# 研究題目

## 統計モデルの基礎とその応用—Rを利用した解析例

指導教員　　　　今野　良彦

上田　亜衣　　小野澤　良子　　数馬　早紀　　小菅　奈奈

鈴木　菜つ美　　樽木　美里　　半藤　雅子　　六川　咲美

## 発表内容

☆ R

☆ 統計モデル

☆ 新記録の誕生と競技水準の向上

☆ O —157による集団食中毒

☆ 今後の課題

☆R とは

- データを効率的の操作し、保管する機能

- 配列と行列の計算のための演算子セット

- データ解析のための道具の大規模で一貫した集まり

- データ解析のためのグラフィカルな機能と画面または、印刷物へ出力

- 条件分岐、ループ、ユーザー定義の再帰的関数や入出力機能を含む、効果的なプログラミング言語

1

☆統計的な手法
- 記述統計的な手法
- 推測統計的な手法

☆統計モデルとは

データの背後に確率的な構造を考え、それを分析するのが推測統計的な手法であり、データの背後に想定される確率的な構造のこと

# 新記録の誕生と競技水準の向上

②

**帰無仮説 H0** : 競技の水準に変化がなく、各年の最高記録が同じ分布に従う

**k 年目の最高記録が新記録である場合の確率変数を１、新記録でない場合の確率変数の値を０とする。**

## 確率変数

$$X_k = \begin{cases} 1\text{、}k\text{年目の１年間の最高記録が新記録である} \\ 0\text{、}k\text{年目の１年間の最高記録が新記録でない} \end{cases}$$

**帰無仮説$H_0$の下では$X_1$から$X_k$は互いに独立で、**

$$\Pr\{X_k = 1\} = \frac{1}{k}$$

$$\Pr\{X_k = 0\} = 1 - \frac{1}{k}$$

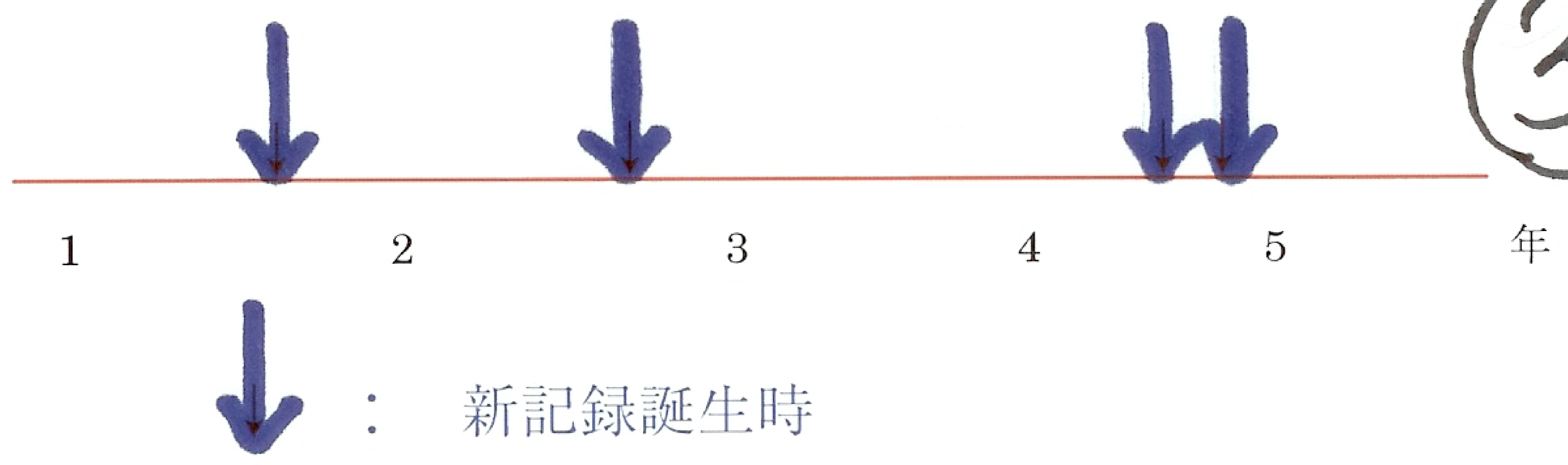
n 年間の新記録の数$S_n$

$$S_n = X_1 + X_2 + \cdots + X_n$$

とする.

n=4

3




1 年目：新記録の数　＝　1

2 年目：新記録の数　＝　1

3 年目：新記録の数　＝　0

4 年目：新記録の数　＝　2

$$X_k = \begin{cases} 1\text{、}k\text{年目の１年間の最高記録が新記録である} \\ 0\text{、}k\text{年目の１年間の最高記録が新記録でない} \end{cases}$$

上のグラフをこの確率変数で表すと

$X_1=1,\quad X_2=1,\quad X_3=0,\quad X_4=1$　　となる.

