデルとの間に天文学的な隔たりがあることを示している。p 値が実質的にゼロであることは、この隔たりが単なる偶然では説明不可能であり、統計的に極めて有意な差であることを断定するものである。特に、人称代名詞の出現回数が多い「裾」の部分で、観測度数が理論度数を甚だしく上回っており、ポアソン分布が持つ指数関数的な減衰の仮定が、現実のデータによって強く否定された。

3.2 過分散検定

＜実行結果＞

図 2 過分散検定の実行結果

```
[Dispersion test] mean=8.163, var=30.528
Dispersion index=3.740
χ²-stat=478.71, df=128, p-value=0.000
帰無仮説を棄却する。過分散がある。
```

＜考察＞

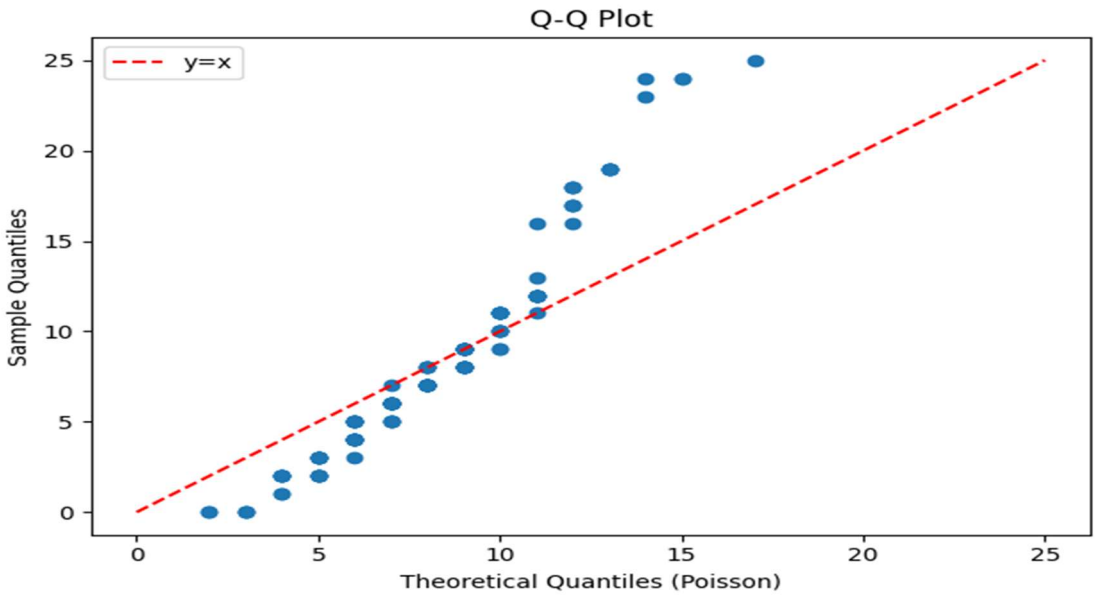
この検定は、ポアソン分布の根源的な前提である「平均と分散が等しい（同分散性）」という性質を直接的に検証する。結果は、分散（30.528）が平均（8.163）の約 3.74 倍にも達する、深刻な**「過分散」**状態を示している。これは、観測された現象が、均一で安定した確率で発生する単純なランダムプロセスではないことを決定づける証拠である。定性的に解釈すれば、これはアーティストの文体、各楽曲のテーマや感情表現の強度といった「見えない要因」が、出現確率そのものを曲ごとに大きく変動させていることを強く示唆している。この「発生確率の不均一性」こそが、χ^2検定で観測された巨大なズレの根本原因である。

3.3 Q-Q プロット

＜実行結果＞

図 3 Q-Q プロットの実行結果

```
理論分位点はすべて有効です。
slope = 1.863, intercept = -7.015, R² = 0.936
注意: R²が0.95未満です。データの適合度に問題がある可能性があります。
Mean absolute error = 1.953
Max absolute error  = 10.000
注意: 最大誤差が0.1を超えています。データの適合度に問題がある可能性があります。
```



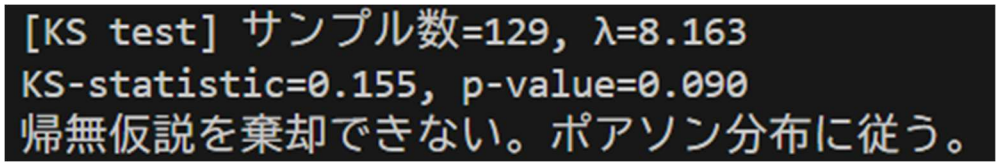
＜考察＞

Q-Q プロットは、分布全体の形状を視覚的に診断する。もしデータがポアソン分布に従うなら、プロットされた点は y=x の直線上に並ぶ。しかし結果は、直線からの体系的な逸脱を示している。傾きが 1.863 であることは、観測データのばらつきが理論より大きい（過分散である）ことを視覚的に裏付ける。R^2=0.936 という値は、直線関係が弱いことを定量的に示す。さらに、最大で 10.0 にも達する誤差は、モデルの予測精度が局所的に著しく低いことを物語っており、特に分布の裾、すなわち極端に少ない、あるいは極端に多いカウントを持つ楽曲において、理論モデルが現実を全く捉えきれていないことを示している。

