Bayesian ML Project

Daria Demidova

October 2018

1 State space formulation

Equations from the paper

$y_t = \mu_t + \gamma_t + z_t^a o_t + (1 - z_t^a)\varepsilon_t$

$\mu_t = \mu_{t-1} + \delta_{t-1} + z_t^c r_t + (1 - z_t^c) u_t$

$\delta_t = \delta_{t-1} + v_t$

$\gamma_t = -\sum_{s=1}^{S} \gamma_{t-s} + w_t$

State space equations

$y_t = Z_t \alpha_t + A_t o_t + (1 - A_t)\varepsilon_t$

$\alpha_{t+1} = T_t \alpha_t + R_t C_t \eta_t + R_t (1 - C_t)\ \xi_t$

$o_t \sim \mathcal{N}(0, H_t^o)$

$\varepsilon_t \sim \mathcal{N}(0, H_t^\varepsilon)$

$\eta_t \sim \mathcal{N}(0, Q_t^\eta)$

$\xi_t \sim \mathcal{N}(0, Q_t^\xi)$

Where

$\alpha_t = (\mu_t, \delta_t, \gamma_t, \gamma_{t-1}, \ldots, \gamma_{t-S+2})^T$

$A_t = z_t^a \sim \text{Bernoulli}(p_a)$

$C_t = z_t^c \sim \text{Bernoulli}(p_c)$

$\eta_t = (r_t, v_t, w_t)$

$\xi_t = (u_t, v_t, w_t)$

$Z_t = (1, 0, 1, 0, \ldots, 0)$

$T_t = \text{diag}(T_\mu, T_\gamma)$

$T_\mu = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$

$$T_\gamma = \begin{pmatrix} -1 & -1 & \dots & -1 & -1 \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{pmatrix}$$

$R = \mathrm{diag}(R_\mu, R_\gamma)$

$R_\mu = I_2$

$R_\gamma = (1, 0, \dots, 0)^T$

$H_t^o = \sigma_o^2$

$H^\varepsilon = \sigma_\varepsilon^2$

$Q^\eta = \mathrm{diag}(\sigma_r^2, \sigma_v^2, \sigma_w^2)$

$Q^\xi = \mathrm{diag}(\sigma_u^2, \sigma_v^2, \sigma_w^2)$

2 Kalman Filter

$a_{t|t} = \mathbb{E}(\alpha_t|Y_t)$

$a_{t+1} = \mathbb{E}(\alpha_{t+1}|Y_t)$

$P_{t|t} = \mathrm{Var}(\alpha_t|Y_t)$

$P_{t+1} = \mathrm{Var}(\alpha_{t+1}|Y_t)$

$v_t = y_t - \mathbb{E}(y_t|Y_{t-1}) = y_t - \mathbb{E}(Z_t\alpha_t + A_t o_t + (1-A_t)\varepsilon_t|Y_{t-1}) = y_t - Z_t a_t$

$a_{t|t} = \mathbb{E}(\alpha_t|Y_t) = \mathbb{E}(\alpha_t|Y_{t-1}, v_t)$

$a_{t+1} = \mathbb{E}(\alpha_{t+1}|Y_t) = \mathbb{E}(\alpha_{t+1}|Y_{t-1}, v_t)$

Lemma 1

If $(x, y) = \mathcal{N}((\mu_x, \mu_y), \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{xy}^T & \Sigma_{yy} \end{pmatrix})$

Then $p(x|y) = \mathcal{N}(\mathbb{E}(x|y), \mathrm{Var}(x|y))$,

$\mathbb{E}(x|y) = \mu_x + \Sigma_{xy}\Sigma_{yy}^{-1}(y - \mu_y)$

$\mathrm{Var}(x|y) = \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{xy}^T$

Apply Lemma 1 for $x = \alpha_t|Y_{t-1}$ and $y = v_t|Y_{t-1}$

Then $a_{t|t} = \mathbb{E}(\alpha_t|T_{t-1}) + \mathrm{Cov}(\alpha_t, v_t)[\mathrm{Var}(v_t)]^{-1}v_t = a_t + P_t Z_t^T F_t^{-1} v$

$\mathrm{Cov}(\alpha_t, v_t) = \mathbb{E}[\alpha_t(Z_t\alpha_t + A_t o_t + (1-A_t)\varepsilon_t - Z_t a_t)^T|Y_{t-1}] = \mathbb{E}[\alpha_t(\alpha_t - a_t)^T Z_t^T|Y_{t-1}] = P_t Z_t^T$

$P_t = \mathbb{E}[\alpha_t(\alpha_t - a_t)^T|Y_{t-1}] = \mathbb{E}[(\alpha_t - a_t)(\alpha_t - a_t)^T|Y_{t-1}] + a_t\mathbb{E}[(\alpha_t - a_t)^T] = \mathrm{Var}(\alpha_t|Y_{t-1})$

