
ПОГОДИ КТО ЭТО
НА НАС ПЯЛИТСЯ
КАКОЙ-ТО НЕКИЙ
НЕДОТЁПА ИЛИ КТО

Сем 2. ETS

$$\begin{cases} y_t = l_{t-1} + b_{t-1} + s_{t-s} + u_t \\ l_t = l_{t-1} + b_{t-1} + \alpha u_t \\ b_t = b_{t-1} + \beta u_t \\ s_t = s_{t-s} + \gamma u_t \end{cases} \qquad u_t \sim N(0, \sigma^2)$$

$(\alpha, \beta, \gamma, \sigma^2)$ – хотим оценить.

Заметим, что $y_t | F_{t-1} \sim N(l_{t-1} + b_{t-1} + s_{t-s}, \sigma^2)$, тогда

$$p(y_t | y_{t-1}, \theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}[y_t - \mu_t]^2\right\}$$, где $\mu_t = l_{t-1} + b_{t-1} + s_{t-s}$.

Значит.

$$p(y_1 \dots y_T) = p(y_1) \cdot p(y_2|y_1) \cdot p(y_3|y_2) \dots p(y_T|y_{T-1}) =$$

$$= \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left\{-\frac{(y_t - (l_{t-1} + b_{t-1} + s_{t-s}))^2}{2\sigma^2}\right\}$$

Тогда $\ln p = -\frac{T}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{t=1}^{T} (y_t - l_{t-1} - b_{t-1} - s_{t-s})^2$

$$\hat{\sigma}^2 = \frac{1}{T} \sum_{t=1}^{T} (y_t - l_{t-1} - b_{t-1} - s_{t-s})^2$$

А дальше всё.

При фикс. σ работаем с

$$\ln p = \sum_{t=1}^{T} u_t^2 \sim \sum_{t=1}^{T} (\underbrace{y_t - l_{t-1} - b_{t-1}}_{u_t})^2$$

, сразу берем $-\ln L$ и на S забиваем т.к. с ним все аналогично

Тогда: $$-\frac{d l_{t-1}}{d\alpha} - \frac{d b_{t-1}}{d\alpha} = \frac{d u_t}{d\alpha}$$

Рассмотрим l_t для примера

1) $l_t = l_{t-1} + b_{t-1} + \alpha u_t \Big| \frac{d}{d\alpha}$, тогда

$$\frac{d l_t}{d\alpha} = \frac{d l_{t-1}}{d\alpha} + \frac{d b_{t-1}}{d\alpha} + u_t + \alpha \frac{d u_t}{d\alpha}$$

$$\frac{d\ell_t}{d\alpha} = \frac{d\ell_{t-1}}{d\alpha} + \frac{db_{t-1}}{d\alpha} + u_t - \alpha\left(\frac{d\ell_{t-1}}{d\alpha} + \frac{db_{t-1}}{d\alpha}\right)$$

Аналогично получаем формулы для ост. ур. т.е находим $\frac{db_t}{d\alpha}$ и $\frac{ds_t}{d\alpha}$. Считаем $\frac{d\ell_0}{d\alpha} = \frac{db_0}{d\alpha} = \frac{ds_0}{d\alpha} = 0$ т.к ℓ_0 и ост. это константы. В итоге получаем

$$\frac{d\ln L}{d\alpha} = -2\sum_{t=1}^{T} u_t\left(\frac{d\ell_{t-1}}{d\alpha} + \frac{db_{t-1}}{d\alpha} + \frac{ds_{t-s}}{d\alpha}\right)$$

$$\frac{d\ln L}{d\beta} = \ldots$$

$$\frac{d\ln L}{d\gamma} = \ldots$$

— эту тему мы и накапливаем при обратном проходе

В итоге алгоритм оценки по шагам (по ша- гам)

I Инициализация:

Есть ряд $y_1 \ldots y_T$ - наблюдаемый, мы задаём начальные состояния ℓ_0, b_0 и s_0 и параметры: α^0, β^0, Стартовые копилки производных:

$\frac{\partial \ell_0}{d\alpha} = 0$, $\frac{\partial \ell_0}{\partial \alpha} = 0$ и так далее остальные.

