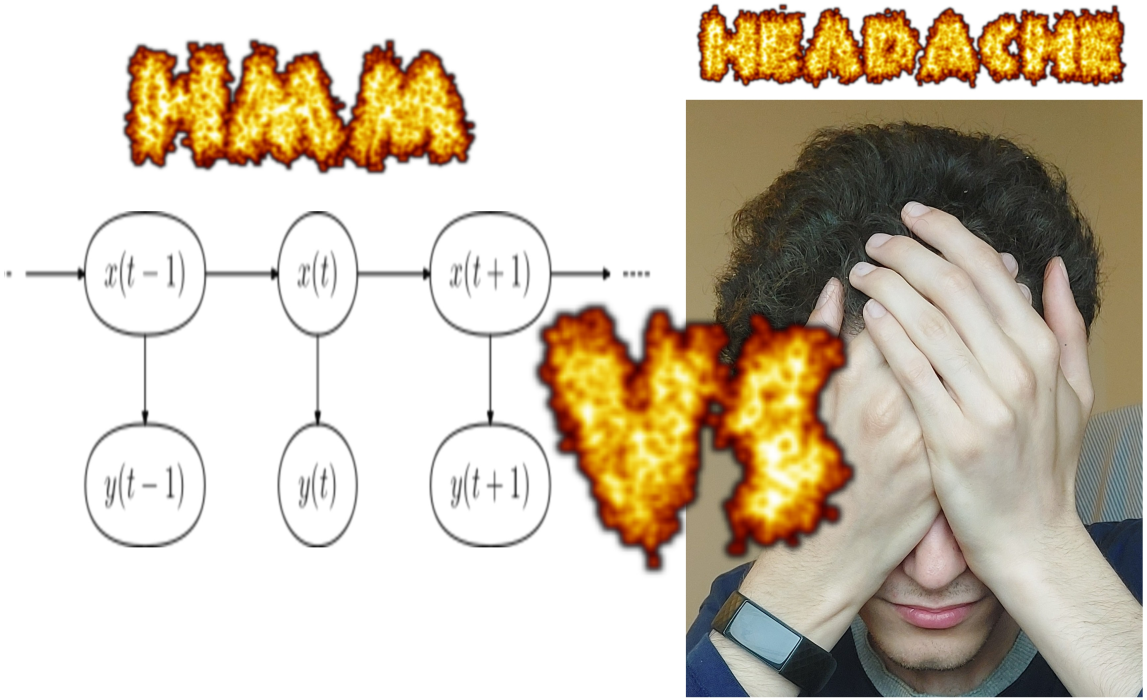
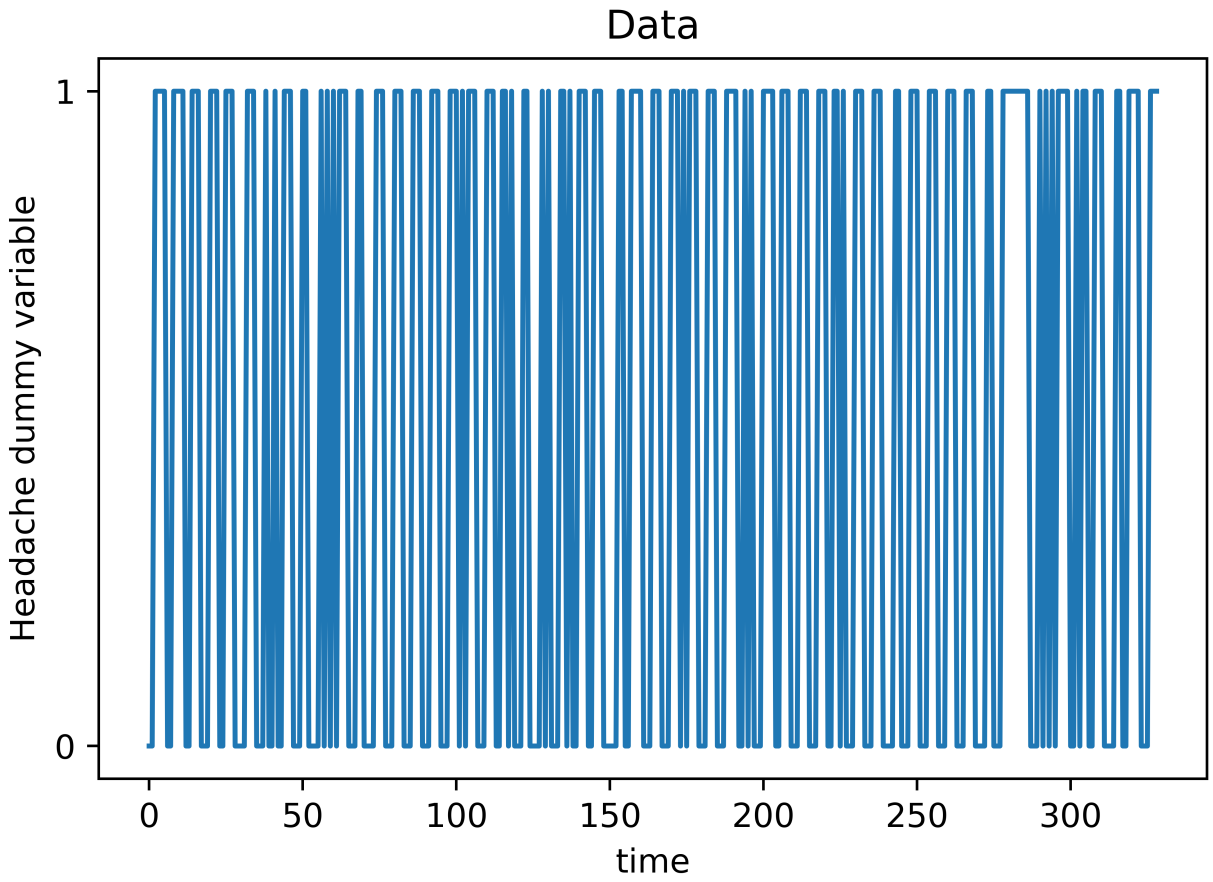


