



<第5回>
統計的時系列モデリング

統計数理研究所

川崎 能典

kawasaki@ism.ac.jp

統計的時系列モデリング

概要

データサイエンティストとして時系列解析を学ぶ際に知っておくべき体系について、定常・非定常、データの変換、平滑化事前分布、状態空間モデルをキーワードに学ぶ

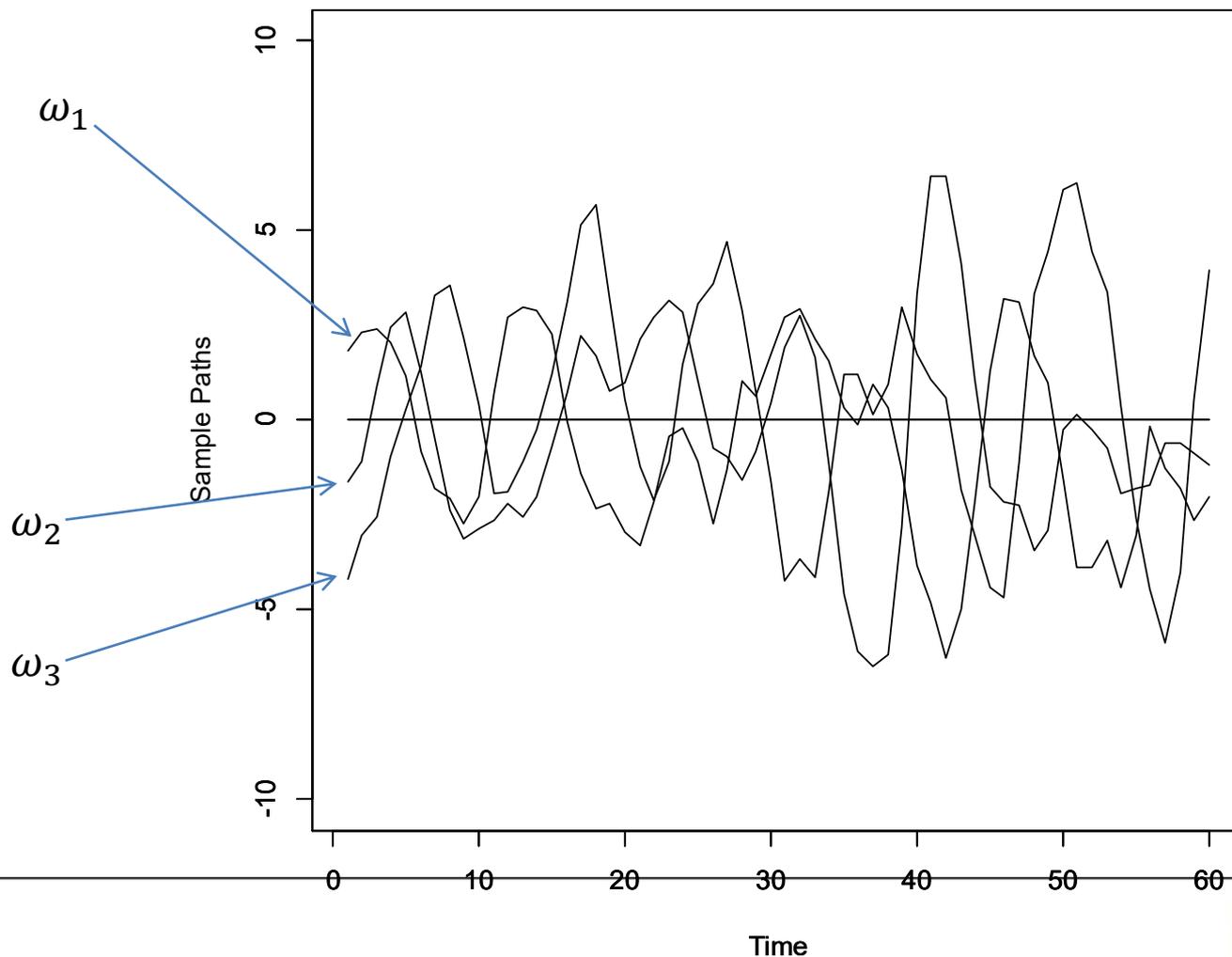
1. 定常時系列
2. 定常から非定常へ: 変換の利用
3. 定常から非定常へ: 平滑化事前分布
4. 状態空間モデル

1. 定常時系列

- 時系列データ=確率過程の実現値(見本過程)
- $X(t, \omega)$: 時間 t を止めれば確率変数、根源事象 ω を固定すれば数列
- 時系列解析の目的: $X(t, \omega)$ の「道の分布」を知りたい
- 大きな制約: 時系列解析では殆どの場合 ω のドロウは一回きり(同一条件下での繰り返し実験に基づく観測が困難)
- 一時点一個のデータしかないのに、各時点での値の出方を支配する確率法則はもちろん、道全体が従う確率法則も導き出したい(かなり無理な注文)→時間軸方向でのある種の一様性の仮定が必要 = 定常性 (stationary)

