Sequential Learning in Univariate DCMM

0. Set m_0, C_0, G_t (see example 4.2), W_t (see supplementary material).

Bernoulli logistic DGLM

1. At any time $t-1$, current information is summarized as $(\theta_{t-1}|D_{t-1}, I_{t-1}) \sim (m_{t-1}, C_{t-1})$
2. Evolutionary equation induces 1-step ahead prior: $(\theta_t|D_{t-1}, I_{t-1}) \sim (a_t, R_t), a_t = G_t m_{t-1}, R_t = G_t C_{t-1} G_t^T + W_t$
3. Conjugate prior: $\pi_t \sim Beta(\alpha_t, \beta_t)$ (no action)
4. Solve (α_t, β_t) from $F_t' a_t =: f_t = \psi(\alpha_t) - \psi(\beta_t), F_t' R_t F_t =: q_t = \psi'(\alpha_t) + \psi'(\beta_t)$,

   where ψ and ψ' are digamma and trigamma function
5. 1-step ahead forecast: $\mathrm{Pr}[z_t = 1|\mathcal{D}_{t-1}, \mathcal{I}_{t-1}] = \alpha_t / (\alpha_t + \beta_t)$. Here we make a prediction with $I(Ez_t > 0.5)$.
6. On observing z_t, we have the posterior. (no action)
7. Calculate $g_t = \psi(\alpha_t + z_t) - \psi(\beta_t + 1 - z_t), p_t = \psi'(\alpha_t + z_t) + \psi'(\beta_t + 1 - z_t)$ (z_t is observed!)
8. Posterior update $(\boldsymbol{\theta}_t|\mathcal{D}_t) \sim (\mathbf{m}_t, \mathbf{C}_t)$:
   $\mathbf{m}_t = \mathbf{a}_t + \mathbf{R}_t \mathbf{F}_t (g_t - f_t) / q_t \quad \text{and}$
   $\mathbf{C}_t = \mathbf{R}_t - \mathbf{R}_t \mathbf{F}_t \mathbf{F}_t' \mathbf{R}_t' (1 - p_t / q_t) / q_t$

(similar to Kalman filtering)

Poisson log DGLM

1. At any time $t-1$, current informsation is summarized as $(\theta_{t-1}|D_{t-1}, I_{t-1}) \sim (m_{t-1}, C_{t-1})$
2. Evolutionary equation induces 1-step ahead prior: $(\theta_t|D_{t-1}, I_{t-1}) \sim (a_t, R_t), a_t = G_t m_{t-1}, R_t = G_t C_{t-1} G_t^T + W_t$
3. Conjugate prior: $\mu_t \sim Gamma(\alpha_t, \beta_t)$ (no action)
4. Solve (α_t, β_t) from $F_t' a_t =: f_t = \psi(\alpha_t) - \log(\beta_t), F_t' R_t F_t =: q_t = \psi'(\alpha_t)$,

   where ψ and ψ' are digamma and trigamma function

   (use scipy.special.polygamma(1,x) to calculate trigamma in python)
5. 1-step ahead forecast: $(y_t|D_{t-1}, I_{t-1}) \sim NB(\alpha_t, \frac{1}{1+\beta_t})$. (The notation here is same as Wikipedia for Nb, which is different from appendix of BerryWest2019)

   We can simply use the fact that $(y_t|z_t = 1) \sim Poi(\mu_t) + 1$, so our 1-step ahead forecast is simply $E(y_t|z_t = 1, D_{t-1}, I_{t-1}) = E(\mu_t) + 1 = \alpha_t / \beta_t + 1$ by iterated expectation.

6. On observing z_t, we have the posterior. (no action)
7. Calculate $g_t = \psi(\alpha_t + z_t) - \psi(\beta_t + 1 - z_t), p_t = \psi'(\alpha_t + z_t) + \psi'(\beta_t + 1 - z_t)$ (z_t is observed!)
8. Posterior update $(\boldsymbol{\theta}_t | \mathcal{D}_t) \sim (\mathbf{m}_t, \mathbf{C}_t)$:
   $\mathbf{m}_t = \mathbf{a}_t + \mathbf{R}_t \mathbf{F}_t (g_t - f_t) / q_t \quad \text{and}$
   $\mathbf{C}_t = \mathbf{R}_t - \mathbf{R}_t \mathbf{F}_t \mathbf{F}_t' \mathbf{R}_t' (1 - p_t / q_t) / q_t$

Filtering and Forecasting

In DGLM, The positive count model component will be updated only when $z_t = 1$. When $z_t = 0$, the positive count value is treated as missing. I.e., update both Poisson and Bernoulli when $z_t = 1$, and update only Bernoulli when $z_t = 0$.

Forward filtering: Bernoulli and Poisson are updated separately.

Forecasting:

$t+k$: Implied mixture of Bernoulli and shifted Poisson (step 5 above, Appendix 3)

At time *t*, the *k*-step ahead forecast distribution has a pdf of the compositional form

$p(y_{t+k} | \mathcal{D}_t, \mathcal{I}_t, \pi_{t+k}) = (1 - \pi_{t+k}) \delta_0 (y_{t+k}) + \pi_{t+k} h_{t,t+k} (y_{t+k})$

where

- $(\pi_{t+k} \mid \mathcal{D}_t, \mathcal{I}_t) \sim Be(\alpha_t^0(k), \beta_t^0(k))$ and $\delta_0(y)$ is the Dirac delta function at zero.
- $h_{t,t+k}(y_{t+k})$ is the density of $y_{t+k} = 1 + x_{t+k}$ where x_{t+k} is negative binomial,

$$(x_{t+k} \mid \mathcal{D}_t, \mathcal{I}_t) \sim \mathrm{Nb}\left(\alpha_t^+(k), \frac{\beta_0^+(k)}{1 + \beta_t^+(k)}\right).$$

- The defining parameters $\alpha_t^0(k)$, $\beta_t^0(k)$, $\alpha_t^+(k)$, $\beta_t^+(k)$ are computed from the binary and positive count DGLMs, respectively.

That is, the mixture places probability $1 - \pi_{t+k}$ on $y_{t+k} = 0$, and probability π_{t+k} on the implied shifted negative binomial distribution.

Full joint forecast of $y_{t+1:t+k}$: Monte Carlo samples from $p(y_{t+1:t+k} | D_t, I_t)$ (Appendix 3)

Next: Multivariate case

Try: (Maybe we could include a "future possibilities" part in the presentation on Nov 19)

1. Normal DLM for simulated continuous data

2. performance of this model when it is not the true model (via simulation). e.g., when the true data generating process is just poisson
3. performance of this model when the counts are not sparse or counts are large, and compare it to normal DGLM or something else (maybe even just a Poisson). As stated in the paper, when counts are large, the series "can often be well-modelled using a normal DLM as an approximation", and, "for a time series with few or no zeros, the binary model will play a relatively limited role".
4. W_t: no need to specify it carefully according to supplementary material. Try different specifications and look at their influence on the outcome (maybe in the univariate case)
5. different loss functions