

時系列解析 (12)

— 非線形・非ガウス型状態空間モデル —
配布用

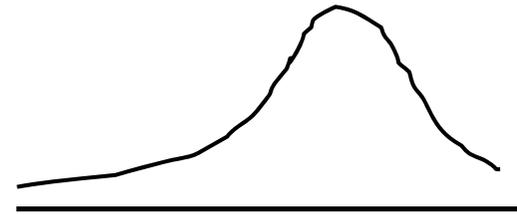
東京大学 数理・情報教育研究センター
北川 源四郎

非ガウス型モデリングの必要性

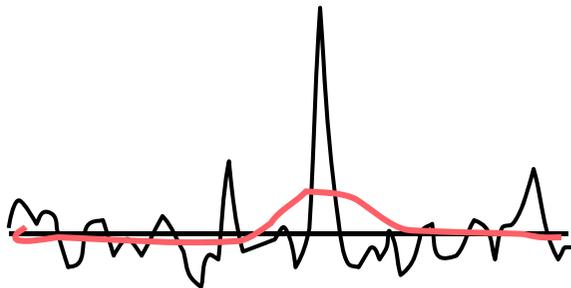
構造変化



非対称分布



異常値 (外れ値)



非線形性

$$x_n = f(x_{n-1}) + v_n$$

離散過程

ポアソン過程
二項過程

状態空間モデルの拡張

線形・ガウス型

$$x_n = Fx_{n-1} + Gv_n$$

$$y_n = Hx_n + w_n$$

非線形・非ガウス型

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

$$y_n = h(x_n, w_n)$$

関数：非線形
分布：非ガウス型

非ガウス型フィルタ・平滑化

一期先予測

$$p(x_n | Y_{n-1}) = \int_{-\infty}^{\infty} p(x_n | x_{n-1}) p(x_{n-1} | Y_{n-1}) dx_{n-1}$$

フィルタ

$$p(x_n | Y_n) = \frac{p(y_n | x_n) p(x_n | Y_{n-1})}{p(y_n | Y_{n-1})}$$

平滑化

$$p(x_n | Y_N) = p(x_n | Y_n) \int_{-\infty}^{\infty} \frac{p(x_{n+1} | x_n) p(x_{n+1} | Y_N)}{p(x_{n+1} | Y_n)} dx_{n+1}$$

Kitagawa(1987)

分布の近似

0. 線形・正規モデル近似

カルマンフィルタ・平滑化

1. 正規分布近似

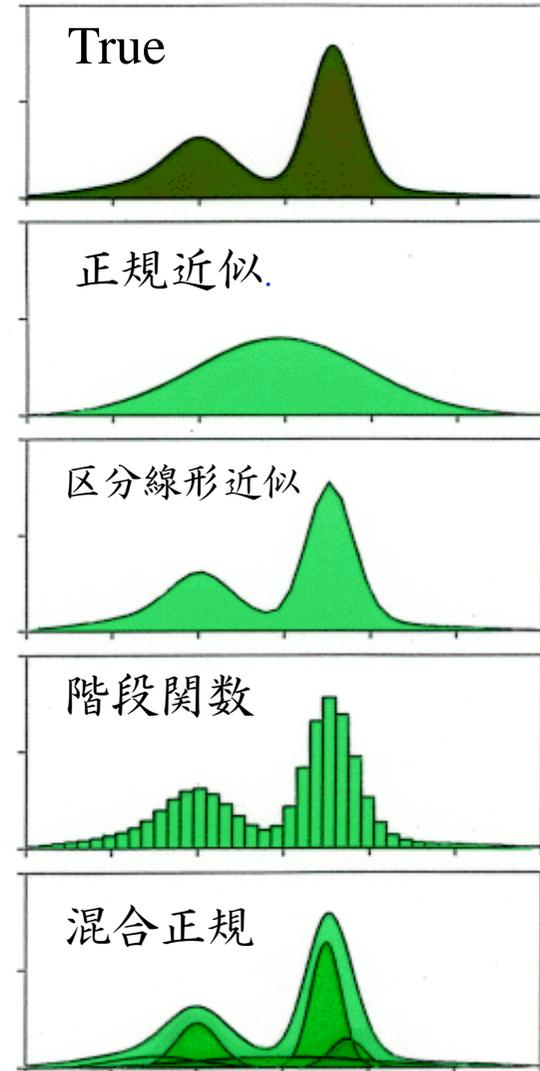
拡張カルマンフィルタ・平滑化

2. 区分線形（階段）近似

非ガウス型フィルタ・平滑化

3. 混合正規分布近似

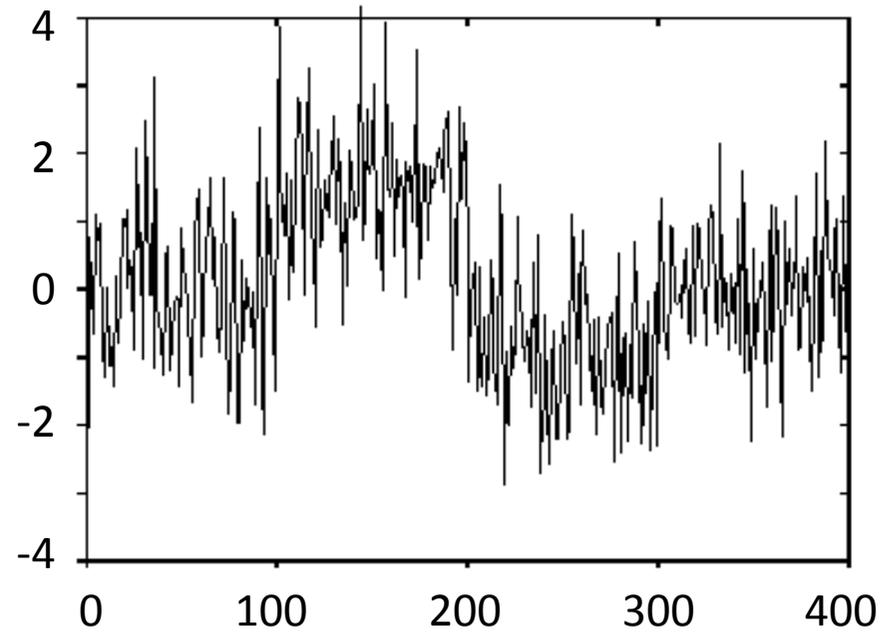
ガウス和フィルタ・平滑化



数值例

$$y_n \sim N(\mu_n, 1)$$

$$\mu_n = \begin{cases} 0 & 1 \leq n \leq 100 \\ 1 & 101 \leq n \leq 200 \\ -1 & 201 \leq n \leq 300 \\ 0 & 301 \leq n \leq 400 \end{cases}$$



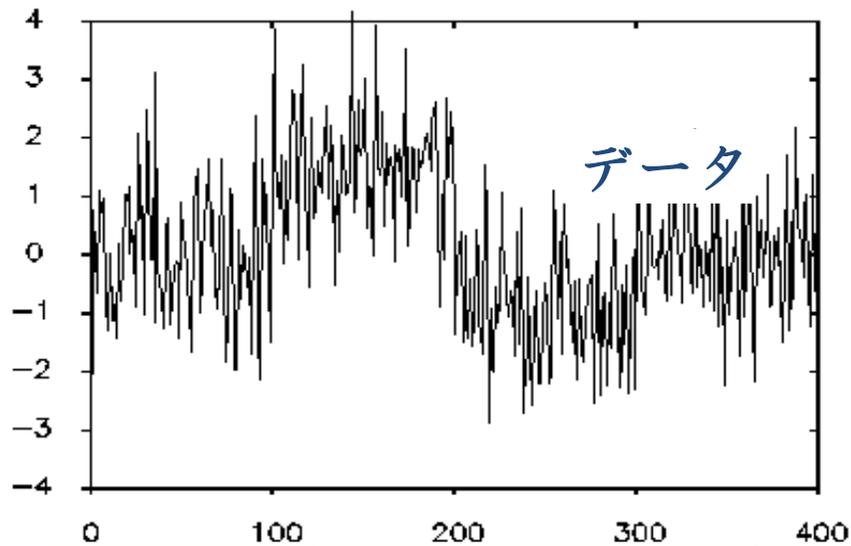
ピアソン分布族の比較

b	τ^2	σ^2	対数尤度	AIC	分布形
0.60	0.211×10^{-9}	1.042	-597.19	1198.38	
0.75	0.299×10^{-7}	1.043	-597.39	1198.78	
1.00	0.353×10^{-4}	1.045	-597.99	1198.98	Cauchy
1.50	0.303×10^{-2}	1.045	-599.13	1202.26	t_2
3.00	0.406×10^{-1}	1.046	-600.40	1204.80	t_5
∞	0.140×10^{-1}	1.048	-600.69	1205.38	Normal

$$b = \frac{3}{m} \quad (m = 0, \dots, 5)$$

$$b = \frac{k+1}{2} \quad (k: t \text{分布の自由度})$$

非ガウスモデルによる構造変化の自動検出



トレンドモデル

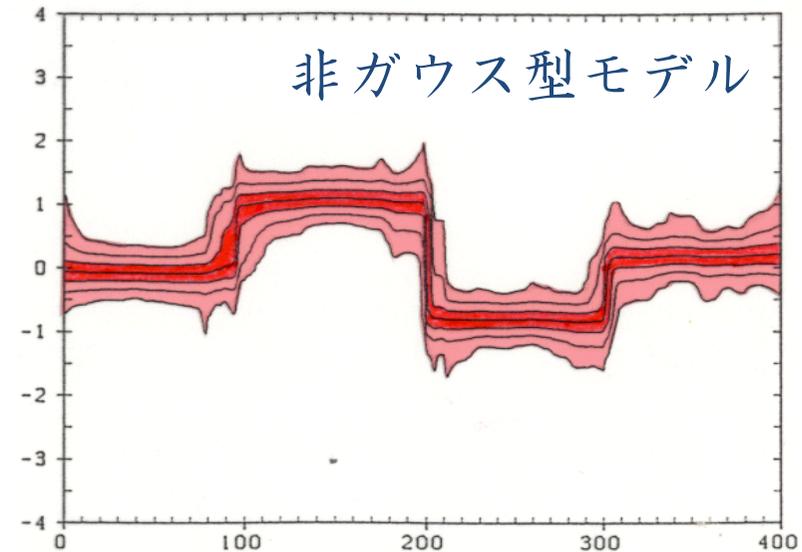
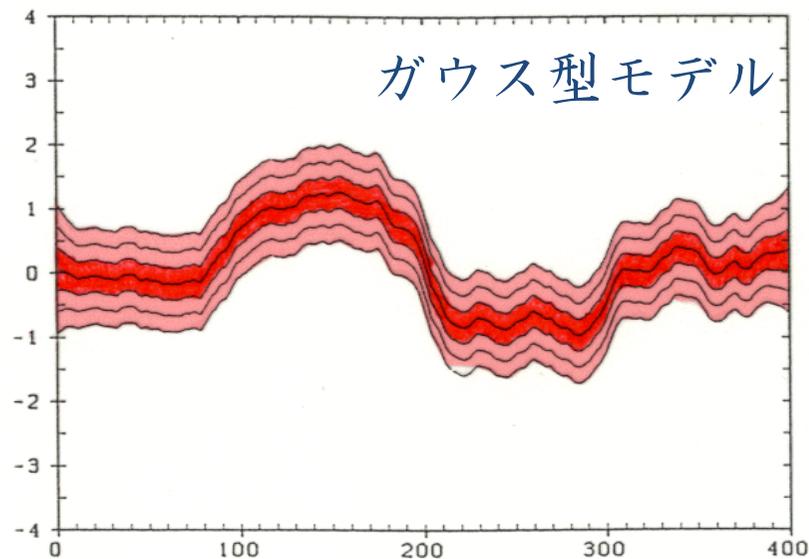
$$t_n = t_{n-1} + v_n$$