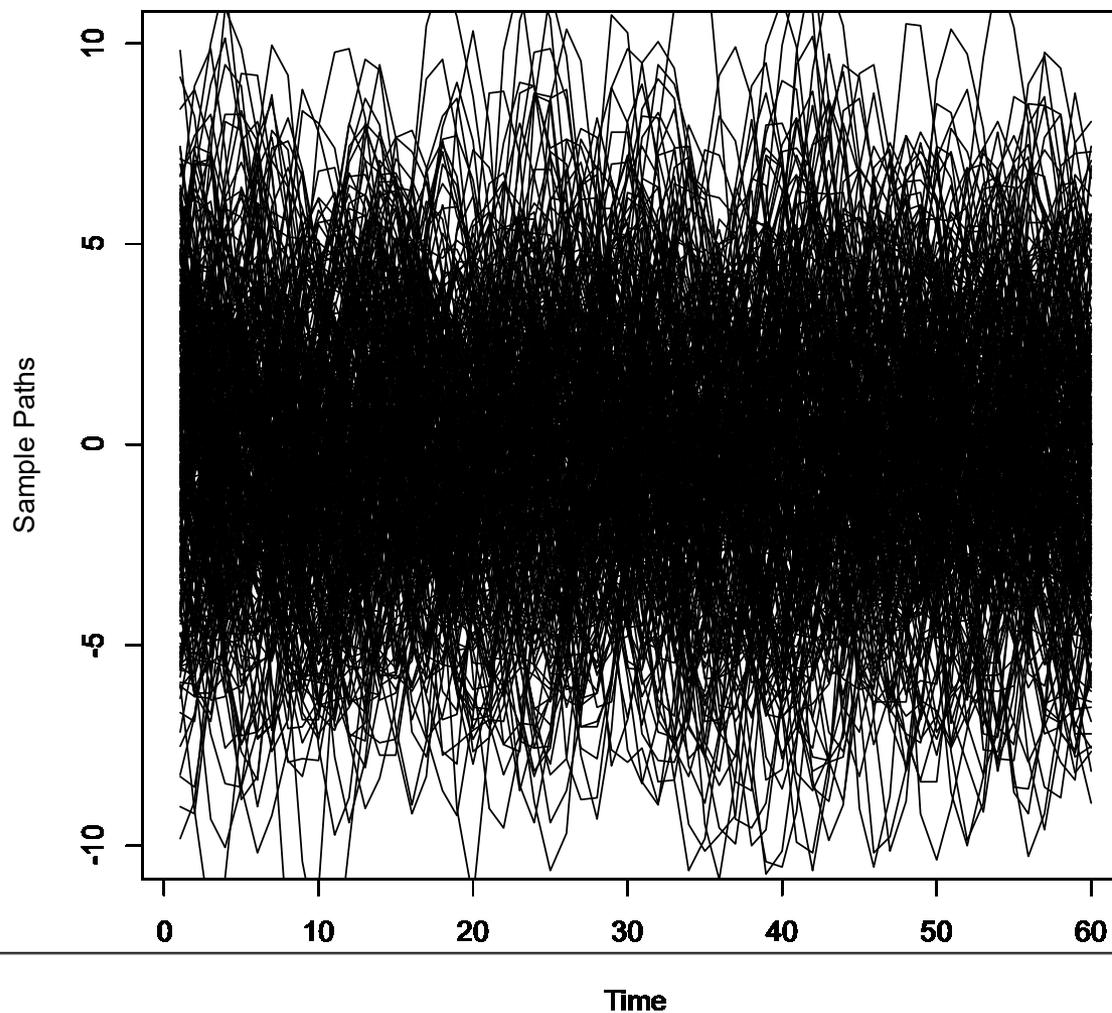
確率過程の見本過程(sample path)

Sample Paths from AR(2)



300本の見本過程を重ね描き

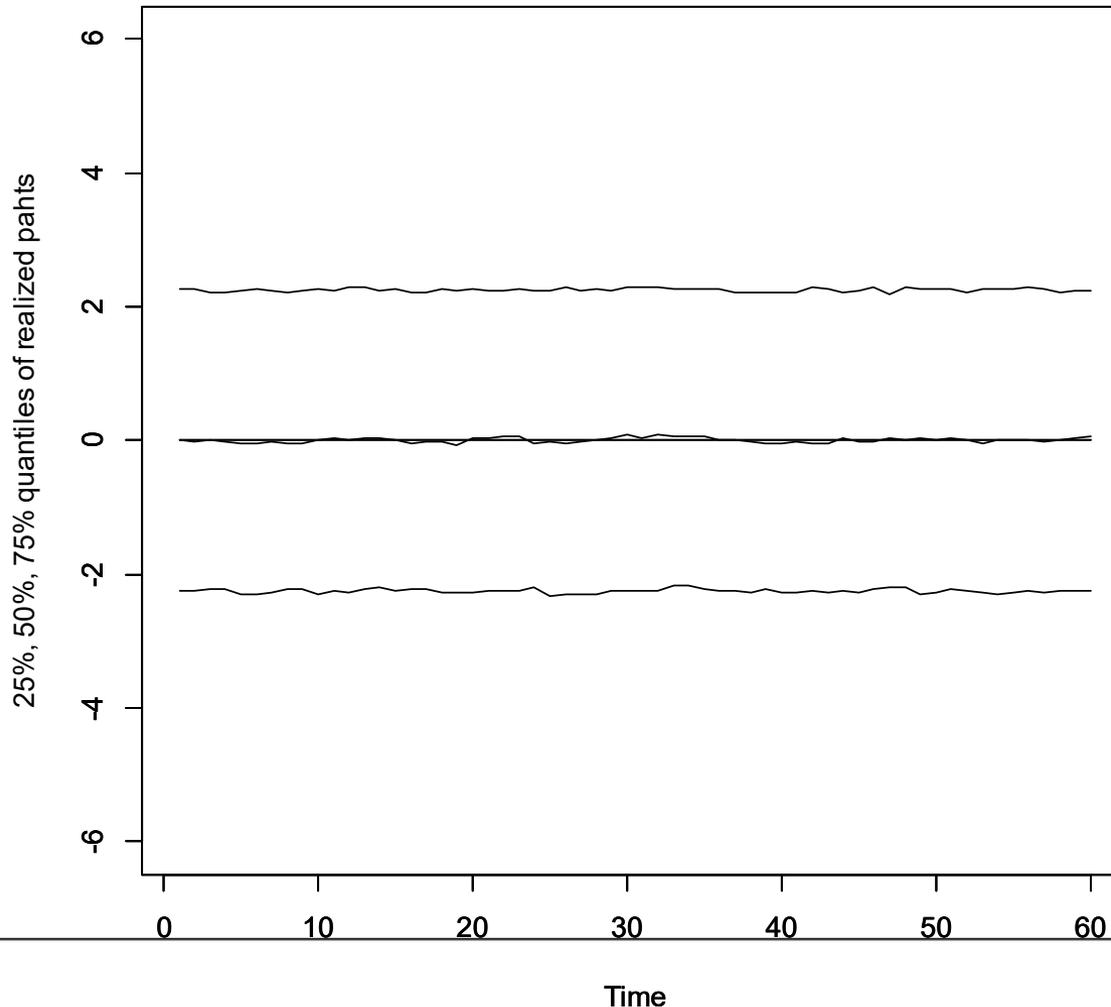
300 Sample Paths from AR(2)