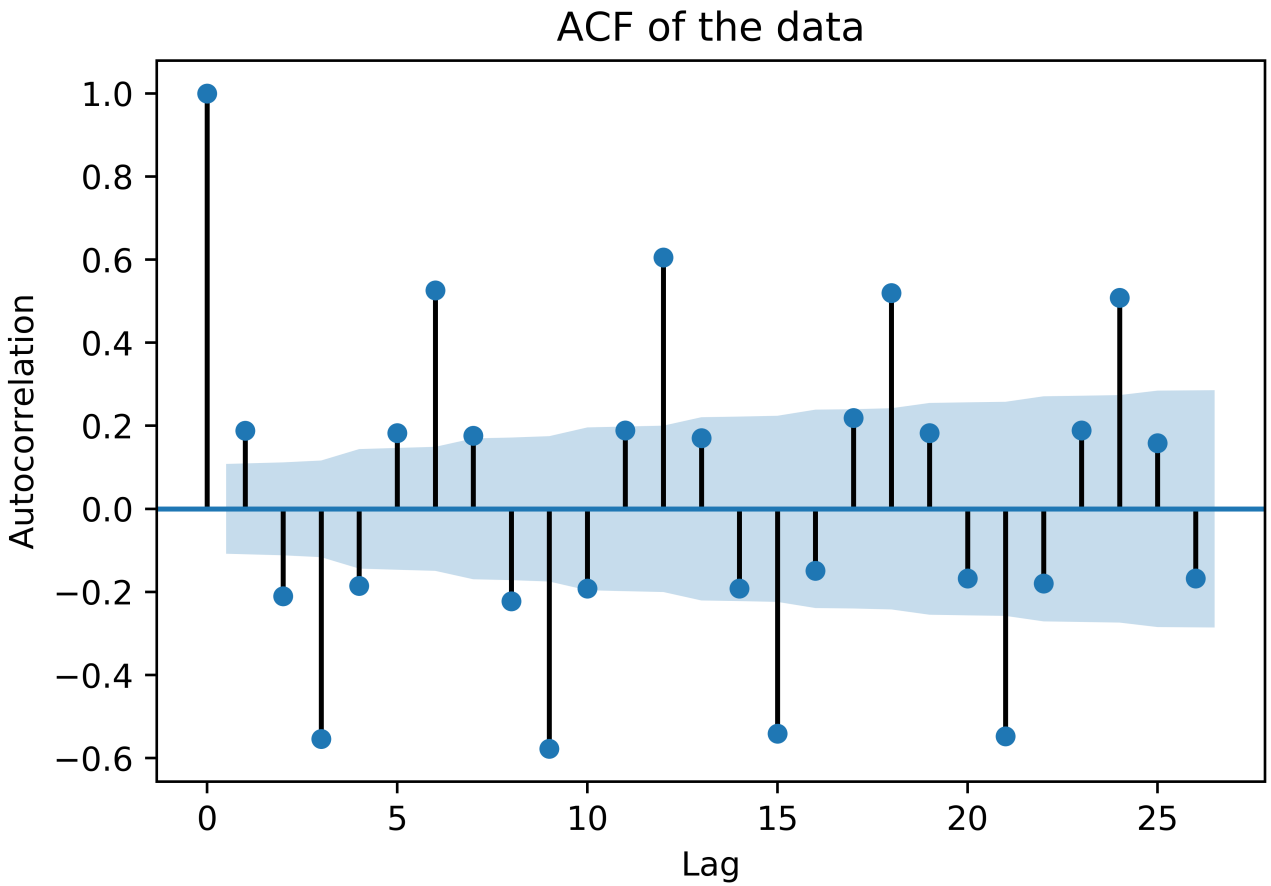
Valeria Insogna
Guglielmo Padula
Università degli Studi di Trieste

