

精密累積データの セミパラメトリックポアソンモデル

統計数理研究所
椿 広計

提供する話題

- 追悼シンポジウムとしての序言
 - 田口玄一先生の統計学の講義のインパクト
 - 私自身のこと(1975-1992)
 - 田口(1962)の「新版実験計画法下、丸善」について
- 精密累積法のセミパラメトリックモデルによる再定式化
 - 精密累積データという考え方
 - 統計的批判と再評価の必要性
 - 精密累積法にリンクする統計モデル
 - 0,1データ取扱いの非対称性について
- おわりに

私自身のこと(1975-1992)

田口玄一先生の 統計学の講義のインパクト

1975年4月東大教養学部202教室「統計学」

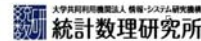
- 田口玄一先生は、1975年4月から1年間東京大学教養学部理科I類において、統計学の講義を担当されました
- 私は、その講義の受講生でした
 - 教科書は、丸善から出版された統計解析第二版でした
 - 当時、田口先生は、青山学院大学に所属されていました。
 - 朗々とした口調で、
 - 「いいですか、良い技術者というものは、xxxのように考えるのです。」
 - 「ペル研では、yyy」
 - 「プリンストン大学のTukeyは、zzz」
 - といった声が朝の教室に響きました。
 - 講義内容もさることながら、高校まではエンジニアリング・センスなどという言葉に全く疎かったので、新鮮でした。
- 私は、その講義を受講し、はじめて統計解析という分野に興味を持ちました。
 - 特に分散分析、回帰分析、寄与率の計算の虜になりました。
 - この夏休みに、統計学を学べる学科はどこかということを調べ、計数工学科という学科に統計学の講座があることを知り、そこに進学することを決意しました。

1975年4月23日の講義

- R. A. Fisher 実験統計学
 - 技術情報の獲得効率を高めるための共通技術の全体
 - 実験計画とデータ解析
 - 行列, 2次形式, 直交展開(分散分析, 直交多項式)
 - 統計ではランクの代わりに自由度という用語を用いる
- N. Wiener サイバネティクス
 - 通信と制御との共通理論
- C. Shannon : 通信理論
 - 情報の伝達効率を上げるための共通技術の全体
 - 情報量を数学的に定義

2013/05/13

田口玄一先生一周忌シンポジウム

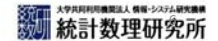


秋休みの宿題

- レポート
- 自分又は他人による実験又は調査データの解析
 - 評価の視点
 - テーマの選定能力
 - 解析能力
 - 結論の出し方
 - データの回帰, 平方根変換, 対数変換
- 自然地理・理科年鑑などの資料を基に, 地球上の各地の月別平均気温と緯度(緯度の余弦変換)についての回帰分析を実施し, 田口先生に提出しました.
 - 自分自身で合理的経験法則を作り出せることに感激.
 - 後年, 田口先生と矢野宏先生との対話本に, この学生のレポートのことが記載されており嬉しかったです

2013/05/13

田口玄一先生一周忌シンポジウム

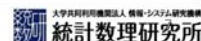


後期1975年11月頃開始

- 先生の講義では, 11月に入り, 累積法などが教えられました
 - 何がなんだか分からず, 後期期末試験では失敗したと思いました.
- **研究とは**
 - 目的: 個性・着想
 - 手段: 技術力・創造力
 - 評価: 実験の計画・確実性・能率
- **関数形による研究技法の分類**
 - 開発調査研究
 - 関数形がある領域で不明: 調査・実験計画
 - 理論の評価
 - 関数形は予想がついているが未知母数がある: 回帰分析
 - 設計計算
 - 関数形が完全に所与: 数値解析, 摂動解析
 - SN比の考え方です。今日の数値実験計画法を予見されていました

