

目次

第1章 イベント時系列の記述

- 1.1 イベント時系列とは
- 1.2 イベント時系列の表現
- 1.3 記述統計
- 1.4 データの記述から統計モデリングへ

第2章 一様ポアソン過程

- 2.1 一様ポアソン過程の性質：無記憶性
- 2.2 なぜ一様ポアソン過程が大事なのか
- 2.3 ポアソン分布：一定時間内のイベントの個数
- 2.4 アーラン分布：一定イベント数の待ち時間
- 2.5 アーラン分布とポアソン分布の関係
- 2.6 一様ポアソン過程の同時確率密度関数
- 2.7 イベント生成率による一様ポアソン過程の定義

第3章 リニューアル過程

- 3.1 イベント生成率（ハザード関数）
- 3.2 イベント間隔分布
- 3.3 イベント個数の分布
- 3.4 リニューアル過程の実現
- 3.5 リニューアル過程の推定

第4章 非一様ポアソン過程

- 4.1 イベント時刻の密度関数
- 4.2 時間伸縮によるイベント間隔分布の導出
- 4.3 非一様ポアソン過程の同時確率密度関数
- 4.4 非一様ポアソン過程の実現（シミュレーション）

- 4.5 ヒストグラムを用いたイベント生成率の推定
- 4.6 カーネルを用いたイベント生成率の推定

第5章 点過程の一般論

- 5.1 過去のイベントの影響を受ける点過程
- 5.2 マーク付き点過程
- 5.3 時間伸縮理論
- 5.4 点過程の実現方法

第6章 カウント時系列モデル

- 6.1 カウント時系列データのモデリング
- 6.2 負の二項分布

第7章 状態空間モデルによるイベント時系列解析

- 7.1 状態空間モデル
- 7.2 逐次ベイズ推定
- 7.3 ガウス近似アルゴリズム
- 7.4 粒子フィルタ・平滑化
- 7.5 補足

第8章 応用

- 8.1 脳情報デコーディング
- 8.2 イベント生成パターンを特徴付ける
- 8.3 イベント発生の内因と外因の寄与を読み取る
- 8.4 感染症の実効再生産数の推定

図 1.1 神経細胞とスパイクの模式図。スパイクが発生した時刻に着目すると神経活動はイベント時系列に見える。

イベント時系列になる (図 1.2)。SNS 上のコミュニケーションの特徴は、投稿がさらなる投稿や書き込みを誘発する連鎖反応である。特にある言語に対してコメントが集中的に寄せられる状態は“炎上”と呼ばれる。8.3節では、コメントが投稿されるタイミングの時系列データから連鎖反応を引き起こす影響力を定量的に評価する方法を解説する。

図 1.2 SNS 上の投稿をイベントと見なし、そのタイミングに着目するとイベント時系列になる。

感染症 コロナウイルス感染症 2019 (COVID-19) は 2019年 12 月中国・武漢で初めて報告され、瞬く間に世界中に流行が広がった。各国は都市閉鎖などの緊急の政策を実施したが、感染拡大を抑える一方で社会経済への影響も懸念された。感染症への対策は、公衆衛生および社会経済の両方の課題である。感染をイベントと

豊富なビジュアルによって直感的に理解できる！

図 5.10 (a) イベント時系列のラスタープロット。(b) イベント間隔分布。(c) 累積分布関数のプロットとモロゾフ・スミスノブ検定の有意水準 $\alpha = 0.05$ の棄却領域線 (点線)。

$$z_i = 1 - \exp(-t_i) \quad (5.87)$$

は $[0, 1]$ 上の一様分布に従う。よって、 χ^2 の検証は $\{z_i\}$ と一様分布の比較に帰着される。この比較は累積分布関数をプロットすることで視覚化できる。具体例を用いて説明しよう。図 5.10 は、イベント時系列データ (図 5.10a) に対して、ガンマ分布と指数分布 (図 5.10b) をそれぞれイベント間隔分布にフィットするリニューアル過程 (指数分布の場合は一様ポアソン過程) を当てはめた結果である³⁰⁾。それぞれのモデルに対して、変換された n 個のサンプルからなるデータ $\{z_1, \dots, z_n\}$ をその大きさが小さい順に並べたものを $\{z_{(1)}, \dots, z_{(n)}\}$ とし、これに対して $b_i = \frac{z_{(i)}}{i}$ ($i = 1, \dots, n$) を縦軸にプロットする (図 5.10c)。 $\{z_i\}$ が一様分布に従っていればプロットした点は横軸 1 の対角線上に分布するので、対角線からの乖離の度合いでモデルの当てはまり具合を視覚化することができる。コロモゴロフ・スミスノブ検定に基づいてプロットの棄却域を定めることができ、サンプル数が大きいときに有意水準 $\alpha = 0.05$ の棄却領域線は $z_{(i)} \pm 1.36\alpha^{1/2}$ で与えられる (図 5.10c 点線)。プロットした点が領域線を超えたと、設定した有意水準でモデルは棄却される。この例では、指数分布のプロットは領域線を超えているので棄却される。つまり、このデータは一様ポアソン過程では説明できないということだ³¹⁾。

³⁰⁾ それぞれのモデルのパラメータ推定方法は 3.5 節を参照。
³¹⁾ さらに、指数分布に当てはめられないサンプルが検出されないことから、イベント時系列は一様ポアソン過程よりも複雑なものであることが読み取れる。

イベント生成の具体的な推定方法を丁寧に解説！

図 8.4 上: コメントが寄せられた時刻別のヒストグラム。下: 1月29日から2月5日の1週間 (a) と、2月7日から2月14日までの1週間 (b)。縦軸はすべてのコメントのヒストグラムであり、右側は投稿に対する返信コメントのヒストグラムである。両図の縦軸は推定した再生産数の 95% 信頼区間を表す。

8.4 感染症の実効再生産数の推定

感染が拡大しているかもしくは収束しているかは実効再生産数という指標で判定される。実効再生産数は、“すでに感染が広がっている状況で 1 人の感染者が平均何人にうつが”を表す指標であり、1 より大きければ感染は拡大し、小さければ収束していく。ここでは、新規感染者数の時系列から実効再生産数を推定する方法

直近の事例である COVID-19 拡大や SNS 拡散の推定方法を詳述！