Numerical Methods, Spring 2023
Assignment 9 [State Space Models - I]
Total: 85, Deadline: 6 May

I. SUGGESTED READING

- [1], Chapter 18 State space models

II. EXERCISES

1. (10) Assume you are flipping a biased coin and the outcome of 100 flips is 75 Heads and 25 Tails. Compute the MLE for the Head probability p_0 and its error δp (the error is defined such that the probability of p to lie in the interval $[p_0 - \delta p, p_0 + \delta p]$ is 95%).

2. (25) Multivariate normal distribution over vectors of length k is defined as follows:

$$\mathcal{N}(\mathbf{x}|\mathbf{\Sigma}) = (2\pi)^{-k/2} \det(\mathbf{\Sigma})^{-1/2} \exp\left(-\frac{1}{2}\mathbf{x}^T\mathbf{\Sigma}^{-1}\mathbf{x}\right). \tag{1}$$

   Your goal is to compute the following integral:

$$I(\mathbf{x}, \mathbf{\Sigma_1}, \mathbf{\Sigma_2}) = \int d^k\mathbf{y}\mathcal{N}(\mathbf{x}-\mathbf{y}|\mathbf{\Sigma}_1)\mathcal{N}(\mathbf{y}|\mathbf{\Sigma}_2) \tag{2}$$

   in two ways. Before proceeding to the multivariate case, you may consider $k = 1$ for simplicity.

   - (5) The integral $I(\mathbf{x}, \mathbf{\Sigma_1}, \mathbf{\Sigma_2})$ has a straightforward probabilistic interpretation of a convolution of two probability distribution functions. Having this in mind, the result of this integration can be written down immediately (consult Eq. (1)).
   - (20) The integral $I(\mathbf{x}, \mathbf{\Sigma_1}, \mathbf{\Sigma_2})$ can be reduced to the standard integral:

$$\int e^{-\frac{1}{2}x^\mathsf{T}\mathbf{A}x + B^\mathsf{T}x} d^kx = \sqrt{\frac{(2\pi)^k}{\det A}} e^{\frac{1}{2}B^\mathsf{T}\mathbf{A}^{-1}B}. \tag{3}$$

     Use this fact to re-derive the result of the previous item.

3. (35) Consider the model, described on the lecture:

$$x_t = x_{t-1} + u_{t-1} + \epsilon_{t-1}, \quad z_t = x_t + \eta_t \tag{4}$$

   where z_t is observed state, x_t is hidden state and, for simplicity, $u_t \equiv u$. Assume that the measurement and the control noise have the following distributions:

$$P(\epsilon_t) = \mathcal{N}(0, \sigma), \quad P(\eta_t) = \mathcal{N}(0, \rho = 2\sigma). \tag{5}$$

   Compute, analytically, the following distributions:

   (a) (5) $P(x_0|z_0)$: distribution of x_0 after the first 1 observation

   (b) (10) $P(x_0|z_0, z_1)$: distribution of x_0 after the first 2 observations

   (c) (20) $P(x_1|z_0, z_2)$: distribution of x_1 after the first 3 observations, when the second observation is missing.

   Note that each of these distributions is Gaussian, so its enough to specify mean and variance in each case. Feel free to use Eq. (3) (`Mathematica` or `sympy` may also be useful for algebraic manipulations.)

4. (15) Consider the model, described by the following state and observation equation

$$x_t = x_{t-1} + u_{t-1} + \epsilon_{t-1}, \quad z_t = x_t + \eta_t \tag{6}$$

with $P(\epsilon_t) = \mathcal{N}(0, \sigma)$, $P(\eta_t) = \mathcal{N}(0, \rho)$. As was shown on the lecture, the max-likelihood estimate of the hidden state conditioned on all information observed by the moment t reads:

$$\hat{x}_t = K_t z_t + (1 - K_t)(\hat{x}_{t-1} + u_{t-1}). \tag{7}$$

Compute $\lim_{t\to\infty} K_t$ and $\lim_{t\to\infty} \left\langle (\hat{x}_t - x_t)^2 \right\rangle$.

5. (50*) Consider the filtering and smoothing problems for the model in Eqs (6) and the set of observations z_t (the model is specified by $\sigma = 1, \rho = 2$)

- (10) Compute the filtered trajectory for x_t.
- (25) Compute the most likely full trajectory of the hidden variable x_t. In order to do so, consider the likelihood $P(x_0, x_1 ..., x_T | z_0, ..., z_T)$ and maximize it over $\mathbf{x}$. This maximization reduces to solving a system of linear equations. Solve this system numerically and plot the resulting $\hat{\mathbf{x}}$ as a function of time. Compare computation complexity (as function of T) of this approach to smoothing and the filtering which you did in the previous step.
- (15) Compute the error bar around each point and plot the 2σ band around the max-likelihood estimation (it should be a function of time, too!). The error bar can be computed as follows:

$$\mathrm{err}_t^2 = \int (x_t - \hat{x}_t)^2 P(x_0, x_1 ..., x_T | z_0, ..., z_T) \Pi_{i=0}^T dx_i \tag{8}$$

REFERENCES

[1] Kevin P Murphy. *Machine learning: a probabilistic perspective.* MIT press, 2012.