2013/05/13

田口玄一先生一周忌シンポジウム

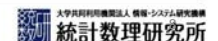


その後の勉強

- 講義が終わってしまったので, 学部の2-4年はひたすら, 統計解析の実践, 田口先生が書かれている文献を勉強しようと思いました.
- 当時, 田口先生は雑誌「現代数学(後にベーシック数学)」に連載をもたれていました
 - 割引係数法, 実験的回帰分析, 分割表の検定に対するFisherの厳密確率計算批判など興味深いテーマが紹介されました.
- 何と言っても, 丸善「実験計画法第三版」上巻, 下巻が順次刊行されていた時期で, 出版を心待ちにしました.
 - 統計解析が大好きだったので, **割引係数法, 精密累積法, 分割表の0-1法(度数法)**などに大変興味を持ちました.
 - 講義で分からなかった累積法が, 本来マハラノビス距離の計算が大変なので, 近似的方法として提唱している.
 - 同様に分散分析よりはエントロピー(情報量)に注目されていることなど, わくわくしながら読みました.

2013/05/13

田口玄一先生一周忌シンポジウム

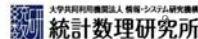


工学部における統計学の講義：懐疑と激賞

- 計数工学科
 - 3年生と教員との談話会
 - 田口先生の講義と上野正先生の講義とに感激してこちらに進学しました。
 - 奥野忠一先生からの質問
 - 何故、2つは正反対
 - 田口さんの言うことを信じているの？
 - 廣津千尋先生の「統計工学(分散分析)」
 - 難解を極め、必死についてゆく
 - 吉村功先生の「数理統計学」を読む
 - 累積カイ二乗法の紹介
 - 累積法が分かったような気になる
 - 統計学の通常の講義では、累積法は必修との誤解
- 高橋馨郎先生
 - 有限体の応用(3年)
 - 実験計画(直交表)の構成
 - 田口玄一先生の功績を激賞
- 久米均先生
 - プロセス解析(3年), 品質管理(4年)
 - 回帰分析の講義
 - 田口先生のことをやはり激賞
 - あらためて田口先生の統計学を勉強して良かったと再確認。

2013/05/13

田口玄一先生一周忌シンポジウム

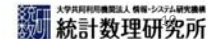


学部卒論のこと：逐次臨床試験への精密累積法適用

- 1978年10月：卒論研究室配置
 - 伏見正則先生(当時, 助教授)
 - 逐次臨床試験の計画と解析
 - 離散データ(分割表型データ)の逐次解析
 - 01010101
 - 00001111:最終時点での成功率はこちらが高い?!
 - 割引係数法と変数選択とは類似的な関係?
 - 0-1法, 精密累積法, 割引係数法を複合した統計解析**
 - 伏見先生の研究とは全く異なることをやらせていただきました
 - 君の研究はこういうことならば、藤野和健先生に相談すれば良かったのに
 - 田口先生の仰ることを信じるばかりではなく、批判的に検討しなさい
 - ちなみに、1982年の修士論文のテーマは、割引係数法(縮小推定)でした
 - 1987年の博士論文の一部は、SN比解析のことでした
 - 1992年ころまでは、**パラメータ設計の思想**を追いかけていました

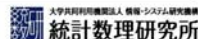
2013/05/13

田口玄一先生一周忌シンポジウム



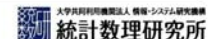
品質誌の特集(1986)16巻2号 「田口の方法によるQC」企画

- 久米均品質誌編集委員長の下で企画担当
- 討論会「田口の方法を巡って」
 - 竹内:田口の実験:その論理と哲学
 - Fisherの実験計画は天上の実験計画、田口さんはこれを地上に引き摺り下ろした
 - 討論:田口、竹内、久米、矢野、廣津、三輪
 - そんなことができたらいかまでですよ
 - 第三号の討論
 - 講義:田口 SN比とその効用
 - 質疑
 - 小生は、統計側参加者に田口先生の最近の仕事をコピーし、キットとして送付
 - テープ起こし:尾島、宮川、椿
 - 宮川先生がタグチメソッドに引きずり込まれた瞬間
- 久米:特集にあたって
- 田口:品質工学への道
- 久米:品質と品種に関する若干の問題
- 田口の方法によるQCへの期待と要望:山本、岩崎
- 田口の品質技術体系概説:中村
- 田口のデータ解析手法「累積法」:森
- オフライン品質工学:矢野
- オフライン品質管理の事例:横山



