平成18年3月10日　公開講座時系列解析入門

応用例：状態空間モデルによる時系列データ解析

樋口知之

情報・システム研究機構　統計数理研究所
JST, CREST

1

ISM

1．線形状態空間モデル（復習をかねて）

2

マイクロマーケティング：
居酒屋店日次売上高の予測

＜共同研究＞

山口類 （九州大学大学院・数理）

文献：

山口類, 土屋映子, 樋口知之, 状態空間モデルを用いた飲食店売上の要因分解, オペレーションズ・リサーチ誌, Vol.**49** No.5, 316-324, 2004

ISM

One-to-Oneへ

Mass Marketing

Segmentation

消費者の多様化、市場の成熟。
広告費用対効果の低下。

Micro Marketing

すべての消費者を対象として

大量生産・大量流通・大量プロモーション。

地理的、**人口学的**に比較的**小さな区画**に、
ターゲットを分割。

例えば、チェーン店舗を一括して同様にオペレーションするのではなく、個々の店舗の置かれている環境を考慮して個別にオペレーションする経営手法。地域の気候的特性、競合状態、地域住民の年齢構成、所得階層、人種構成、趣味嗜好の特性などを十分に研究した上で、商品構成、価格レイジン、店員の対応をきめ細かく行う。

Micro Marketing

突き詰めると

個への細分化

One-to-one

ISM

マイクロマーケティングとは：ITとともに

- 背景：

  各種大量データベースが利用可能

  次のスライドでさらに詳しく

  －詳細なセンサス・データ

  －POSデータ

  膨大な情報を的確に分析

- 最適なマーケットの発見，出店計画，DM戦略（含：Eメール．HTMLメールに個人の趣向に合わせた広告のみ載せる）

5

ISM

ある飲食店の日々売上データの予測モデル構成を通じて，

- 飲食店売り上げを説明する状態空間モデルの構築
- “現場の勘”の数量化
- 各項を考察することにより新たな知見の抽出

- 店舗固有の情報に基づくマーケティング

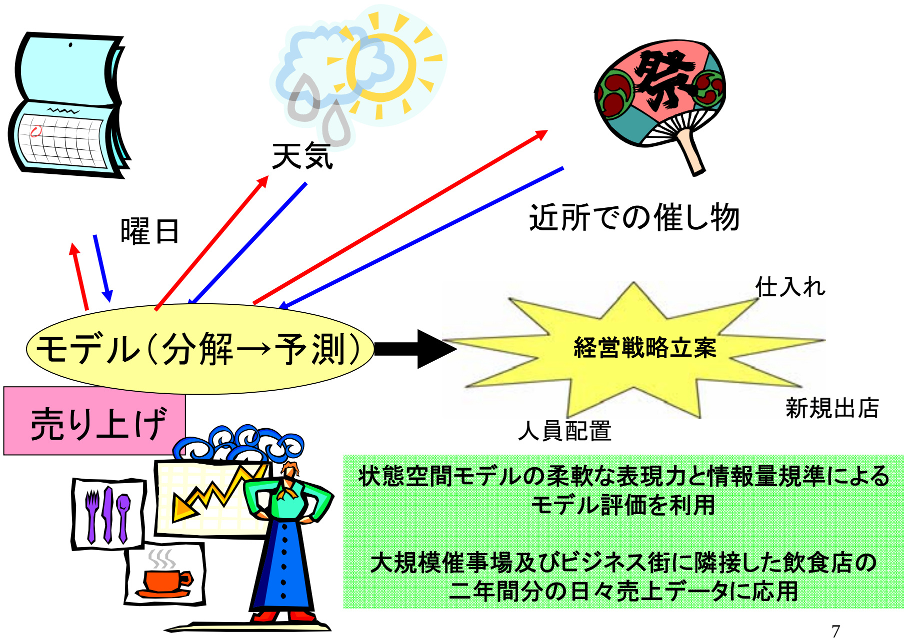
  （**マイクロマーケティング**）

  ー個人経営の単体店舗へのソフト

  ー新規出店の際の，店舗の売り上げのシミュレーション

6

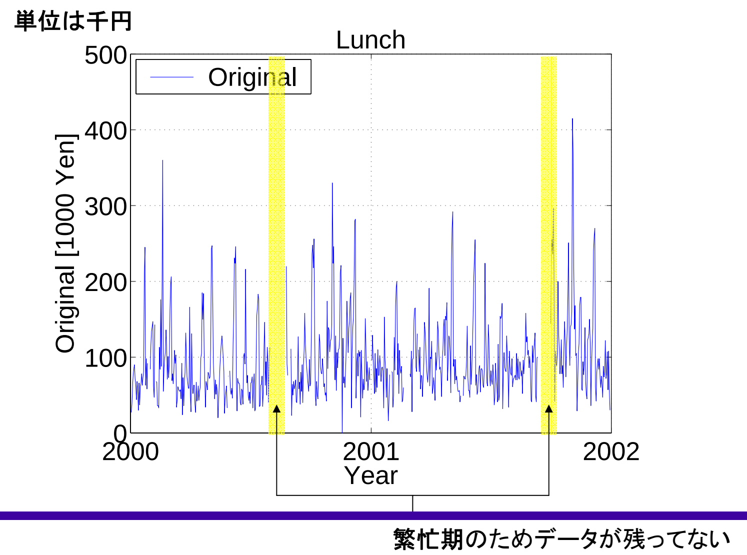
ISM



**状態空間モデルの柔軟な表現力と情報量規準による
モデル評価を利用**

**大規模催事場及びビジネス街に隣接した飲食店の
二年間分の日々売上データに応用**

売上げ原系列（ランチ）



ISM

日データの分解モデル

$$y_n = t_n + W_n + R_n + E_n + r_n + \varepsilon_n$$

売上げ

トレンド

$$t_n = 2t_{n-1} - t_{n-2} + v_{t,n}$$

$$v_{t,n} \sim N(0, \tau_t^2)$$

週効果

雨効果

イベント効果

AR項

$$r_n = \sum_{i=1}^{2} a_i r_{n-i} + v_r$$

$$v_r \sim N(0, \tau_r^2)$$

残差

n： 2000年1月1日からの日数 $(n = 1, \ldots, 731)$

ISM

モデル(トレンド)

長期的変動　　二次トレンドモデル

$$y_n = \boxed{t_n} + W_n + R_n + E_n + r_n + \varepsilon_n$$

$$\boxed{t_n} = 2t_{n-1} - t_{n-2} + v_{t,n}$$

$$v_{t,n} \sim N(0, \tau_t^2)$$

分散の値は最尤法で決める

階層	ベイズモデル	状態空間モデル
レベル1	パラメータ(潜在変数)	状態ベクトル
レベル2	ハイパーパラメータ	パラメータ

ISM

モデル（週効果）

$$y_n = t_n + \boxed{W_n} + R_n + E_n + r_n + \varepsilon_n$$

Indicator functions

	1	0
h_{1n}	月～金の祝日	それ以外
h_{2n}	祝日でない月～木かつ翌日が祝日	それ以外

$$\boxed{W_n} = \boxed{w_n} + h_{1n}\boxed{\beta_1(w_{日,n} - w_n)}$$
$$+ h_{2n}\boxed{\{\beta_2(w_{金,n} - w_n) + \beta_3(w_{土,n} - w_n)\}}$$

基本パターン 季節成分モデル（周期7）		$\sum_{j=0}^{6} w_{n-j} = v_{w,n},\ v_{w,n} \sim N(0, \tau_w^2)$
祝日効果	祝日の売上げがどのくらい直近の日曜の売上げに似ているか	類似度 $0 \le \beta_1 \le 1$
祝前日効果	祝日の売上げがどのくらい直近の金曜または土曜の売上げに似ているか	$0 \le \beta_2, \beta_3 \le 1$ $0 \le \beta_2 + \beta_3 \le 1$