各時点での四分位点

Quantiles based on 10001 sample paths

平均と分散が時間によらず一定



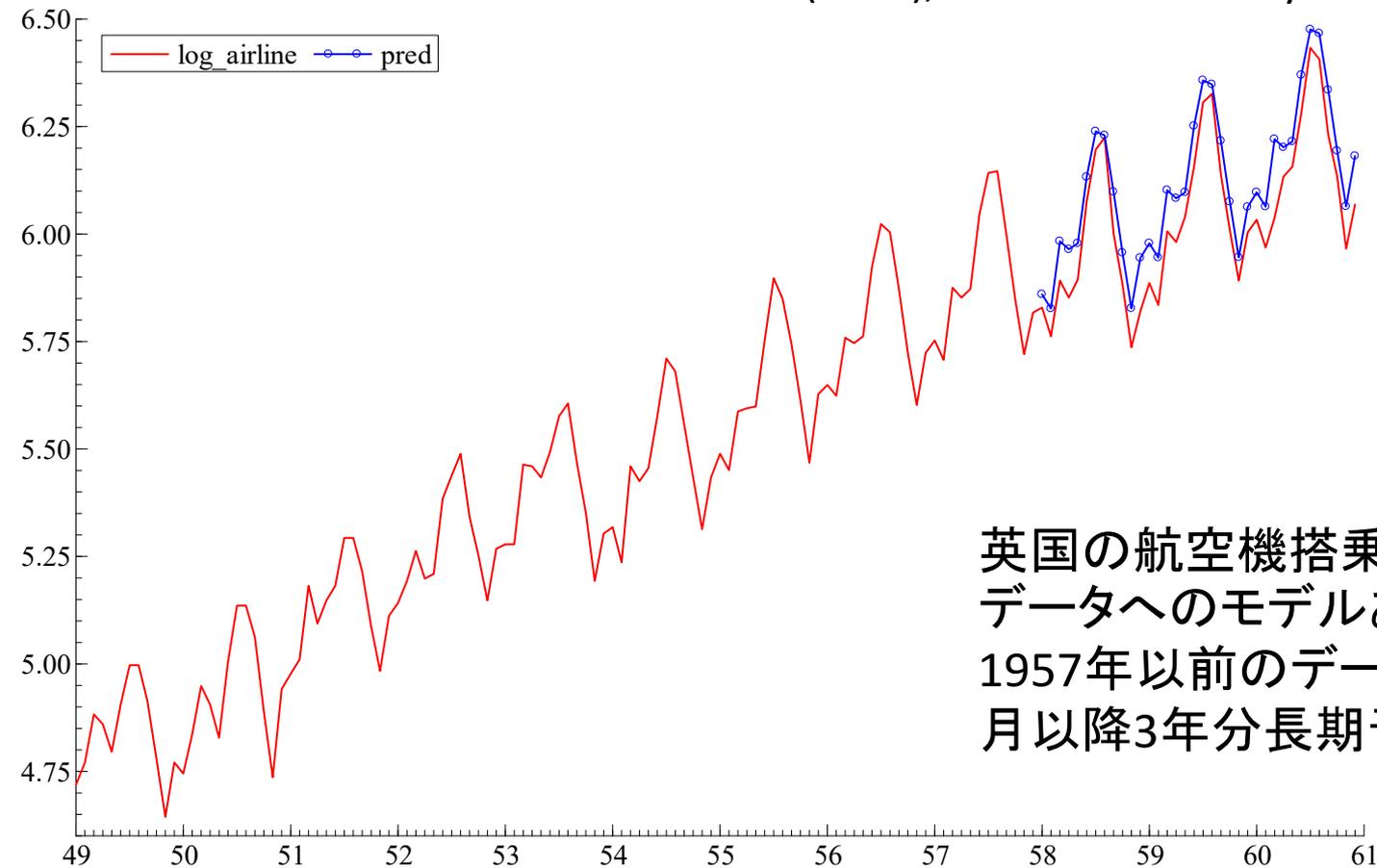
ここでは $N \rightarrow \infty$ として実験しているが、これと同じ効果が $T \rightarrow \infty$ で得られる。(エルゴード性)

2. 定常から非定常へ:変換の利用

- 定常時系列モデルの代表: ARMAモデル→現実にはこのクラスで対処できる範囲は限られている
- どうする? → **データを変換**して定常時系列モデルに帰着
 - 対数変換: 指数的トレンドを線形化
 - Box-Cox変換: 対数変換を含む巾変換のクラス
 - 階差: トレンドの除去
 - 季節階差: 季節変動の除去
- Box-Jenkins法: 標本自己相関、偏自己相関を手がかりにしたモデルの特定化、推定、事後評価(診断検定)を含む一連の手続き
 - 原著 Box and Jenkins (1970)あるいは最新版Box et al. (2008)にあたらなくても、類書多数

