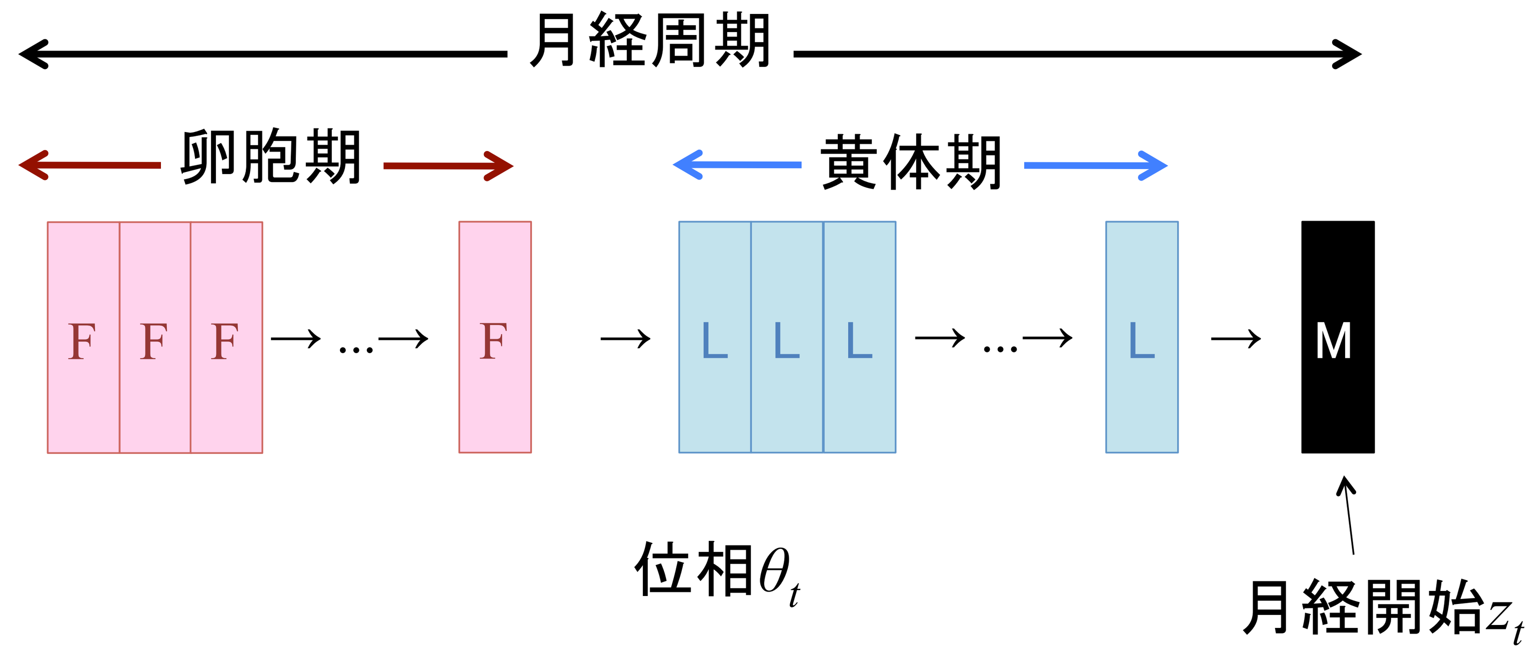
2016年6月17日 統計数理研究所 オープンハウス

自己励起型状態空間モデルを用いた月経周期解析

川森 愛　　データ科学研究系 特任研究員

[月経周期]

- 1周期は約28日
- 排卵を境に卵胞期[F]と黄体期[L]に分かれる
- 卵胞期の日数長はばらつきが大きい
- 黄体期の日数長はばらつきが小さい（約14日）
- 平均体温は卵胞期で低く，黄体期には高い（0.3〜0.5℃差）



[目標]

体温データから次の月経開始日を予測する

[状態空間モデル]

