

## 創立記念講演会要旨

## 複雑すぎる現象をどう捉えるか

## ——大規模なパラメトリックモデル——

統計数理研究所 石 黒 真 木 夫

(昭和60年10月26日, 統計数理研究所 講堂)

## 1. 例1: 地球潮汐データの解析

## 1.1 地球潮汐

太陽と月の引力によって潮の干満が引き起こされることはよく知られているが、地球そのものもごくわずかながら同じ力による変形を起こしている。地球潮汐と呼ばれる現象である。この現象は、地面の傾きの変動、2地点間の距離の変動、あるいは重力の変動として、それぞれ傾斜計、歪計、重力計によって観測される。地球潮汐は、太陽および月の引力を入力する地球というシステムの出力であり、入出力の関係、特に周波数応答を知ることによって地球の内部に関する知識を得ることができる。

## 1.2 データの特徴

入力である太陽、月の引力は理論的に精密に計算できるので、分かっているものとしてよい。これに対して出力の測定にはいろいろと問題がある。たとえば重力は、おもりを下げたバネの長さを測ることによって、原理的には、測定できる。ただしその測定を非常に精密に行なわなくてはならないことから、さまざまな困難が生ずる(装置全体を数センチ持ち上げて地球の中心からの距離を変えたことが測定にかかる程度の測定が必要である)。とりわけ大きな問題の一つとして、測定の原点が時間とともに変動する(たとえば重力計のバネが伸びる)ことなどによるトレンドの存在が挙げられる。正弦波の重ね合わせで表現される地球潮汐に時間的にゆっくりした変動  $t_i$  と観測誤差  $\varepsilon_i$  が加わった、

$$y_i = \sum_n a_n R_n^* \cos(\omega_n^* i - p_n^* + q_n) + t_i + \varepsilon_i$$

が実際に観測されるデータである。ここで  $i$  は時刻を示すインデックスである。右辺第一項が地球潮汐であり、

$$\sum_n R_n^* \cos(\omega_n^* i - p_n^*)$$

で定義される「理論潮汐」を線形システムを通して観測した形をしている。理論潮汐は時刻  $i$  における太陽と月の地球に対する位置関係から計算される。

トレンドを含まない

$$y_i = \sum_n a_n R_n^* \cos(\omega_n^* i - p_n^* + q_n) + \varepsilon_i$$

の形でデータが得られるなら、観測誤差に分散  $\sigma^2$  の正規白色雑音を仮定して導かれるデータ分

布

$$l = \prod_i \frac{1}{\sqrt{2\pi}\sigma} \exp \left\{ (y_i - \sum_n a_n R_n^* \cos(\omega_n^* i - p_n^* + q_n)) \right\}$$

のパラメータを最尤法と AIC 最小化によって決定する方法で「地球潮汐」を構成する正弦波の振幅などが推定できる。地球潮汐の項を三角関数の公式を使って書き直すと

$$\sum_n \{ a_n R_n^* \cos(\omega_n^* i - p_n^*) + \beta_n R_n^* \sin(\omega_n^* i - p_n^*) \}$$

となり最尤法は線形最小二乗法に帰着するので計算は簡単である。

トレンドが単純な形をしている場合には、これを多項式で近似して観測値を

$$y_i = \sum_n a_n R_n^* \cos(\omega_n^* i - p_n^* + q_n) + \sum_k \gamma_k i^k + \varepsilon_i$$

と表現し、多項式の係数も推定すべきパラメータとみなすことによって AIC 最小化で処理できる。しかし実際のトレンドは複雑で多項式などではうまく表現できない。

### 1.3. 大規模パラメトリックモデルのあてはめ

トレンドを多項式で近似するのをやめて各時点の値そのものをパラメータとみなしてモデルを構成する。この場合のパラメータが多すぎて最尤法では推定できないのでトレンドの変動がゆるやかであることを表現する事前分布

$$\pi = \left( \frac{v}{\sqrt{2\pi}\sigma} \right)^c \exp \left\{ -\frac{v^2}{2\sigma^2} \sum_i |t_i - 2t_{i-1} + t_{i-2}|^2 \right\}$$