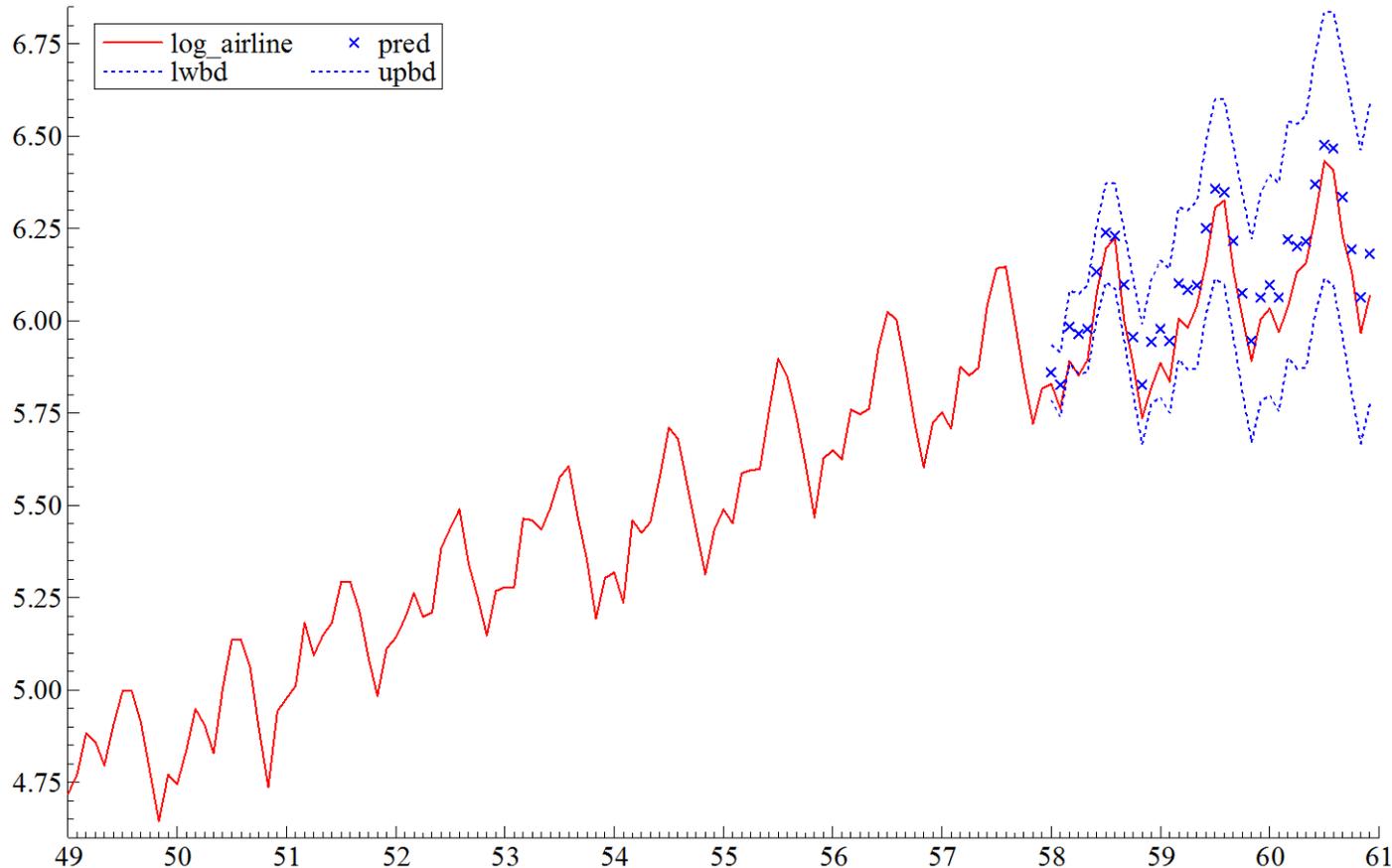
エアラインモデル：推定と予測

Box and Jenkins (1970), "Time Series Analysis"