$$y_n = t_n + w_n$$

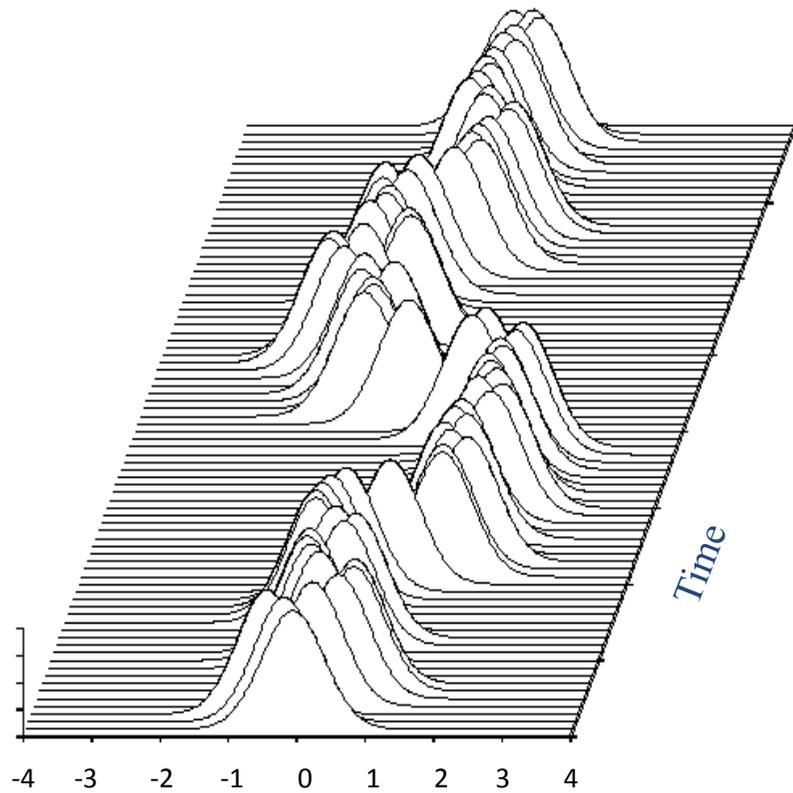
ノイズ分布

$v_n \sim p(0, \tau^2, b)$ ピアソン分布

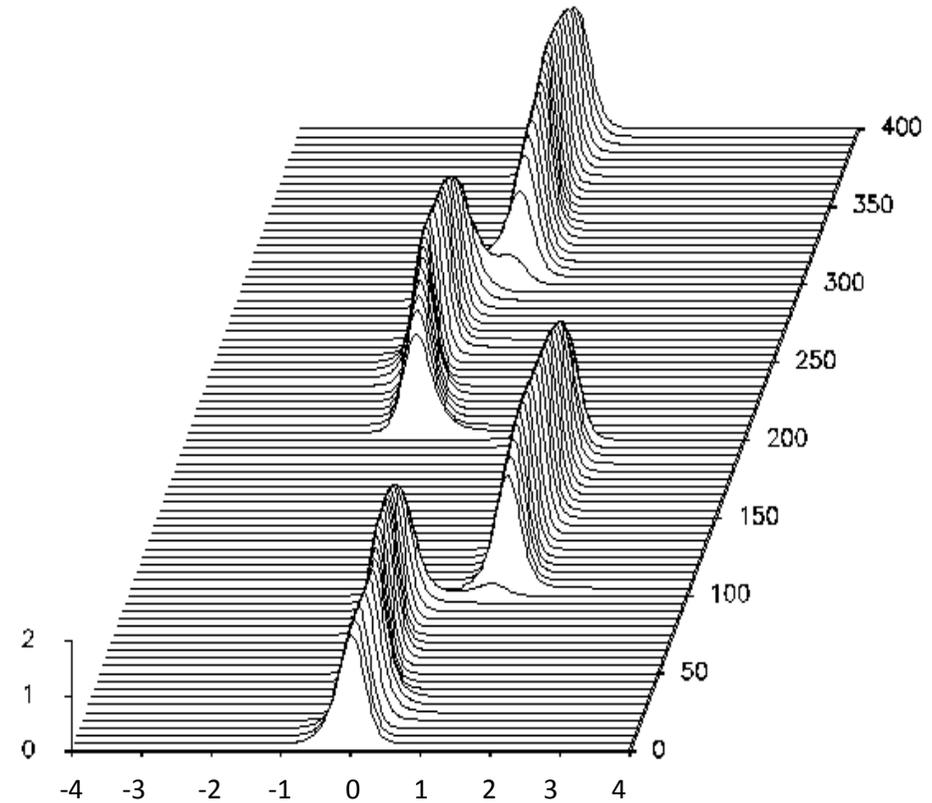
$w_n \sim N(0, \sigma^2)$ 正規分布



非ガウス型平滑化による事後分布

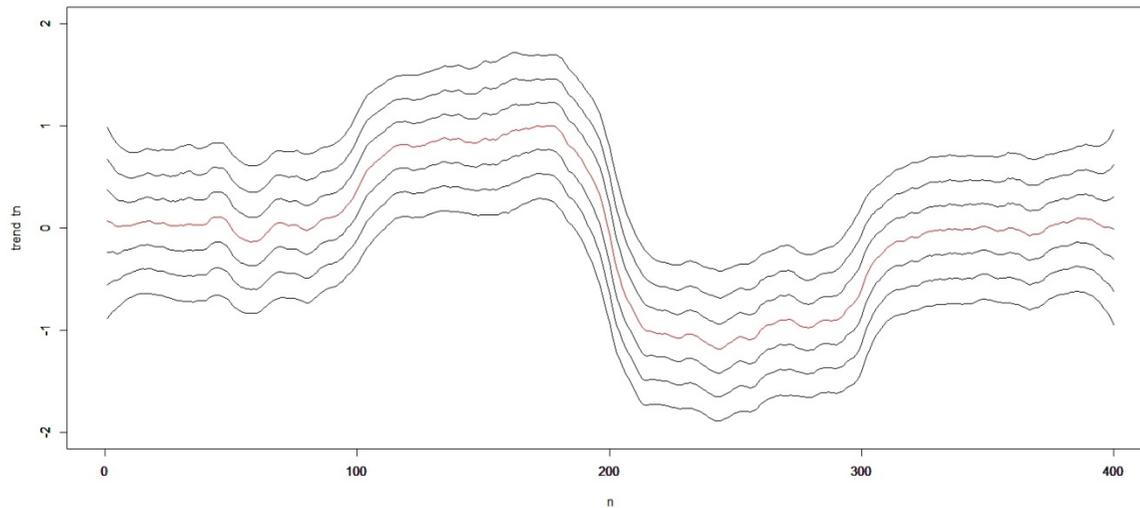


正規分布モデル

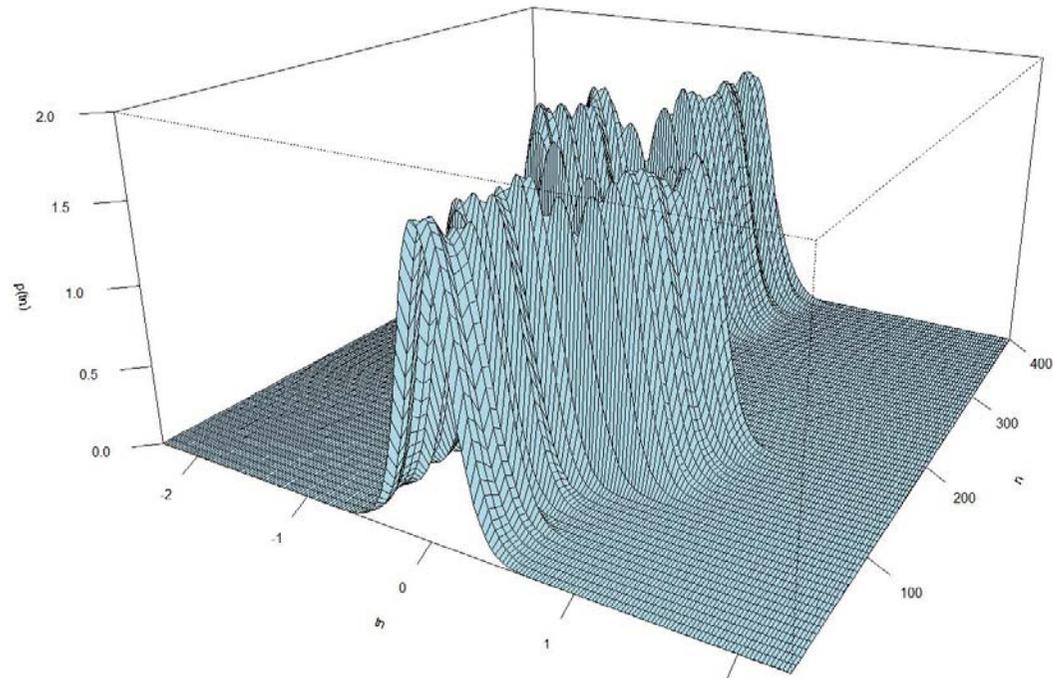


ピアソン分布族モデル

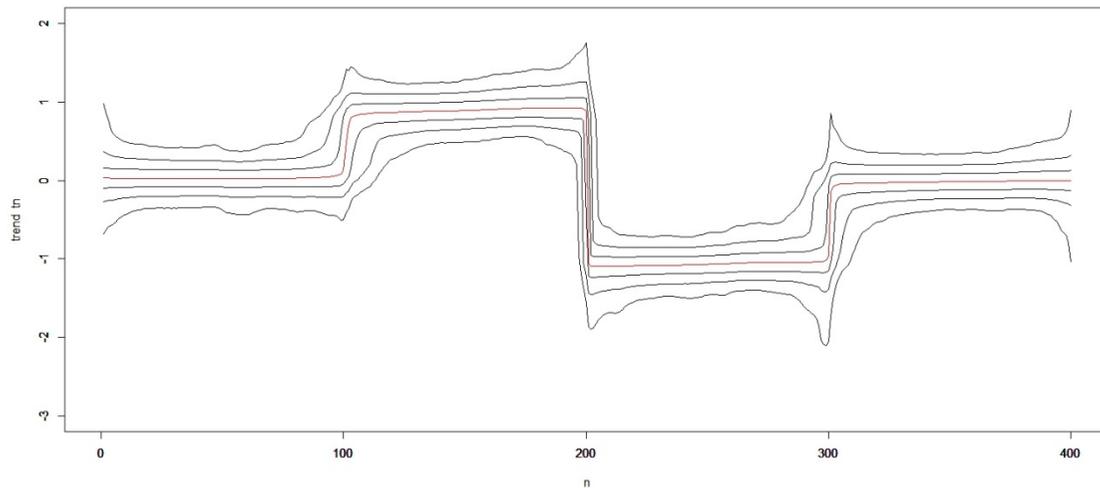
システムノイズ：正規分布



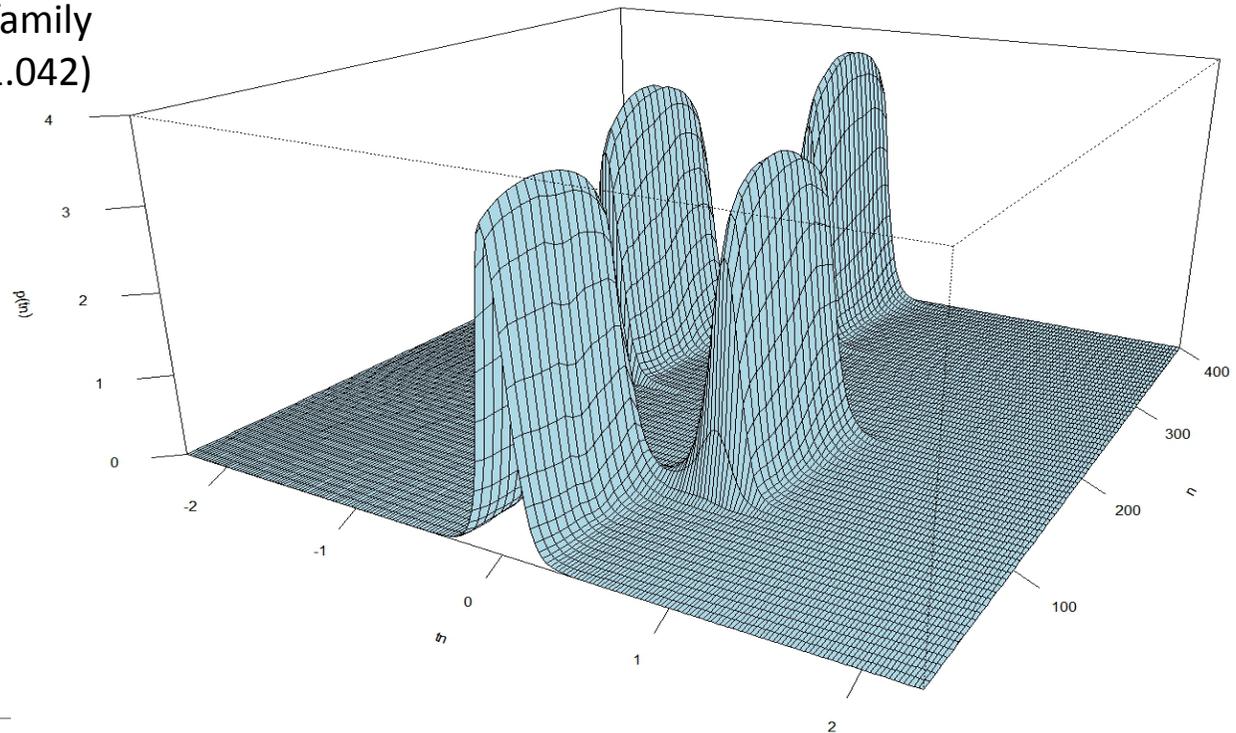
system noise density : Gaussian (normal)