ISM

モデル(雨効果）

$$y_n = t_n + W_n + \boxed{R_n} + E_n + r_n + \varepsilon_n$$

$$\boxed{R_n} = \gamma_R f_R(X_{2,n})$$

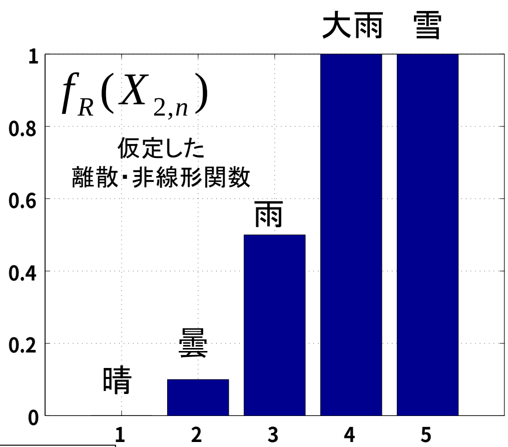
γ_R ：定数[千円]

$X_{2,n}$ ：天気（1,2,3,4,5）

天候が悪いとき

$\gamma_R > 0$ ならば売り上げ増

$\gamma_R < 0$ ならば売り上げ減



ローカルかつ時間単位（午前，午後）の情報を利用する予定

店長によってカテゴライズされた情報

$X_{2,n}$ ：天気

12

ISM

モデル(イベント効果):基本項

$$y_n = t_n + W_n + R_n + \boxed{E_n} + r_n + \varepsilon_n$$

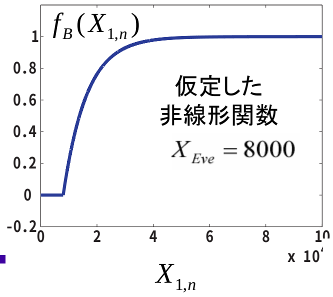
$$\boxed{E_n} = \underbrace{\alpha_{B,n} f_B(X_{1,n})}_{\text{基本項}} + I(f_B(X_{1,n}))(\alpha_w \underbrace{f_w(d_n) + h_{3n}\alpha_h)}_{\text{補正項}}$$

ランダムウォークモデル

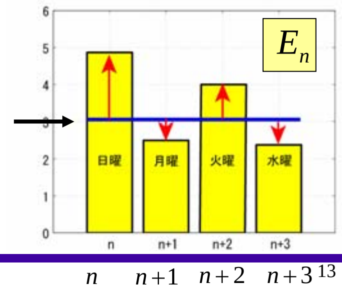
$$\alpha_{B,n} = \alpha_{B,n-1} + v_B, v_B \sim N(0, \tau_B^2)$$

$X_{1,n}$:予想平均入場者数

例: 予想延入場者:10000人
会期:4日(日曜~水曜)
$X_{1,n+i} = 2500$人;$i = \{0,1,2,3\}$



基本項 →



13

ISM

モデル(イベント効果):補正項

$$y_n = t_n + W_n + R_n + \boxed{E_n} + r_n + \varepsilon_n$$

$$\boxed{E_n} = \underbrace{\alpha_{B,n} f_B(X_{1,n})}_{\text{基本項}} + I(f_B(X_{1,n}))(\alpha_w \underbrace{f_w(d_n) + h_{3n}\alpha_h)}_{\text{補正項}}$$

状態空間表現
$\alpha_{W,n} = \alpha_{W,n-1}$

$I(\cdot)$:階段関数

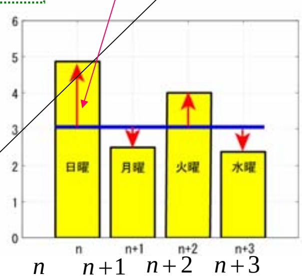
$f_w(d_n)$:曜日ごとの補整項

α_w :定数

α_h :休日嵩上げ項(定数)

h_{3n} Indicator function
1:休日{土日祝}、0:それ以外

E_n



14

ISM

分解の結果

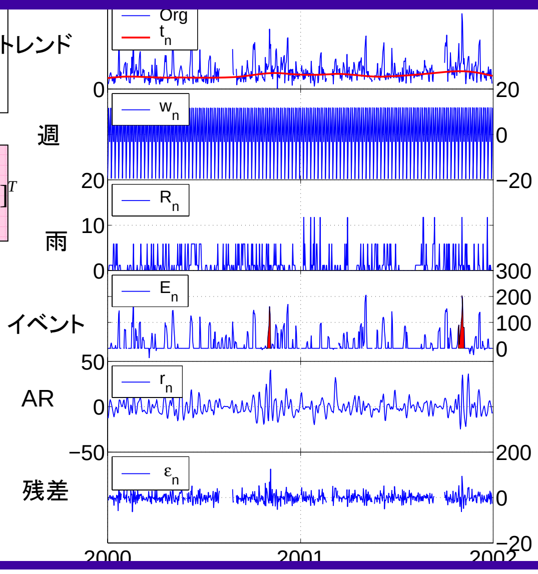
信号が定常となるよう
Parcorへ変換
さらにロジット変換

パラメータベクトル

$$\theta = [\sigma^2, \tau_t^2, \tau_w^2, \tau_B^2, \tau_r^2, \beta_1, \beta_2, \beta_3, \boxed{a_1, a_2}]^T$$

グリッドサーチにより
パラメータの最尤推定値
を求め、時系列を分解

各々，適切に非線形変換
された空間で最適化



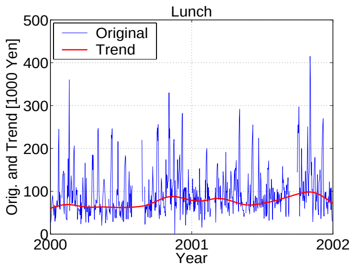
ISM

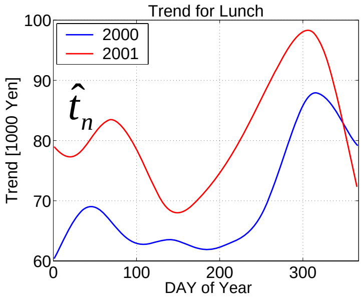
結果：原系列、トレンド

$$t_n = 2t_{n-1} - t_{n-2} + v_{t,n}$$

長期的変動

$\hat{t}_n$





2年間

1年間

以後，状態変数推定値は平滑化による値

16

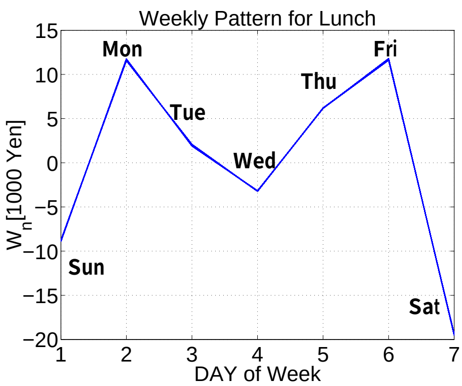
ISM

結果：週効果（基本パターン）

$$W_n = \boxed{w_n} + h_{1n}\beta_1(w_{日,n} - w_n) + h_{2n}\{\beta_2(w_{金,n} - w_n) + \beta_3(w_{土,n} - w_n)\}$$

$$\boxed{w_n} = -\sum_{j=1}^{6} w_{n-j} + v_{w,n}$$

$\hat{w}_n$



週末に売り上げが減る

ビジネス街からの顧客の動向

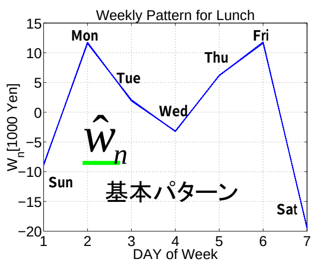
結果：祝日効果、祝前日効果

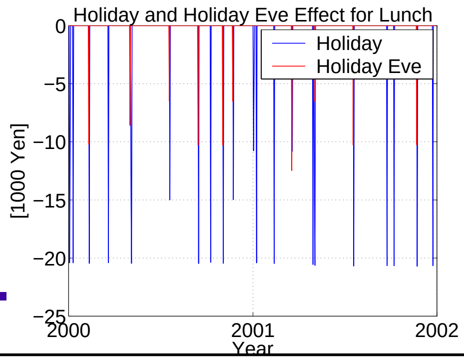
$$W_n = w_n + h_{1n}\beta_1(w_{日,n} - w_n) + h_{2n}\{\beta_2(w_{金,n} - w_n) + \beta_3(w_{土,n} - w_n)\}$$

$\hat{\beta}_1 = 1, (h_{1n}, h_{2n}) = (1,0)$ $\longrightarrow$ $\hat{W}_n^{祝} = \hat{w}_{日,n}$