Short description of the dataset and the problem

The dataset contains a binary sequence of size 296. Each element has value 1 if in day i I had some headache and 0 if I had not. We want to make predictions on the future.



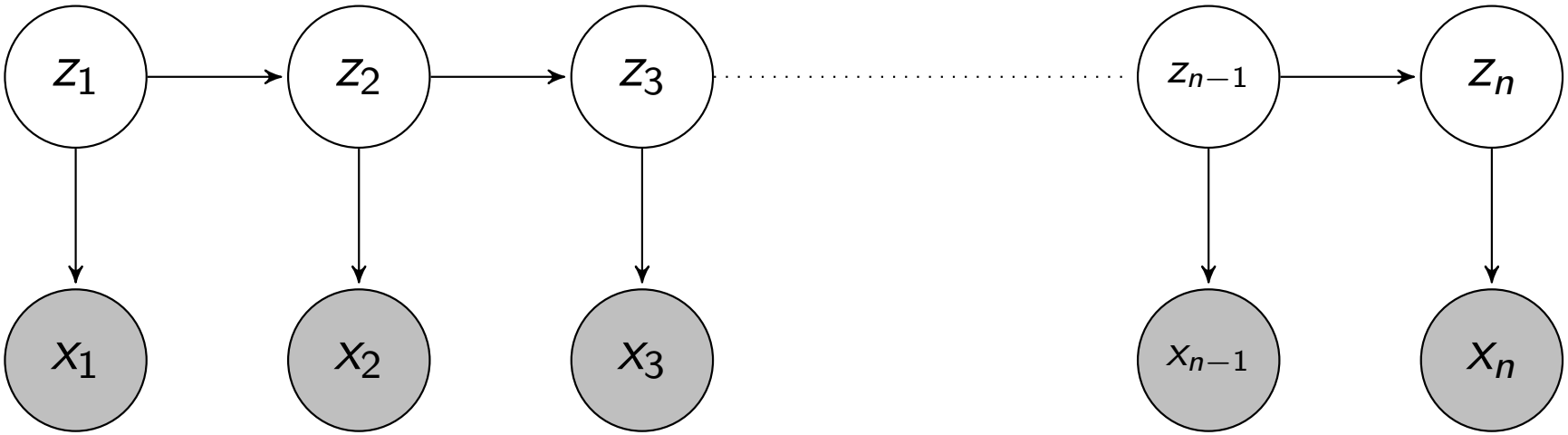
Autocorrelation



First three lags are statistically significant, so it makes sense to model the data as a time series. We decide to use short memory HMM models.

General model structure

Z_t Markov process (maybe with memory) parametrized by θ
$X_t|Z_t = z_t \sim Ber(g(z_t))$ where $g : A \subseteq \mathbb{R} \to [0, 1]$ eventually parametrized by k
Z_t and g characterize the model.



Inference

For parameter inference we will use a Black Box Expectation Maximization algorithm (implemented in the Pyro library), which is similar to the classical EM algorithm with the difference that in the E step the posterior distribution is approximated with a parameterized family of distribution using Black Box inference.

Performance measures

In cases for which internal model selection will be needed we will use the BIC and the autocorrelation distance in the training data. For choosing the best average model we will use the BIC in the test data and the following loss:

$$\sum_{i=1}^{25} \frac{1}{25}(\rho_i^D - \rho_i^M)^2$$

where ρ_i^D is the autocorrelation of the data at lag i and ρ_i^M is the estimated autocorrelation at lag i of the model. This loss will be called autocorrelation distance. We use the BIC over other likehood based performance indicators because we prefer simpler models.

