

# MIS参加者に関する 計量情報学的分析

小野寺夏生（筑波大学）  
**onodera@slis.tsukuba.ac.jp**

**2009. 7. 4**

**第26回医学情報サービス研究大会**

# 本発表の背景と目的

- ◆ 医学情報サービス研究大会(MIS)は、25回大会までの**実参加者2,057名、延べ参加者数4,691名**を数える。
- ◆ MIS幹事会により、**第1回以降の参加者記録が保存**されている。
- ◆ これを用いて、**参加回数分布について統計的な分析**を行うことを目的とする。
- ◆ 特に、参加回数分布が**Lotkaの法則**によく従うことを示し、**累積優位性モデル**によりこれを検討する。

# 1. 参加者についての 基本的な統計

# 第25回までの参加回数の分布

参加回数	人数	比率(%)	累積比率	参加回数	人数	比率(%)	累積比率
1	1292	62.8	62.8	14	2	0.1	98.9
2	298	14.5	77.3	15	5	0.2	99.1
3	145	7.0	84.3	16	5	0.2	99.4
4	82	4.0	88.3	17	3	0.1	99.5
5	64	3.1	91.4	18	0	0.0	99.5
6	36	1.8	93.2	19	1	0.0	99.6
7	32	1.6	94.7	20	3	0.1	99.7
8	19	0.9	95.7	21	2	0.1	99.8
9	20	1.0	96.6	22	0	0.0	99.8
10	14	0.7	97.3	23	1	0.0	99.9
11	11	0.5	97.9	24	2	0.1	100.0
12	11	0.5	98.4	25	1	0.0	100.0
13	8	0.4	98.8				

実参加者数: **2,057** 延べ参加者数: **4,691**

平均参加回数: **2.28**

# 性別の参加統計

	男	女	全体
実参加者数	721	1336	2057
平均参加回数	2.28	2.28	2.28
1回のみ参加者比率(%)	64.1	62.1	62.8
5回以上参加者比率(%)	10.5	12.3	11.7

# 地域別の参加統計

	北海道	東北	関東	中部	近畿	中国・四国	九州
実参加者数	113	64	998	265	321	158	82
平均参加回数	1.79	2.14	2.61	2.11	2.25	1.75	1.46
1回のみ参加者比率(%)	71.7	64.1	56.3	63.0	67.6	72.8	73.2
5回以上参加者比率(%)	6.2	12.5	13.7	12.5	13.4	5.7	2.4