$(\hat{\beta}_2, \hat{\beta}_3) = (0, 0.4)$ $\longrightarrow$ $\hat{W}_n^{祝前} = 0.6\hat{w}_n + 0.4\hat{w}_{土,n}$

$(h_{1n}, h_{2n}) = (0,1)$

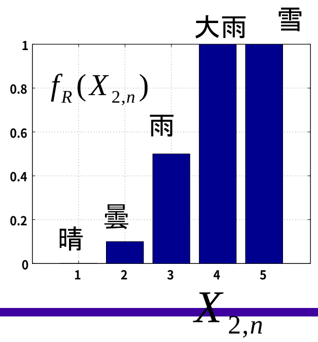




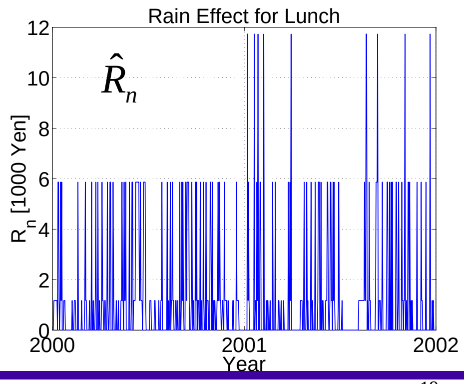
結果：雨効果

$$R_n = \gamma_R f_R(X_{2,n})$$

$$\hat{\gamma}_R = 11.7 > 0$$



天候が悪いと売り上げが伸びる。
ビジネス街の顧客の動向の反映。



19

ISM

結果：イベント効果　（基本項）

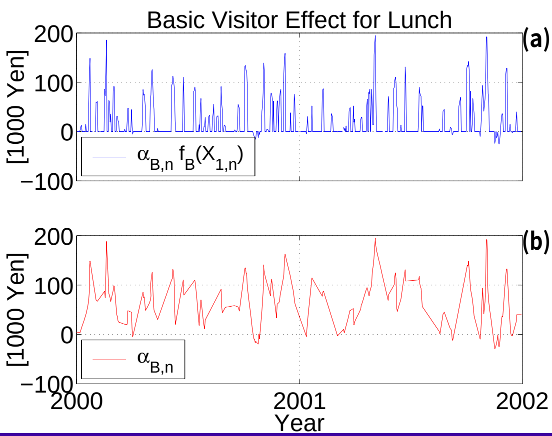
$$E_n = \boxed{\alpha_{B,n} f_B(X_{1,n})} + I(f_B(X_{1,n}))(\alpha_w f_w(d_n) + h_{3n}\alpha_h)$$

$\hat{\alpha}_{B,n} f_B(X_{1,n})$

イベント予想入場者数に応じて変動

他項に比べて大きく売上げを説明

$\hat{\alpha}_{B,n}$



20

ISM

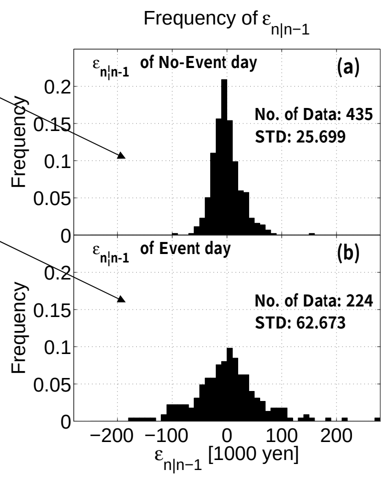
予測誤差分布 $\varepsilon_{n|n-1} = y_n - H_n x_{n|n-1}$

イベント効果なしの日
標準偏差~26000円

イベント効果ありの日
標準偏差~63000円

予測誤差の大きなイベント
（例）子供向け、物産展
有名人講演

イベントの分類



予測誤差の大きいもの

- 最近はイベントの情報が過去のものも含めてかなりネット上にある．予測誤差の大きい日のイベント情報をネットから調べてみた．
- 予測誤差の大きいものは，
  - イベント参加予想人数が実際とぜんぜん違う（実人数は，後日開催者から報告されているものなどで追調査できる）
  - イベントのタイプが，特殊

など，ネットで情報を検索すると，概して**定性的**なレベルにおいては原因を**すぐ類推**しやすいものも多いが，．．．

観測機器

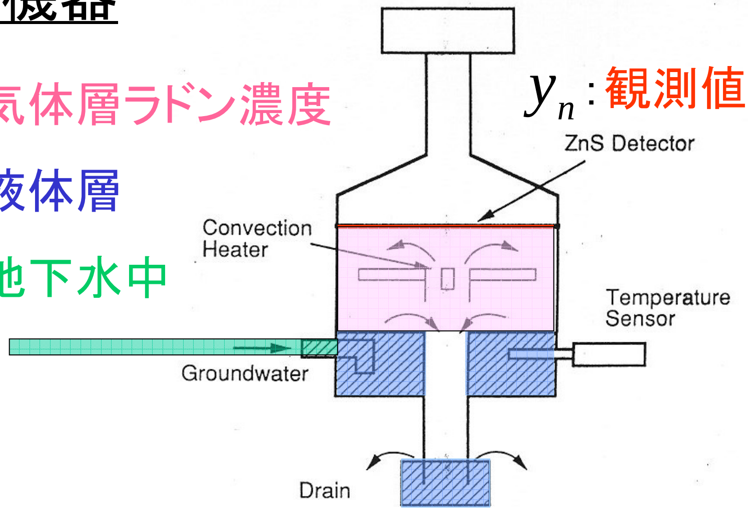
(Nakatani et al., 1991)

C_g : 気体層ラドン濃度

C_l : 液体層

C_0 : 地下水中

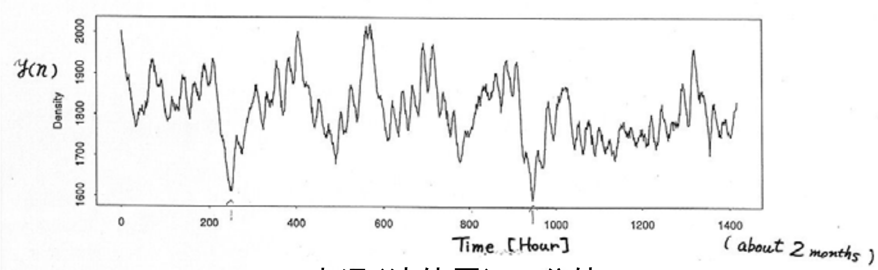
y_n : 観測値



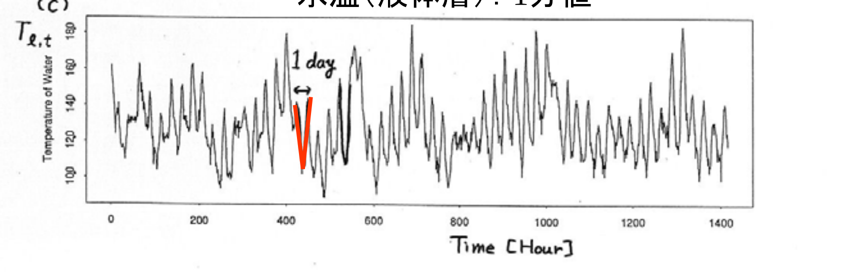
Schematic figure of the chamber for detecting α ray from groundwater radon.