ngsmth(y, 1, 1.4e-02,, 2, 1.048)



システムノイズ：ピアソン分布族 ($b=0.6$)



system noise density : Pearson family
ngsmth(y, 2, 2.11e-10, bv=0.6, 2, 1.042)

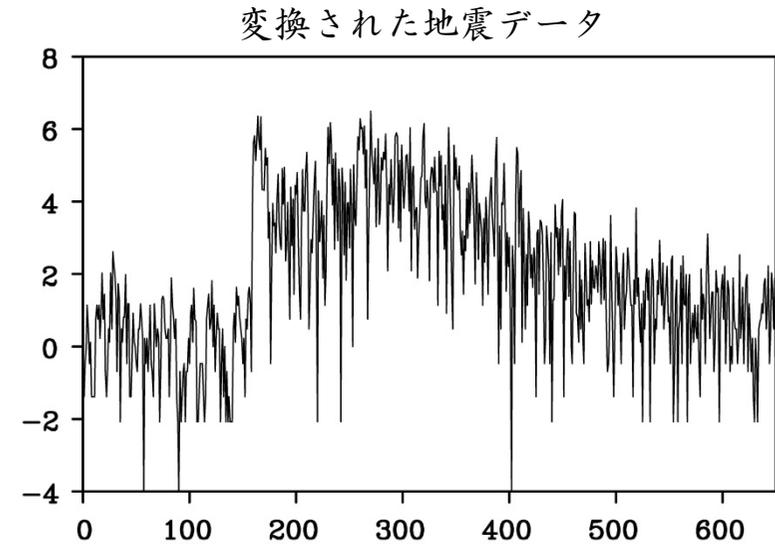


時変分散の推定

$$u_m = \log\{(y_{2m-1}^2 + y_{2m}^2) / 2\}$$

$$t_m = t_{m-1} + v_m$$

$$u_m = t_m + w_m$$

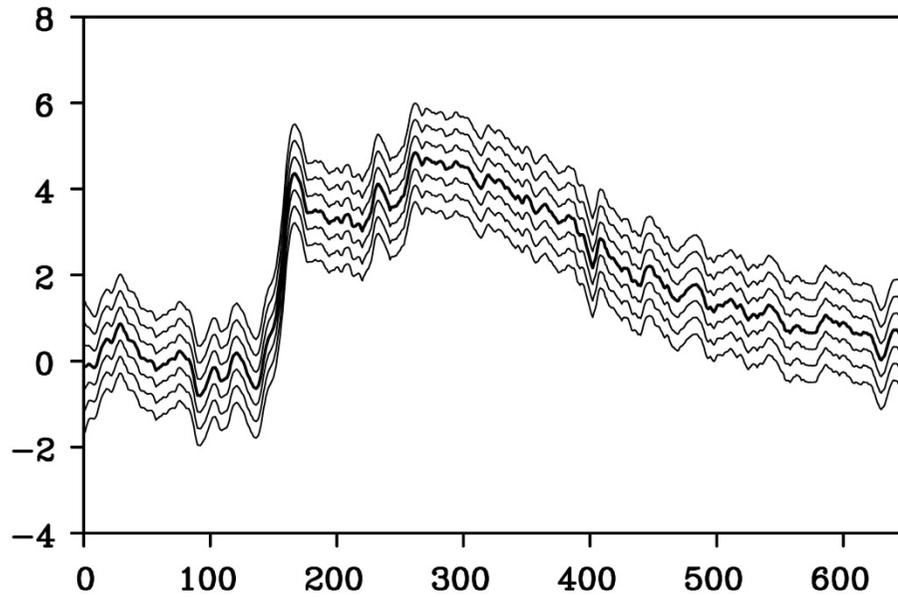


システムノイズ	観測ノイズ	AIC
$N(0, 0.04909)$	$N(-\zeta, \pi^2/6)$	2165.10
$C(0, 0.000260)$	$\exp\{w - e^w\}$	2056.97