# 館種別の参加統計

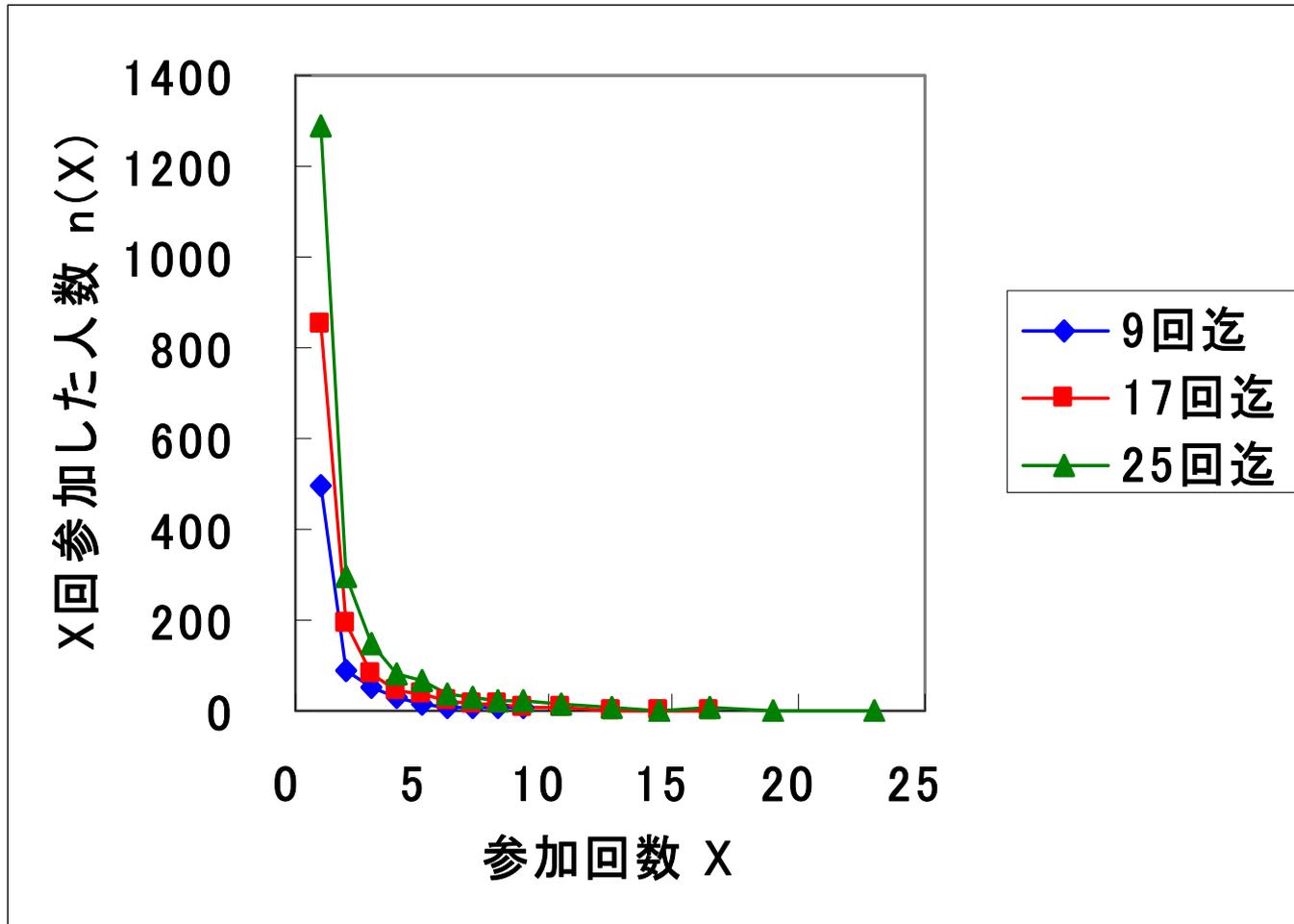
	医図	歯図	看図	薬図	病図	大図
実参加者数	409	69	166	117	388	178
平均参加回数	2.47	2.52	2.37	2.33	2.90	2.05
1回のみ参加者比率(%)	56.0	53.6	57.2	59.8	52.1	66.3
5回以上参加者比率(%)	12.5	14.5	13.3	12.0	18.3	10.1
	研図	公共	教職	学生	企業	その他
実参加者数	19	37	59	60	406	149
平均参加回数	1.63	1.73	2.64	1.20	1.89	1.72
1回のみ参加者比率(%)	68.4	78.4	69.5	90.0	70.4	79.2
5回以上参加者比率(%)	5.3	8.1	15.3	1.7	8.1	4.7