観測値

気体層中のラドン濃度: 1時間値



水温(液体層): 1分値



Differential Equations

これだけでは問題はとけない

$$\frac{dC_g}{dt} = a\,C_g + b\,C_l$$

$$\frac{dC_l}{dt} = e\,C_g + f\,C_l + g\,C_0$$

$$a = -(\lambda + \frac{k_l S}{\boxed{H_T} V_g}), \quad b = \frac{k_l S}{V_g}$$

$$e = \frac{k_l S}{\boxed{H_T} V_l}, \quad f = -(\lambda + \frac{Q}{V_l} + \frac{k_l S}{V_l}), \quad g = \frac{Q}{V_l}$$

$$\frac{1}{H_T} = 0.50774 - 2.836\times10^{-2} T_l^1 + 4.683\times10^{-4} T_l^2 - 4.058\times10^{-6} T_l^3$$

$$T_{l,n} = T_{surface,n} = T_{obs,n} + (\frac{\beta}{T_{obs,n}} - 1)$$

25

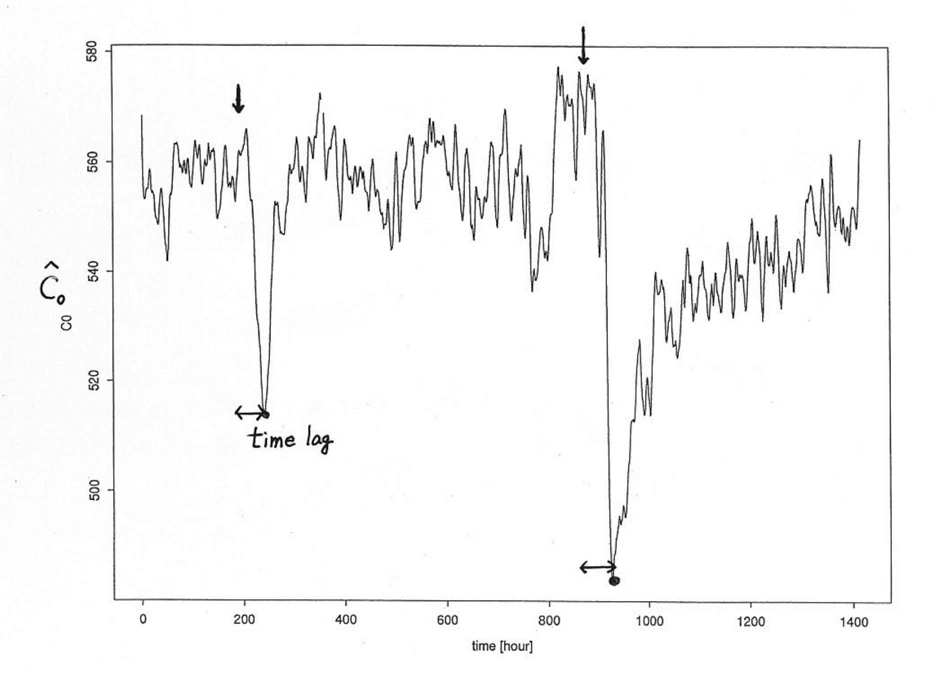
Stochastic Differential Equations

- Assumption $\frac{dC_0}{dt} = w_t$ ← 事前情報
  - w_t : scaler white noise with the auto-correlation function of $E[w_t w_{t'}] = \tau^2 \delta(t - t')$
- State Space Representation

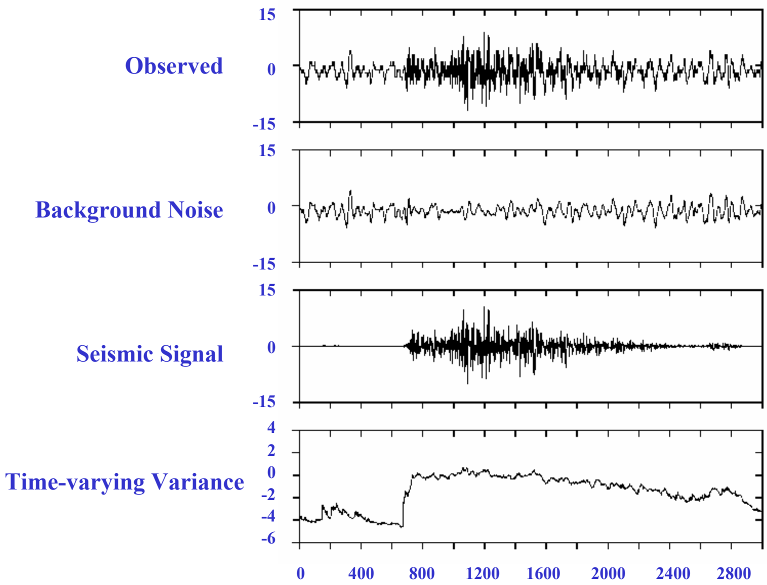
$$\frac{dx_t}{dt} = F_t x_t + G_t w_t$$

$$x_t = \begin{bmatrix} C_g \\ C_l \\ C_0 \end{bmatrix}, \quad F_t = \begin{bmatrix} a & b & 0 \\ e & f & g \\ 0 & 0 & 0 \end{bmatrix}, \quad G_t = G = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

26



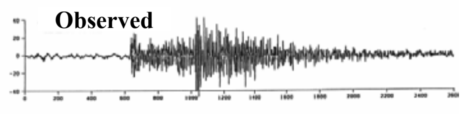
微小信号の抽出



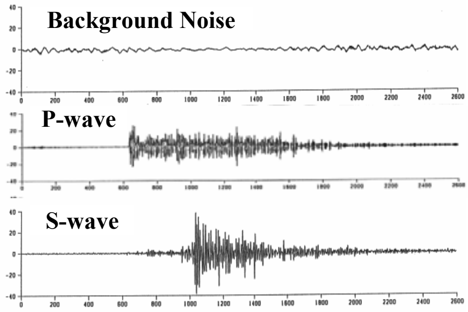
28

★このスライドは，北川　統計数理研究所長から借用

P−波とS−波の分離



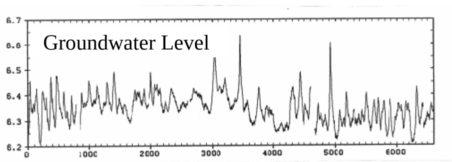
$$y_n = BN_n + P_n + S_n + w_n$$



29

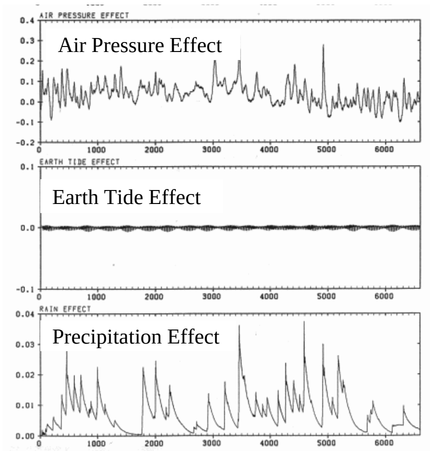
★このスライドは，北川 統計数理研究所長から借用

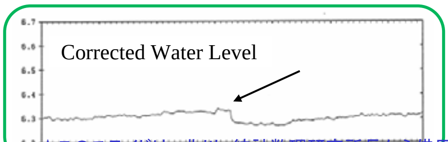
Extraction of Coseismic Effects



M=4.8, D=48km

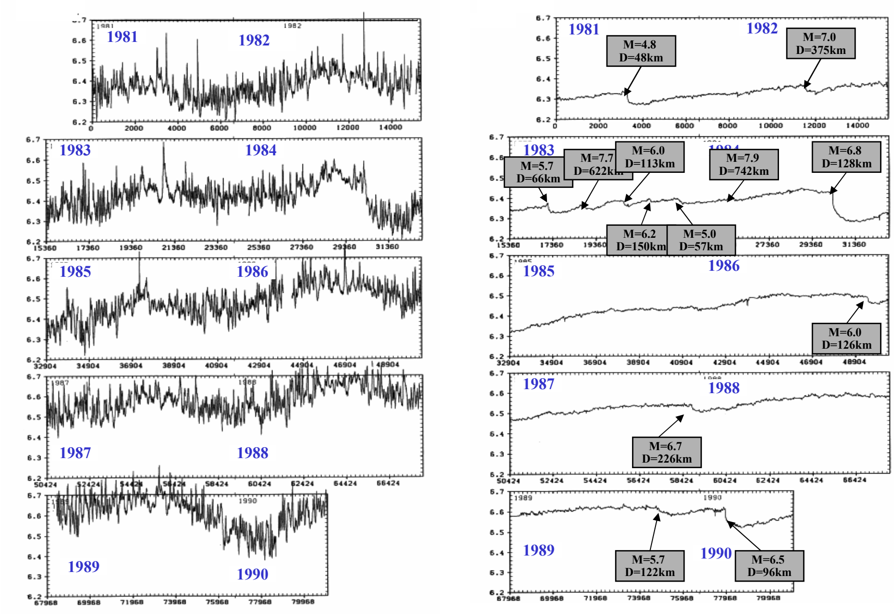
min AIC model
m=25, l=2, k=5





★このスライドは，北川 統計数理研究所長から借用

地震の影響の検出



制御問題

[線形2次評価問題]

$$x_t = Fx_{t-1} + Eu_{t-1} + Gv_t$$

$$y_t = Hx_t \boxed{+ e_t}$$

x_t ：状態ベクトル

u_{t-1} ：操作変数 (与えられる入力)

y_t ：被制御変数（出力）

$$u_{t-1} \equiv 0$$

32

2．一般状態空間モデル

33

観測されない非価格プロモーション実施の有無のPOSデータからの統計的推測法

（財）流通経済研究所
佐藤　忠彦

統計数理研究所
樋口　知之
北川源四郎

文献：
T.Sato, T.Higuchi, and G.Kitagawa, Statistical Inference using Stochastic Switching Models for the Discrimination of Unobserved Display Promotion from POS Data, *Marketing Letters*, Vol.**15** No.1, 37-60, 2004