θ_t :時刻tにおける位相
y_t :時刻tにおける体温データ
z_t :時刻tに月経開始したか否か

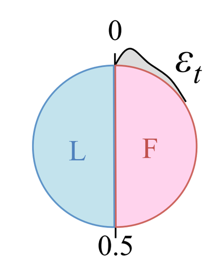
1. 過程モデル

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

$\varepsilon_t \sim \mathrm{Gamma}(\alpha_t, \beta_t)$

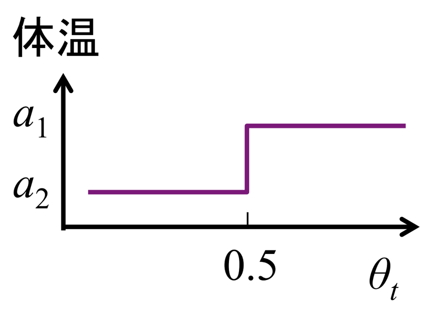
$(\alpha_t, \beta_t) = (\alpha_1, \beta_1)$ when $\theta_t < 0.5$
$\quad = (\alpha_2, \beta_2)$ when $\theta_t \geq 0.5$

自己励起モデル



2. 観測モデル

(体温)



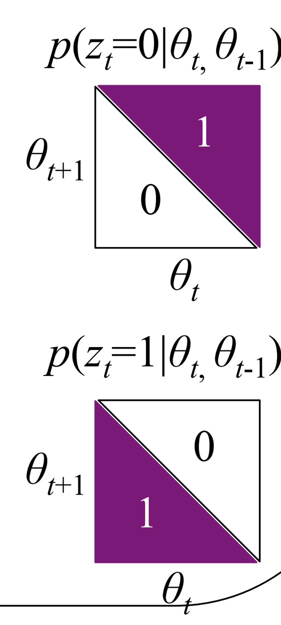
$y_t \sim \mathrm{Norm}(a_t, \sigma_t)$

$(a_t, \sigma_t) = (a_1, \sigma_1)$ when $\theta_t < 0.5$
$\quad = (a_2, \sigma_2)$ when $\theta_t \geq 0.5$

(月経開始)

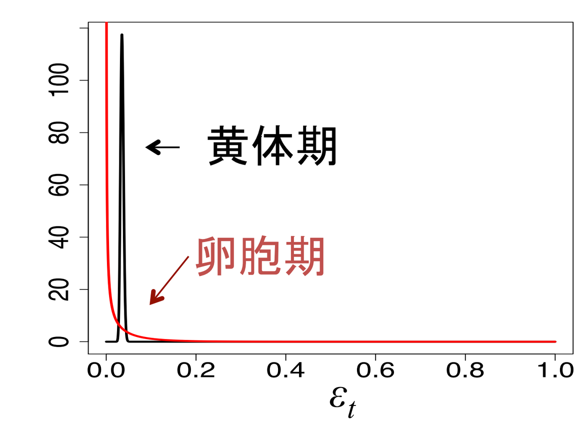
$z_t = 0$ when $\theta_t > \theta_{t-1}$
$z_t = 1$ when $\theta_t \leq \theta_{t-1}$

$p(z_t|\theta_t, \theta_{t-1}) = (1-z_t)I(\theta_t > \theta_{t-1}) + z_t I(\theta_t \leq \theta_{t-1})$



[パラメータ推定値]