## 2. 参加回数の分布は Lotkaの法則によく従う

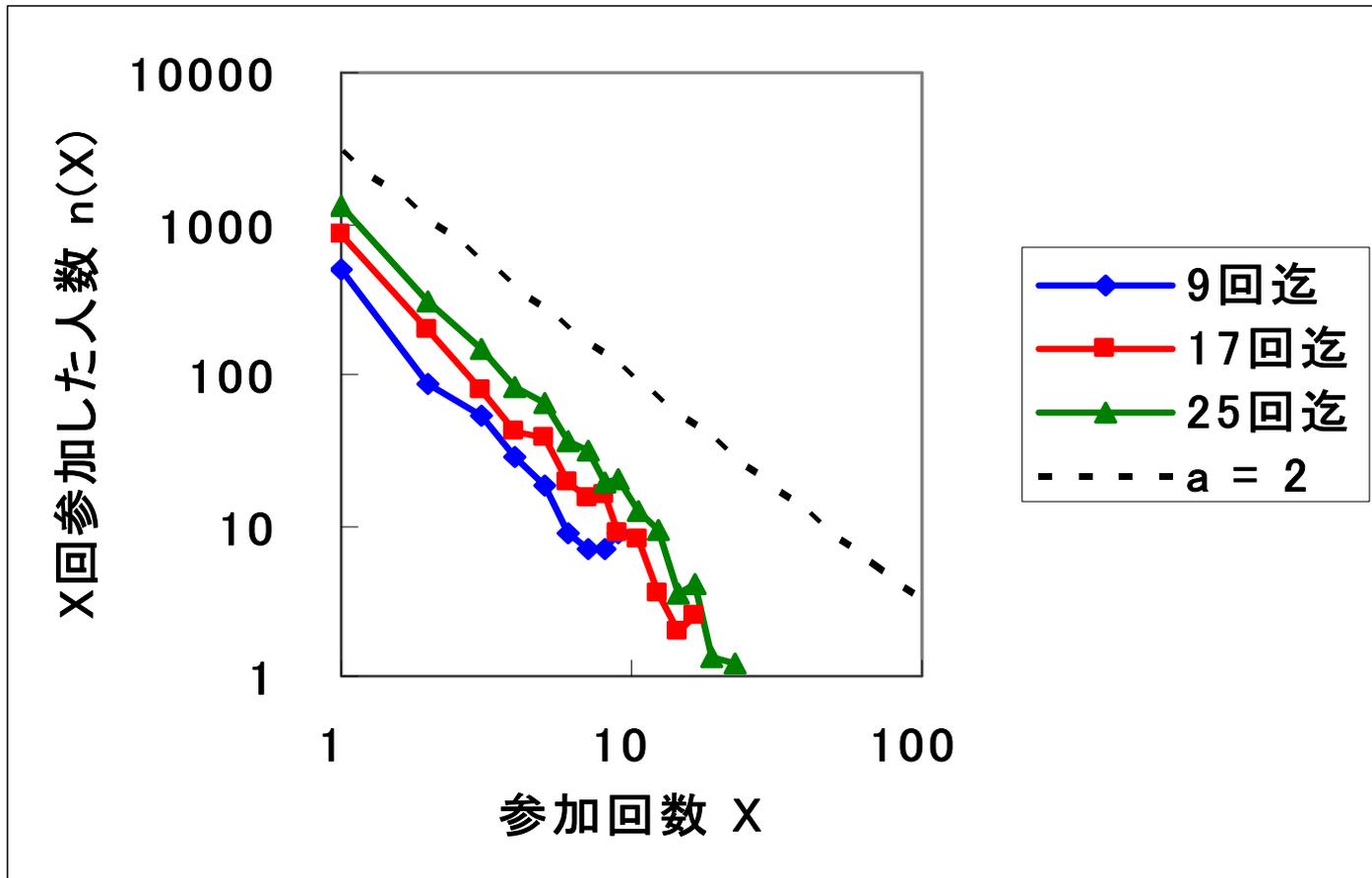
# 第25回までにX回参加した人の数

参加回数 X	人数 n(X)	$\frac{n(X)}{n(1)}$	$1/X^2$
1	1292	1.000	1.000
2	298	0.231	0.250
3	145	0.112	0.111
4	82	0.063	0.063
5	64	0.050	0.040
6	36	0.028	0.028
7	32	0.025	0.020
8	19	0.015	0.016
9	20	0.015	0.012
10	14	0.011	0.010

