非線形・非ガウス状態空間モデルによる 月経周期のモデル化と予測

統計思考院　特任助教　深谷肇一

概要

月経周期に関連した基礎体温の概周期変動を状態空間モデルを用いてモデル化し、月経周期の位相と次回月経日を逐次的に予測するための枠組みを提案する。

状態空間モデル

データ

- 日次基礎体温
- 月経開始日

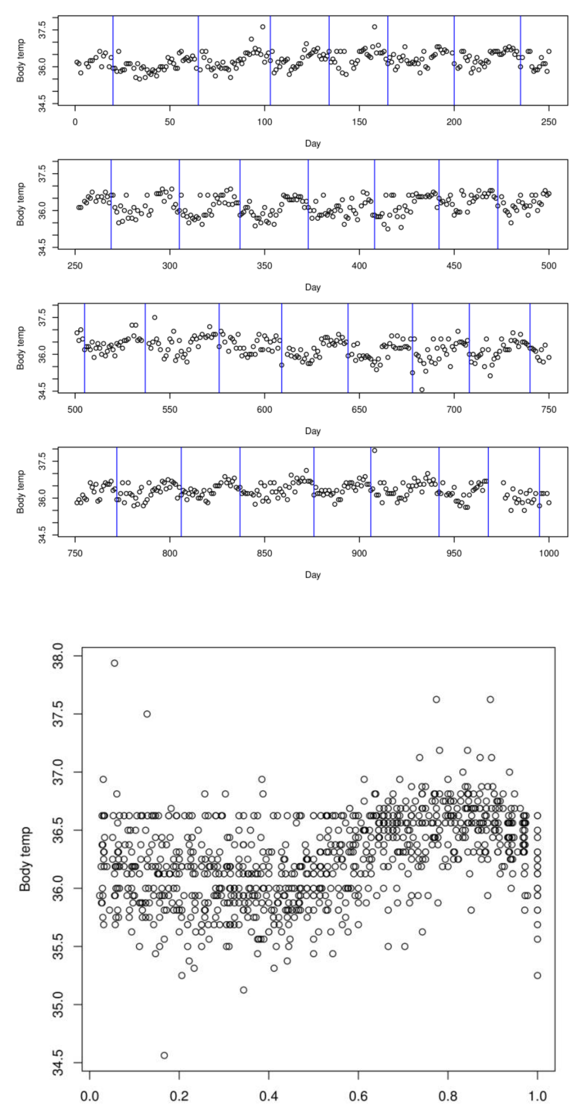

Fig. 1. 日次基礎体温データ

Notation

- y_t : t 日目に観測された温度
- z_t : t 日目に観測された位相
  - ➢ t 日目が月経開始日のとき $z_t = 1$
  - ➢ それ以外のとき $z_t = \mathrm{NA}$
- θ_t : t 日目の位相 （周期1の潜在変数）
  - ➢ $\theta_t \in (0, 1]$
- Y_t, Z_t, Θ_t: 時点 t までのデータ/ 状態時系列
  - ➢ $Y_t = (y_1, y_2, ..., y_t)$
  - ➢ $Z_t = (z_1, z_2, ..., z_t)$
  - ➢ $\Theta_t = (\theta_0, \theta_1, ..., \theta_t)$

観測モデル（体温）

- 温度の観測方程式

位相に関連した温度期待値の周期変動をフーリエ級数で近似する。

$$y_t = a + \sum_{m=1}^{M} [b_m \cos 2m\pi\theta_t + c_m \sin 2m\pi\theta_t] + e_t$$

$$e_t \sim \mathrm{Normal}\left(0, \sigma^2\right)$$

- y_tの条件付き分布

$$p(y_t|a, \mathbf{b}, \mathbf{c}, \theta_t, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{(y_t - a - \sum_{m=1}^{M}[b_m \cos 2m\pi\theta_t + c_m \sin 2m\pi\theta_t])^2}{2\sigma^2} \right\}$$