この 4 年間の新記録の数：$S_4$ は,次のように計算する.

$$S_4 = X_1 + X_2 + X_3 + X_4$$
$$=1+1+0+1$$
$$=3$$

グラフから読み取れる新記録の数：4

$S_4$ の値：3

“同じ年に新記録が複数出ても,その一年間の最高記録を,その年の新記録と見ていくので,新記録の数は１つとみなす！！” ④

$S_n$の確率分布は次のようにして求める.

$$p(q,n) = \Pr\{S_n = q\}$$

とおく.

$\{S_n = q\} = \{S_{n-1} = q - 1 \text{かつ} X_n = 1\} \cup \{S_{n-1} = q \text{かつ} X_n = 0\}$ (1.1)

より，n 年間で q 回新記録を出した確率の関係式

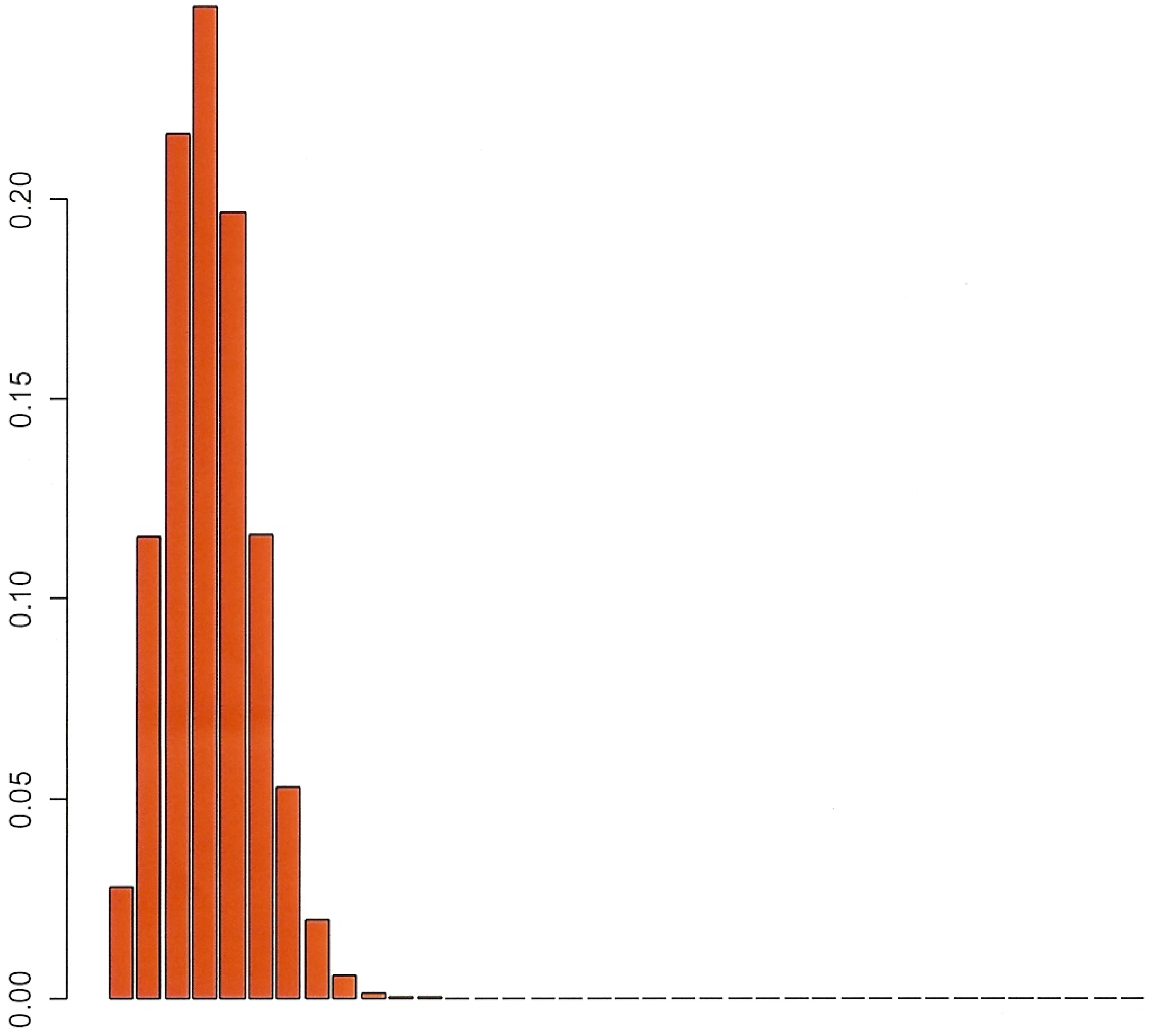
$$p(q,n) = \frac{1}{n} p(q-1, n-1) + \frac{n-1}{n} p(q, n-1)$$
$$q = 1,2,\ldots,n \; ; n = 2,3,\ldots$$