Model 1: structure

Wiener Process

Z_t is a Wiener process if it is a onedimensional Gaussian process with:

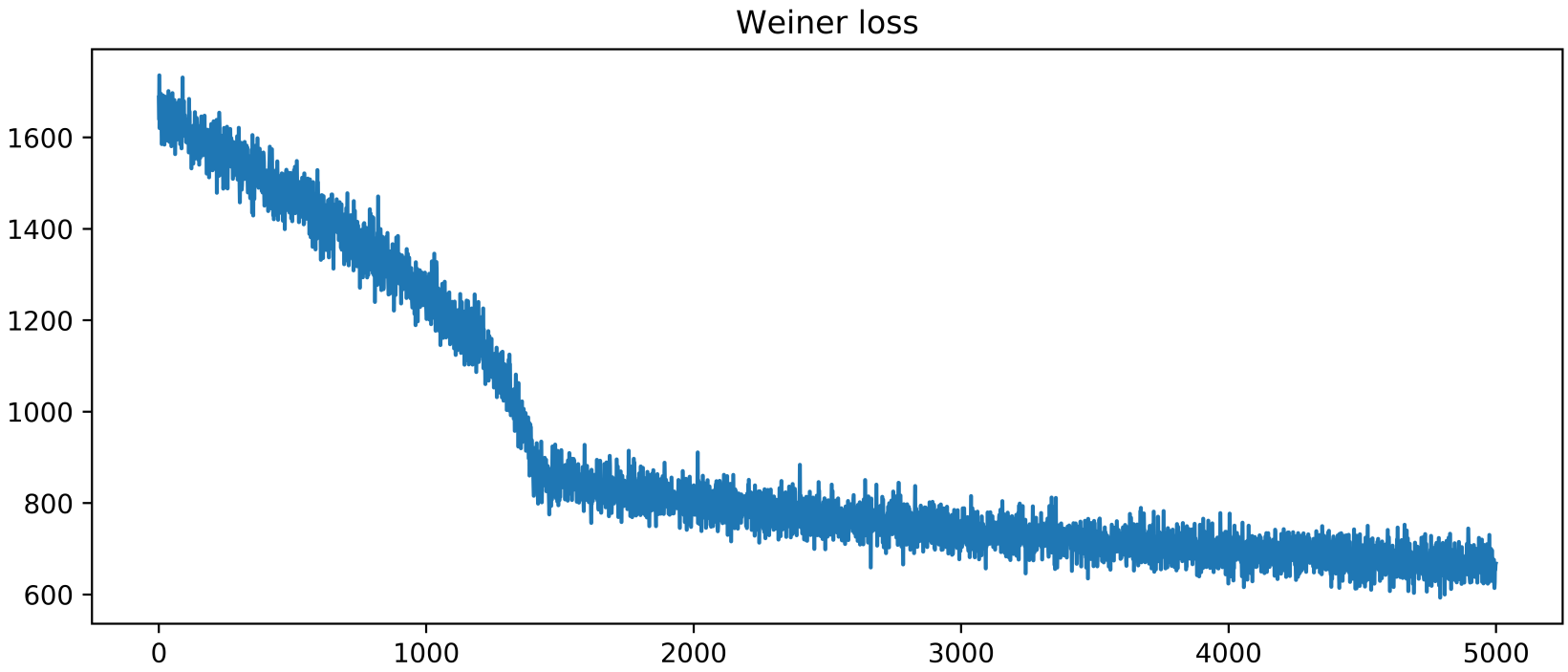
- $\mu(x) = 0 \quad \forall x \in \mathbb{R}$
- $K(x, y) = \min\{x, y\} \quad \forall x, y \in \mathbb{R}^+$

Our first model consists of:

- Z_t Wiener process
- $g(z) = \frac{1}{1+e^{-kz}}$

Model 1: estimate and diagnostic

For SVI we chose to approximate the posterior with a Normal guide (Autonormal pyro function). The optimal k is 2.69. The BIC is 1277 and the autocorrelation distance is 0.12.



Stagional Arima

Lag operator

The lag operator L is defined as:

$$L^s X_t = X_{t-s}$$

White noise

The white noise process is a stochastic process defined as:

$$\epsilon_t \sim N(0, \sigma)$$

Stagional ARIMA

Z_t is a $ARIMA((P, D, Q) \times (p, d, q)^s)$ with degree s if:

- $(1 - \sum_{i=1}^{P+sp} \alpha_i L^i)(1 - L)^{D+sd} X_t = (1 + \sum_{i=1}^{Q+sq} \theta_i L^i)\epsilon_t$

