Exercise session of the course "Managing Uncertainties"

frederic.pennerath@centralesupelec.fr

December 5, 2017

1 Exercise on Bayesian estimators:

Suppose one wants to measure the allergy risk of a new drug. A number n of people test the drug and a of them appear to be allergic.

1. What is the MLE estimator for this problem?

2. Show that the conjugate prior of a Bernouilli or binomial distribution is a Beta distribution. What are the hyperparameters of the posterior distribution of p? A Beta distribution $Beta(\alpha, \beta)$ is parameterized by two coefficients $\alpha > 0$ and $\beta > 0$. It defines the distribution of a variable X whose support is $[0, 1]$ with a density function:

$$Beta(\alpha, \beta) \sim \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)}$$

   The normalization factor is the beta function:

$$B(\alpha, \beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1} dx$$

3. Some prior tests in laboratory have estimated that the allergy risk is about 0.15 with a standard deviation of 0.1. What is the MAP estimation?

   One gives (but it can easily be checked) the mode x_{max}, expected value μ and variance σ^2 of a beta distribution $Beta(\alpha, \beta)$:

$$x_{max} = \frac{\alpha - 1}{\alpha + \beta - 2}, \mu = \frac{\alpha}{\alpha + \beta}, \sigma^2 = \frac{\alpha\beta}{(\alpha + \beta)^2 (\alpha + \beta + 1)}$$

4. One assumes to commercialize the drug only if the probability of having an allergic risk greater than 0.05 is less than 0.01. How could we make the decision?

2 Exercise on Bayesian classification: tax evasion

The tax authorities want to optimize their activities by focusing controls on taxpayers who are the most likely to evade taxes. They have collected millions of descriptions of citizens in a large database. Each description contains up to a hundred attributes (social and professional category, types of revenues, marital status, house(s) (type, size,localization), car(s) (type), etc). Some taxpayers (a hundred thousand) have already been controlled. Some of them (ten thousand) have been charged for cheating the tax authorities.

1. One assumes in a first stage that all the attributes take discrete values. One also assumes the authorities want to control every people that is expected to evade taxes. Help the authorities by proposing a simple method to select the subset of people to be controlled.

2. Is this problem biased in some way? How can we take into account this bias ?

3. How do you integrate numerical attributes such as declared income, asset value, etc?

4. What are the strong and weak points of the method?

5. Suppose now that the authorities are only able to control up to a hundred people a day in average. How does it affect the method?

6. Suppose controlling one taxpayer costs in average 1000 € and that the discovery of a tax cheater yields 20000 € in average. How can we integrate this input?

3 Exercise on Bayesian filtering: an aircraft tracking system

One considers a radar based on a ground station (airport) monitoring the airspace. The principle of a radar is to emit electromagnetic waves in some direction. If the wave hits some flying object, some wave energy is reflected towards the emitting antenna. The echo, when received by the radar, allows to measure approximatively the bearing angle and range of the target. The radar is rotating at a low frequency of 0.5 Hz.

Because the echo can be missed (especially when the target is far) and because the radar is pointing to the target only every 2 seconds, it is important to track continuously the positions of the airplanes in real-time between the instants of reception. It is also important to understand that the measures only provide an approximated estimation of the aircraft position: the uncertainty over the beam angles are fixed while the uncertainty over the measured range is proportional with the real range.

1. Propose a system to track the position of the airplane that takes into account the variable accuracy on the measures.
2. Give the state representation model of your system.
3. Propose a version of it that can be easily implemented and visualized on the screen of the operator.
4. Write down the equations to predict the expected state of your system.
5. Write down the equations to update the state of your system when receiving a new measure.
6. Most aircrafts have now GPS systems to localize accurately their positions. These positions can then be radio transmitted to the ground stations. How do you improve your system with this new pieces of information?