を導入し、データ分布とあわせてベイズモデルとする。事前分布の広がりを規定する超パラメータ  $v^2$  の値を

$$ABIC = -2 \log \int l \pi dt$$

の最小化によって決定したときの事後分布（これは  $l\pi$  に比例する）のモードを推定値として採用することによってトレンドが適度に平滑化された図 1 のような地球潮汐データの分解が得られる。

## 2. 例 2: 電波望遠鏡データに基づく像の回復

### 2.1. 電波天文学

地球に届く星の光りを観測する天文学はよく知られているが、電波の観測も行われている。電波天文学と呼ばれる分野である。光の場合と同じように金属板を鏡とした反射望遠鏡、電波望遠鏡によって観測される。電波の輝度分布を知ることによって遠い宇宙に関する知識を得ることができる。

### 2.2. データの特徴

同じ電磁波ではあるが、電波の波長が光にくらべてはるかに長いことから、困難が生ずる（望遠鏡の分解能は口径と波長の比で決まるので、波長 1 cm 前後の電波に対して口径 1 m の光学望遠鏡と同じと同じ分解能を得ようとすれば口径 20 km の電波望遠鏡が必要である）。直径 20 km の望遠鏡を作るわけにはいかないので電波干渉計というテクニックが用いられる。簡単な

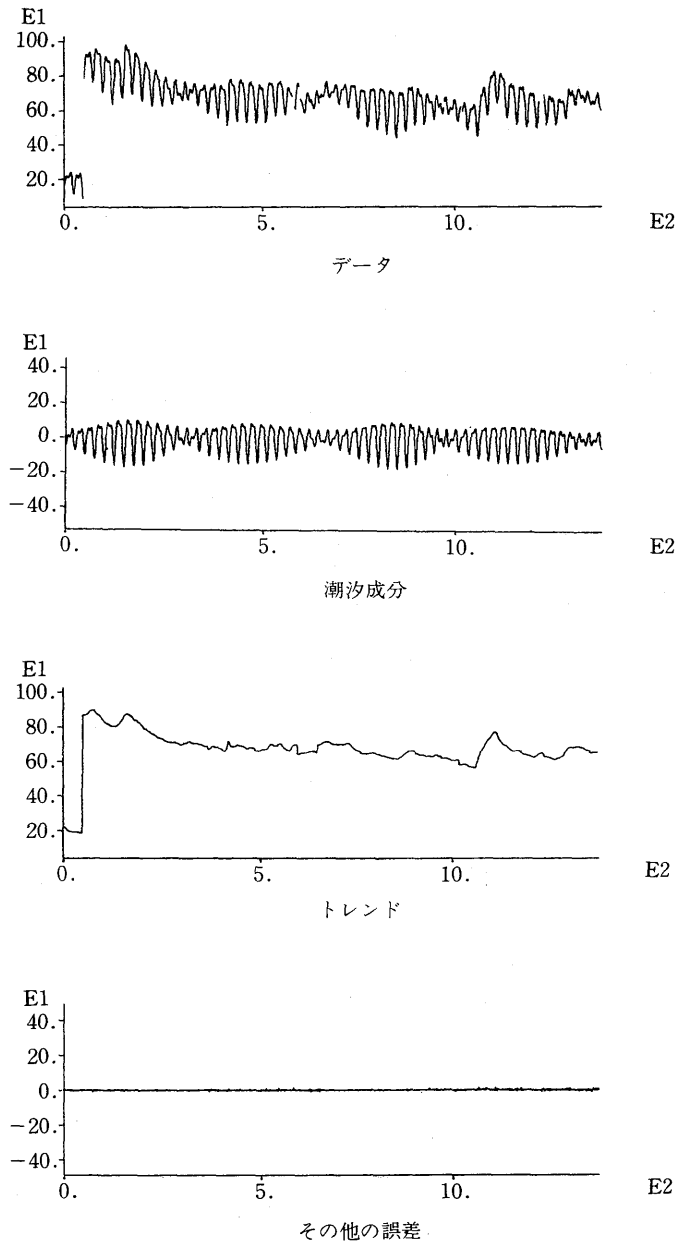


図 1. 地球潮汐データの分解

ために(黄道に沿った電波源のような)一次元の単色電波源の強度分布を  $f(x)$  で表すことにする。距離  $s$  だけ離れた 2 台のアンテナに入ってくる電波の複素相関をとると  $f(x)$  のフーリエ変換の  $s$  成分が得られる(図 2 を参照)。