# X回参加者の人数 $n(X)$ をグラフにすると (第9, 17, 25回までの参加統計)

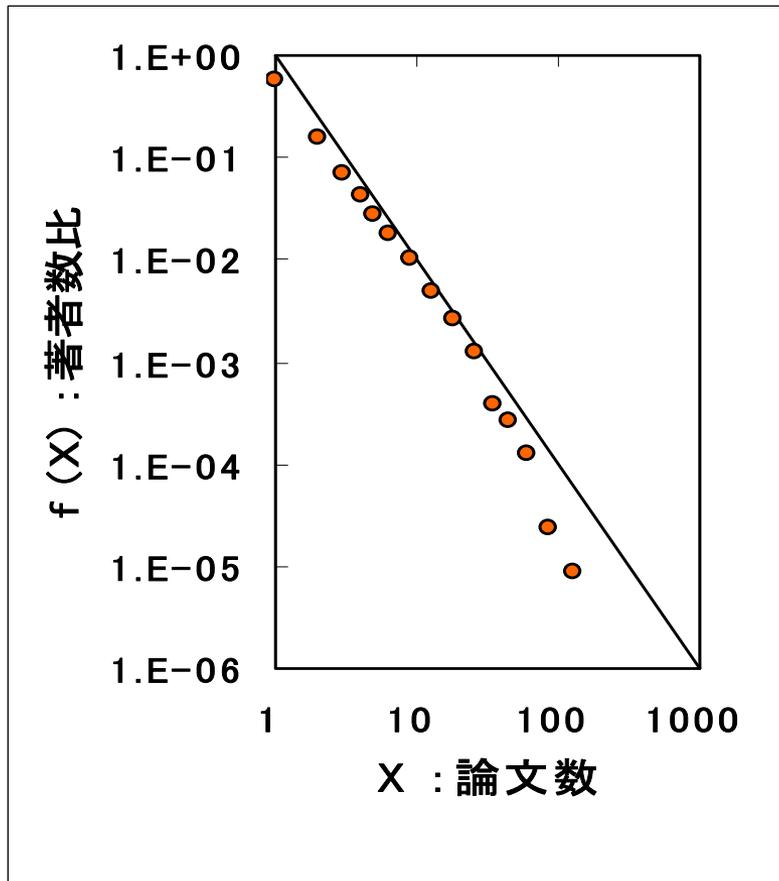


# グラフの両軸を対数目盛に変換すると



両対数メモリにすると、 $y = C / X^a$  のグラフは、傾きが  $-a$  の直線になる。

# Lotkaの法則とは？



ある分野である期間に $X$ 編の論文を発表する著者の比率  $f(X)$  は次式に従う。

$$f(X) = C / X^a$$

$$X = 1, 2, \dots$$

(左図の直線は  $a=2$  の場合)

上式を一般化し、 $f(X)$  を、 $X$  個のitemを生産(または所持)するsourceの確率ということがある。