34

ISM

研究目的

通常POSデータより獲得できるデータから、**教師データが存在しない下で**、店頭における非価格プロモーション実施状況を推定すること

研究対象のデータの特徴

本研究で使用するデータには、**通常はPOSデータでは獲得できない非価格プロモーション実施の有無のデータがある**

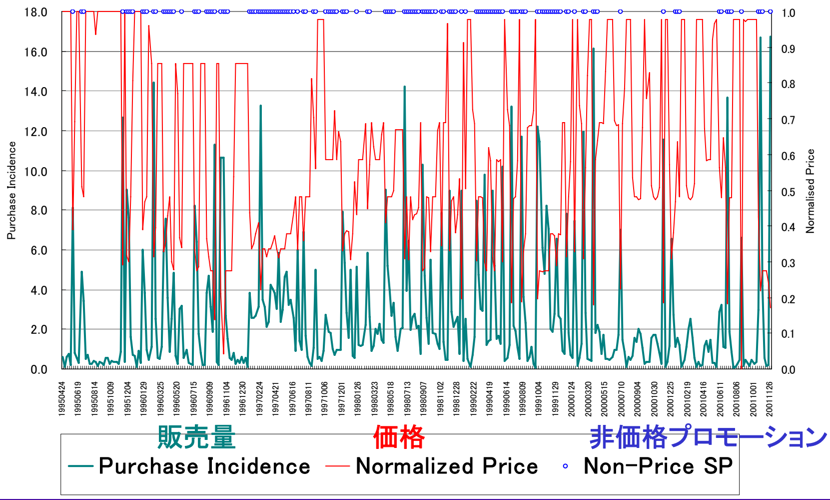
- POSデータのみから、非価格プロモーション実施の有無を推定するモデルを構築し、そのモデルの精度を検証できる！！
- 非価格プロモーション実施の有無はモデルの検証の際にしか使用しない！！

35

ISM

使用したデータ：点数PIと正規化価格の系列

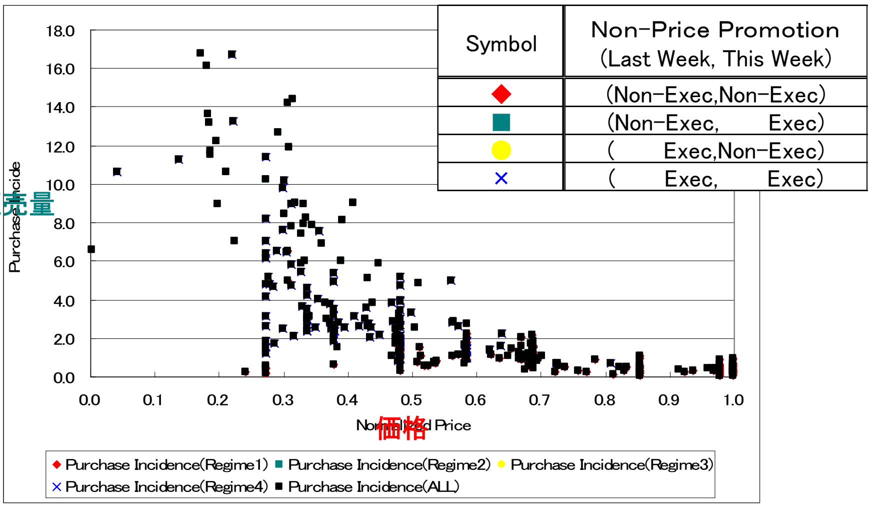
週データ



36

ISM

使用したデータ：
点数PI×価格掛け率散布図



37

ISM

モデル化

- ✓ 非価格プロモーション実施の有無を<u>潜在的な状態</u>として捉え、その状態が切り替わると仮定する
- ✓ 販売点数は売価に影響され変動する
- ✓ 各状態毎に売価帯の分布には差がある

<u>**潜在的な状態**</u>を、非価格プロモーション実施の有無の<u>**先週と今週の組み合わせ**</u>で、以下のように仮定してモデル化

（先週、今週）＝（非実施、非実施）→状態1
（先週、今週）＝（非実施、　実施）→状態2
（先週、今週）＝（　実施、非実施）→状態3
（先週、今週）＝（　実施、　実施）→状態4

38

ISM

一般状態空間モデル

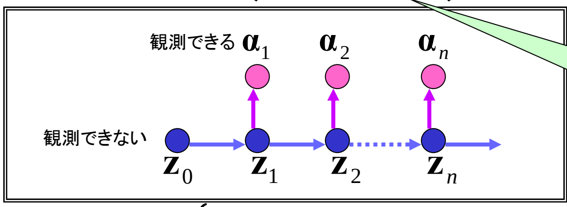
遷移確率行列: $q_{ij} \equiv q(Z_n = j \mid Z_{n-1} = i)$

信号の発生メカニズムを表すモデル

$$Z_n \sim q(Z_n \mid Z_{n-1})$$ ← システムモデル

$$\alpha_n \sim r(\alpha_n \mid Z_n)$$ ← 観測モデル

その信号を実際に観測するとき信号が変換されノイズが加わる様子を表すモデル



$$Z_n = \{regime1, regime2, regime3, regime4\}$$

regime1 = 1, regime2 = 2, regime3 = 3, regime4 = 4

39 ISM

観測モデル(4状態)

x_n ：正規化価格

y_n ：点数PI

分布モデル

$$g_j(x_n \mid Z_n = j) = \frac{x_n^{p_j-1}(1-x_n)^{q_j-1}}{B(p_j, q_j)}, j = 1, \cdots, 4$$

回帰モデル

$$f_j(y_n \mid C_j + \alpha_j x_n, Z_n = j) \sim N(C_j + \alpha_j x_n, \sigma_j^2), j = 1, \cdots, 4$$

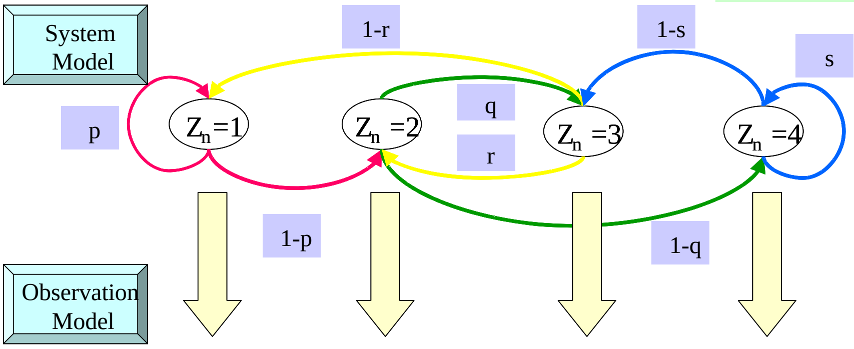
2変量モデル

$$h_j(y_n, x_n \mid Z_n = j) = f_j(y_n \mid C_j + \alpha_j x_n, Z_n = j) g_j(x_n \mid Z_n = j), j = 1, \cdots, 4$$