II Forward pass: пусть мы на итерации τ, т.е. имеем $\hat{\alpha}_\tau$, $\hat{\beta}_\tau$ и т.д.

Для $t = 1 \ldots T$ считаем прогноз $\hat{y}_t = \ell_{t-1} + b_{t-1}$, находим ошибку:

$u_t = y_t - \hat{y}_t$, затем обновляем состояния:

$$\begin{cases} \ell_t = \ell_{t-1} + b_{t-1} + \hat{\alpha}_\tau u_t \\ b_t = b_{t-1} + \hat{\beta}_\tau u_t \end{cases}$$

Тип тут просто считаем ETS по формулам

III Накопление градиентов.

$$\frac{du_t}{d\alpha} = -\left(\frac{d\ell_{t-1}}{d\alpha} + \frac{db_{t-1}}{d\alpha}\right)$$

Обновляем производные состояний

$$\frac{d\ell_t}{d\alpha} = (1-\hat{\alpha}_\tau)\left(\frac{d\ell_{t-1}}{d\alpha} + \frac{db_{t-1}}{d\alpha}\right) + u_t$$

$\frac{d\ell_t}{d\alpha} = \frac{d\ell_{t-1}}{d\alpha} - \hat\beta_\tau\left(\frac{d\ell_{t-1}}{d\alpha} + \frac{db_{t-1}}{d\alpha}\right)$

и так далее по формулам для всех параметров

Тогда итоговый град обновляется как:

$$\frac{d \ln L_t}{d\alpha} = \frac{1}{\sigma^2} u_t\left(\frac{d\ell_{t-1}}{d\alpha} + \frac{db_t}{d\alpha}\right)$$ – копим антиград.

и так для всех α, β, σ

IV. Сделали проход, посчитали град. теперь обновляем параметры, например: $\hat\alpha_{\tau+1} \leftarrow \hat\alpha_\tau + lr \cdot \frac{dL}{d\alpha}$. Но т.к. задача у нас с огр. т.е $\alpha + \beta < 1$, то обычно используют что-то типа L-BFGS.

Всё, обновили параметры и имеем $\hat\alpha_{\tau+1}, \hat\beta_{\tau+1}$ и так далее. Теперь идём к шагу 1.

Предсказание на h шагов вперёд:

Пусть $x_t = \begin{pmatrix} \ell_t \\ b_t \end{pmatrix}$, тогда:

$$\begin{cases} x_t = A x_{t-1} + k u_t \\ y_t = C x_{t-1} + u_t \end{cases}$$, где $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$, $k = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$, $C = (1, 1)$

Факт: $x_t = A^t x_0 + \sum_{j=1}^{t} A^{t-j} k u_j$, заметим, что $A^k = \begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix}$

Подставим в y_t:

$y_t = C\left(A^t x_0 + \sum_{j=1}^{t} A^{t-j-1} k \varepsilon_j\right) + \varepsilon_t$, заметим, что

$C A^{t-1-j} k = (1, 1)\begin{bmatrix} 1 & t-j-1 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \alpha + (t-j)\beta$, а $C A^{t-1} x_0 = \ell_0 + t b_0$

Так как

$x_t = A x_{t-1} + k \varepsilon_t$

$x_t = A(A x_{t-2} + k_{t-1}) + k u_t = A^2 x_{t-2} + A k u_{t-1} + k u_t$

и т.д. короч всё видно

В итоге получаем, что

$$y_t = \ell_0 + t b_0 + \sum_{j=1}^{t-1} (\alpha + (t-j)\beta) u_j + u_t$$

Тогда предсказание на h шагов с момента s $(y_{s+h}|y_s)$ имеет вид.