英国の航空機搭乗者数の時系列データへのモデルあてはめと予測。
1957年以前のデータで推定、58年1月以降3年分長期予測。

エアラインモデル: 予測の信頼区間



おおよそ95%の
信頼区間を付け
ると左の通り

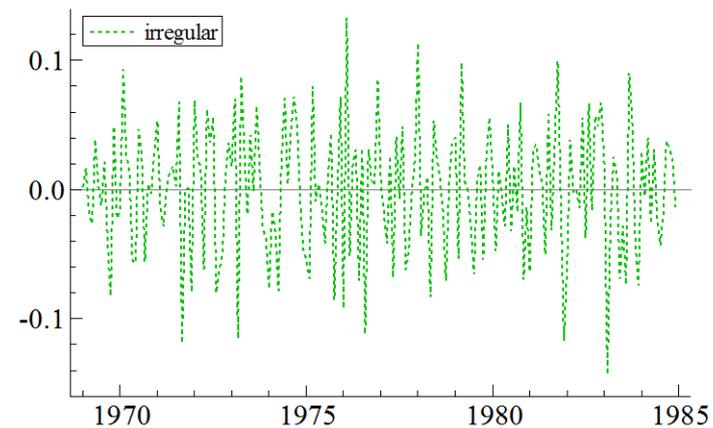
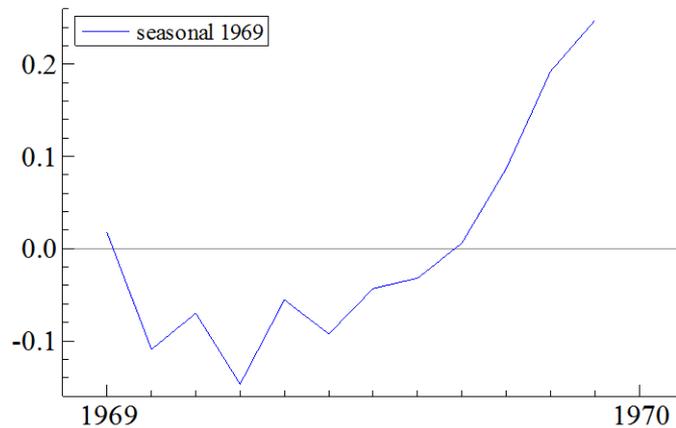
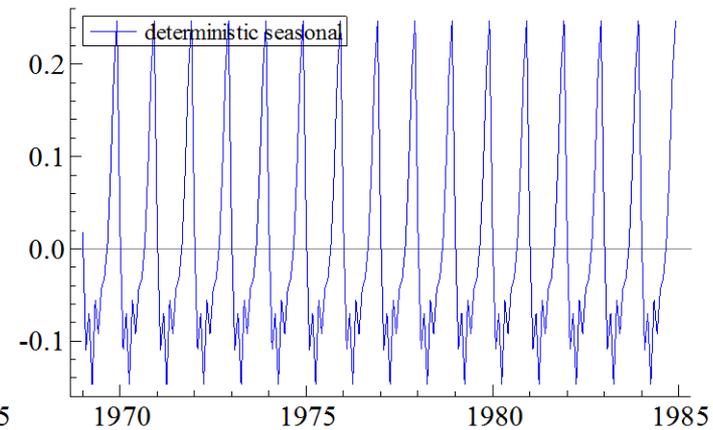
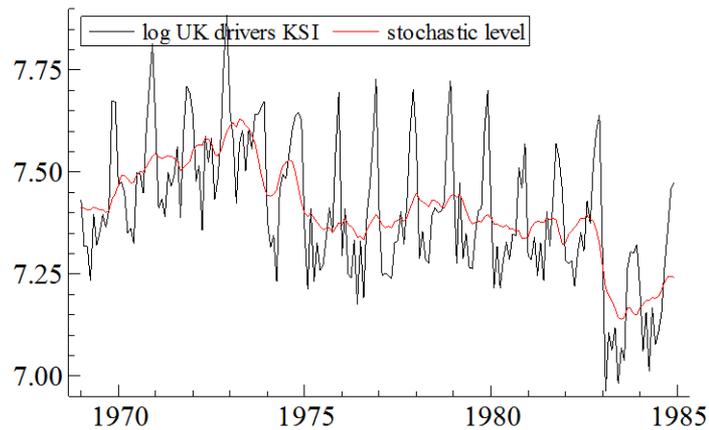
多くの問題では
每期每期データ
更新の度にモデ
ル更新と予測更
新を繰り返せる
はずで、これよ
り誤差幅は小さ
いはず。

3. 定常から非定常へ：平滑化事前分布

- 例：季節調整モデル
- 観測値 = トレンド + 季節成分 + 不規則成分
- 1個の観測値から3つ(本質的には2つ)の未知量をひねり出す必要。無理難題の度合いが増しているように見える。
- → 時点の近い成分どうしは値が近い、という制約を置けば推定可能。(平滑化事前分布)
- 観測不能な時系列要素(パラメータ)に対する確率的な制約を置くことで実効パラメータ数を減らす。パラメータに確率分布を仮定することから、ベイズモデリングの一種と言える。

成分分解の例

すぐ後に事例紹介で再掲



なぜ分けてモデル化するか？

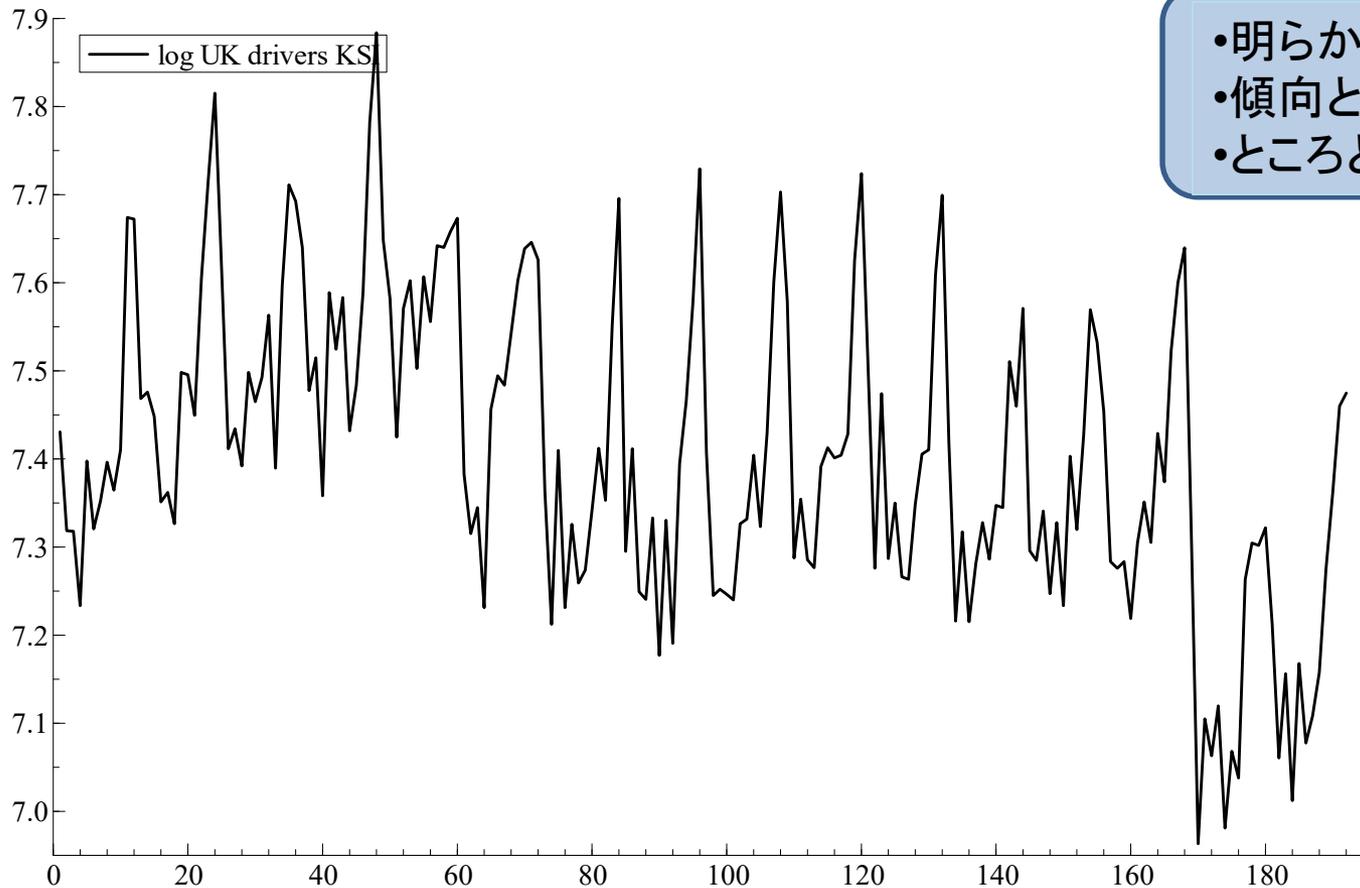
- 不確実性の度合いが違うものを別々にモデル化することで、全体として推定・予測の不確実性を低減できる。
- Box-Jenkins法(季節ARIMAモデル)では、成分分解はしないが、実質的に同様の効果を内包している。(階差、季節階差)
- 大まかな動きをモデル化した後で残る、一見残り滓のような残差系列から構造を見つけるのが統計学の真骨頂。