40 ISM

コンセプトモデル（4状態モデル）

x_n：正規化価格
y_n：点数PI



MSD	$g(x_n \mid Z_n=1,\psi^x_{n-1})$	$g(x_n \mid Z_n=2,\psi^x_{n-1})$	$g(x_n \mid Z_n=3,\psi^x_{n-1})$	$g(x_n \mid Z_n=4,\psi^x_{n-1})$
MSR	$f(y_n \mid x_n, Z_n=1,\psi^y_{n-1})$	$f(y_n \mid x_n, Z_n=2,\psi^y_{n-1})$	$f(y_n \mid x_n, Z_n=3,\psi^y_{n-1})$	$f(y_n \mid x_n, Z_n=4,\psi^y_{n-1})$
MSB	$h(y_n, x_n \mid Z_n=1,\psi^{xy}_{n-1})$	$h(y_n, x_n \mid Z_n=2,\psi^{xy}_{n-1})$	$h(y_n, x_n \mid Z_n=3,\psi^{xy}_{n-1})$	$h(y_n, x_n \mid Z_n=4,\psi^{xy}_{n-1})$

RBFとMSモデルとの判別力の比較

- RBF：データの<u>前半部分</u>で学習、<u>後半</u>で判別（←現実には使えないことに注意）
- MSモデル：後半のみで教師なし学習、かつ判別

Model	Valiable	Non-Execution	Execution	Total
RBF	Misclassification Rate	0.15	0.19	0.17
			15%	
	Number of Misclassification Sample	15	14	29
MSB	Misclassification Rate	0.15	0.22	0.18
			15%	
	Number of Misclassification Sample	15	16	31

本研究のマーケティング上の活用領域

- **非価格プロモーションの効果測定の側面**

→店頭の大量陳列の実施の有無を判別できるため、効果的なプロモーション計画・価格戦略の立案が可能になる

- **店頭需要予測の側面**

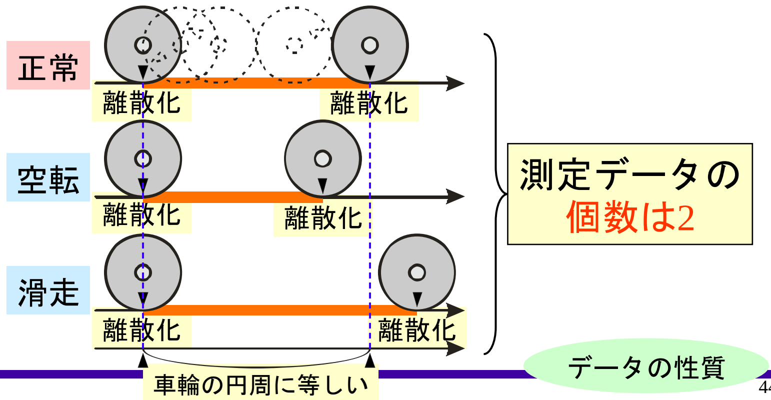
→非価格プロモーション実施時の効果が定量的に評価できるため、精度の高い需要予測が実現できる

等

43 ISM

サンプリング間隔の伸縮の原因

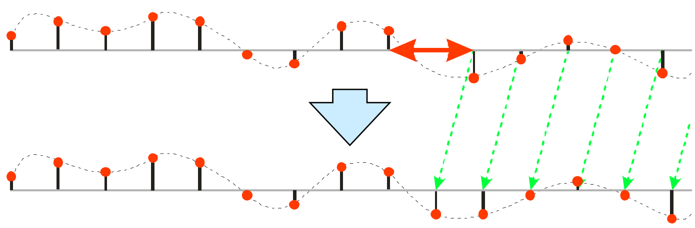
- 車輪の空転・滑走による距離パルスの伸縮
  - データの個数の増減にもつながる
  - 検出できない



44 ISM

問題設定

- 不均一サンプリング（位置は不明）が混在する空間系列データ

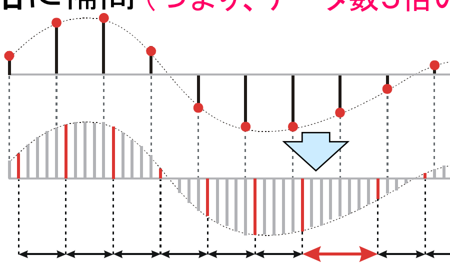


- 不均一サンプリング区間を推定し，
  - サンプリング間隔を均一に修正

45

補正方法

- 学習データセットのサンプリング間隔を，**1/5** 倍に補間（つまり、データ数5倍の学習データを作成）



- $n_t - n_{t-1} > 5$：空転
- $n_t - n_{t-1} = 5$：通常
- $n_t - n_{t-1} < 5$：滑走

t 番目の教師データセットに合わせて，データポイントn_tを選択

DPと最尤法による解法

46

ISM

$\boldsymbol{n}_{1:T}(=n_1,\ldots,n_T)$を評価する関数

- 評価関数を下記3要素の和によって定義

1) 残差系列をARモデルで予測した誤差の2乗和

$$\sum_{t=1}^{T}\left[Y_t - X_{n_t}\right]^2$$

2) 1階差分≠5 のペナルティ

$$\mu_1 \sum_{t=2}^{T} \xi(n_t - n_{t-1} - 5, 4)$$

3) 2階差分≠0 のペナルティ

$$\mu_2 \sum_{t=3}^{T} \xi(n_t - 2n_{t-1} + n_{t-2}, 8)$$

Y_t：t番目の教師データの値
n_t：t番目の教師データに対応する学習データの番号
X_{n_t}：n_t番目の学習データの値
T：教師データの数

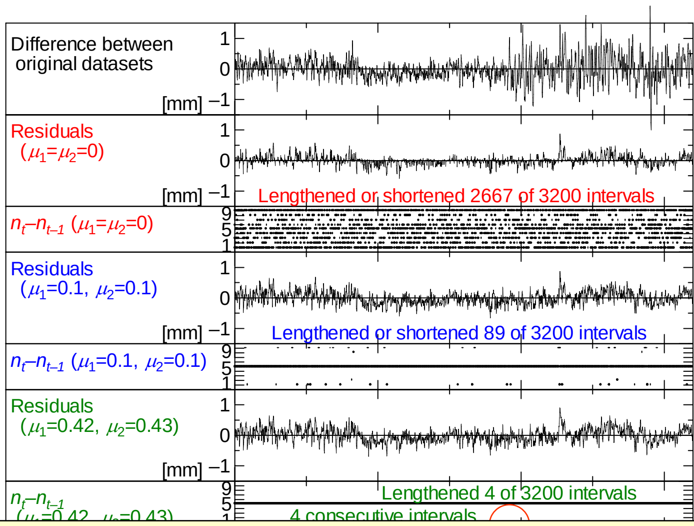
$$\xi(n,\gamma) = \begin{cases} 0 & (n = 0, |n| > \gamma) \\ 1 & (1 \le |n| \le \gamma) \\ \infty & (\gamma < |n|) \end{cases}$$

- 動的計画法（DP）を用いて最小化できる
- パラメータを合理的に決める必要がある

ISM

Optimization

Solution Depends on μ_1 and μ_2



The parameters should be carefully selected.

48

ISM

$\boldsymbol{n}_{1:T}$の最適解の解釈

最適化問題→統計モデルへ

min

$$\sum_{t=1}^{T}\left[Y_t - X_{n_t}\right]^2$$

残差系列をARモデルで予測した誤差の2乗和

$$\mu_1 \sum_{t=2}^{T} \xi(n_t - n_{t-1} - 5{,}4)$$

1階差分≠5 のペナルティ

$$\mu_2 \sum_{t=3}^{T} \xi(n_t - 2n_{t-1} + n_{t-2}{,}8)$$

2階差分≠0 のペナルティ

3つの項の和

変形

max

$$\prod_{t=1}^{T} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}\left[Y_t - X_{n_t}\right]^2\right\}$$

n_t が全て($\boldsymbol{n}_{1:T}$)与えられたときの y_t 全て($\boldsymbol{Y}_{1:T}$)の尤度 $p(\boldsymbol{Y}_{1:T}|\boldsymbol{n}_{1:T})$

$$\times \exp\left[-\frac{\mu_1}{2\sigma^2} \sum_{t=2}^{T} \xi(n_t - n_{t-1} - 5{,}4)\right]$$

$$\times \exp\left[-\frac{\mu_2}{2\sigma^2} \sum_{t=3}^{T} \xi(n_t - 2n_{t-1} + n_{t-2}{,}8)\right]$$