$F_t = \mathrm{Var}(v_t|Y_{t-1}) = \mathrm{Var}(Z_t\alpha_t + A_t o_t + (1-A_t)\varepsilon_t - Z_t a_t|Y_{t-1})$

$F_t = \mathbb{E}(Z_t(\alpha_t - a_t)(\alpha_t - a_t)^T Z_t^T|Y_{t-1}) + \mathbb{E}[(A_t o_t + (1-A_t)\varepsilon_t)(A_t o_t + (1-A_t)\varepsilon_t)^T] = Z_t P_t Z_t^T + p_t^a H_t^o + [1-p_t^a]H_t^\varepsilon$

$a_{t|t} = a_t + P_t Z^t F_t^{-1} v_t$

$P_{t|t} = \mathrm{Var}(\alpha_t|Y_t) = \mathrm{Var}(\alpha_t|Y_{t-1}, v_t) = \mathrm{Var}(\alpha_t|Y_{t-1}) - \mathrm{Cov}(\alpha_t, v_t)[\mathrm{Var}(v_t)]^{-1}\mathrm{Cov}(\alpha_t, v_t)^T = P_t - P_t Z_t^T F_t^{-1} Z_t P_t$

$a_{t+1} = \mathbb{E}(T_t\alpha_t + R_t C_t \eta_t + R_t(1-C_t)\xi_t|Y_t) = T_t\mathbb{E}(\alpha_t|Y_t) = T_t a_{t|t} = T_t a_t + K_t v_t$

$K_t = T_t P_t Z^t F_t^{-1}$

$P_{t+1} = \mathrm{Var}(T_t\alpha_t + R_t C_t\eta_t + R_t(1-C_t)\xi_t|Y_t) = T_t\mathrm{Var}(\alpha_t|Y_t)T_t^T + R_t[p_t^c Q_t^\eta + (1-p_t^c)Q_t^\xi]R_t^T = T_t P_{t|t} T_t^T + R_t[p_t^c Q_t^\eta + (1-p_t^c)Q_t^\xi]R_t^T = T_t P_t(T_t^T - Z_t^T F_t^{-1} Z_t P_t T_t^T) + R_t[p_t^c Q_t^\eta + (1-p_t^c)Q_t^\xi]R_t^T$

$K_t Z_t = T_t P_t Z_t^T F_t^{-1} Z_t$

$(K_t Z_t)^T = Z_t^T (F_t^{-1})^T Z_t P_t^T T_t^T$

KALMAN FILTER RECURSION

$v_t = y_t - Z_t a_t$

$a_{t|t} = a_t + P_t Z^t F_t^{-1} v_t$

$a_{t+1} = T_t a_t + K_t v_t$

$F_t = Z_t P_t Z_t^T + p_t^a H_t^o + [1-p_t^a]H_t^\varepsilon$

$P_{t|t} = P_t - P_t Z_t^T F_t^{-1} Z_t P_t$

$P_{t+1} = T_t P_t(T_t^T - K_t Z_t) + R_t[p_t^c Q_t^\eta + (1-p_t^c)Q_t^\xi]R_t^T$

3 State estimation

$y_t = Z_t\alpha_t + A_t o_t + (1-A_t)\varepsilon_t$

$x_t = \alpha_t - a_t$

$P_t = \mathrm{Var}(v_t)$

$v_t = y_t - \mathbb{E}(y_t|Y_{t-1}) = y_t - Z_t a_t$

v_t's are innovations (cannot be predicted from the past)

$v_t = y_t - Z_t a_t = Z_t\alpha_t + A_t o_t + (1-A_t)\varepsilon_t - Z_t a_t = Z_t x_t + A_t o_t + (1-A_t)\varepsilon_t$

$a_{t+1} = T_t a_t + K_t v_t$ - from Kalman filter equations

$x_{t+1} = \alpha_{t+1} - a_{t+1} = T_t\alpha_t + R_t C_t\eta_t + R_t(1-C_t)\xi_t - T_t a_t - K_t v_t =$

$= T_t(\alpha_t - a_t) + R_t C_t\eta_t + R_t(1-C_t)\xi_t - K_t(Z_t x_t + A_t o_t + (1-A_t)\varepsilon_t) =$

$= T_t x_t + R_t C_t\eta_t + R_t(1-C_t)\xi_t - K_t Z_t x_t - K_t(A_t o_t + (1-A_t)\varepsilon_t) =$

$= L_t x_t + R_t C_t \eta_t + R_t(1 - C_t)\xi_t - K_t[A_t o_t + (1 - A_t)\varepsilon_t]$

$K_t = T_t P_t Z_t^T F_t^{-1}$ - from Kalman filter

$L_t = T_t - K_t Z_t$

Innovation analigue:

$v_t = Z_t x_t + A_t o_t + (1 - A_t)\varepsilon_t$

$x_{t+1} = L_t x_t + R_t \eta_t - K_t[A_t o_t + (1 - A_t)\varepsilon_t]$

Recursion for P_{t+1}:

$x_t = \alpha_t - a_t$

$P_{t+1} = \mathrm{Var}(x_{t+1}) = \mathbb{E}[(\alpha_{t+1} - a_{t+1})x_{t+1}^T] = \mathbb{E}(\alpha_{t+1} x_{t+1}^T) =$

$= \mathbb{E}[(T_t \alpha_t + R_t p_t^c \eta_t + R_t(1 - p_t^c)\xi_t)(L_t x_t + R_t p_t^c \eta_t + R_t(1 - p_t^c)\xi_t - K_t \varepsilon_t)^T] =$

$= T_t \mathbb{E}(\alpha_t x_t^T)L_t^T + p_t^c R_t \mathbb{E}(\eta_t x_t^T)L_t^T + (1 - p_t^c)R_t \mathbb{E}(\xi_t x_t^T)L_t^T + p_t^c T_t \mathbb{E}(\alpha_t \eta_t^T)R_t^T +$
$(1 - p_t^c)T_t \mathbb{E}(\alpha_t \xi_t^T)R_t^T + p_t^c R_t Q_t^\eta R_t^T + (1 - p_t^c)R_t Q_t^\xi R_t^T - T_t \mathbb{E}(\alpha_t \varepsilon_t)K_t^T =$

$= T_t \mathbb{E}(\alpha_t x_t^T)L_t^T + p_t^c R_t Q_t^\eta R_t^T + (1 - p_t^c)R_t Q_t^\xi R_t^T - T_t \mathbb{E}(\alpha_t \varepsilon_t)K_t^T =$

$= T_t P_t L_t^T + p_t^c R_t Q_t^\eta R_t^T + (1 - p_t^c)R_t Q_t^\xi R_t^T$

$\mathrm{Cov}(\eta_t, x_t) = 0$

$\mathrm{Cov}(\xi_t, x_t) = 0$

$\mathrm{Cov}(\eta_t, \alpha_t) = 0$

$\mathrm{Cov}(\xi_t, \alpha_t) = 0$

$\mathrm{Cov}(\varepsilon_t, \alpha_t) = 0$

$\mathbb{E}(\alpha_t x_t^T) = \mathbb{E}[(\alpha_t - a_t)(\alpha_t - a_t)^T] + a_t \mathbb{E}[\alpha_t - a_t] = \mathrm{Var}(\alpha_t) = P_t$

So:

$v_t = Z_t x_t + A_t o_t + (1 - A_t)\varepsilon_t$

$x_{t+1} = L_t x_t + R_t \eta_t - K_t[A_t o_t + (1 - A_t)\varepsilon_t]$

$P_{t+1} = T_t P_t L_t^T + p_t^c R_t Q_t^\eta R_t^T + (1 - p_t^c)R_t Q_t^\xi R_t^T$

4 Kalman Smoothing

$\hat{\alpha}_t = \mathbb{E}(\alpha_t | Y_n)$

$V_t = \mathrm{Var}(\alpha_t | Y_n)$

$\alpha_1 = \mathcal{N}(a_1, P_1)$ - our assumption

$v_t = y_t - Z_t a_t$

$v_{t:n} = (v_t^T, \ldots, v_n^T)^T$

If Y_{t-1} and $v_{t:n}$ are fixed, then Y_n is fixed

Apply Lemma 1 to $x = \alpha_t | Y_{t-1}$ and $y = v_{t:n} | Y_{t-1}$

$v_t, \dots v_n$ are independent of Y_{t-1} and of each other with zero means

$x_t = \alpha_t - a_t$

$\hat{\alpha}_t = \mathbb{E}(\alpha_t | Y_n) = \mathbb{E}(\alpha_t | Y_{t-1}, v_{t:n}) = a_t + \sum_{j=t}^{n} \mathrm{Cov}(\alpha_t, v_j) F_j^{-1} v_j$

$F_j = \mathrm{Var}(v_j | Y_{t-1})$

$v_t = Z_t x_t + A_t o_t + (1 - A_t)\varepsilon_t$

$\mathrm{Cov}(\alpha_t, v_j) = \mathbb{E}(\alpha_t v_j^T | Y_{t-1}) = \mathbb{E}(\alpha_t (Z_j x_j + A_j o_j + (1 - A_j)\varepsilon_j)^T | Y_{t-1}) = \mathbb{E}(\alpha_j x_j^T | T_{t-1}) Z_j^T$

5 Kalman Smoothing Recurrent Equations

$r_n = 0$

$N_n = 0$

$r_{t-1} = Z_t^T F_t^{-1} v_t + L_t^T r_t$

$\hat{\alpha}_t = a_t + P_t r_{t-1}$

$N_{t-1} = Z_t^T F_t^{-1} Z_t + L_t^T N_t L_t$

$V_t = P_t - P_t N_{t-1} P_t$

where

v_t, F_t, K_t, a_t, P_t - from Kalman filter

$v_t = y_t - Z_t a_t$ - from Kalman filter

$L_t = T_t - K_t Z_t$