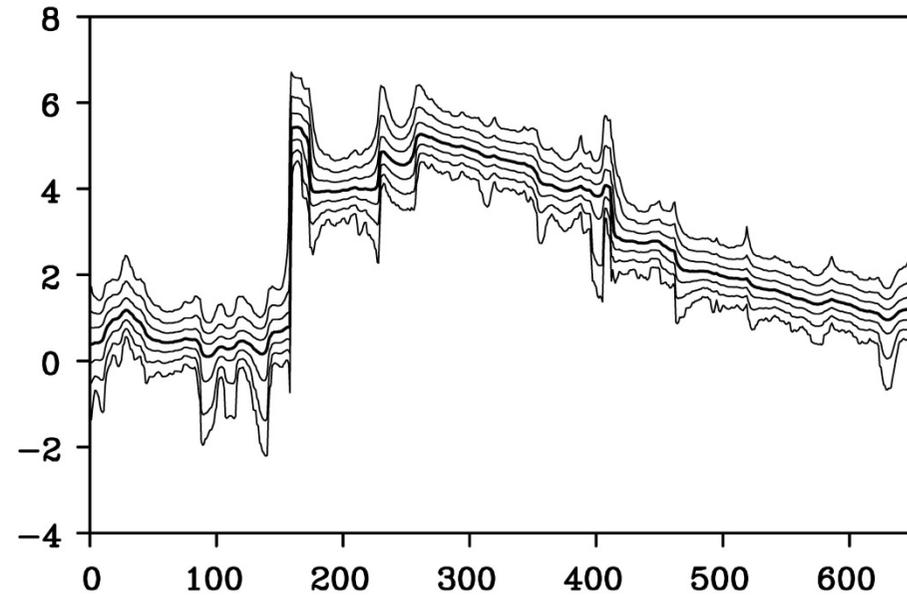
$h(w) = \exp\{w - e^w\}$: 二重指数分布

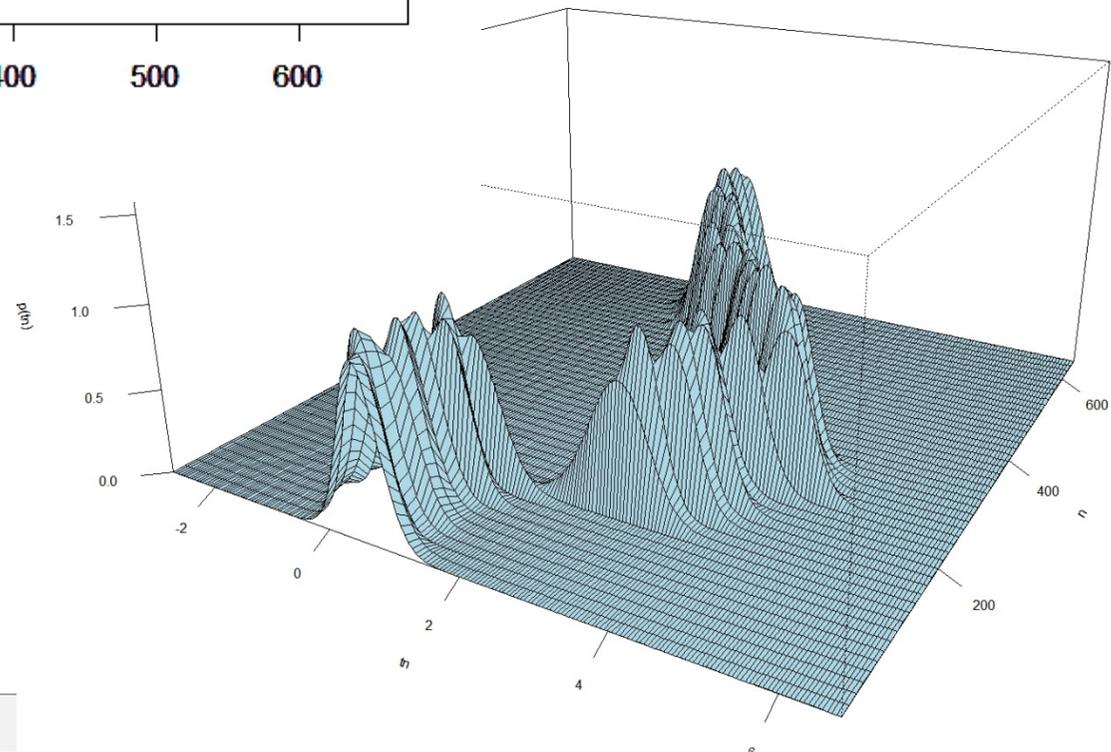
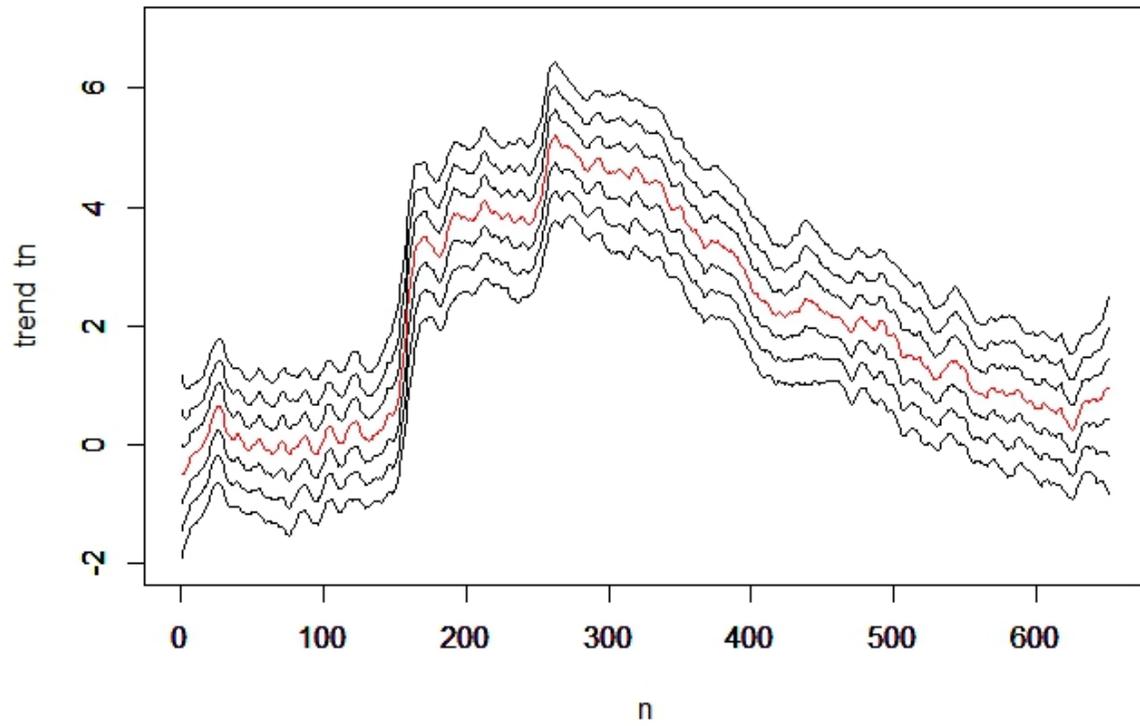
時変分散の推定 (平均, $\pm\sigma$, $\pm2\sigma$, $\pm3\sigma$)

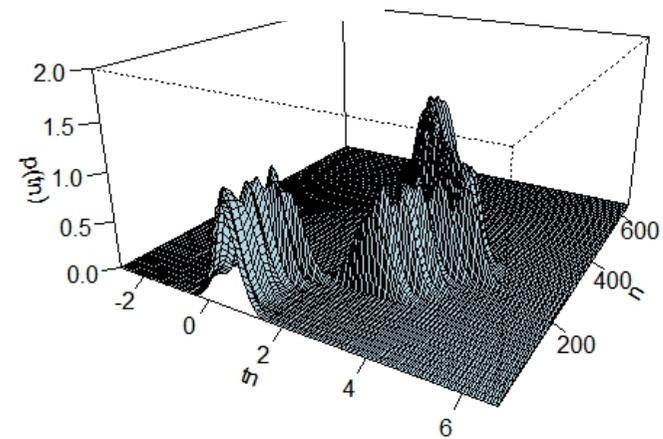
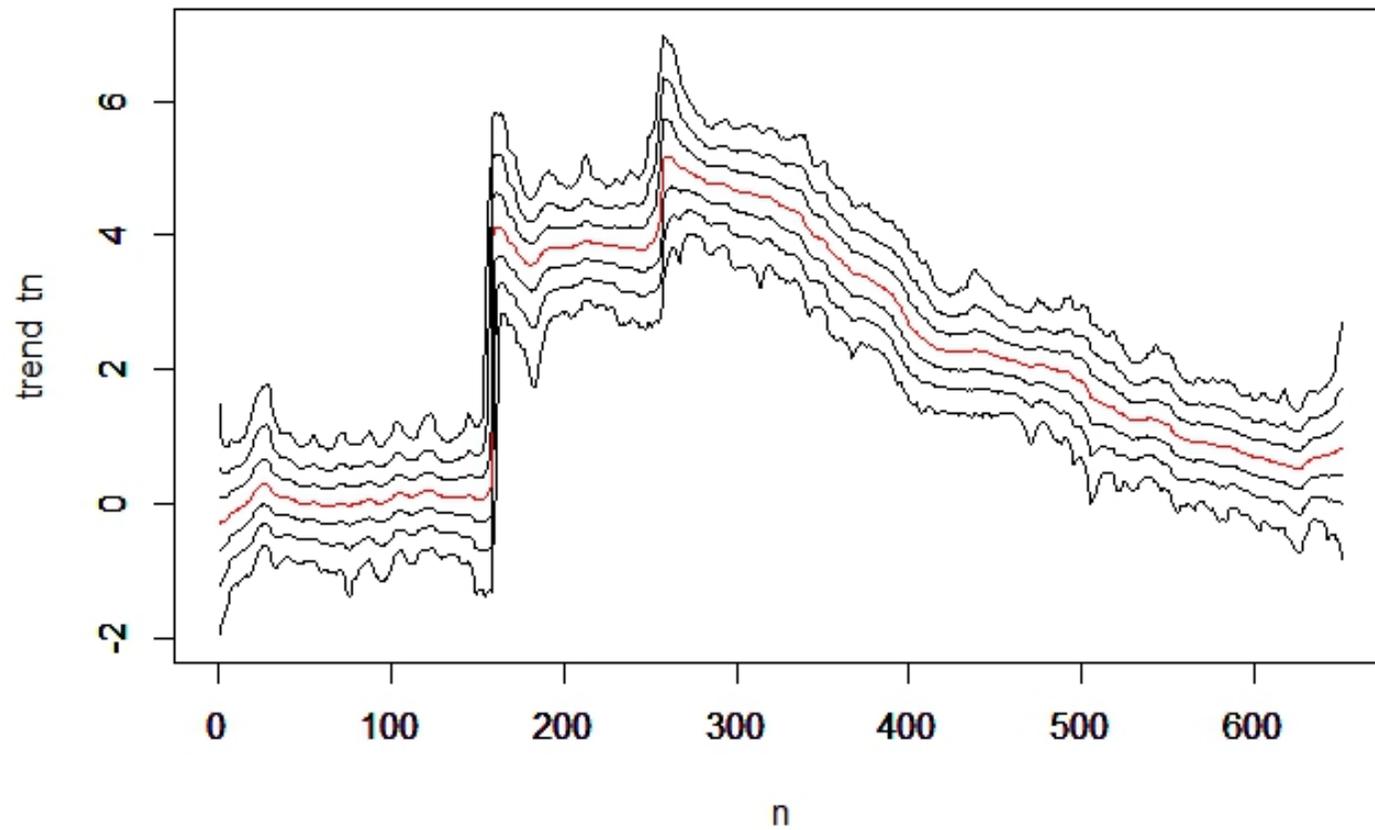
ガウスモデル