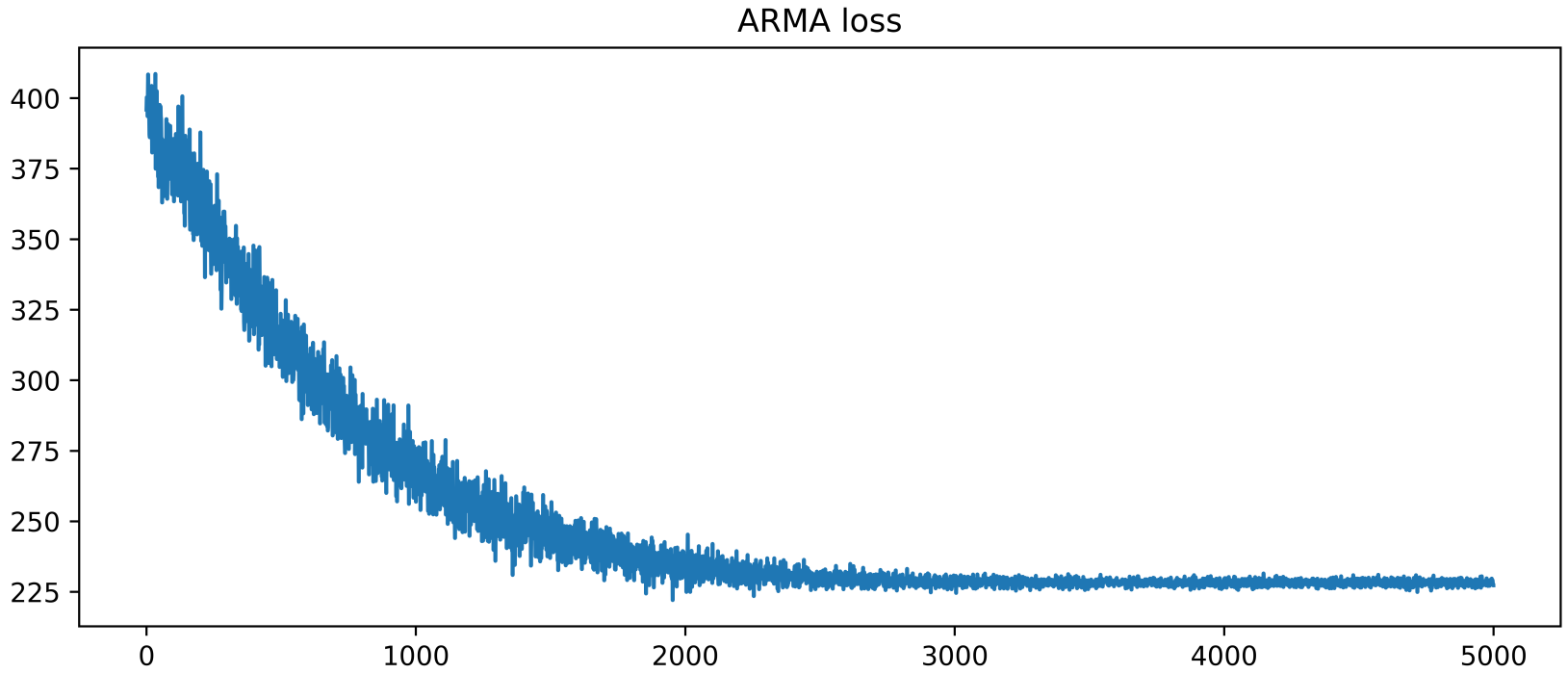
Model 2:structure

If the acf would be of a classical would series, that would be the graph of an $sARIMA(0,0,0) \times (0,1,0)^3$, so we decide to an HMM with memory:

- $Z_t \sim sARIMA(0,0,0) \times (0,1,0)^3$
- $g(z) = \frac{1}{1+e^{-kz}}$

Model 2:estimate and diagnostic

For SVI we chose to approximate the posterior with a Normal guide (Autonormal pyro function).
Optimal k is 2.92. The Bic is 1023 and the autocorr distance is 0.08.



Model 3: structure

We will now try a fully bayesian normal mixture model:

$\lambda \sim Beta(\alpha_1, \beta_1)$

$\mu_i \sim Norm(\mu_\mu, \sigma_\mu), i = 1...K$

$\sigma_i \sim InverseGamma(\alpha_2, \beta_2), i = 1...K$

$h \sim Categorical(\phi)$

$Z_t | Z_{t-1} \sim Normal(\lambda * z_{t-1} + (1 - \lambda)\mu_h, \sigma_h)$

$g(z) = g(z) = \frac{1}{1+e^{-kz}}$

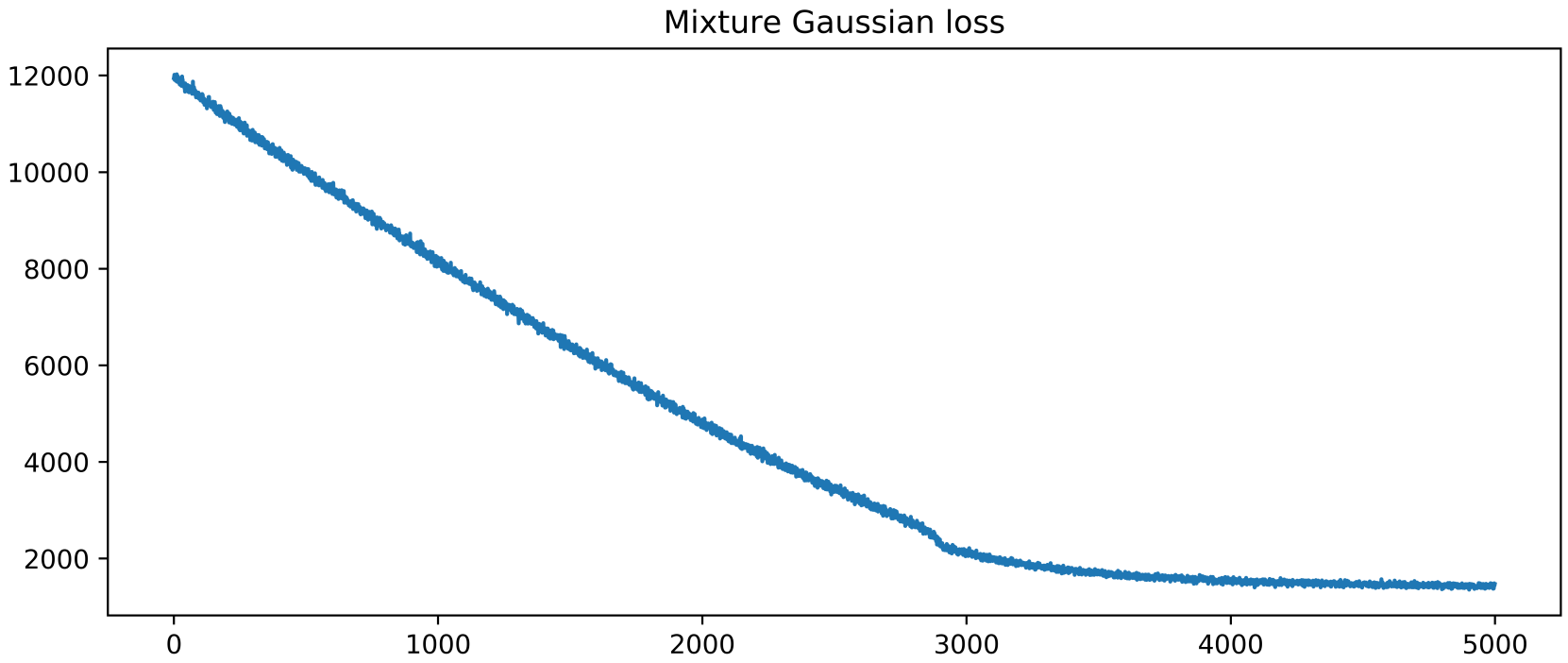
Model 3: estimate and diagnostic

For SVI we chose to approximate the posterior with a Normal guide (Autonormal pyro function). Estimates:

- $\alpha_1 = 1.7$
- $\beta_1 = 1.9$
- $k = 0.07$
- $\mu_\mu = 0.43$
- $\mu_\sigma = 1.8$
- $\alpha_2 = 0.9$
- $\beta_2 = 0.86$
- $\phi = [0.2, 0.2, 0.1, 0.1, 0.4]$

BIC on the test data is 662 and autocorrelation distance is 0.12.

Model 3: estimate and diagnostic



Model 4: structure

We also decided to try some discrete models.

- $Z_t \sim DTMC(A, x_0)$
- $X_t \sim Bernoulli(BZ_t)$

where A is an NxN stochastic matrix and B is a $Nx2$ stochastic matrix. N must be determined by model selection.

Model 4 : model selection

N	autocorrelation distance	BIC
1	0.025	856
2	0.027	217
3	0.027	232
4	0.027	311
5	0.027	308
6	0.008	287
7	0.03	440
8	0.026	530
9	0.026	615
10	0.27	720

Best autocorrelation distance is achieved with $N = 6$.

