Sorry arima, I'm going Bayesian

Pierre Gauthier

École des Mines de Nancy

May 2019

Tuteur : Denis Villemonais

Sommaire

Approche bayésinne pour la modélisation des séries temporelles

- Les modèle espace-états

		bruit blanc gaussien
equation d'observation	$y_t = Z_t^T \alpha_t + \epsilon_t$	$\epsilon_t \sim N(0, H_t)$
equation de transition	$\alpha_{t+1} = T_t \alpha_t + R_t \eta_t$	$\eta_t \sim N(0, Q_t)$

- y_t obervations
- α_t variables d'états / latentes / cachées
- Z_t matrice de mesure
- T_t matrice de transitions

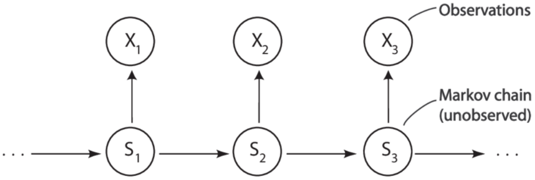

Figure – [researchgate.net]

Approche bayésinne pour la modélisation des séries temporelles

→ Bayesian structural time series (BSTS)

		bruit blanc gaussien
observation	$y_t = \mu_t + \beta^T x_t + \tau_t + \varepsilon_t$	$\varepsilon_t \sim N\left(0, \sigma_\varepsilon^2\right)$
regression	$\beta^T x_t$	
tendance + marche aléatoire	$\mu_t = \mu_{t-1} + \delta_{t-1} + u_t$	$u_t \sim N\left(0, \sigma_u^2\right)$
marche aléatoire	$\delta_t = \delta_{t-1} + v_t$	$v_t \sim N\left(0, \sigma_v^2\right)$
sainsonnalité	$\tau_t = -\sum_{s=1}^{s-1} \tau_{t-s} + w_t$	$w_t \sim N\left(0, \sigma_w^2\right)$

→ Bayesian structural time series (BSTS)

observation	$y_t = Z_t^T \alpha_t + \epsilon_t$	$\epsilon_t \sim N(0, H_t)$
	Z_t^T $\begin{pmatrix} 1 & 0 & \beta^T \mathbf{x}_t \end{pmatrix}$	α_t^T $\begin{pmatrix} \mu_t & \delta_t & 1 \end{pmatrix}^T$
equation de transition	$\alpha_{t+1} = T_t \alpha_t + R_t \eta_t$	$\eta_t \sim N(0, Q_t)$
α_t $\begin{pmatrix} \mu_t \\ \delta_t \\ 1 \end{pmatrix}$	T_t $\begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$	$N_t \eta_t$ $\begin{pmatrix} u_t \\ v_t \\ w_t \end{pmatrix}$

→ estimation des paramètres

Loi à postériori états cachés α_t : Le filtre de Kalman

Itérations sur l'estimation $p(\alpha_t | y_{1:t}) \sim \mathcal{N}(\hat{\alpha}_t, P_t)$

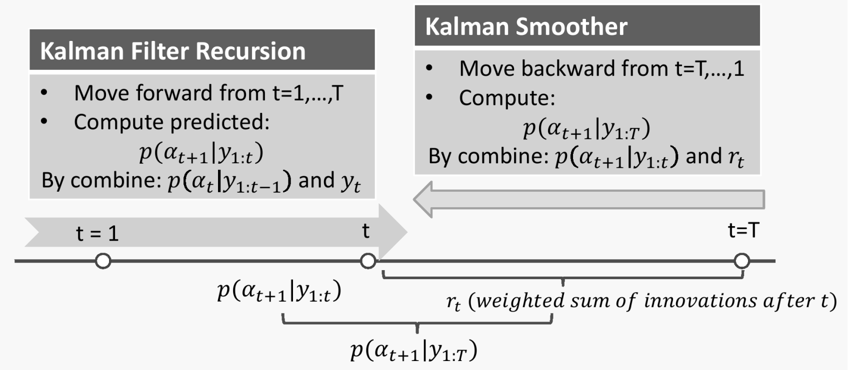

Figure – [github : anhdanggit/nowcasting-google-queries/]

Loi à postériori de β : *spike-and-slab prior*

- On prend la partie regression $y_t^* = y_t - \mu_t$
- On utilise pour β une distribution à priori *spike-and-slab* :
  - $p(\gamma) = \prod_{k=1}^{N} \pi^{\gamma_k}(1-\pi)^{1-\gamma_k}, \; \gamma_k \in \{0,1\} \;\; N = Card(\mathbf{x})$
  - Á priori : $p\left(\beta, \gamma, \sigma_\varepsilon^2\right) = p\left(\beta_\gamma | \gamma, \sigma_\varepsilon^2\right) p\left(\sigma_\varepsilon^2 | \gamma\right) p(\gamma)$
  - $\beta_\gamma \left| \sigma_\epsilon^2, \gamma \sim \mathcal{N}\left(b_\gamma, \sigma_\epsilon^2 \left(\Omega_\gamma^{-1}\right)^{-1}\right) \quad \sigma_\epsilon^2 \right| \gamma \sim IG\left(\frac{\nu}{2}, \frac{ss}{2}\right)$

  papramètres à priori : v nombre de paramètres, $\frac{ss}{v} = \left(1 - R^2\right) s_y^2$, $\Omega^{-1} \propto X^T X$

- On utilise les propriété des lois conjugé pour obtenir les loi à postériori

  $\beta_\gamma | \sigma_\epsilon, \gamma, \mathbf{y}^* \qquad \gamma_\epsilon^2 | \gamma, \mathbf{y}^* \qquad \gamma | \mathbf{y}^*$
- Intérêt

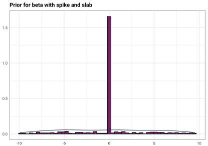

Figure – [batisengul.co.uk]

Échantilloneur de Gibbs pour BSTS : SSVS algorithm

$\Theta = \left(\gamma, \beta, \sigma_\varepsilon^2, \sigma_v^2, \sigma_u^2\right)$

- Choisir paramètres à priori v, R^2, s_y^2

- Tirer $\gamma, \beta, \sigma_\varepsilon^2, \sigma_v^2, \sigma_u^2$ σ_u^2,σ_u^2,σ_w^2 sont tiré selon la loi $.|\gamma \sim IG\left(\frac{\nu}{2}, \frac{ss}{2}\right)$ Sur $1, \ldots$:

1. Avec le filtre de Kalman on simule les état latents α depuis $p\left(\alpha | y, \gamma, \beta, \sigma_\varepsilon^2, \sigma_v^2, \sigma_u^2\right)$
2. On simule σ_u^2 et σ_v^2 avec la distribution postérieure $p\left(\frac{1}{\sigma_u^2}, \frac{1}{\sigma_v^2} | y, \alpha, \beta, \sigma_\varepsilon^2\right)$
3. On simule β et σ_ϵ^2 avec la distribution postérieure $p\left(\beta, \sigma_\varepsilon^2 | y, \alpha, \sigma_u^2, \sigma_v^2\right)$
4. Retour à la première étape.

σ_u^2,σ_u^2,σ_w^2 sont tiré selon la loi $.|\gamma \sim IG\left(\frac{\nu}{2}, \frac{ss}{2}\right)$
Le modèle final est la moyenne des modèles $\left(\gamma, \beta, \sigma_\varepsilon^2, \sigma_v^2, \sigma_u^2\right)_t$ ainsi tirés.

Conclusion

- Auteur prèfère mettre incertitude dans la prior que sur l'estimation des coéfficientss

Merci de votre attention