非ガウスモデル

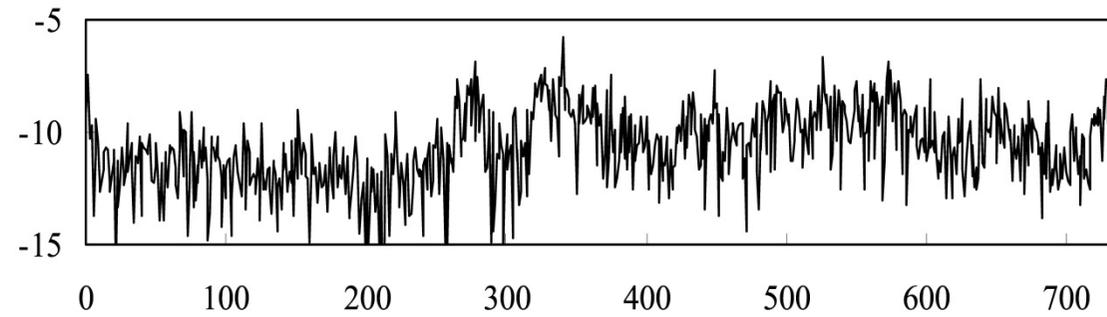




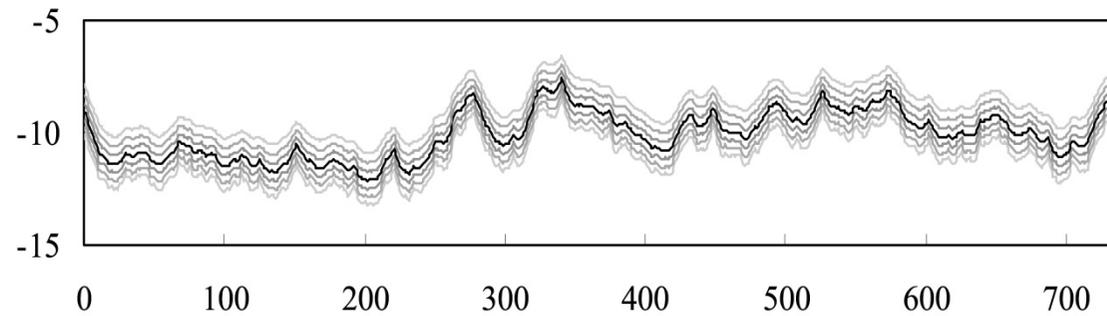


日経225平均株価データ

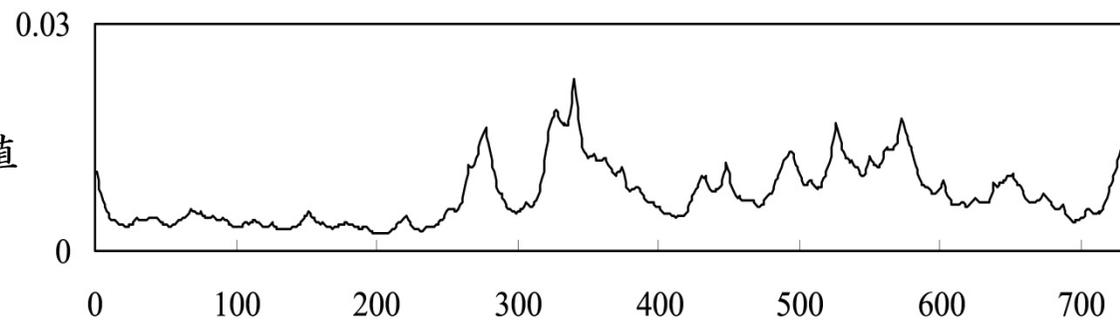
変換後



$\log \sigma_n^2$ の事後分布



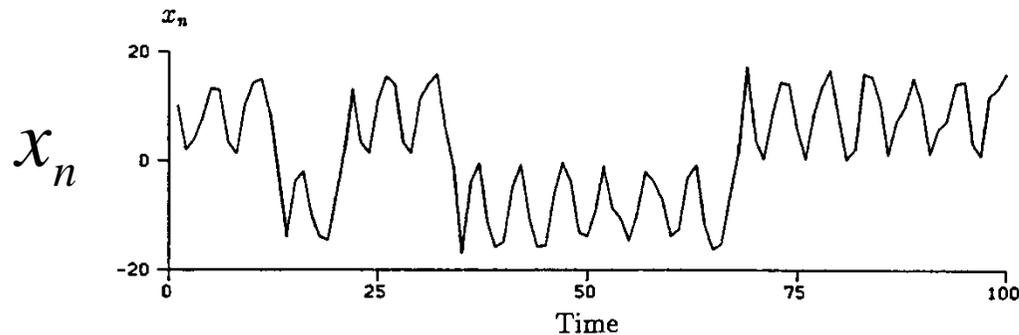
ボラティリティの推定値



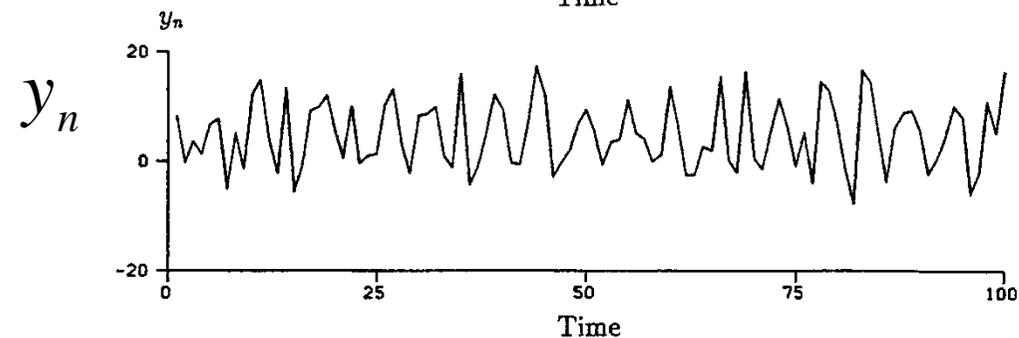
非線形モデル

$$x_n = \frac{1}{2} x_{n-1} + \frac{25x_{n-1}}{1 + x_{n-1}^2} + 8 \cos(1.2n) + v_n$$