G. E. P.Boxらとの討論会

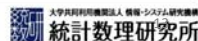
- この討論会には、1987年頃、日本規格協会が実施した幻の第二ラウンドがありました。
 - 出席者、田口、竹内、久米、矢野、廣津、三輪
 - Box, Nair, Wu, Shoemaker
 - ベル研究所のグループのタグチメソッドに関する日本視察で、田口伸先生も来日され参加されていました
 - Leon, Shoemaker, Kacker(1987)PerMIA
 - Performance Measures Independent of Adjustment
 - 加法性はエネルギーに対して成立する
- このテープは残念ながら起こされていません。どこかに眠っているはず！



田口(1962)の 「新版実験計画法下、丸善」

2013/05/13

田口玄一先生一周忌シンポジウム

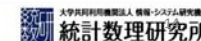


増山元三郎(1956)「実験計画法」岩波全書

- 田口「実験計画法」第1版、1957年、1958年、丸善に対する評価
 - 「独創的なものを多く含み、考えさせることの少ない書物である。」
 - 「ここに含まれた思想を明確に定式化し、これを発展させることも、また日本の若い人たちに残された問題である。」
- 今回のシンポジウムのために遡れたのは1962年版です
 - 統計的アイデアは下巻に満載されています

2013/05/13

田口玄一先生一周忌シンポジウム

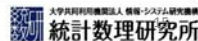


田口(1962)下巻前書き

- 旧版下巻への批判
 - 「上巻と異なって、整理されていない。雑然としていると批判された。全くその通りであった。新版を書き終わって、雑然さが更に倍化したようである」
- 「**雑然さの第一の理由は、統計理論の基礎が確立されていないためではないだろうか**」
- 「**実験家であった実験計画法の創始者R.A. Fisherの立場の中には、私拭しきれないあいまいさがあることは確かな事実である。Fisherの神秘性のヴェールを剥いで、確率的計算の上に理論的基礎を立てようとしたNeyman-Waldの理論のなかには華麗さは認められても、実際家を十分納得させる大地のような実質さがはたして実在しているのだろうか。**」
- 「**統計的計算技術については、本書はNeyman-Waldの立場を中心にしている。確かにt検定やF検定に多くの問題があっても、それはある意味でミニマックス的解になっていると思われるし、近似的なものとして考えるならば実用にかかなりの程度まで堪えられるものと云って良いだろう**」
- 「**理論と実際との間に橋をかけることを自分の任務と考えている著者にとって、統計理論についての橋渡しが十分出来なかったことは、慙愧に堪えないものと云わねばなるまい。統計の数学的基礎について、新しい道が開かれたとき、そのときその下巻はもっと整理された形にまとめることができるだろう。**」

2013/05/13

田口玄一先生一周忌シンポジウム

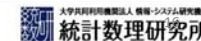


田口(1962)新版実験計画法下巻

- 基礎概念
 - 母数模型と変量模型
- 関数形の推定
- 分散分析
 - 推測過程論的考察
- 欠測値の処理法
 - EMアルゴリズム
- 補助測定値のある実験
 - 共分散分析と逐次修正法
- 逐次分類法
 - AID、CART
- 分散比とその応用
 - SN比、判別分析
- 数量化の方法
 - 変数変換
 - 林の数量化理論、Tobitなどへ
- 数値計算の方法
- 実験回数の決め方
 - 統計的決定理論
- 許容限界
- 不良率の信頼限界
- **累積法**
 - 小数自由度
 - 含む**精密累積法**
- 有効反復数の公式
- 寿命試験
- 直交表の作り方とその応用
 - 最適実験計画
 - 上巻に確率対応法
- 統計理論入門
 - 尤度論
- 特殊問題
 - 割引係数、極値の推定、時系列解析
- 乱数
 - マルコフ性の検定