$$y_{s+h} = \ell_s + h b_s + \sum_{k=1}^{h-1} (\alpha + (h-k)\beta) u_{s+k} + u_{s+h}$$

- Значит $\mathbb{E}[y_{s+h}|y_s] = \ell_s + h\, b_s$
- $Var(y_{s+h}|y_s) = \sigma^2 \left[1 + \sum_{k=1}^{h} (\alpha + (h-k)\beta)^2\right]$ – кубич. рост по h

- $Cov(y_t, y_s) = \mathbb{E}[y_t - Ey_t)(y_s - Ey_s)]$, очев: $\mathbb{E}[y_t] = \ell_0 + t b_0$,
  $\mathbb{E}[y_s] = \ell_0 + s\, b_0$

Тогда

$$Cov(y_t; y_s) = \mathbb{E}\left[\left(\sum_{j=1}^{t-1} (\alpha + (t-j)\beta) u_j + u_t\right)\left(\sum_{i=1}^{s-1} (\alpha + (s-i)\beta) u_i + u_s\right)\right]$$

Заметим, что $\mathbb{E}[u_i u_j] = 0$, $\mathbb{E}[u_i^2] = \sigma^2$, опять заметим

- $\mathbb{E}\left[\varepsilon_t \sum_{i=1}^{s-1} (\alpha + (s-i)\beta) u_i\right] = 0$
- $\mathbb{E}\left[\varepsilon_s \sum_{j=1}^{t-1} (\sim) u_j\right] = \sum_{j=1}^{t-1} (\sim\sim)\, \mathbb{E}[u_s u_j] = (\alpha + (t-s)\beta)\sigma^2$
- $\mathbb{E}\left[\sum_{j=1}^{t-1} \sum_{i=1}^{s-1} (\sim)(\sim) u_j u_i\right]$, не нуль при $i = j$, тогда имеем

$$\sum_{j=1}^{s-1} (\alpha + (t-j)\beta)(\alpha + (s-j))\sigma^2$$

В итоге получаем: $\sigma^2 \sum_{j=1}^{s-1} (\alpha + (t-j)\beta)(\alpha + (s-j)\beta) + \sigma^2(\alpha - (t-s)\beta$

CI для ед. ETS

Т.к. y_{s+h} – лин. комб. шумов, то

$$y_{s+h}|F_s \sim \mathcal{N}(\hat{y}_{s+h|s}, Var(y_{s+h}|F_s))$$

Значит $CI = \ell_s + h b_s \pm z_{1-\alpha/2} \cdot \sigma \sqrt{1 + \sum_{k}^{h-1} (\alpha + (h-k)\beta)^2}$

Приколы: Сведение к RNN

Рассмотрим ETS(AAN), тогда

$$\begin{cases} y_t = l_{t-1} + b_{t-1} + \varepsilon_t \\ l_t = l_{t-1} + b_{t-1} + \alpha\varepsilon_t \\ b_t = b_{t-1} + \beta\varepsilon_t \end{cases},$$

Пусть hid-state $h_t = \begin{pmatrix} l_t \\ b_t \end{pmatrix}$, тогда $\hat{y}_t = W_y h_{t-1}$, $W_y = [1, 1]$, а

$h_t = A h_{t-1} + k\varepsilon_t$, где $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$, $k = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$

$h_t = A h_{t-1} + k(y_t - W_y h_{t-1})$, тогда

$h_t = (A - kW_y) h_{t-1} + k y_t$, пусть $W_h = A - kW_y$, $W_x = k$

Значит

$$\begin{cases} h_t = W_h h_{t-1} + W_x y_t, \text{ а} \\ y_t = W_y \cdot h_{t-1} \end{cases}$$

– линейная RNN

$$\left\{ \begin{aligned} h_t &= \phi(W_h h_{t-1} + W_x x_t + b) \\ y_t &= \varphi(W_y h_t + c) \end{aligned} \right\}$$