統計的モデリングとは?

- 現象のモデル化にあたって、不確実性の表現に確率分布を仮定。
- 物理モデルのような意味でのデータの「真の生成構造」にこだわらず、データの動きを模倣するモデルや入出力関係をうまく近似してくれるモデルを探索する。
- データのバラツキの元となる事象を特定化し、不断にモデル改善につなげる。

例：英国交通事故死傷者数（月次時系列）

1969年1月～1984年12月

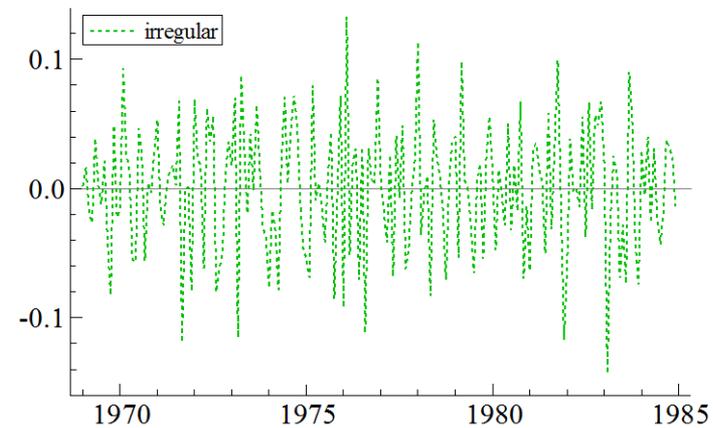
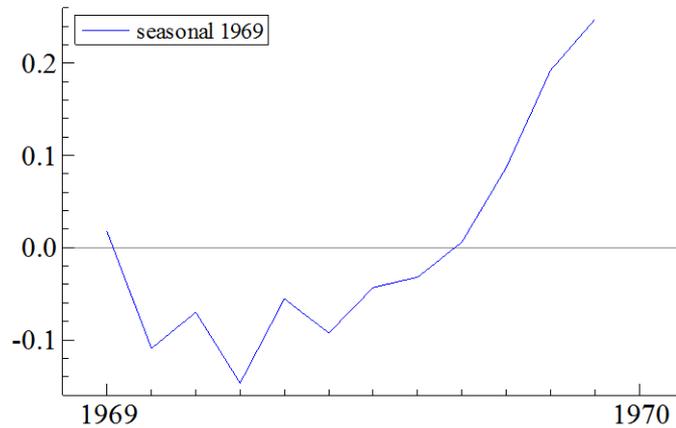
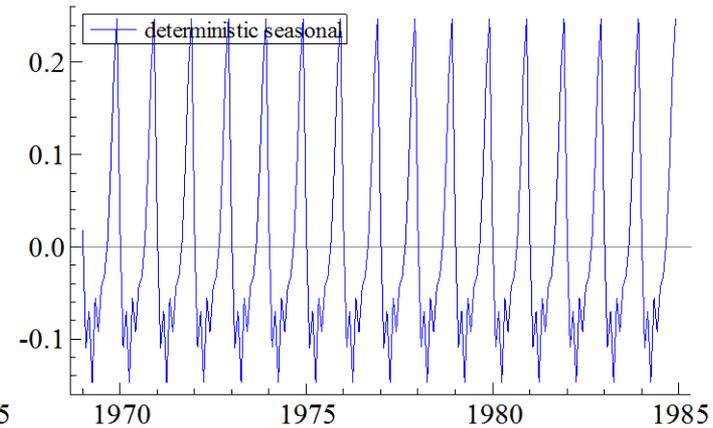
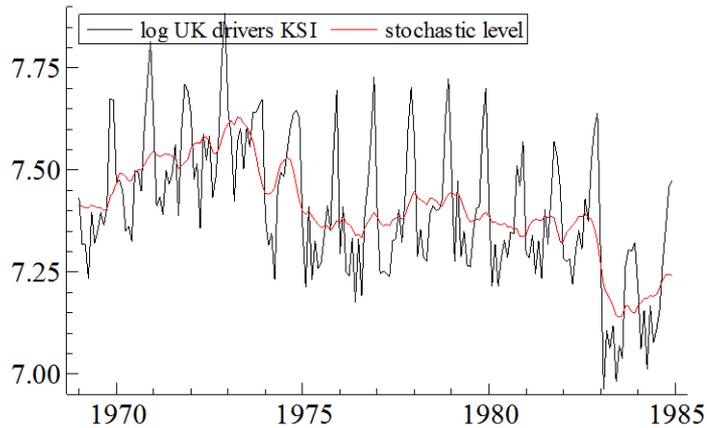