2013/05/13

田口玄一先生一周忌シンポジウム



田口(1986)品質工学への道

- 「1951~1952年頃は尤度理論に基づく統計的推論の信頼限界や検定の数表作成を行っていたが、偶然とは関係のない誤差因子に起因するバラつきを小さくするには、重要でない参考情報にしかならないことを認識し、尤度法は公表しなかった」

精密累積データという考え方
統計的批判と再評価の必要性
精密累積法にリンクする統計モデル
0,1データ取扱いの非対称性について

精密累積法のセミパラメトリックモデルによる再定式化

田口(1962)の精密累積データと精密累積法

- 1955年インドでの実験からの問題意識？
 - すべてのデータは座標変数 (ω 因子、座標因子) 上で、0-1表現することが可能
 - 田口(1988)品質設計のための実験計画法、JSA
 - データ解析上、情報量大きい！！**
 - 累積法は平均とばらつきを総合的に判断
 - オメガ(座標)因子を加えた精密累積法は、平均と分布系を分離して把握可能！！
 - 分析は度数法で行われている(0-1データ分散分析)
 - » **オメガ因子と制御因子との交互作用**が扱われている

田口(1962)の寿命試験データ

サイクル	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
A1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	3	0	1	0	2	3
A2	0	0	0	0	0	0	0	0	0	0	0	1	2	0	1	1	1	2	1	0	1

- 寿命試験データで20サイクルは目標値
- そこまで実験を行ったと記載
 - 20サイクルに記載されているA1の3件、A2の1件は、以下では右側打ち切りデータと見なす
 - 実際の寿命が20サイクルであった可能性はある
 - 打ち切りデータのある分析事例を示すために単純化
 - 要因Aは2種類のマグネットワイヤー
 - A1はポリエチレン被覆、A2はシリコン被覆

欧米統計学者の批判

- Box, Bisgard and Fung(1988), Pignatiello and Ramberg (2007)
 - 累積法・精密累積法は統計的には有効な方法ではない
 - 田口の負の貢献
- 批判は0.1データに対して、正規線形モデル、分散分析モデルを当てはめを推奨したことに起因
- Box, Bisgard and Fung(1988)
 - 精密累積法を、Cochran(1950)の方法を独立でないデータに対して誤用
- 精密累積法は打ち切りデータに対する尤度上の取り扱いを考慮しない
- Box、Wuらの代替手法提案
 - 精密累積法の解釈ないしは発展というよりは既存の統計的生存時間データ解析の適用
 - 田口の精密累積法のミッションを正当化する統計モデルが存在するかという疑問に答えたものではない。
 - 田口の問題意識⇒Hamada and Wu(1990)の共感
- **タグチメソッドのグローバル化後、1987年の遅すぎた翻訳書に基づく批判**
 - 1970年代の尤度論の成熟がもたらしたGLIMもCoxモデルも開発された後に、普遍的・先鋭の問題意識に基づく1950年代の方法を見せられた欧米

精密累積データのための統計理論モデル

- **準備** : Lindsay(1997)の修正精密データ
 - 正の値をとる確率変数(期間、寿命)のモデル化
 - パラメトリックモデル
 - 確率変数Y
 - » 累積確率分布 $F(x)=Pr(X \leq x)$
 - » 密度関数 $f(x)=dF/dx$ が最尤法における主要な役割
 - 正值確率変数
 - 生存関数: $S(x)=Pr(X \geq x)=1-F(x)$
 - Hazard function(瞬間故障率, 死亡率)
 - » $\lambda(x)=Pr(x \leq X < x+dx | X \geq x)/dx=f(x)/S(x)$
 - 累積ハザード関数 $\Lambda(x)=\int_0^x \lambda(s)ds = -\log S(x)$
 - 従って、ハザード関数を与えれば分布関数が定まる
 - » $S(x)=\exp[-\Lambda(x)], f(x)=\lambda(x) \exp[-\Lambda(x)]$