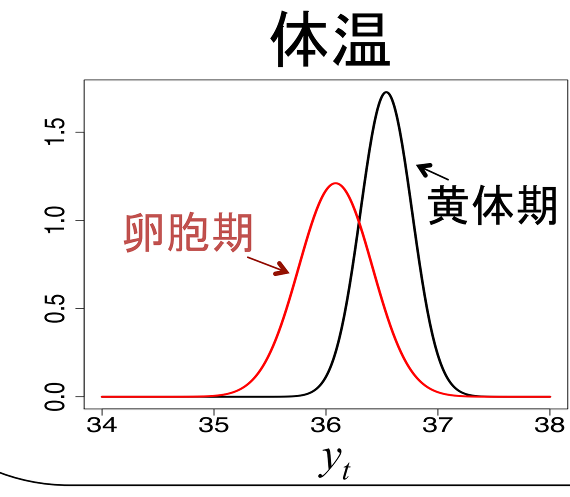
システムノイズ



$\alpha_1 = 0.3255499$
$\beta_1 = 12.4047963$

$\alpha_2 = 106.9998992$
$\beta_2 = 3036.2378939$

体温



$a_1 = 36.0861182$
$\sigma_1 = 0.3295040$

$a_2 = 36.5387787$
$\sigma_2 = 0.2309690$

[位相の推移]

逐次ベイズフィルタ

予測分布 ・・・・・・・・・

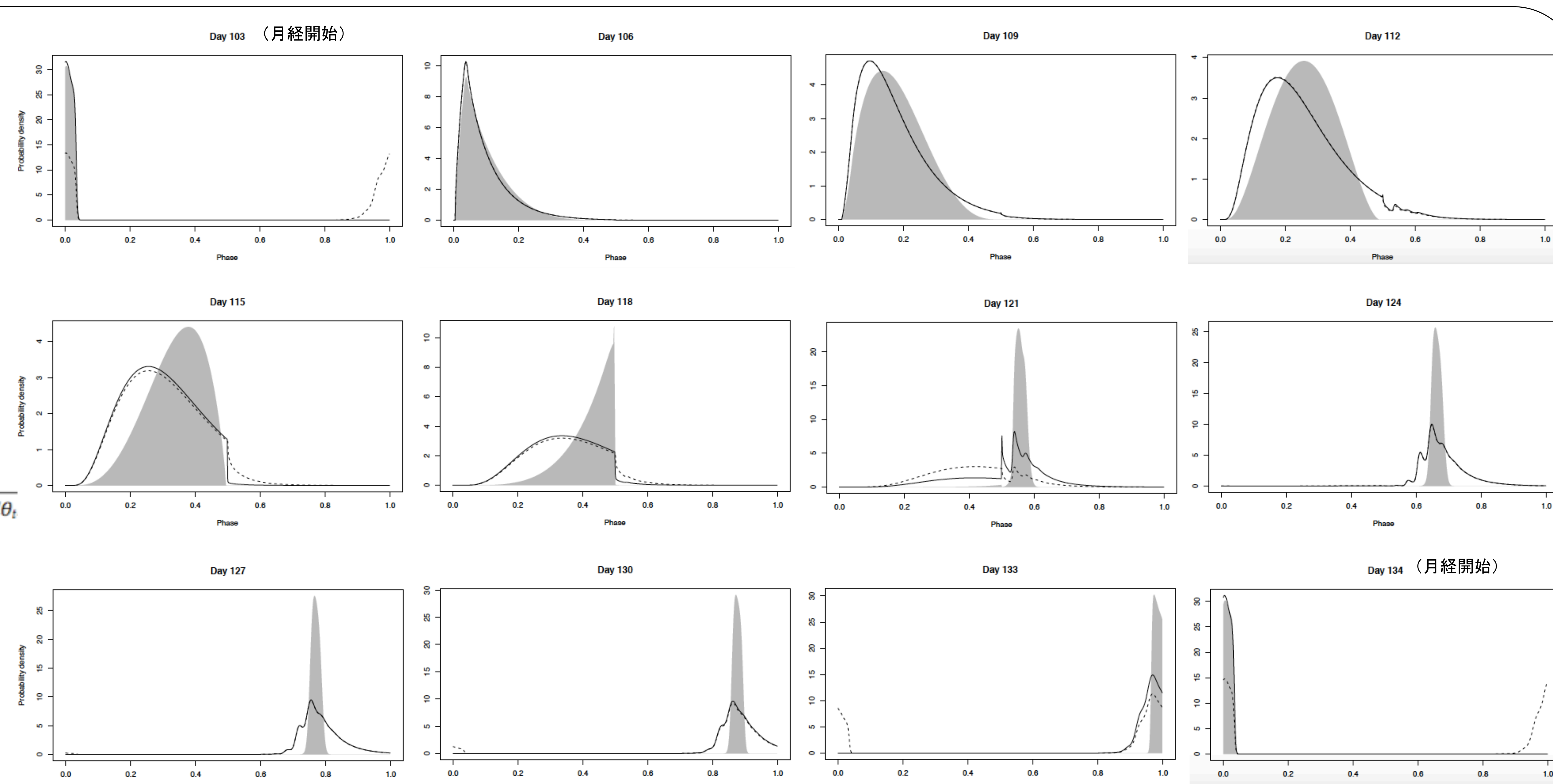
$p(\theta_{t+1}, \theta_t|Y_t, Z_t) = p(\theta_{t+1}|\theta_t)p(\theta_t|Y_t, Z_t)$