最初の年の最高記録が新記録である確率

$$p(1,1) = 1$$

36 年間の新記録の確率分布

⑤

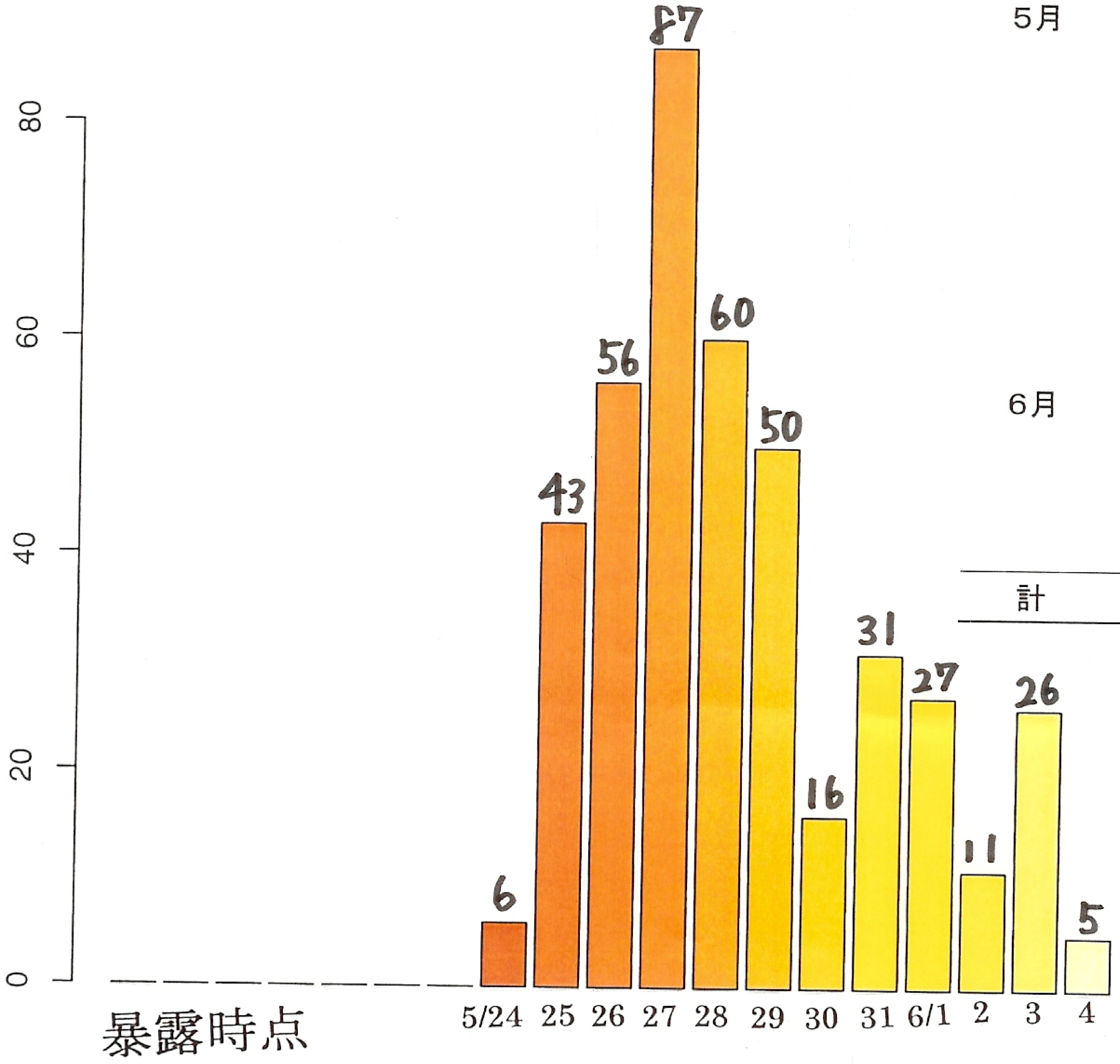



| $S_n$ の値 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 以上 |
|---|---|---|---|---|---|---|---|---|
| 確率 | 0.028 | 0.115 | 0.216 | 0.248 | 0.197 | 0.116 | 0.053 | 0.027 |

<table>
<tr><th colspan="3">女子1500m世界記録</th></tr>
<tr><th>結果</th><th>データ</th><th>$X_k$</th></tr>
<tr><td>4:17.3 03</td><td>1967年6月</td><td rowspan="2">$X_1=1$</td></tr>
<tr><td>4:15.6 24</td><td>1967年10月</td></tr>
<tr><td>4:12.4 02</td><td>1969年7月</td><td rowspan="2">$X_3=1$</td></tr>
<tr><td>4:10.7 20</td><td>1969年9月</td></tr>
<tr><td>4:09.6 15</td><td>1971年8月</td><td>$X_5=1$</td></tr>
<tr><td>4:06.9 18</td><td>1972年7月</td><td rowspan="4">$X_6=1$</td></tr>
<tr><td>4:06.5 04</td><td>1972年9月</td></tr>
<tr><td>4:05.1 07</td><td>1972年9月</td></tr>
<tr><td>4:01.4 09</td><td>1972年9月</td></tr>
<tr><td>3:56.0 28</td><td>1976年6月</td><td>$X_{10}=1$</td></tr>
<tr><td>3:55.0 06</td><td>1980年7月</td><td rowspan="2">$X_{13}=1$</td></tr>
<tr><td>3:52.47 13</td><td>1980年8月</td></tr>
<tr><td>3:50.46 11</td><td>1993年9月</td><td>$X_{26}=1$</td></tr>