修正精密データによる分布の再現

- 小区間 n
 - $(n-1)d \leq x < nd$
 - $x_n = (n-0.5)d$
 - 区間 n のハザード率を $\lambda(x_n)$
- 確率変数の値 X が第 N 区間に属する確率
 - 第 n 区間に起きる事象の個数: M_n

$$Pr(M_N = 1) = \prod_{i=1}^{N-1} Pr(M_i = 0) = \lambda(x_N) d \exp(-d \sum_{i=1}^{N-1} \lambda(x_i))$$

≒ $d \lambda(x_N) \exp(-\Lambda(x_N)) \rightarrow$ ハザード関数 $\lambda(x)$ の密度関数近似!

- ω 因子を導入(区間に分割)して、
 - 確率変数が観測されていない区間では0
 - 観測された区間には1を対応させる
- その0, 1列を独立確率変数と見なした非定常ポアソンモデルの尤度は、確かに X の確率密度を近似、打ち切りデータも最終観測を0にするだけで整合的(生存関数の近似となる)

Lindsay(1997)のPMIM: 精密データ分析統計モデル

$$\log \lambda(x_j | z) = \beta_0 + \sum_{i=1}^p \beta_i g_i(x_j) + \gamma^T z$$

Lindsayの提案: 強度関数をパラメトリックモデル

$p=0$ ならば、 $\lambda(x)$ は定数で、指数分布に対応
 $p=1, g_1(x)=\log x$ とすると、ワイブル分布
 $p=1, g_1(x)=x$ とすれば、ゲンベル分布

修正精密累積法

ω 因子を質的因子とするセミパラメトリックなモデリングが修正精密累積法

GAM(一般化加法モデル)に基づくセミパラメトリックモデリング

強度関数 g をGAMでノンパラメトリックに推定するセミパラメトリックモデル
 藤田正彦(1998)「寿命分布における故障率関数の推定—地震生起間隔への適用」
 慶応大学理工学部数理科学科椿研卒業論文のアイデア
 ハザード関数自体を推定し、分布形情報(ハザード関数)を抽出

Coxの比例ハザードモデルは、ハザード関数を推測しない接近法

田口(1962)のデータの再解析

修正精密累積データ行列 (一部表示)

行番号	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	...	341	342	343	344	345	346
event	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	...	0	0	0	0	0	0
cycle	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	...	15	16	17	18	19	20
A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	...	1	1	1	1	1	1

修正精密累積法

ω 因子: 21水準

要因Aとの交互作用も含めた分散分析モデル

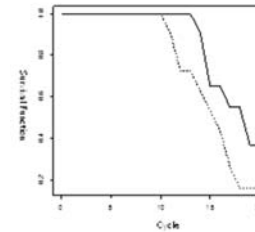
$$Y_{it} \sim \text{Po}(\mu_{it})$$

$$\log \mu_{it} = \mu + t_i + a_j + (t \times a)_{ij} \quad i=1, \dots, 21, j=1, 2$$

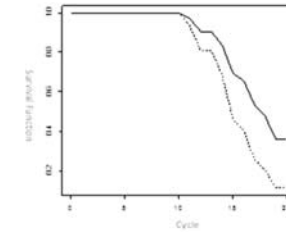
識別のための制約条件としては $i = j = 1$ の添え字を持つ要因効果を0

修正精密累積法: Analysis of Deviance (尤度比検定、AIC)

	Df	Deviance	Resid. Df	Resid. Dev	AIC
NULL			345	98.363	132.363
factor(cycle)	20	43.926	325	54.437	128.437
A	1	2.171	324	52.265	128.265
factor(cycle):factor(A)	20	8.345	304	43.920	159.920