# Sourceが持つItemの分布の例

Source	Item数
著者	発表論文数
単語	テキスト中の出現頻度
キーワード	データベース中の索引頻度
科学雑誌	ある主題に関する掲載論文数
論文	引用される回数
都市	居住者数(人口)
個人、企業	保有資産の額
生物の属	属中の種の数

### 3. MIS参加者の参加確率は 「累積優位性モデル」で 説明できるか？

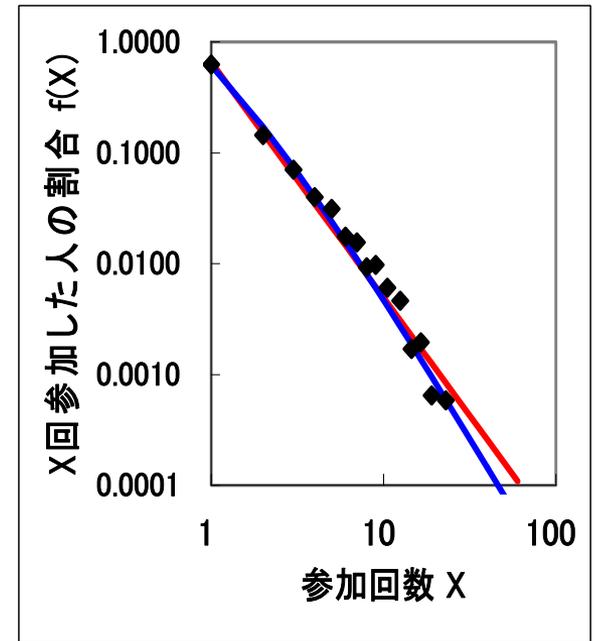
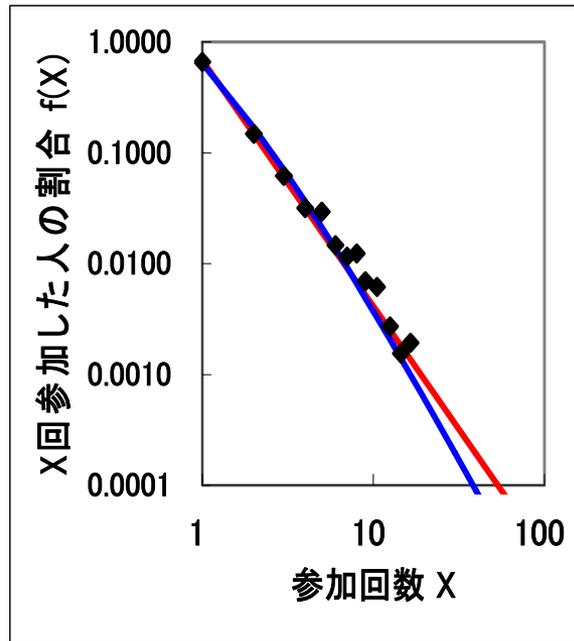
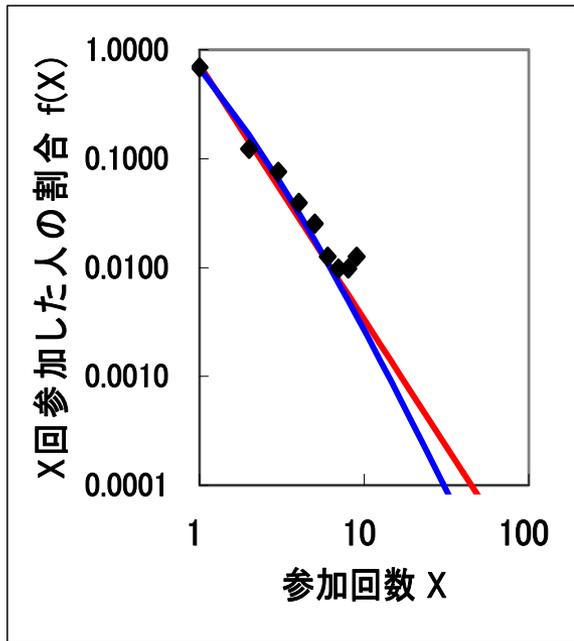
## なぜ多くの事例でLotkaの法則が成り立つか？