$$y_n = \frac{x_n^2}{20} + w_n \quad v_n \sim N(0,0.1), w_n \sim N(0,1)$$

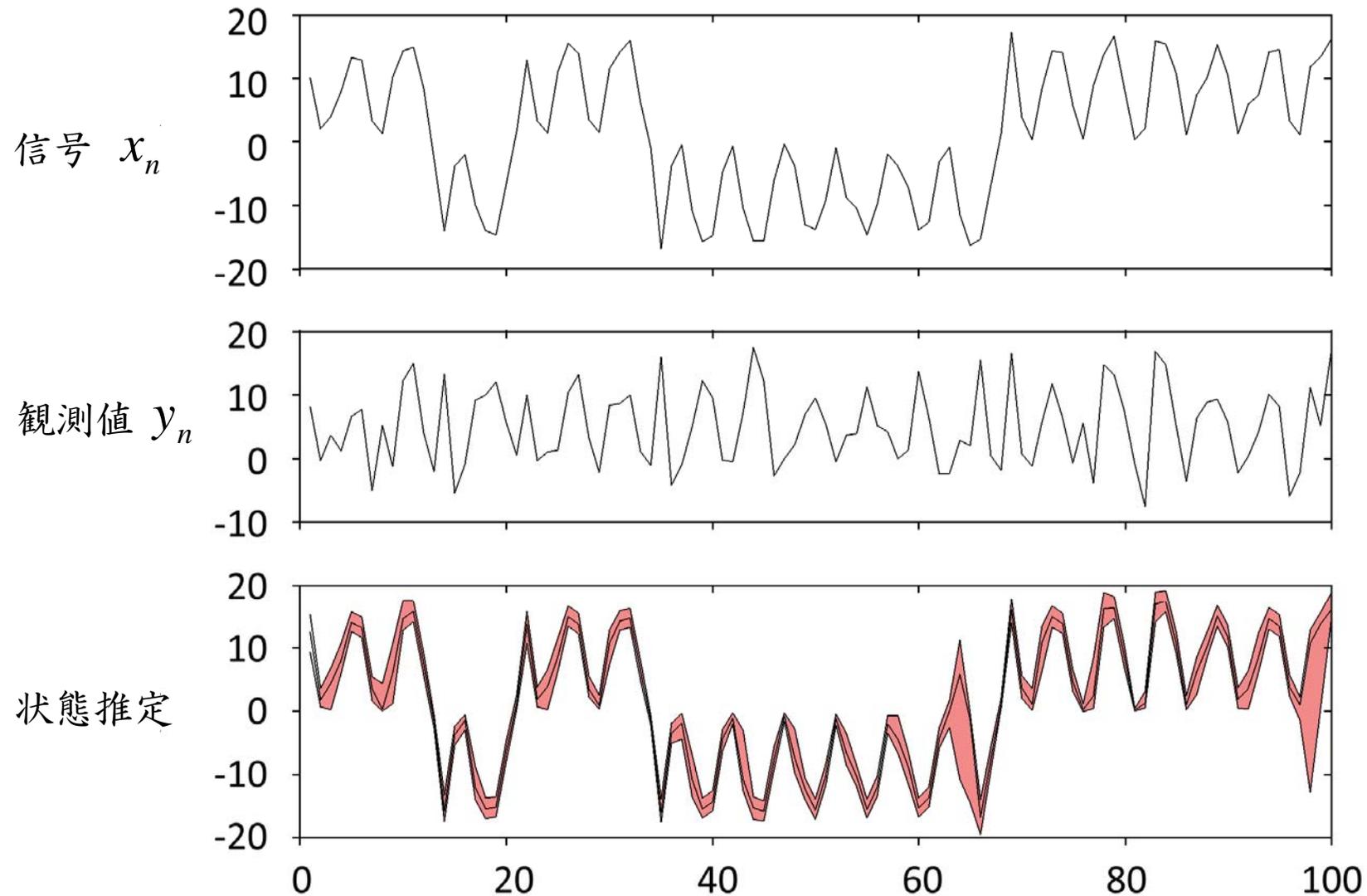


信号



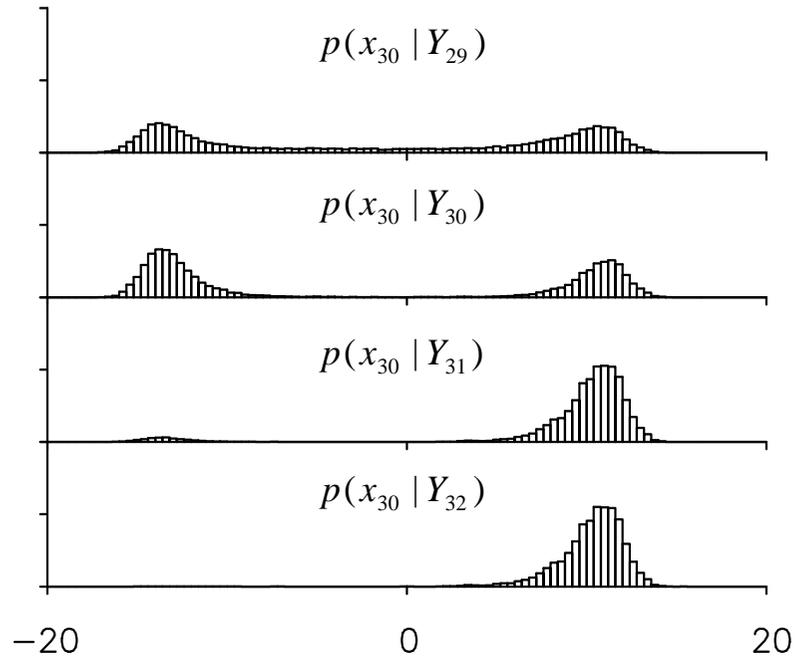
観測値

非線形平滑化による信号復元



非ガウス型分布の効果

x_{30} の推定

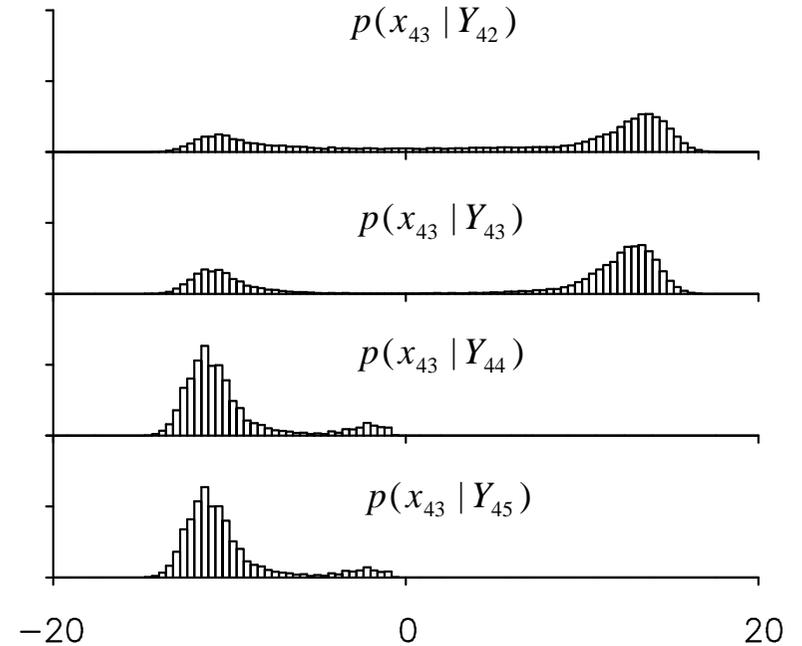


予測分布

フィルタ

平滑分布

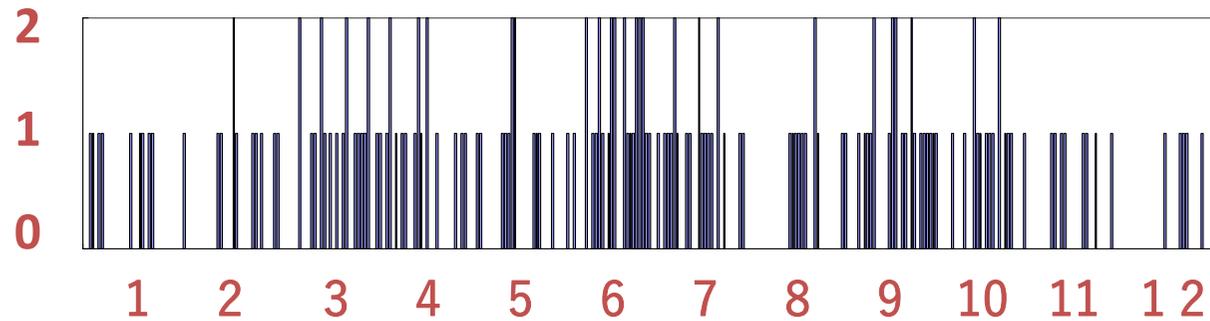
x_{43} の推定



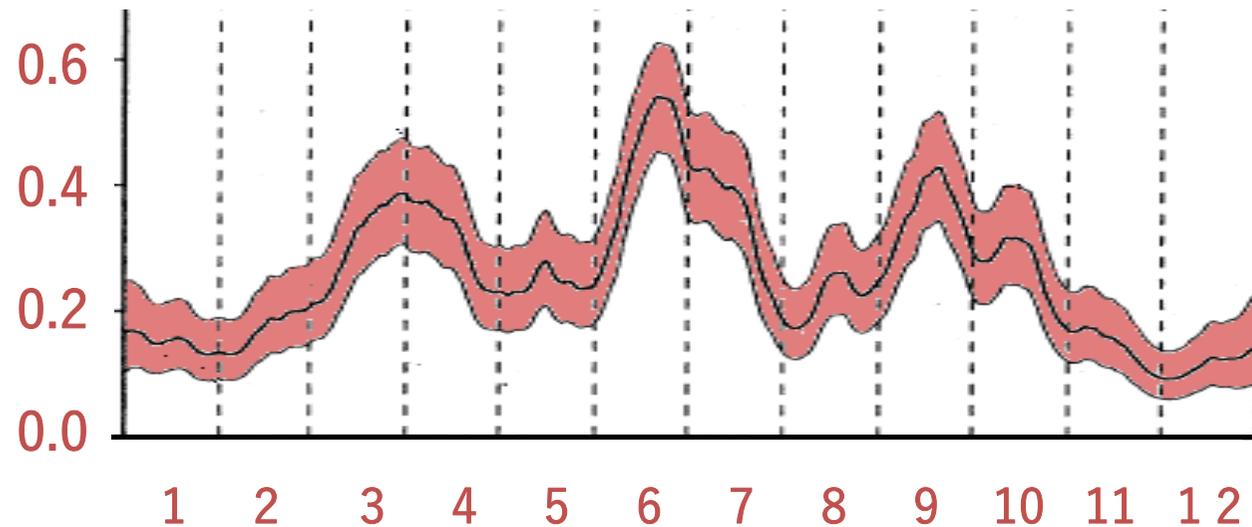
降雨データ (東京)

y_n : 2年間のうち雨が降った回数

降雨度数



降雨確率



自己組織型フィルタ・平滑化

状態空間モデル

$$x_n = x_{n-1} + v_n$$

$$y_n = x_n + w_n$$

コーシー分布

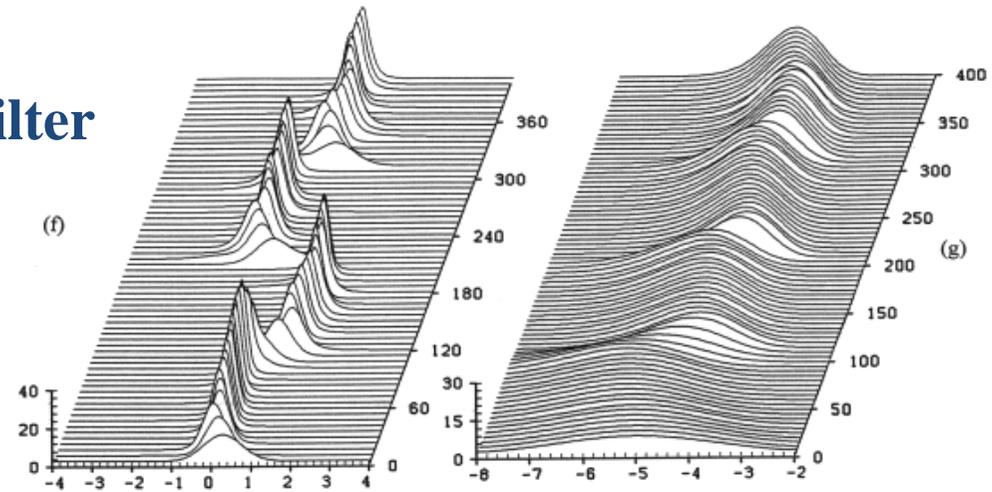
$$p(v_n) = \frac{\tau}{\pi} \frac{1}{(v_n^2 + \tau^2)}$$

Augmented State Space Model

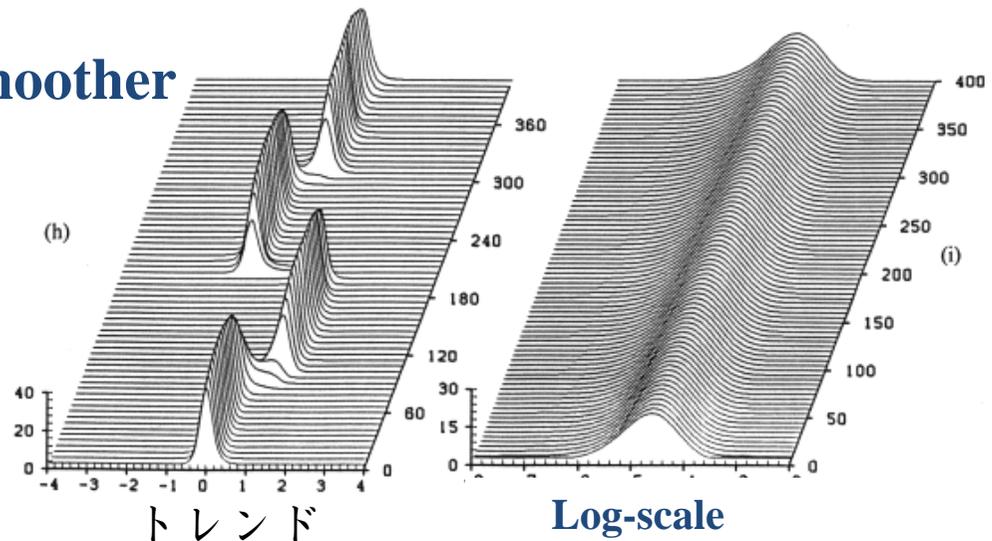
$$\begin{bmatrix} x_n \\ \log \tau_n^2 \end{bmatrix} = \begin{bmatrix} x_{n-1} \\ \log \tau_{n-1}^2 \end{bmatrix} + \begin{bmatrix} \tau_{n-1} \\ 0 \end{bmatrix} v_n$$

$$y_n = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} x_n \\ \log \tau_n^2 \end{bmatrix} + w_n$$