交互作用A × ω を含む生存関数



Aの主効果のみの生存関数

ワイブル型PMIM

$$\log \mu_{it} = \mu + \beta_1 \log(t+0.5) + \beta_2 \{1 \log(t+0.5)\}^2 + a_j \quad t=0, \dots, 20, j=1, 2$$

$\log(\text{Cycle}+0.5)$ の多項式項を導入した PMIM の ANODEV の結果

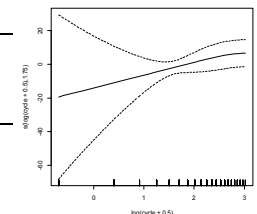
	Df	Deviance	Resid. Df	Resid. Dev	AIC
NULL			345	98.363	132.363
$\log(\text{cycle} + 0.5)$	1	31.788	344	66.575	102.575
$I(\log(\text{cycle} + 0.5)^2)$	1	3.569	343	63.006	101.006
A	1	2.024	342	60.982	100.982
$A * \log(\text{cycle} + 0.5)$	1	0.861	341	60.121	102.121
$A * \log(\text{cycle} + 0.5)^2$	1	0.910	340	59.211	103.211

セミパラメトリックモデル(GAM)

$$\log \mu_{it} = \mu + \lambda(\log(t+0.5)) + a_j \quad t=0, \dots, 20, j=1, 2$$

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-8.2616	4.0093	-2.061	0.0393
A	0.7595	0.5113	1.486	0.1374

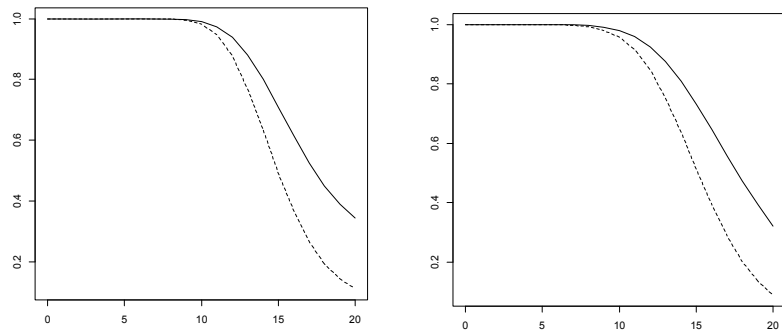
	edf	Ref. df	Chi. sq	p-value
$s(\log(\text{cycle} + 0.5))$	1.748	2.188	3.341	0.212



偏回帰関数

λ はスプライン回帰で推定

左図PMIM, 右図GAMの生存関数



2013/05/13

田口玄一先生一周忌シンポジウム

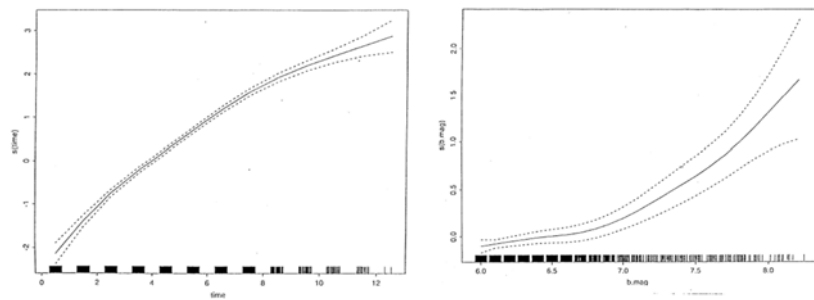
慶応大学工学部数理科学科樁研究室(1987/04-1997/03)
最後の卒業論文: 藤田(1998)の分析: 修正精密データ