- ◆ Lotkaの法則で示される分布  $f(X) = C / X^a$  を説明する有力な確率モデルに「**累積優位性**」(cumulative advantage)仮説がある。
- ◆ このモデルは、「**あるsourceが新たなitemを得る確率は、そのsourceが既に持つitemの数に比例する**」と仮定する。
- ◆ その解として得られる**Yule分布**は、Lotkaの式より複雑な数式で表されるが、**Xが十分に大きいところで両者は一致する**。



# MISの参加回数分布はLotka型? Yule型?

◆: obs —: Lotka —: Yule



左:第1~9回

中:第1~17回

右:第1~25回

# 過去の参加回数による 次のMISへの参加確率予測

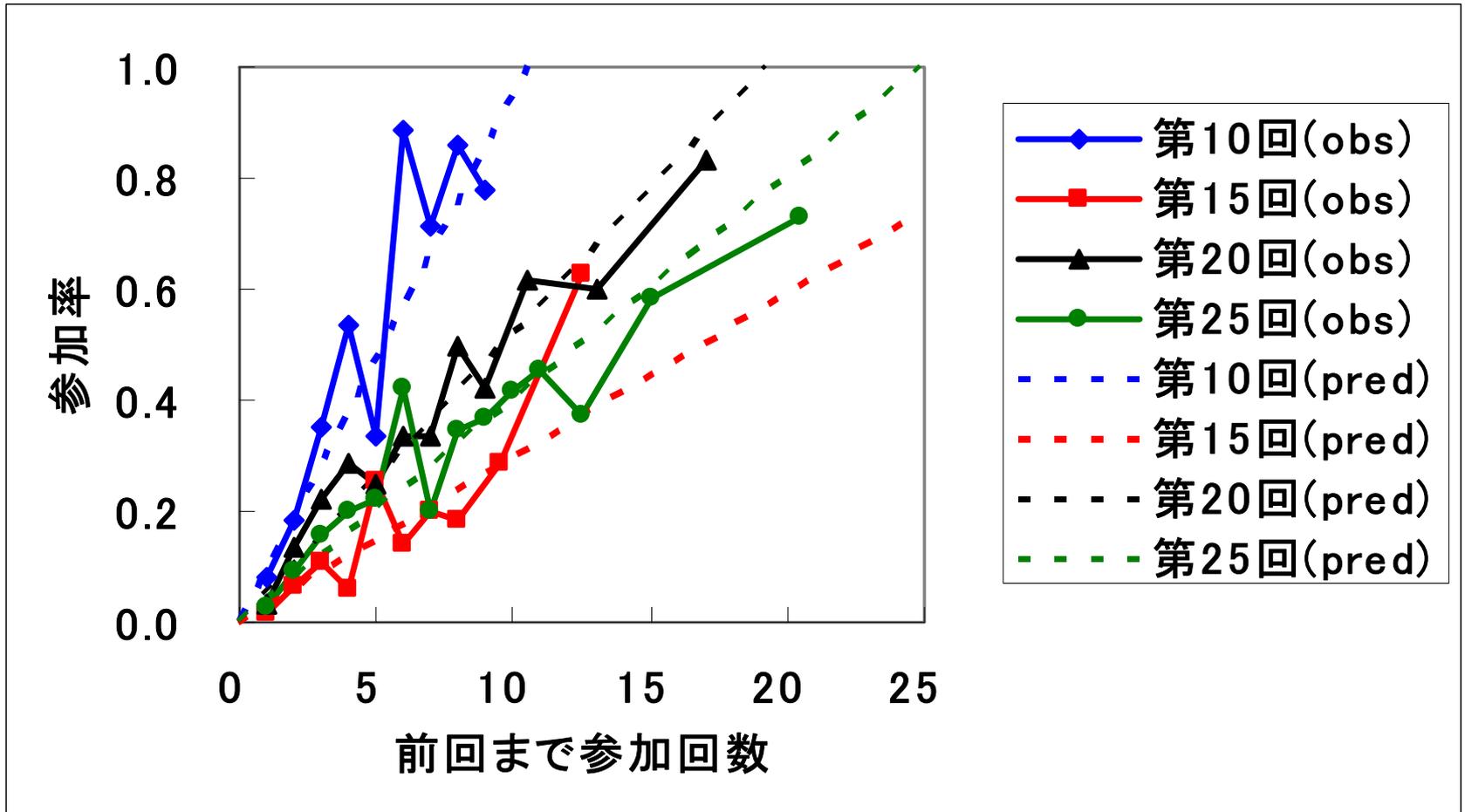
- ◆ 累積優位性モデルによれば、ある人が第 $k$ 回MISに参加する確率は、その人が過去( $k-1$ )回までに参加した回数 $X$ に比例する。
- ◆ この関係は次式で表される。  
$$Pr(X, k) = [\Delta N(k) / N(k-1)] X$$

