Sepsis in the ICU: Using AR Models to Help Save Lives

Anita Mehrotra*, advised by **Dr. Shamim Nemati**

CS 281, School of Engineering & Applied Sciences, Harvard University, Cambridge MA, USA.

*anitamehrotra@fas.harvard.edu

Introduction

Every year, **sepsis**, a medical condition caused by an immense immune response to infection, strikes over 750,000 Americans. 28% to 50% of these patients die -- more than the number of U.S. deaths from prostate cancer, breast cancer and AIDS *combined*. Over a short period of time, sepsis can impair blood flow, damage organs, and cause sepsis shock. It is the number one cause of patient death in the ICU.

A sepsis patient's physical state is patently observed by physicians and nurses while in the ICU. However, current literature suggests that additional information, like vital signs and hospital intervention and treatments, can be used to dynamically model and better understand patient state behavior over time.

- "Vital signs" consist of continuous blood pressure (BP) and heart rate (HR).
- "Hospital intervention" or "treatments" are general terms used to encompass a range of intervention treatments taken by hospital staff. They include administration of antibiotics and drugs like vasopressors or corticosteroids, provision of additional oxygen, intravenous fluids and mechanical ventilation, and dialysis.

In this project, I learn and apply an autoregressive (AR) model to bedside monitor data for sepsis patients. In the first portion, I learn the theory behind SSM, SAR and SKF, then, following [2], generate a random matrix coefficient for the exogenous input vector. In the second portion, I build an AR model in Python and use EM to infer the matrix coefficients on the vital sign and exogenous input vectors.

Objectives

Over-arching goal:
Model observed sepsis patients' states as a function of vital signs and hospital intervention/treatments.

Mini-objectives:

- Learn Expectation-Maximization for Autoregressive models and how to create simulated cardiovascular data
- Learn Switching Autoregressive (SAR) models
- Have a general understanding of Switching Kalman Filters (SKF)
  - Use existing SKF code to cross-validate on simulated data
  - Use existing SKF code to cross-validate on real data
  - Generate a random coefficient matrix for exogenous input vector
- Do EM on AR model, inferring coefficient matrix on exogenous input vector
- Test model on real (MIMC II) and simulated data

The Theory

Sepsis patients can be classified as being in certain physiological "states;" for example, "alive," "onset of sepsis," "sepsis shock," or "not alive." We are trying to characterize the state patients are in, using dynamical behaviors (vital signs and exogenous inputs like hospital treatments).

To model this data, I use an autoregressive time series model – a form of a Linear Dynamical System (LDS). LDS models are State Space Models (SSM) where the observation model is continuous and typically Gaussian. SSM's are a form of Hidden Markov Models. These models consist of a discrete-time, discrete-state Markov chain with hidden states, plus the afore-mentioned continuous observation model.

The Data

Simulated Cardiovascular Data

- Why use simulated data?

Simulating data allows us to test the accuracy of our "state" predictions (see Preliminary Results below). In a real-world setting, we would like to automatically classify which state patients are in during periods in time.

- What is in the simulated data, and how do you create it?

The simulated data is created using a modified version of Kevin Murphy's `sample_lds` package. We assume that sepsis patients can be classified into two categories for this portion, then use Uniform and Poisson-distributed random samples to create the coefficient matrices.

MIMC II Data

This data consists of 200 patients, with the number of observations per person varying between ~800 to 50000. These observations are collected per minute via bedside monitors.

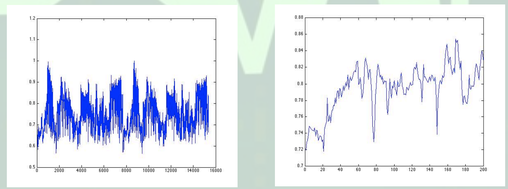

Figure 1: Real data for randomly selected patient (out of 200): entire data set, and trimmed subset (one observation every 5 minutes, 200 observations).

Preliminary Results

Cross-validation on simulated data with random coefficient matrix on exogenous inputs

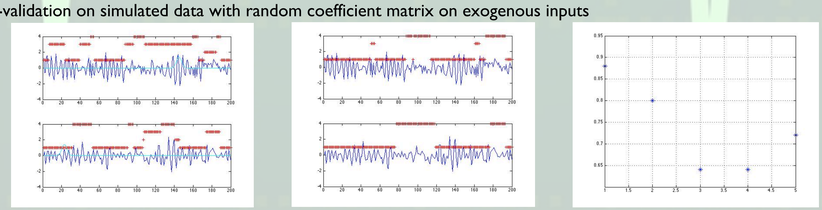

Figure 2: Simulated data for first and last of 20 patients with "true" states, predicted states and AUC measure over 5-fold CV.

Linear State Space Model

For each patient,

$$x_t = Ax_{t-1} + Bu_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, Q) \text{ i.i.d}$$
$$y_t = Cx_t + \omega_t, \quad \omega_t \sim N(0, Q) \text{ i.i.d}$$

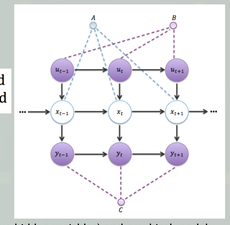

Figure 3: Linear state space model (continuous hidden variables) and graphical model.

Current Work

- Cross-validation on MIMC II ("real") data with random coefficient matrix on exogenous inputs
- Expectation-Maximization on Autoregressive model to infer coefficient matrix on vital sign vector
- Expectation Maximization on Autoregressive model to infer coefficient matrix on exogenous variable vector
- Time permitting, Switching Autoregressive (SAR) with inferred exogenous variable coefficient
  - The parameters {A, B, C, Q, R} of the model have been, until now, assumed to be time-invariant. In an SAR model, is that there is a bank of *different linear* models M, and that we switch between models depending on the time (or take a linear combination of them).

References

[1] Blocker, Alexander W. "An EM algorithm for the estimation of affine state-space systems with or without known inputs." (2008).

[2] Lehman, Li-wei H, et al. "A physiological time series dynamics-based approach to patient monitoring and outcome prediction." IEEE EMBC (2013).

[3] Murphy, Kevin P. "Switching Kalman Filters." 1998.

[4] Murphy, Kevin P. Machine Learning: A Probabilistic Perspective. Palo Alto. Springer 2012.

[5] National Institute of General Medical Sciences. Sepsis Fact Sheet. November 2012. 24 October 2013. <http://www.nigms.nih.gov/Education/factsheet_sepsis.htm.>

[6] Quinn, John A., et al. "Factorial switching linear dynamical systems applied to physiological condition monitoring." IEE TPAMI (2009).