$0 \leq s \leq S$  なるすべての  $s$  について誤差を含まないデータが得られればフーリエ逆変換によって口径  $S$  の反射望遠鏡で見たのと同様な  $f(x)$  の像が回復できることになる。しかし  $S$  を 20 km にとってこれを実行するのは不可能である。実際に得られるのは比較的少ない数の

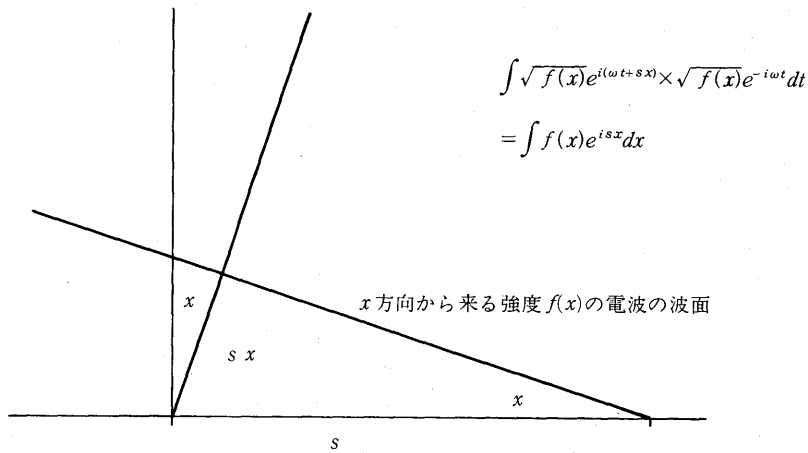


図2. 電波干渉計

データ =  $f(x)$  のフーリエ変換の  $s$  成分 + 観測誤差

でしかない。

電波源分布が正規分布の重ね合わせで表現できるような単純な形をしている場合には、最尤法と AIC 最小化によるモデルあてはめによって構成要素である正規分布の重みなどのパラメータを決定する方法が使える。しかし実際には複雑な電波源が多く正規分布などではうまく表現できない。

### 2.3. 大規模パラメトリックモデルのあてはめ

天球にメッシュを切りその各部分の電波強度そのものの対数をパラメータとみなし、観測誤差に正規分布を仮定するとデータ分布

$$l = \prod_j \frac{1}{2\pi\sigma^2} \exp \left\{ -\sum_n l^n \int I_n(x) e^{i s_j x} dx \right\}^2$$

が構成される。ここで  $I_n$  は  $n$  番目のメッシュでのみ 1、その外では 0 となる関数、 $s_1, s_2, \dots$  は電波望遠鏡の配置から決まる定数である。これに電波源が星雲状に広がって分布していることを表現する事前分布

$$\pi = \left( \frac{v}{\sqrt{2\pi}\sigma} \right)^c \exp \left\{ -\frac{v^2}{2\sigma^2} \sum_n |t_n - 2t_{n-1} + t_{n-2}|^2 \right\}$$

をあわせてベイズモデルとする。ABIC を最小化する超パラメータで決まる事前分布で決定されるパラメータの事後分布から適度に平滑化された図 3 のような電波源分布の像が得られる。

## 3. 例 $n$ : $\times\times\times\times$ データに基づく $\times\times\times\times$ の解析

### 3.1. $\times\times\times\times$ 学

$\times\times\times\times$  学の分野で  $\times\times\times\times$  の観測が行われている。 $\times\times\times\times\times\times$  を知ることで  $\times\times\times\times$  に関する知識を得ることができる。