$Pr(X, k)$ : 第( $k-1$ )回までに $X$ 回参加した人が、第 $k$ 回に参加する確率

$N(k-1)$ : 第( $k-1$ )回までの延べ参加者数

$\Delta N$ : 第 $k$ 回の参加者数(新規参加を除く)
- ◆  $k = 10, 15, 20, 25$ の各大会のデータについてこの検証を試みた。

# 前回(第 $k-1$ 回)までの参加回数と 第 $k$ 回の参加率の関係 ( $k=10, 15, 20, 25$ )



# 過去の参加回数データから 次のMISへの新規参加者数予測

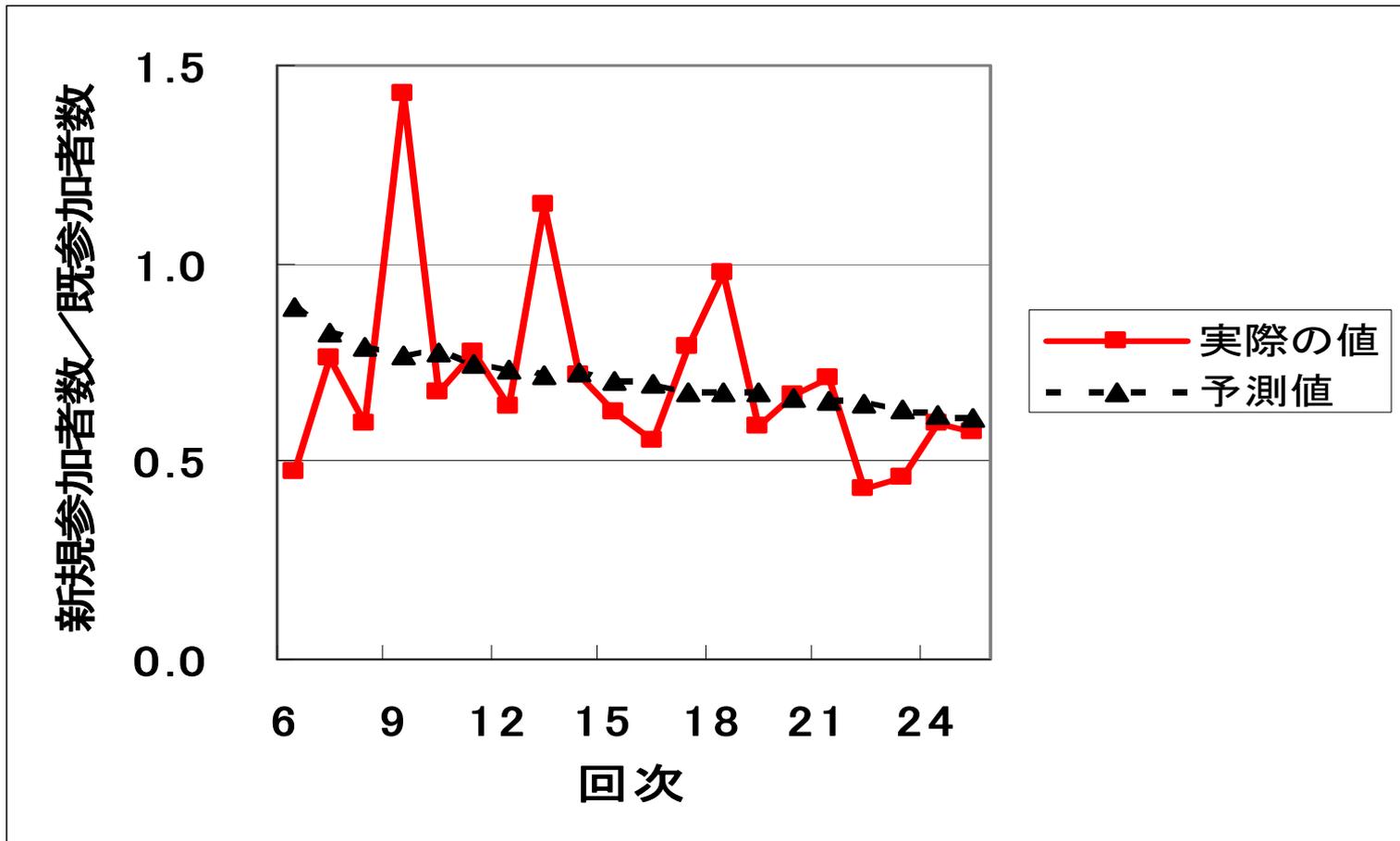
- ◆ 演者の累積優位性モデルでは、第 $k$ 回大会における新規参加者と既参加者の比率 $\beta$ は、 $(k-1)$ 回大会までの参加者の平均参加回数 $\mu$ に反比例する。

すなわち、

$$\beta = C / \mu \quad (C \text{はどの大会にも共通の定数})$$

- ◆ 第6回から第25回までの各大会のデータから実際に得られる $\beta$ と $\mu$ の値から、上式に従ってそれぞれに対する $C$ を求め、この平均値 $C'$ を上式の比例定数として、これと $\mu$ から各大会の $\beta$ を予測した。実際の $\beta$ と予測の $\beta$ を比較した。

# 各大会の新規参加者率の予測



特に外れが大きい大会

新規が予測より過剰: 9回(京都)、13回(名古屋)、18回(愛媛)

新規が予測より過小: 6回(大阪)、22回(愛知医)

# まとめ

1. MIS25までの参加者実数は2,057名、平均参加回数は2.28回である。1回のみ参加者、5回以上参加者はそれぞれ63%、12%で、20回以上の参加者は9名(0.4%)である。
2. 参加者の男女比は35:65で、参加回数分布はごく類似している。一方、地域別、館種別の参加回数分布には差異が見られる。
3. 参加回数分布はLotkaの法則によく従い、期間を変えても安定している。冪指数パラメータ  $a$  は近似的に2に等しい。

# まとめ

4. しかし、両対数グラフは僅かに直線から湾曲しており、Lotka分布よりYule分布によく合う。
5. ある参加者の第 $k$ 回の参加確率は、第 $(k-1)$ 回までのその参加者の参加回数に概ね比例しており、累積優位性モデルの仮定に従う。
6. 第 $k$ 回大会における新規参加者と既参加者の比率は、第 $(k-1)$ 回までの平均参加回数に反比例するという仮定を、実際のデータから確かめた。しかし、実測値と予測値の間にはややばらつきが見られた。

# 謝 辞

- ◆この研究は、第1～25回の参加者データなしには不可能でした。それを提供していただいた医学情報サービス研究会幹事会に感謝致します。
- ◆なお、提供されたデータは、性別、地域、所属の館種、及び各会の参加の有無のみを含み、それ以上の個人を特定する情報は含まれていません。