$\boldsymbol{n}_{1:T}$の事前分布 $p(\boldsymbol{n}_{1:T})$

- $p(\boldsymbol{n}_{1:T}|\boldsymbol{Y}_{1:T}) \propto p(\boldsymbol{Y}_{1:T}|\boldsymbol{n}_{1:T})\ p(\boldsymbol{n}_{1:T})$であるから，
- 事後分布を最大化する$\boldsymbol{n}_{1:T}$ に等しい

解法

ISM

一般状態空間表現への変換

$$y_t \overset{\text{def}}{=} Y_t - \sum_{i=1}^{k} a_i Y_{t-i}$$

$$x_t \overset{\text{def}}{=} X_{n_t} - \sum_{i=1}^{k} a_i X_{n_{t-i}}$$

観測モデル

$$y_t = x_t + w_t \overset{\text{def}}{=} h(\boldsymbol{x}_t) + w_t$$

$$w_t \sim N(0, \sigma^2)$$

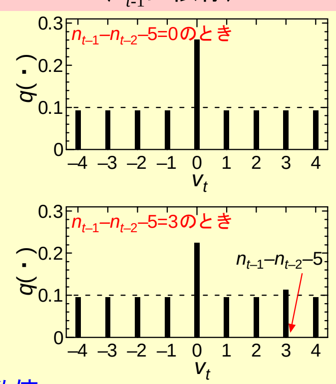
観測ノイズ：正規分布に従う実数値

システムモデル

$$\boldsymbol{x}_t \overset{\text{def}}{=} \begin{bmatrix} n_t \\ n_{t-1} \end{bmatrix} = \begin{bmatrix} n_{t-1} + 5 + v_t(\boldsymbol{x}_{t-1}) \\ n_{t-1} \end{bmatrix}$$

$$v_t \sim q(\cdot|n_{t-1}, n_{t-2})$$ システムノイズ：整数値

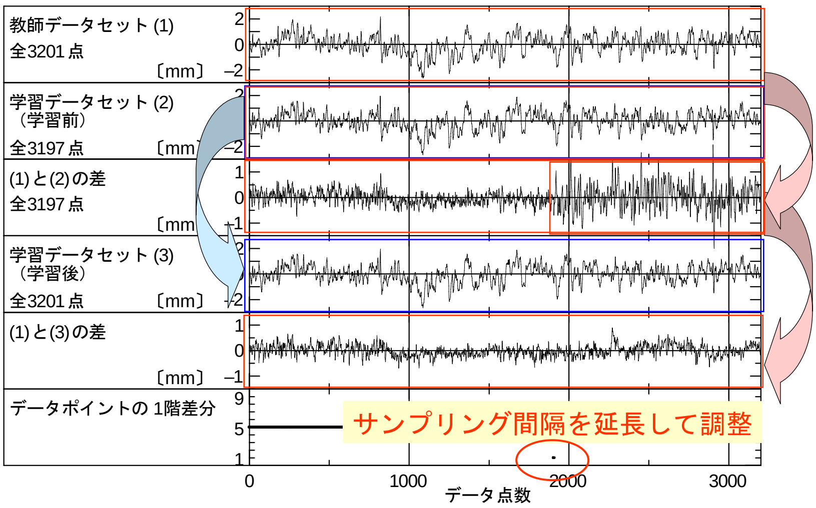
システムノイズの確率分布
（$\boldsymbol{x}_{t-1}$に依存）



非線形フィルタ演算が直接計算できる と最尤法による解法

ISM

計算例

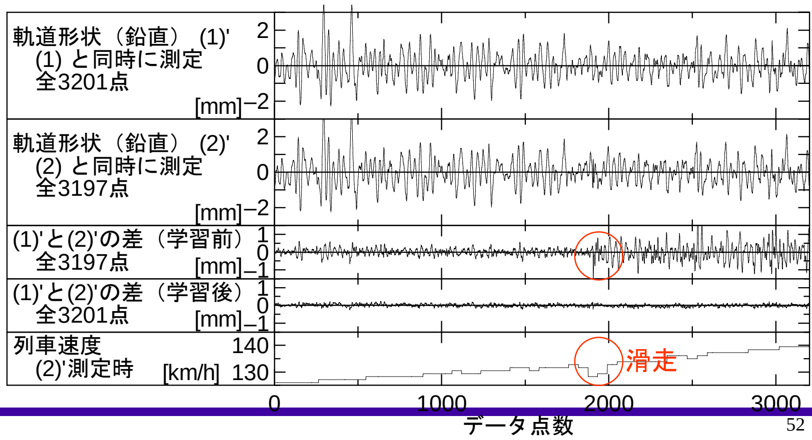


51

ISM

実測値に対する推定結果

•軌間データによって推定した位置補正結果を，同時に測定した鉛直方向の軌道形状に対して適応した例



52

ISM

少数カウントデータの季節成分推定

(Higuchi, CSDA 1999)

交通事故月別死者・重傷者数

$$y_t \sim Poisson(\lambda_t)$$
$$\log \lambda_t = \mu_t + s_t + \alpha \cdot I_t$$

μ_t : トレンド成分

s_t : 季節変動成分

I_t : Indicator function

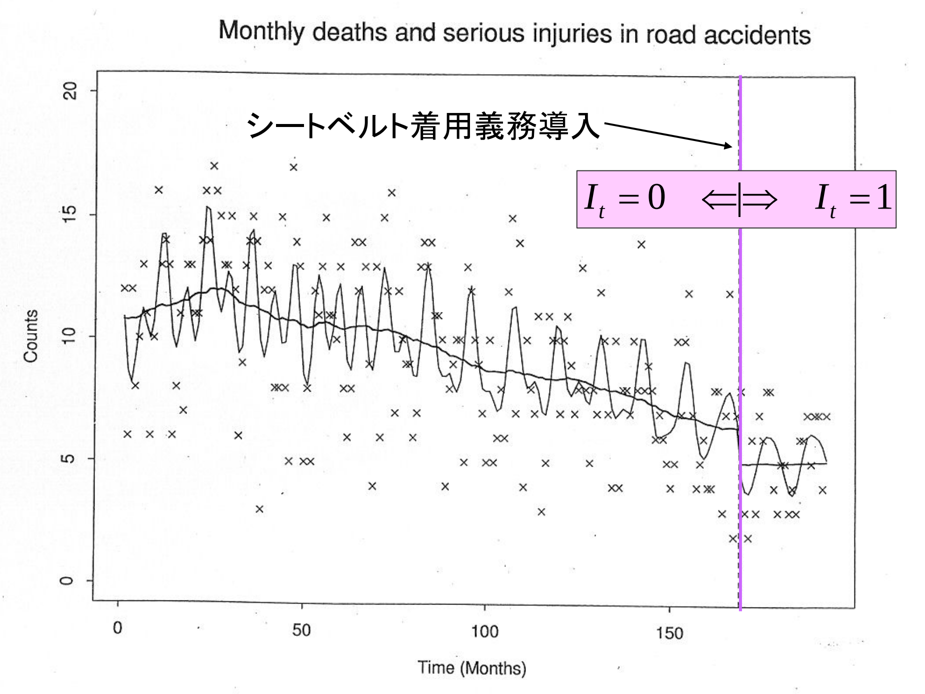
・Deterministic (parametric) Model:

$$s_t = \sum_{j=1}^{J} \gamma_{j,c} \cos(2\pi t f_j) + \gamma_{j,s} \sin(2\pi t f_j), \quad f_j = \frac{1}{12}, \frac{1}{6}, \frac{1}{4}, \ldots$$

・Stochastic Model:(form-free monthly seasonal pattern)