where $\mathbf{b} = (b_1, \ldots, b_M), \mathbf{c} = (c_1, \ldots, c_M)$

観測モデル（位相）

- 位相の観測方程式

月経開始日 $\theta_t = 1$ に位相の観測値 $z_t = 1$ が観測されると仮定する。

$$z_t = \theta_t \quad \text{when} \quad \theta_t = 1$$

- z_tの条件付き分布

$$p(z_t|\theta_t) = \delta(\theta_t - 1)\delta(z_t - 1)$$

システムモデル

- 位相のシステム方程式

位相θ_tは加法的なシステムノイズ $\epsilon_t > 0$ によって単調に増加すると仮定。

$$\theta_t = \theta_{t-1} + \epsilon_t \mod 1$$

$$\epsilon_t \sim \mathrm{Gamma}(\alpha, \beta)$$

- θ_tの条件付き分布（巻き込みガンマ分布, Coelho 2007）

$$f(\theta_t|\theta_{t-1}, \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \exp\left(-\beta(\theta_t - \theta_{t-1})\right) \Phi\left(\exp(-\beta), 1-\alpha, \theta_t - \theta_{t-1}\right)$$

$$\text{where} \quad \Phi(z, s, a) = \sum_{k=0}^{\infty} \frac{z^k}{(a+k)^s}$$

パラメータ推定値

$a = 36.36, b_1 = 0.12, b_2 = -0.05, b_3 = -0.04,$
$c_1 = -0.28, c_2 = -0.02, c_3 = -0.03, \sigma = 0.29,$
$\alpha = 1.2, \beta = 40.0$

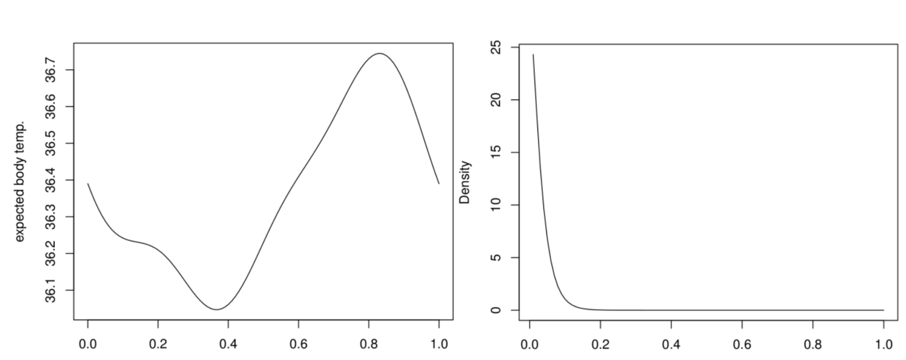

Fig. 2. 期待観測温度（左）とシステムノイズ密度関数（右）

逐次ベイズフィルタ

- 非ガウス型フィルタ(Kitagawa 1987)を用いて、位相の条件付き分布をオンライン推定する

$0 < \theta \leq 1$ を等間隔に分割するN 個の分点 $\{\theta(i); i = 1, \cdots, N, \theta(N) = 1\}$による密度関数 $p(\theta)$の階段関数近似を$\tilde{p}(\theta)$ と表す。

$$\tilde{p}(\theta) = \tilde{p}(\theta(i)) \quad \text{for} \quad \begin{array}{l} \theta \in (0, \theta(i)],\ i = 1 \\ \theta \in (\theta(i-1), \theta(i)],\ i = 2, \cdots, N \end{array}$$

1期先予測分布

$$p(\theta_t|Y_{t-1}, Z_{t-1}) = \int_0^1 p(\theta_t|\theta_{t-1}) p(\theta_{t-1}|Y_{t-1}, Z_{t-1}) d\theta_{t-1}$$
$$= \int_0^1 p(\theta_t - \theta_{t-1}) p(\theta_{t-1}|Y_{t-1}, Z_{t-1}) d\theta_{t-1}$$