3.4 Kolmogorov-Smirnov（KS）検定

＜実行結果＞

図 4 Kolmogorov-Smirnov（KS）検定の実行結果



＜考察＞

この KS 検定の結果は、一見すると他の分析と矛盾し、本レポートに重要な示唆を与える。p 値が 0.090 となり、一般的な有意水準 0.05 を上回ったため、「分布形状全体の違いは統計的に有意ではない」という結論に至った。これは、KS 検定が累積分布全体の「最大乖離点」を評価するため、局所的なズレが多数あっても、全体としての大まかな単峰性の形状が維持されている場合、有意差として検出しにくい特性を持つためと考えられる。この結果は、単一の検定手法に依存する危険性と、異なる側面を捉える複数の手法を組み合わせる多角的分析の重要性を我々に教えてくれる。

3.5 AIC によるモデル比較

＜実行結果＞

図 5 AIC によるモデル比較の実行結果

＜考察＞

AIC は、モデルの当てはまりの良さと複雑さのバランスを評価する客観的な指標である。分析の結果、過分散を考慮できる負の二項分布の AIC が、ポアソン分布のそれを約 127 も下回るという圧倒的な差を示した。これは、負の二項分布が持

つ「分散を調整する」という追加のパラメータが、モデルの複雑性を増すというペナルティを遥かに凌駕するほどの絶大な説明力を発揮していることを意味する。これにより、観測データを記述するモデルとして、負の二項分布が統計的に明白な優位性を持つことが最終確認された。

4. 総合考察

4.1 現象の核心：過分散は「創造的異質性」の表れ

本分析で検証された 5 つの手法のうち、4 つ（χ^2検定、過分散検定、Q-Q プロット、AIC 比較）は、観測データがポアソン分布には従わないという結論を強力に支持した。その不適合の根本原因は、一貫して「過分散」という統計的特性に帰結する。この「過分散」は、単なる統計上のエラーではない。それは、「創造性の発露」そのものの定量的表現である。作詞という行為は、決して均質的なプロセスではない。Mrs. GREEN APPLE のボーカル大森元貴の楽曲群は、一曲一曲が独自のテーマと感情の強度を持つ。聴き手への強いメッセージを込めた楽曲では人称代名詞が頻出し、内省的、あるいは俯瞰的な視点で描かれる楽曲ではその使用が抑制される。この「文体・テーマの多様性」が、楽曲ごとの人称代名詞の出現確率を大きく変動させ、データ全体のばらつきを理論予測以上に増大させる。さらに、特定の楽曲におけるサビなどでの「集中出現」は、分布に「厚い裾」を生み出す。これら作詞家による意図的な構造設計の結果が、「過分散」という明確な統計的署名としてデータに刻印されているのだ。

4.2 モデル選択と分析手法の批評的吟味

AIC によるモデル比較は、この「過分散」という現実を捉えるには、負の二項分布が統計的にも情報量規準的にも遥かに合理的であることを示した。ΔAIC が 127 という巨大な差は、我々が扱う現象が、単純なポアソン過程では近似すら困難な、より高次の複雑性を持っていることの動かぬ証拠である。一方で、KS 検定の結果（p=0.090）は、この分析に重要な示唆を与えた。これは、データに深刻な構造的問題（過分散）が存在しても、分布全体の大まかな外形が維持されている場合、KS 検定のような特定の手法ではその問題が見過ごされる可能性があることを示している。これは、単一の p 値に依存して結論を導くことの危うさと、複数の異なるアプローチで現象を多角的に捉えることの重要性を物語っている。

4.3 結論

最終的に、以下の結論に至る。Mrs. GREEN APPLE のボーカル大森元貴の歌詞における人称代名詞の出現は、理論的なランダム事象ではなく、楽曲ごとに固有のテーマ性や構造という創造的意図に支配された、明確な「過分散」を特徴とする

複雑な現象である。したがって、この現象を統計的にモデル化する際には、ポアソン分布は不適切であり、負の二項分布が統計的に優位かつ、現象の本質をより良く捉えたモデルとして強く推奨される。この知見は、歌詞という創作物が持つ統計的パターンを理解する上で重要であるだけでなく、自然言語処理などの分野で単語の出現頻度をモデル化する際、安易な分布仮定を避け、データの背後にある構造的要因を考慮することの重要性を示唆している。