</table>

⑦

| | 発症日 | 人数 |
|---|---|---|
| 5月 | 24 | 6 |
| | 25 | 43 |
| | 26 | 56 |
| | 27 | 87 |
| | 28 | 60 |
| | 29 | 50 |
| | 30 | 16 |
| | 31 | 31 |
| 6月 | 1 | 27 |
| | 2 | 11 |
| | 3 | 26 |
| | 4 | 5 |
| 計 | | 418 |




$\gamma$

図２．１　　O-157:H7による集団食中毒における発症日別度数分布

OHP

8

$\gamma$：一斉に暴露した時点

$x_i, i=1,\cdots,n$：i番目の食中毒患者の症状の発生時点

尤度関数は、

$$L(\gamma,\mu,\sigma^2)=\prod_{i=1}^{n} f(x_i;\gamma,\mu,\sigma^2) \qquad (x_i>\gamma)$$

$$=(2\pi\sigma^2)^{-\frac{n}{2}}\prod_{i=1}^{n}(x_i-\gamma)^{-1}\cdot\exp\left[-\frac{1}{2}\sum_{i=1}^{n}\left\{\frac{\log(x_i-\gamma)-\mu}{\sigma}\right\}^2\right]$$

$$l(\gamma,\mu,\sigma^2)=\log L(\gamma,\mu,\sigma^2) \qquad \cdots(2.\ 1)$$

対数尤度関数の偏微分を計算した連立方程式

$$\frac{\partial l}{\partial\gamma}=\frac{\partial l}{\partial\mu}=\frac{\partial l}{\partial\sigma^2}=0 \qquad \cdots(2.\ 2)$$

$\gamma$を所与としたときの、$\mu$と$\sigma^2$の最尤推定量

$$\hat{\mu}(\gamma)=\frac{1}{n}\sum_{1}^{n}\log(x_i-\gamma) \qquad \cdots(2.\ 3)$$

$$\hat{\sigma}^2(\gamma)=\frac{1}{n}\sum_{1}^{n}\{\log(x_i-\gamma)-\hat{\mu}(\gamma)\}^2 \qquad \cdots(2.\ 4)$$