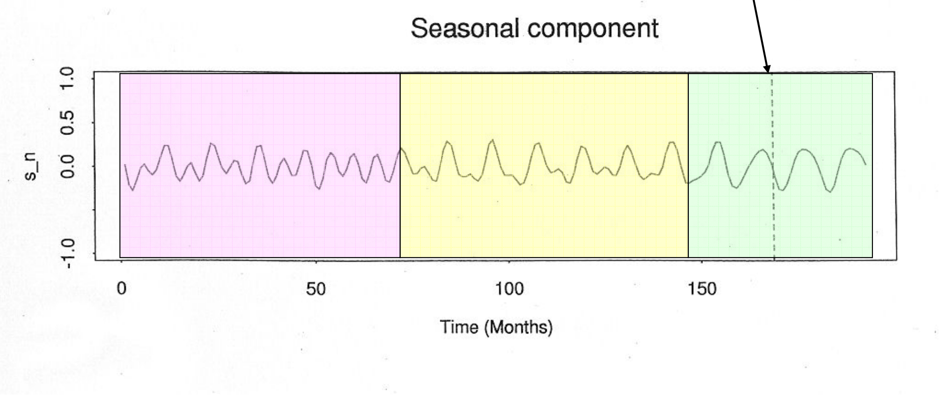
$$s_t \sim s_{t-12}$$
$$s_t = -\left(s_{t-11} + s_{t-10} + \cdots + s_{t-1}\right) + v_{t,s}, \quad v_{t,s} \sim N(0, \tau_s^2)$$

53



季節変動成分

シートベルト着用義務導入



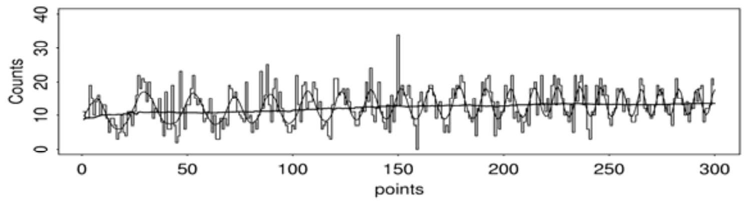
55

季節変動成分の変化

$\exp(\mu_t + s_t)$

56

Prometeus 7:6



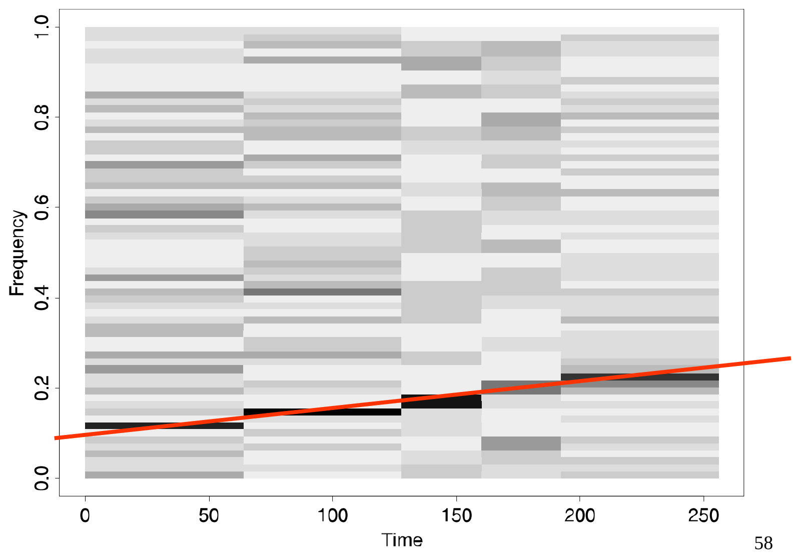
$$y_t \sim Poisson(\lambda_t)$$

$$\log \lambda_t = \mu_t + s_t$$

s_t :wavy component

57

時間・周波数分解



58

時変周波数の推定

- Wavy component: $s_t = 2\cos(2\pi \cdot f_t)\, s_{t-1} - s_{t-2}$

- Frequency: $f_t = \dfrac{0.5}{1+\exp(-\beta_t)}, \quad (-\infty < \beta_t < +\infty)$

$$\beta_t = \beta_{t-1} + v_{t,\beta}, \quad v_{t,\beta} \sim N(0, \tau^2_{t,\beta})$$

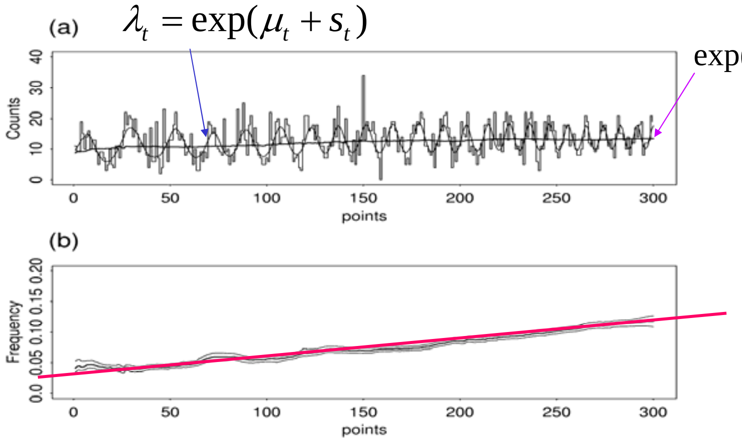
$$\mathbf{x}_t = \begin{bmatrix} \mu_t \\ s_t \\ s_{t-1} \\ \beta_t \\ \log_{10}\tau^2_{t,\mu} \\ \log_{10}\tau^2_{t,\beta} \end{bmatrix} = \begin{bmatrix} \mu_t + v_{t,\mu} \\ 2\cos\left(\dfrac{\pi}{1+\exp(-\beta_{t-1})}\right) s_{t-1} - s_{t-2} \\ s_{t-1} \\ \beta_{t-1} + v_{t,\beta} \\ \log_{10}\tau^2_{t-1,\mu} \\ \log_{10}\tau^2_{t-1,\beta} \end{bmatrix}, \quad \begin{matrix} v_{t,\mu} \sim C(0, \tau^2_{t,\mu}) \\ v_{t,\beta} \sim N(0, \tau^2_{t,\beta}) \end{matrix}$$

59

Prometeus 7:6

$\lambda_t = \exp(\mu_t + s_t)$

$\exp(\mu_t)$



A theory of the spiral density wave suggests
a linear increase in f_t .

60

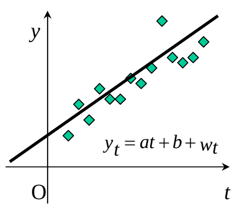
3．データ同化

61

データ同化とは？（1）

要するに、
観測データにモデルをあてはめること

データに直線をあてはめる



データにシミュレーションモデルをあてはめる

＝「シミュレーションモデルにデータを同化する」

データ同化

複雑なモデルを扱うことになるのが特徴

気象・海洋分野での"方言"

62

ISM

データ同化とは？(2)

- 気象学・海洋学の分野で発達
- 物理数値シミュレーションモデルと実際の観測を統合する手法
  - シミュレーションのみでは適切に物理現象を再現できない
    - シミュレーションモデルには，モデルの不完全性や境界条件が正確にはわからないなどの不確かさが存在
    - (たとえば)正確な気象予測には適切な初期条件の構成が必要
  - 観測データは物理的・社会的制約により得られる情報に限界がある

⇒ 観測データを用い数値シミュレーション内の変数を修正

＝データ同化

ISM

シミュレーションモデルの構成

(日本周辺の簡易化した気象モデルの例を用いて説明)

実システムに対応した偏微分方程式
(連続時間・空間)

$$\frac{\partial x}{\partial t} = cx^2 + \cdots$$

数値シミュレーションモデル
(離散時間・空間，有限差分方程式)

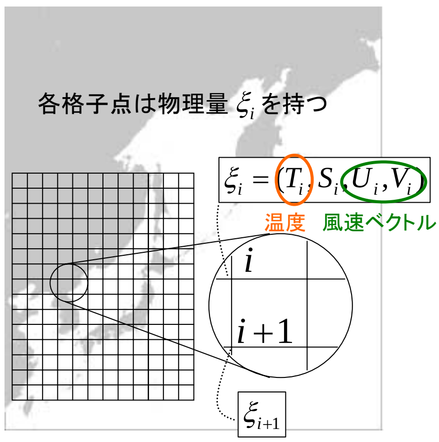
$$x_n = f_n(x_{n-1})$$

境界条件・モデル化誤差
由来の不確かさ

v_n

非線形状態空間表現のシステムモデル
(離散時間・空間，確率差分方程式)

$$x_n = f_n(x_{n-1}, v_n)$$



観測点・観測される物理量の種類は限定

ISM