$\tilde{p}_\mathrm{p}(\theta) := \tilde{p}(\theta_t|Y_{t-1}, Z_{t-1}),\ \tilde{p}_\mathrm{d}(\theta) := \tilde{p}(\theta_t - \theta_{t-1}),$
$\tilde{p}_\mathrm{f}(\theta) := \tilde{p}(\theta_{t-1}|Y_{t-1}, Z_{t-1})$ として

$$\tilde{p}_\mathrm{p}(\theta(i)) \simeq \frac{1}{N} \sum_{j=1}^{N} \tilde{p}_\mathrm{d}(\theta(i-j)) \tilde{p}_\mathrm{f}(\theta(j)) \quad \text{for } i = 1, \cdots, N$$

ただし $\theta(k) := \theta(N+k)$ for $k < 1$

フィルタ分布

$$p(\theta_t|Y_t, Z_t) = \frac{p(y_t|\theta_t) p(z_t|\theta_t) p(\theta_t|Y_{t-1}, Z_{t-1})}{\int_0^1 p(y_t|\theta_t) p(z_t|\theta_t) p(\theta_t|Y_{t-1}, Z_{t-1}) d\theta_t}$$

$\tilde{p}_\mathrm{f}(\theta) := \tilde{p}(\theta_t|Y_t, Z_t),\ \tilde{p}_\mathrm{p}(\theta) := \tilde{p}(\theta_t|Y_{t-1}, Z_{t-1})$ として

$$\tilde{p}_\mathrm{f}(\theta(i)) \simeq \frac{p(y_t|\theta(i)) p(z_t|\theta(i)) \tilde{p}_\mathrm{p}(\theta(i))}{\sum_{j=1}^{N} p(y_t|\theta(j)) p(z_t|\theta(i)) \tilde{p}_\mathrm{p}(\theta(j)) / N}$$
$$\text{for } i = 1, \cdots, N$$

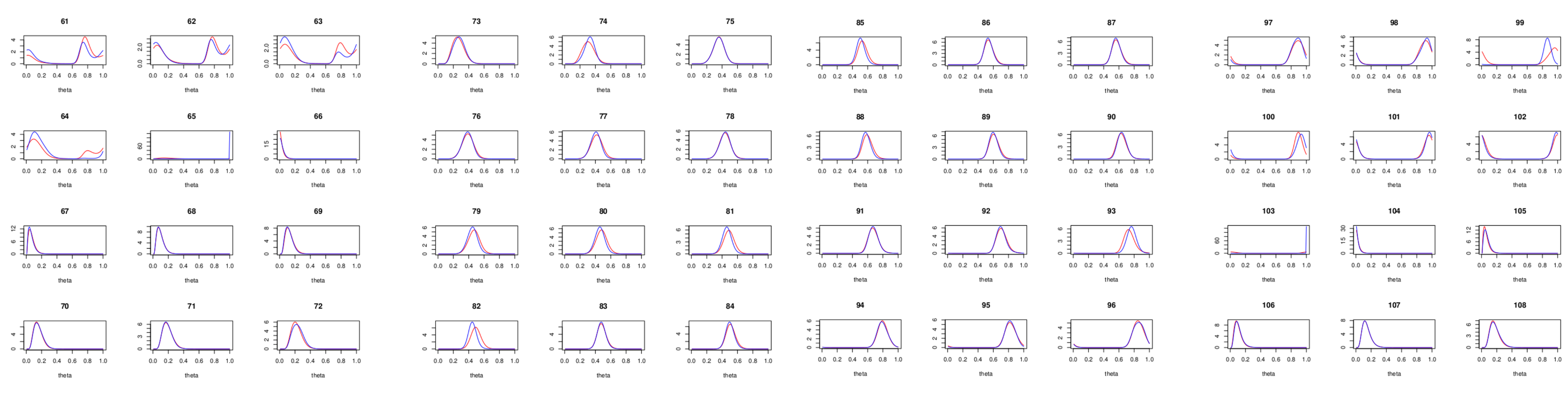

Fig. 3. 推定された条件付き分布の例

月経開始日の予測

- 位相のフィルタ分布から次の月経開始日の予測分布を得る

$\mathrm{ga}(\cdot; s, r)$ と $\mathrm{Ga}(\cdot; s, r)$をそれぞれ形状 s 尺度 r のガンマ分布のPDFおよびCDF、$\Delta_k(t) := \sum_{r=t+1}^{t+k} \epsilon_r, k = 1, 2, \ldots$として、位相が$\theta_t$のとき t から k 日後に月経が起こっている確率 $\Pr(\Delta_k(t) > 1 - \theta_t)$ を考える。