プロファイル対数尤度：

$$l_{**}(\gamma)=n(\hat{\mu}(\gamma)+\log\hat{\sigma}(\gamma))-\frac{n}{2}(1+\log\pi))$$

$$\cdots(2.\ 5)$$

⑨

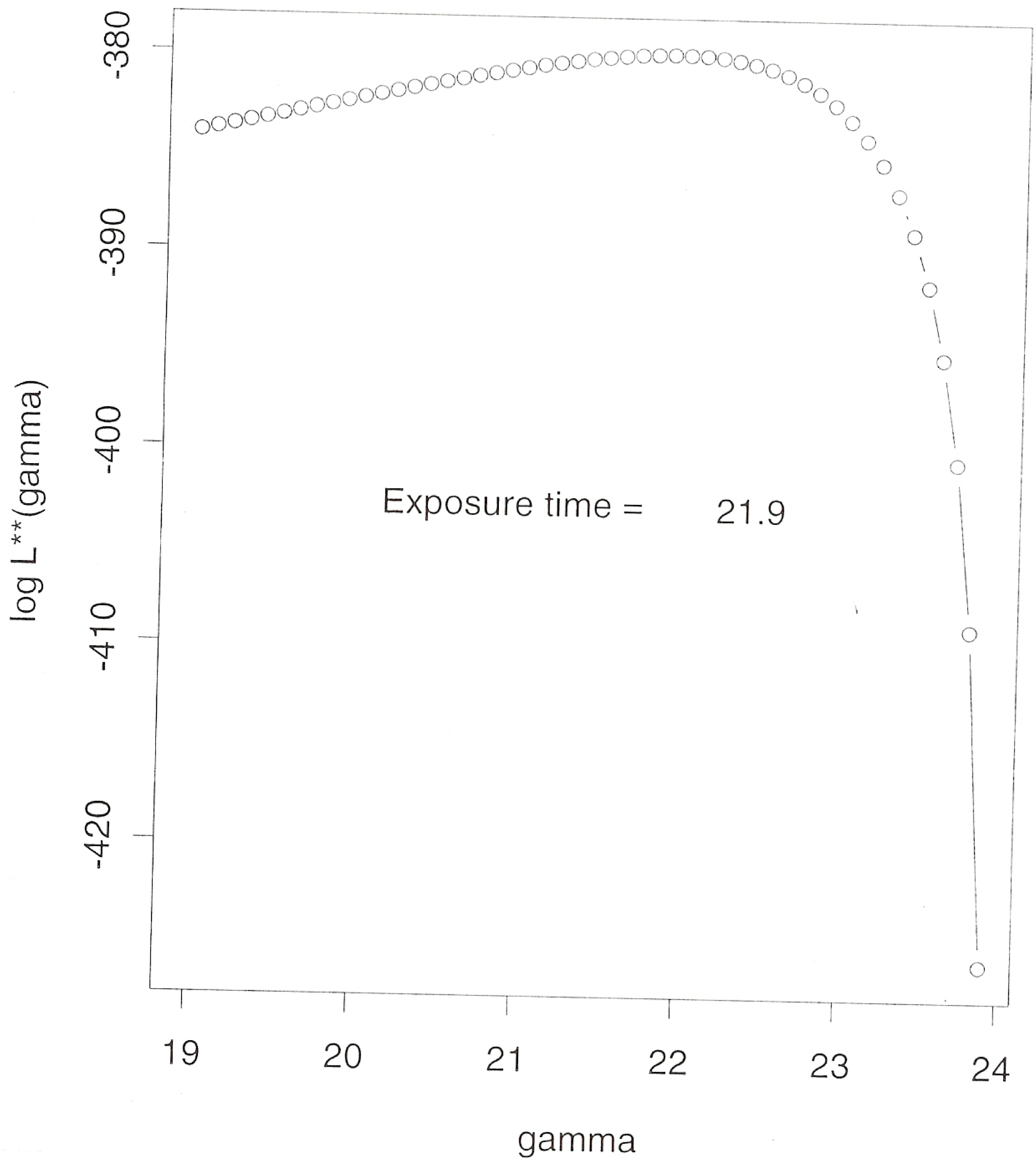



図２．２　　数値列$\gamma = \dfrac{j}{10}(j = 190, 191, \cdots, 239)$における図２．１のデータのプロファイル対数尤度$l**(\gamma)$と推定値