Filter

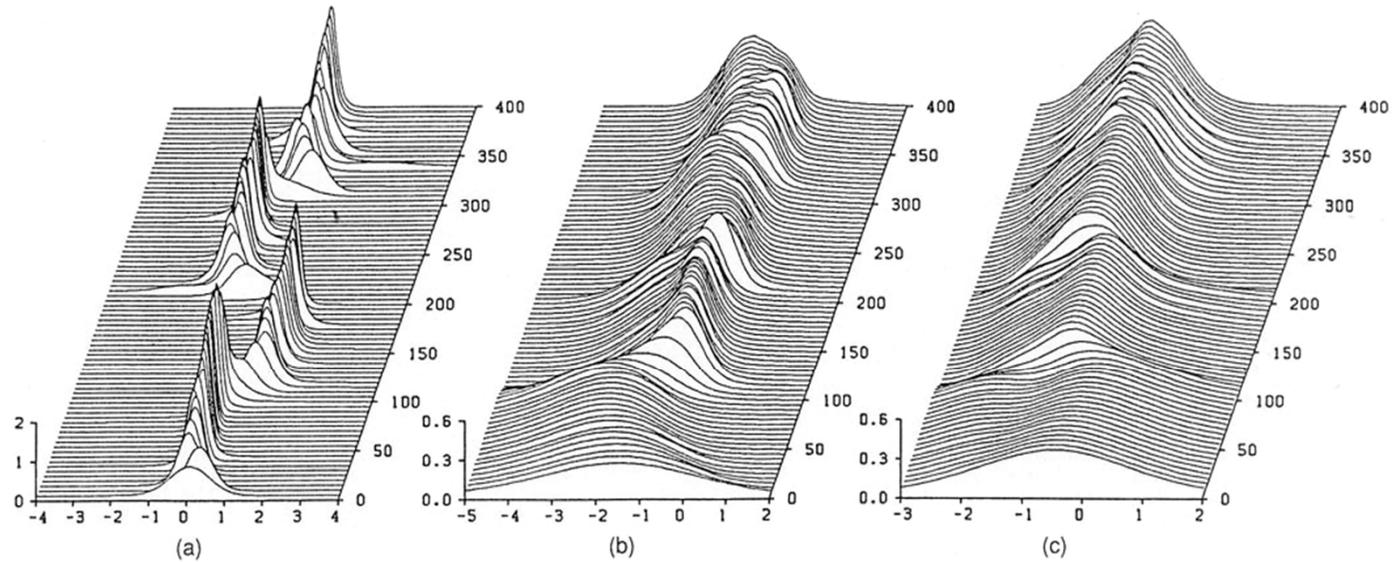


Smoother



自己組織型フィルタ・平滑化

Filter

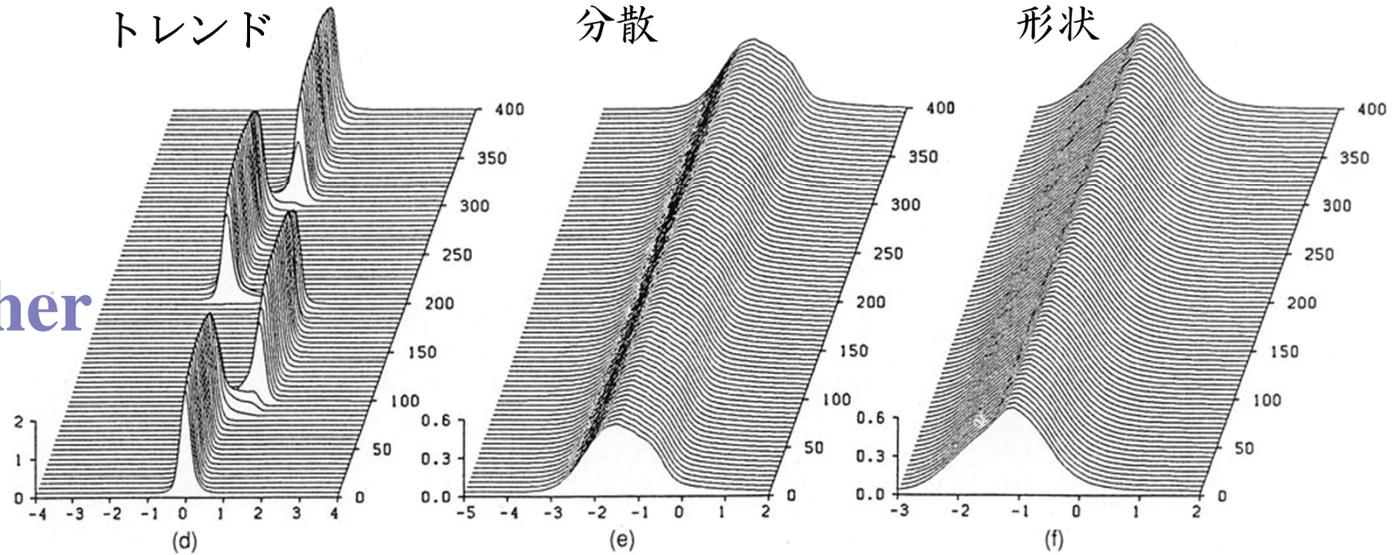


トレンド

分散

形状

Smoother



ガウス和フィルタ・平滑化

モデル

$$x_n = F_n x_{n-1} + G_n v_n$$

$$y_n = H_n x_n + w_n$$

ノイズ分布

$$p(v_n) \sim \sum_{i=1}^{K_v} \alpha_i \varphi_i(v_n)$$

$$p(w_n) \sim \sum_{j=1}^{K_w} \beta_j \varphi_j(w_n)$$

予測分布とフィルタ分布

$$p(x_n | Y_{n-1}) \sim \sum_{k=1}^{L_n} \gamma_k \varphi_k(x_n | Y_{n-1})$$

$$p(x_n | Y_n) \sim \sum_{\ell=1}^{M_n} \delta_\ell \varphi_\ell(x_n | Y_n)$$

$$\varphi_k(x_n | Y_{n-1}) \equiv N(x_{n|n-1}^{(k)}, V_{n|n-1}^{(k)}),$$

$$\varphi_\ell(x_n | Y_n) \equiv N(x_{n|n}^{(\ell)}, V_{n|n}^{(\ell)})$$

ガウス和フィルタ

Prediction $p(x_n | Y_{n-1}) \sim \sum_{k=1}^{L_n} \gamma_k \varphi_k(x_n | Y_{n-1})$

$$L_n = M_{n-1} K_v$$

$$x_{n|n-1} = F_n x_{n-1|n-1}$$

$$V_{n|n-1} = F_n V_{n-1|n-1} F_n^T + G_n Q_n G_n^T$$

Filter $p(x_n | Y_n) \sim \sum_{\ell=1}^{M_n} \delta_\ell \varphi_\ell(x_n | Y_n)$

$$\delta_n = \beta_j \gamma_{kn} \varphi(y_n | Y_{n-1}), \quad M_n = K_w L_n$$

$$K_n = V_{n|n-1} H_n^T (H_n V_{n|n-1} H_n^T + R_n)^{-1}$$

$$x_{n|n} = x_{n|n-1} + K_n (y_n - H_n x_{n|n-1})$$

$$V_{n|n} = (I - K_n H_n) V_{n|n-1}$$

混合ガウス成分の縮減

$$p(x_n | Y_{n-1}) \sim \sum_{k=1}^{L_n} \gamma_k \phi_k(x_n | Y_{n-1}) \Rightarrow \sum_{k=1}^m \gamma'_k \phi'_k(x_n | Y_{n-1})$$

$$p(x_n | Y_n) \sim \sum_{\ell=1}^{M_n} \delta_\ell \phi_\ell(x_n | Y_n) \Rightarrow \sum_{\ell=1}^m \delta'_\ell \phi'_\ell(x_n | Y_n)$$

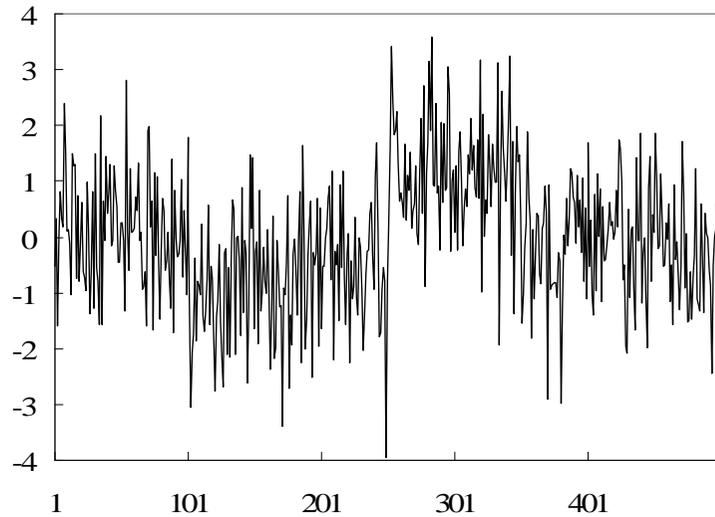
原理的にはKL-情報量最大化

$$I(p; q) \int \log \frac{p(x)}{q(x)} p(x) dx$$

現実的には分布数の逐次削減

$$\begin{aligned} D(\varphi_k, \varphi_m) &= -2\delta_k \delta_m \{I(\varphi_k; \varphi_m) + I(\varphi_m; \varphi_k)\} \\ &= \delta_k \delta_m \{V_k^{-1} V_m + V_m^{-1} V_k + (\mu_k - \mu_m)^T (V_k^{-1} + V_m^{-1}) (\mu_k - \mu_m)\} \\ \mu_{km} &= \frac{\delta_k \mu_k + \delta_m \mu_m}{\delta_k + \delta_m} \\ V_{km} &= \frac{\delta_k \{V_k + (\mu_{km} - \mu_k)(\mu_{km} - \mu_k)^T\} + \delta_m \{V_m + (\mu_{km} - \mu_m)(\mu_{km} - \mu_m)^T\}}{\delta_k + \delta_m} \end{aligned}$$