Model 4:estimate and diagnostic

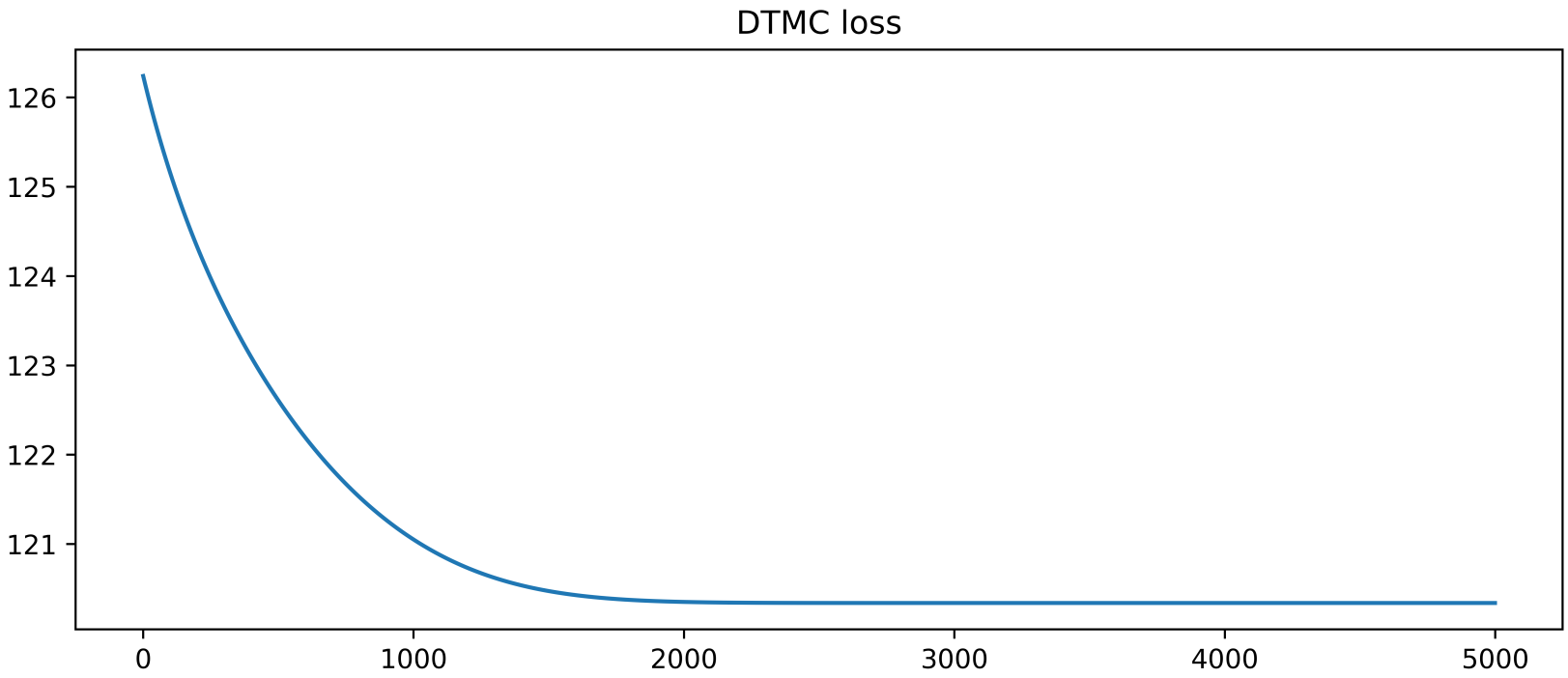
For SVI we chose to approximate the posterior with a Discrete guide (Autodiscreteparallel pyro function).

We have $A \approx \begin{pmatrix} 0.01 & 0.9 & 0.02 & 0.03 & 0.01 & 0.04 \\ 0 & 0 & 0.99 & 0.01 & 0 & 0 \\ 0 & 0.02 & 0 & 0.98 & 0 & 0 \\ 0.03 & 0 & 0.02 & 0 & 0.95 & 0 \\ 0 & 0.02 & 0.01 & 0 & 0.02 & 0.95 \\ 0.95 & 0.03 & 0.02 & 0 & 0 & 0 \end{pmatrix}$

$B \approx \begin{pmatrix} 0.95 & 0.05 \\ 0.88 & 0.12 \\ 0.87 & 0.13 \\ 0.12 & 0.88 \\ 0.13 & 0.87 \\ 0.12 & 0.88 \end{pmatrix}$ $x0 \approx \begin{bmatrix} 0.95 \\ 0.02 \\ 0.03 \\ 0 \\ 0 \\ 0 \end{bmatrix}$

BIC on the test data is 288 and is 0.006.