フィルタ分布 ———

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

平滑化分布 ■

$$p(\theta_t, \theta_{t-1}|Y_T, Z_T) = \frac{p(\theta_t, \theta_{t-1}|Y_t, Z_t)p(\theta_t|Y_T, Z_T)}{p(\theta_t|Y_t, Z_t)}$$



[月経開始日予測]

STEP1．s日後にフェイズ移行し，
k日後に月経が起こる確率$f_s(k|\theta_t)$を求める

STEP2．k日後に月経が起こる確率$f(k|\theta_t)$を求める

$$\sum_{s=1}^{k} f_s(k|\theta_t) = f(k|\theta_t)$$

STEP3．月経予測分布$h(k|Y_t, Z_t)$を求める

$$h(k|Y_t, Z_t) = \int_0^1 f(k|\theta_t)p(\theta_t|Y_t, Z_t)d\theta_t$$

$f_s(k|\theta_t)$の求め方

F → F　$P[\theta^F_{t+1}|\theta^F_t] = P[\theta_t + \epsilon^F_{t+1} < 0.5] = P[\epsilon^F_{t+1} < 0.5 - \theta_t]$

F → L　$P[\theta^L_{t+1}|\theta^F_t] = P[\theta_t + \epsilon^F_{t+1} \geq 0.5] - P[\theta_t + \epsilon^F_{t+1} < 1]$
$= P[\epsilon^F_{t+1} \geq 0.5 - \theta_t] - P[\epsilon^F_{t+1} < 1 - \theta_t]$

L → L　$P[\theta^L_{t+1}|\theta^L_t] = P[\theta_t + \epsilon^L_{t+1} < 1] = P[\epsilon^L_{t+1} < 1 - \theta_t]$

L → M　$P[\theta^M_{t+1}|\theta^L_t] = P[\epsilon^L_{t+1} \geq 1 - \theta_t]$

$f_s(k|\theta_t) = \prod_{i=1}^{s}$ [F→F] × [F → L] × $\prod_{i=1}^{k-s}$ [L → L] × [L → M]

$P[\varepsilon < x]$ 及び $P[\varepsilon \geq x]$の求め方

$\varepsilon^F \sim f(t)$, $\varepsilon^L \sim g(t)$
$\varepsilon = \varepsilon^F + \varepsilon^L$, $\varepsilon \sim m(t)$

ただし，$f(t)$, $g(t)$はガンマ分布の確率密度関数

畳み込みにより，

$$m(t) = f * g(t) = \int_0^t f(\tau)g(t-\tau)\,d\tau$$

累積分布関数$M(t)$

$$M(t) = \int_0^t m(u)\,du = \int_0^t \int_0^u f(\tau)g(u-\tau)\,d\tau du = 1 - \int_t^\infty \int_0^u f(\tau)g(u-\tau)\,d\tau du$$

$P[\epsilon < x] = M(x)$，　$P[\epsilon \geq x] = 1 - M(x)$