月経開始日の条件付き分布関数

$$F(k|\theta_t) := \Pr(\Delta_k(t) > 1 - \theta_t) = \int_{1-\theta_t}^{\infty} \mathrm{ga}(x; k\alpha, \beta) dx \quad = 1 - \mathrm{Ga}(1 - \theta_t; k\alpha, \beta)$$

月経開始日の条件付き確率関数

$$f(k|\theta_t) := F(k|\theta_t) - F(k-1|\theta_t)$$
$$= (1 - \mathrm{Ga}(1 - \theta_t; k\alpha, \beta)) - (1 - \mathrm{Ga}(1 - \theta_t; (k-1)\alpha, \beta))$$
$$= \mathrm{Ga}(1 - \theta_t; (k-1)\alpha, \beta) - \mathrm{Ga}(1 - \theta_t; k\alpha, \beta)$$

ただし $F(0|\theta_t) = 0$

月経開始日の周辺確率関数

$$f(k|Y_t, Z_t) = \int f(k|\theta_t) p(\theta_t|Y_t, Z_t) d\theta_t$$

$\tilde{p}_\mathrm{f}(\theta) := \tilde{p}(\theta_t|Y_t, Z_t)$ として

$$f(k|Y_t, Z_t) = \tilde{p}_\mathrm{f}(\theta(1)) \int_0^{\theta(1)} f(k|\theta_t) d\theta_t + \cdots + \tilde{p}_\mathrm{f}(\theta(N)) \int_{\theta(N-1)}^{\theta(N)} f(k|\theta_t) d\theta_t$$
$$= \tilde{p}_\mathrm{f}(\theta(1)) \int_0^{\theta(1)} \mathrm{Ga}(1 - \theta_t; (k-1)\alpha, \beta) - \mathrm{Ga}(1 - \theta_t; k\alpha, \beta) d\theta_t$$
$$+ \cdots + \tilde{p}_\mathrm{f}(\theta(N)) \int_{\theta(N-1)}^{\theta(N)} \mathrm{Ga}(1 - \theta_t; (k-1)\alpha, \beta) - \mathrm{Ga}(1 - \theta_t; k\alpha, \beta) d\theta_t$$

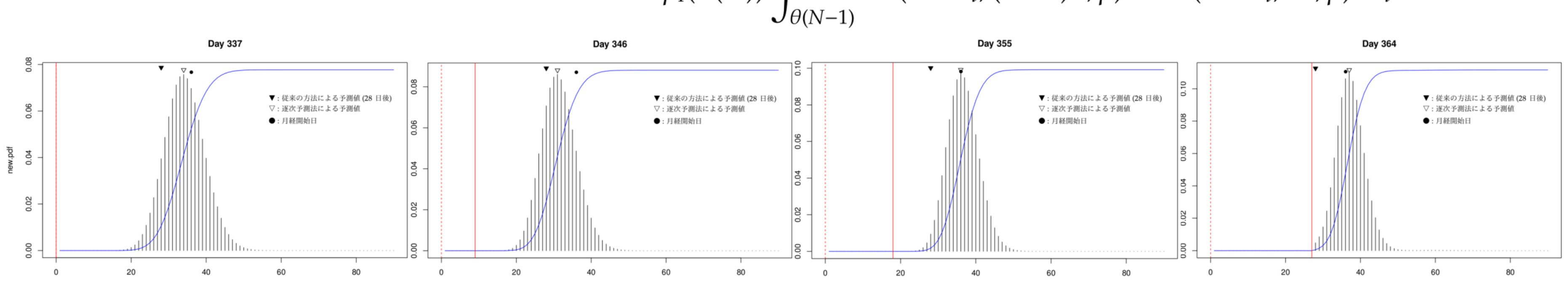

Fig. 4. 推定された月経開始日の予測分布の例