Model 4: estimate and diagnostic

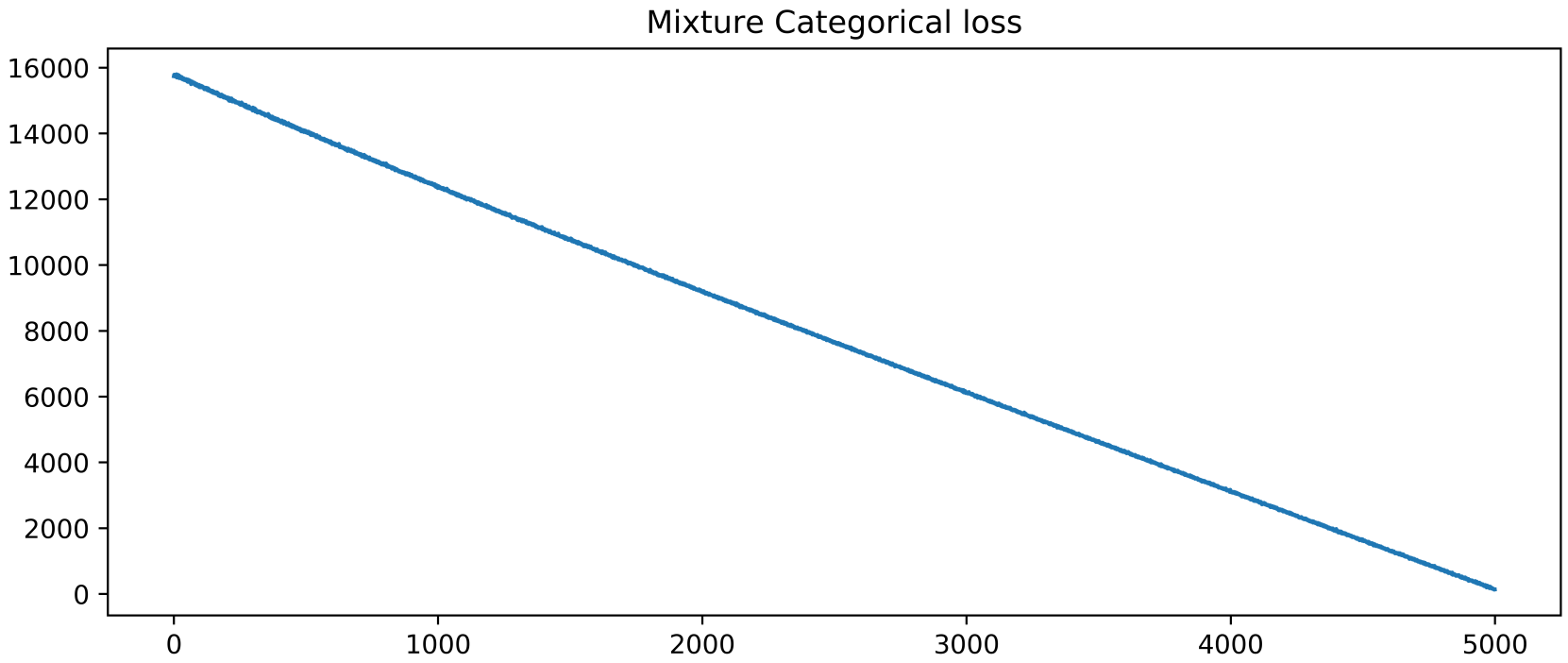


Model 5: structure

To conclude we try a mixture of Categorical variables:

- $\phi \sim Dirichlet(\beta_1)$
- $\theta \sim Dirichlet(\alpha_1)$
- $\lambda \sim Beta(\alpha_2, \beta_2)$
- $h \sim Categorical(\phi)$
- $Y_t \sim Categorical(\theta_h)$
- $Z_t | Z_{t-1} \sim \lambda * Y_t + (1 - \lambda) * z_{t-1}$
- $g(z) = \frac{1}{1+e^{-kz}}$

Model 5: estimate and diagnostic



Model 5: estimate and diagnostic

For SVI we chose to approximate the posterior with its laplace approximation (AutoLaplace pyro function).

- $\beta_1 = [0.8, 1.2, 0.9, 0.88, 0.96, 0.98]$
- $\alpha_1 = [0.8, 0.9, 0.9, 0.8, 0.9, 0.7]$
- $\alpha_2 = 0.8$
- $\beta_2 = 0.7$
- $k = 1.9$

BIC is 252 and autocorrelation distance is 0.04.

Summary

Model	BIC	autocorrelation distance
1	1277	0.120
2	1023	0.080
3	662	0.120
4	288	0.006
5	222	0.040

The best model is the 4th as it has a low BIC and has the lowest autocorrelation distance. Thanks to the simplicity of the DTMC model we can give an interpretation by looking at the matrices A and B: most of the time there is a cycle of three days of high probability of getting the headache followed by other three days of low probability.