- 明らかな季節性
- 傾向としては全体に低減
- ところどころ大きな下落



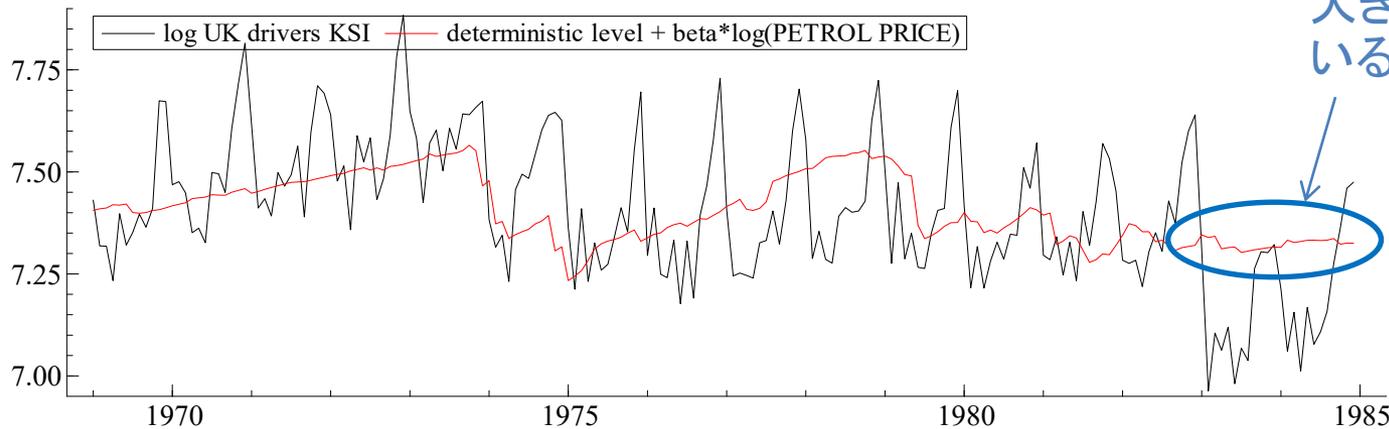
どうモデル化するか？

トレンド、季節性、不規則変動



モデリング事例1:外生時系列の利用

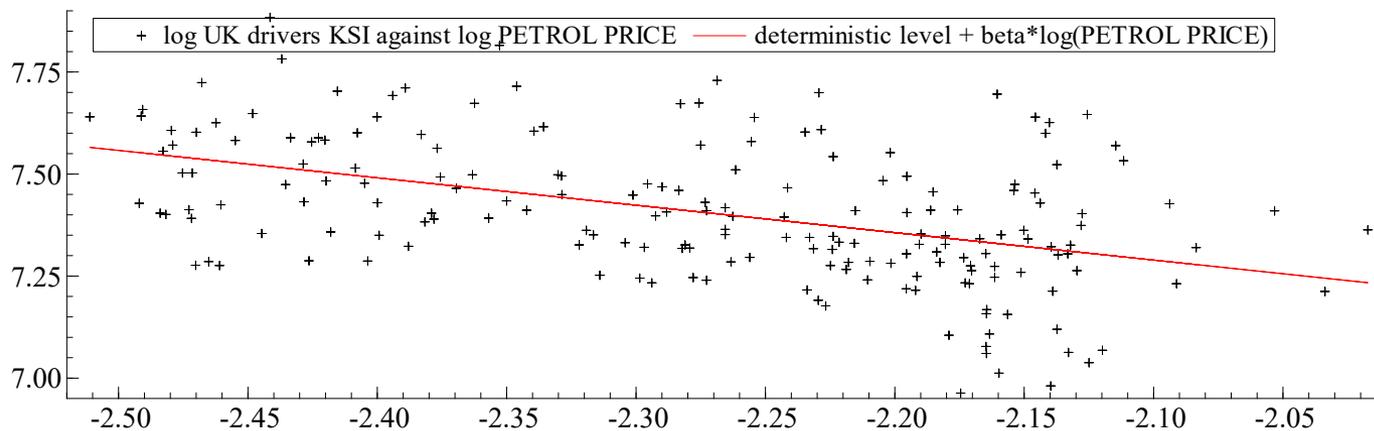
大きく水準線を外して
いる箇所がある...