- 地震発生間隔データ(分単位データ1/5乗変換)
 - 東京大学地震研究所WEBの気象庁決定日本付近地震データ[1926/01/01-1998/08/31]
 - マグニチュード6以上
 - 地域別、全日本T1, T2, T3,
 - その修正精密データ表現(time)
 - 000000100010000010000001
- 要因
 - 直前地震のマグニチュード(b.mag)
 - 現地震のマグニチュード(mag)
 - 前地震と現地震との距離(len)
- モデル[一般化加法モデル]
 - $\text{Log}(\lambda(t)) = f_1(\text{time}) + f_2(\text{b.mag}) + f_3(\text{mag}) + f_4(\text{len})$

2013/05/13

田口玄一先生一周忌シンポジウム

オメガ因子効果と前地震規模効果



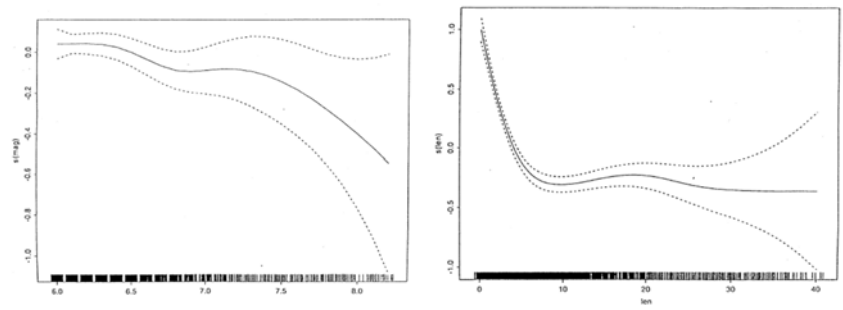
Increasing hazard

巨大地震は余震を誘発

2013/05/13

田口玄一先生一周忌シンポジウム

今回の地震規模と前回地震の距離



今回の地震が大きくなるほど期間は空く

経度・緯度が3以上はなれると殆ど余震震源にはなりにくい

2013/05/13

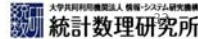
田口玄一先生一周忌シンポジウム

再び精密データについて一謎の部分

- 3つの問題提起
 - 生起した時0, 非生起で1という逆転
 - ハザード関数(故障しやすさ)に対するモデリングではなく、逆ハザード関数(f/F :故障しにくさ)に対するモデリング
 - 結果や解釈は異なる(対称ではない)
 - 死亡、故障などのイベントの精密データは途切れるが、そうでないイベント(スイッチのオンオフ)
 - インパルスのイベントは修正精密データ
 - ステップ的イベント
 - 田口先生はステップ的イベントを考えていらっしゃるのか?
 - 本当に分割実験(Repeated Measurement ANOVA)か?
 - 1次誤差、2次誤差を精密累積法では導入
 - Lindsayの離散尤度近似だけではない、統計的思考?

2013/05/13

田口玄一先生一周忌シンポジウム



大学共同利用機関法人 情報・システム研究機構
統計数理研究所

おわりに

- 増山元三郎先生がご指摘されたように、若手統計科学者にとって、田口先生のデータ分析に対する深い見識と独創に基づいて提示された技法群を紐解けば、今でも実に面白い研究が進展するとともに、既成の統計的観念への惑溺とらわれない自由な発想を涵養できるのではないかと考えます。
- 田口先生は、人間の進歩というものを自由度の増大と考えていたように思います。
 - 私自身が統計実務家を目指したため、田口先生の統計学に実質的貢献をせず、解釈に留まることは、少し残念ですが、自身の性分と力量とに見合ったものでしょう

2013/05/13

田口玄一先生一周忌シンポジウム



大学共同利用機関法人 情報・システム研究機構
統計数理研究所

「なお、本書の内容に関する議論や質問については、巻末にある著者の住所宛に御手紙頂ければ幸甚である。」

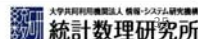
プリンストン大学、ガウスの家にて、1962年3月9日

田口玄一

質問の手紙はもう出せないのが残念です
ご冥福を心よりお祈り申し上げます

2013/05/13

田口玄一先生一周忌シンポジウム



大学共同利用機関法人 情報・システム研究機構
統計数理研究所