非ガウス型平滑化



Trend Model

$$t_n = t_{n-1} + v_n$$

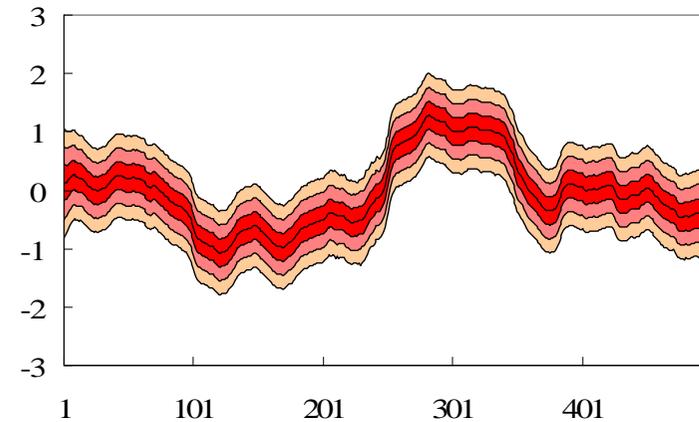
$$y_n = t_n + w_n$$

Noise Distribution

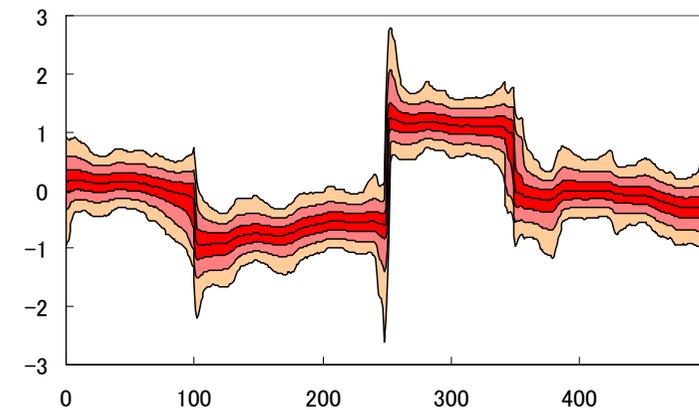
$$v_n \sim N(0, \tau^2) \text{ or } C(0, \tau^2)$$

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

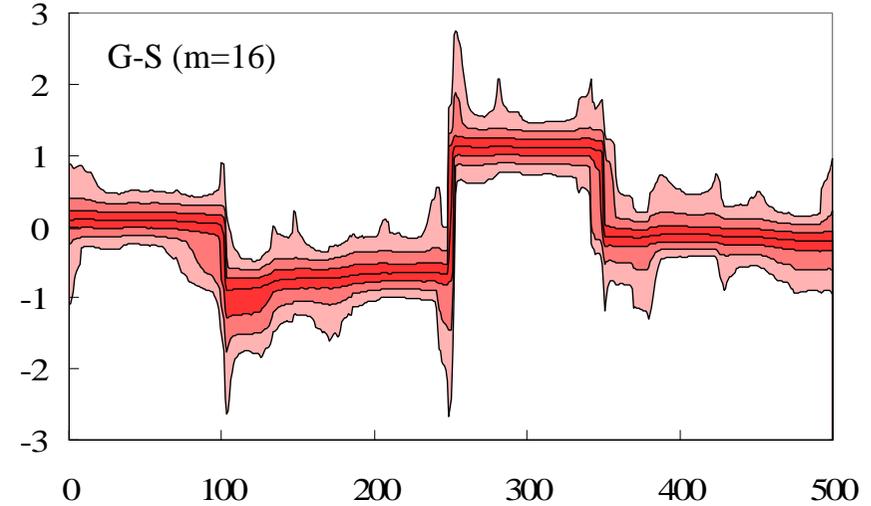
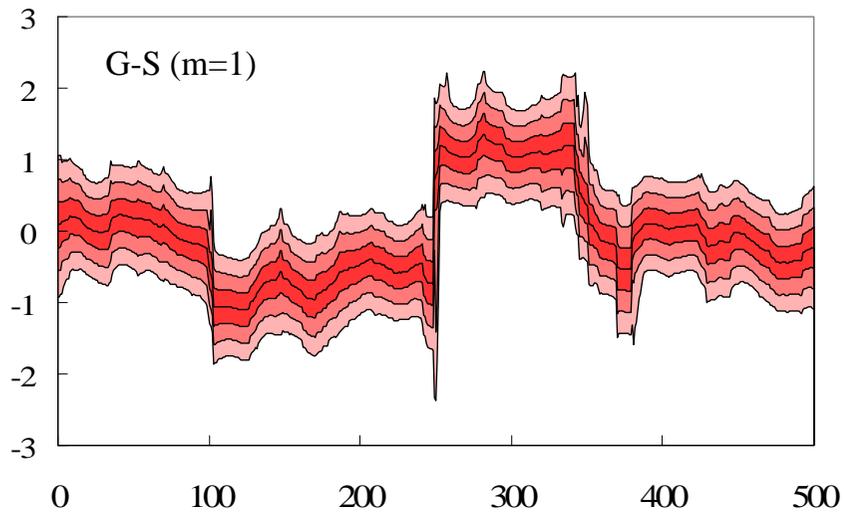
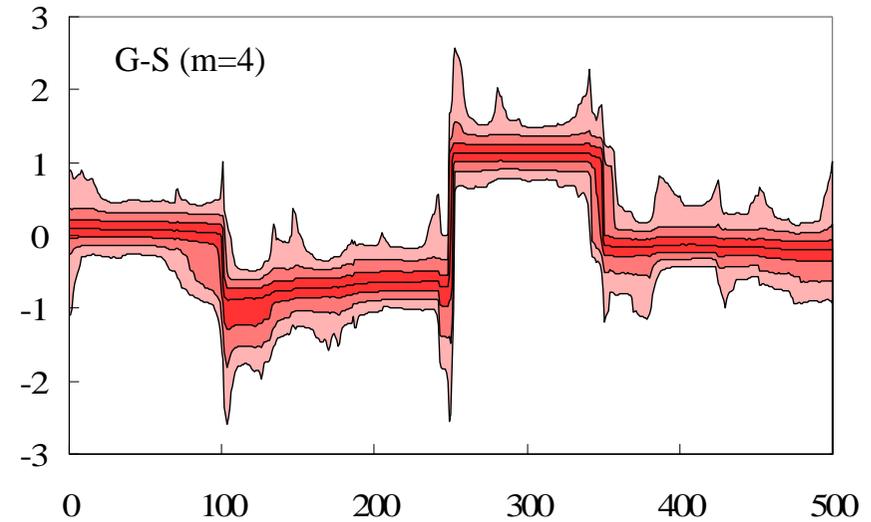
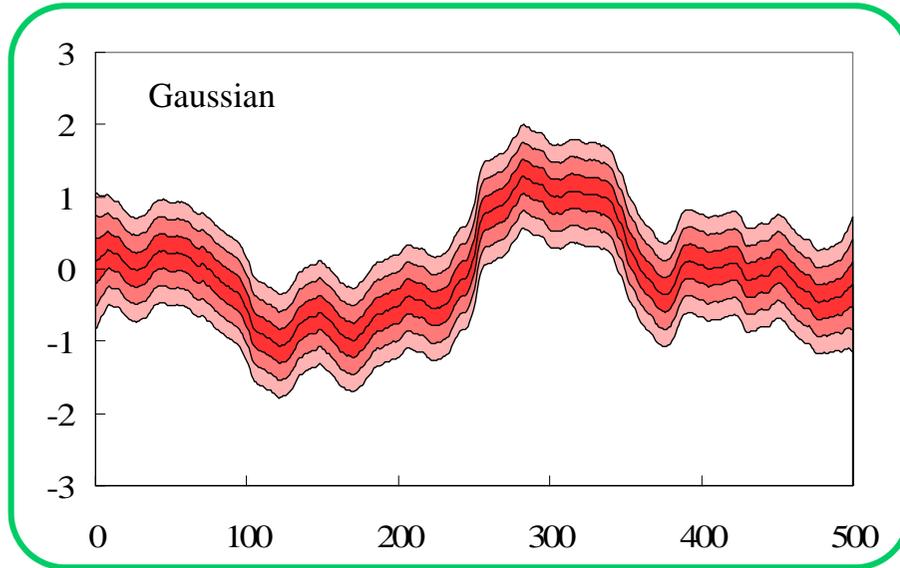
Kalman Smoother



Exact Non-Gaussian Smoother



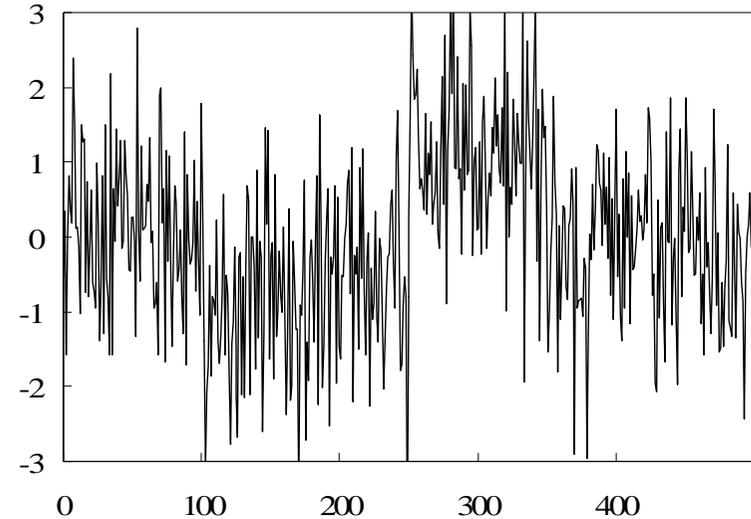
ガウス和フィルタ・平滑化



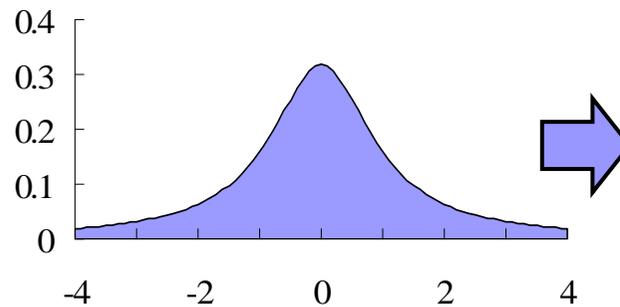
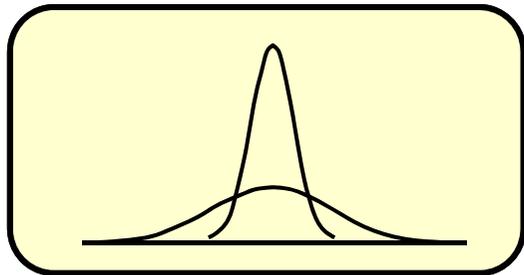
Examples: Gaussian-sum Filter with Resampling

Trend model

$$x_n = x_{n-1} + v_n$$
$$y_n = x_n + w_n$$



Distribution of v_n



Mixture of k distributions