モデルに線形項と
して入れ込むと、
大まかな傾向線を
示してくれる

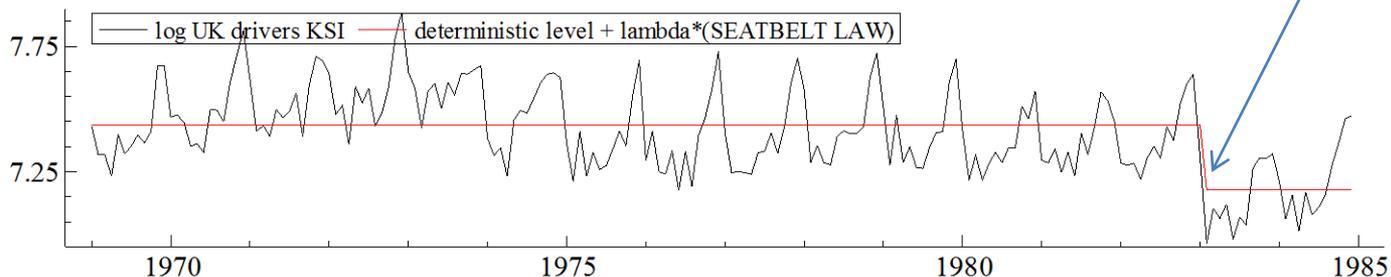


ガソリン価格(横軸)
と交通事故死傷者
数(縦軸)には**負の
相関**

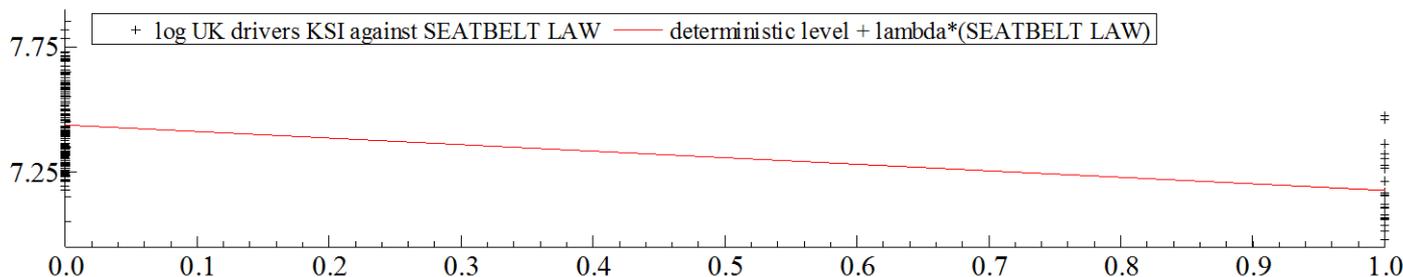


モデリング事例2:構造変化の取り込み

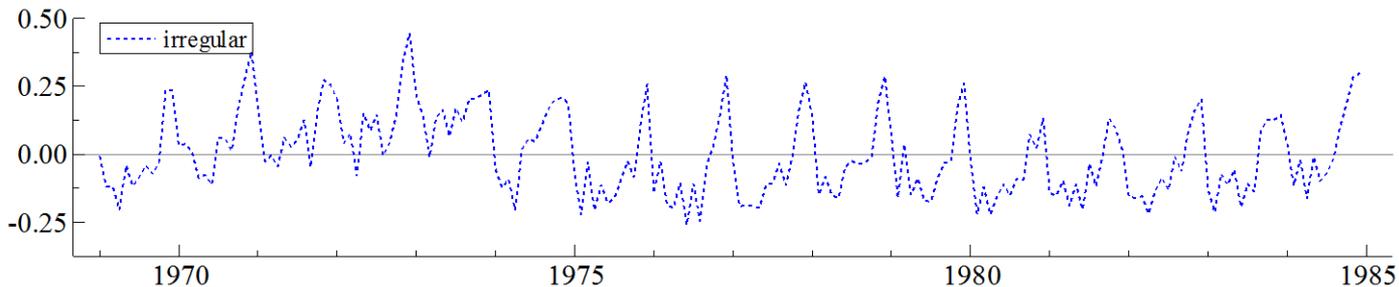
1983年2月、英国でシートベルト着用を義務づける法令が施行



83年2月にレベルシフト



施工前(横軸0.0)と施工後(1.0)での死傷者数の「散布図」



残りは、ガソリン価格、季節性、トレンド項で説明できそう

4. 状態空間モデル

- 前節で示したモデリングは、線形ガウス状態空間モデルの一例。状態空間モデルは「モデルのモデル」あるいは「メタモデル」
- 実用的な時系列モデルの殆どは、時系列の分布を規定する本質的な変数群(状態)のマルコフ遷移と、それら変数群と観測値を結びつける関係式で表現される。
- 状態空間モデルの枠組みでは、状態の推定(予測、フィルタ、平滑化)と尤度計算の方法が統一的に与えられる。つまり、一度この形で習得しておけば応用がきく。
- 適用例としては、おおまかに信号抽出型か時変係数型かに分けられる。

時系列の尤度

- 時系列の最大の特徴は、時間的に近接する観測値に相関があること→独立同一分布の場合のように、個々の密度のかけ算で尤度を定義できない!
- 時系列モデルの場合は「**一期先予測誤差の累積**」として尤度を表現し、それを数値的に最大化する。これを尤度の予測誤差分解という。
- データサイエンティストとして、後々応用力を高めるためには、単にパッケージを利用するだけでなく、予測誤差分解を自らプログラミングで体感しておく必要がある。
- 北川 (2005)の第9章を熟読、第10～13章で応用のイメージを膨らませる。

最前線に向かって

- 最前線：状態遷移を非線形化、応答変数の非ガウス化
 - Particle filter, MCMC
 - 抱えている具体的問題でこれらが要請される場合は、北川 (2005)の第14章、15章まで学習。
- 季節ARIMA, 線形ガウス状態空間モデルに基づく成分分解型時系列モデルは、成熟段階に達した技術で使いやすい。
 - ソフトウェア実装も選択肢多数、需要分析をはじめとして広く浸透
- 多変量の場合：多変量AR(北川 2005, 6.7, 7.6, 7.7節)が基本だが、高次元の場合、闇雲に適用する前に、全ての変数を同時決定するようなモデルが求められているかどうか再検討したほうがよい。

参考文献

1. 北川源四郎『時系列解析入門』(岩波書店、2005年)
2. Box, G. E. P., Jenkins, G. M. and Reinsel, G. C., *Time Series Analysis: Forecasting and Control 4th ed.*, Wiley, 2008.
3. A. C. ハーベイ『時系列モデル入門』(国友直人・山本拓訳、東大出版会、1985年)
4. 田中勝人『現代時系列分析』(岩波書店、2006年)
5. Commandeur, J. J. F. and Koopman, S. J., *An Introduction to State Space Time Series Analysis*, Oxford University Press, 2007. [シーエーピー出版から和訳あり]