Theory of belief functions: Application to machine learning and statistical inference
Projects

Thierry Denœux

Summer 2023

Project 1: Logistic regression

We consider the logistic regression model

$$\log \frac{P(x)}{1 - P(x)} = \beta_0 + \beta_1 x,$$

where $P(x) = P(Y = 1|X = x)$, $Y \in \{0, 1\}$ is the binary response and X is a scalar covariate.

1. Write a function in R that computes the relative likelihood function for a dataset $\{(x_i, y_i)\}_{i=1}^n$. Plot the contours of this function for the dataset `chdage` in package `applore3`.

2. Write a function that computes the profile likelihoods for β_0 or β_1. Plot these functions for the `chdage` dataset.

3. Show that values of the contour function of $P(x)$, defined as $pl : p \mapsto Pl[P(x) = p]$ can computed as the solution of a univariate optimization problem. Write a function that solves this problem. Plot $pl(p)$ as a function of p for different values of x.

4. Write a structural equation of the form $Y = \varphi(P(x), U)$ for this problem, where U is a random variable with known probability distribution. Write a function that computes the predictive mass function of Y for any value of x. Plot the masses $m_Y(\{0\})$, $m_Y(\{1\})$ and $m_Y(\{0, 1\})$ as functions of x.

   (Hint: cf. Section 3.2 of the paper "O. Kanjanatarakul, S. Sriboonchitta and T. Denoeux. Forecasting using belief functions: an application to marketing econometrics. International Journal of Approximate Reasoning, Vol. 55, Issue 5, pages 1113-1128, 2014".)

Project 2: AR(1) process

We consider a zero-mean AR(1) process governed by the equation

$$X_t = \rho X_{t-1} + \epsilon_t, \quad t = 1, 2, \ldots,$$

where $\rho \in (-1, 1)$ is a coefficient and $\epsilon_t \sim \mathcal{N}(0, \sigma^2)$ is a Gaussian error term with zero mean and standard deviation σ. The parameter vector is denoted by $\theta = (\rho, \sigma)$.

1. Write a function that generates a time series of length T from this process.

2. Write a function that computes the relative likelihood function for a dataset $x_1, \ldots, x_T$. Plot the contours of this function for a randomly-generated dataset.

3. Write a function that computes the profile likelihoods for ρ or σ. Plot these functions.

4. One-step-ahead prediction: write a structural equation of the form $X_{T+1} = \varphi(\theta, U)$ for this problem, where U is a random variable with known probability distribution. Write a function that generates focal intervals of the predictive belief function on X_{T+1}. Plot the lower and upper cdfs, as well as the contour function of this belief function.

5. Two-steps-ahead prediction: write a structural equation of the form $(X_{T+1}, X_{T+2}) = \varphi(\theta, U_1, U_2)$ for this problem, where (U_1, U_2) is a random vector with known probability distribution. Write a function that generates boxes of the form $[a_i, b_i] \times [c_i, d_i]$ approximating focal intervals of the predictive belief on X_{T+1}. Estimate the degrees of belief and plausibility of events of the form "$X_{T+1} \leq X_{T+2}$", "$(X_{T+1} \geq 0)\&(X_{T+2} \geq 0)$", etc.

(Hint: cf. "T. Denoeux and O. Kanjanatarakul. Multistep Prediction using Point-Cloud Approximation of Continuous Belief Functions. 2019 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2019), New Orleans, USA, June 23-26, 2019".)

Project 3: Linear regression with Student-t distributed erreor

We consider a linear model of the form

$$Y = \beta_0 + \beta_1 x_i + \epsilon_i,$$

where the error terms are iid from a Student distribution with 5 degrees of freedom.

1. Write a function that generates a data set according to this model (the distribution of X can be arbitrary).

2. Write a function that computes the relative likelihood function for a given dataset. Plot the contours of this function for a randomly-generated dataset.

3. Write a function that computes the profile likelihoods for β_0 or β_1. Plot these functions.

4. Write a structural equation of the form $Y = \varphi(x, \beta, U)$ for this problem, where U is a random variable with known probability distribution. Write a function that generates focal intervals of the predictive belief function on Y. Plot the lower and upper cdfs, as well as the contour function of this belief function for different values of x.

5. We now assume that the value x of the covariate is uncertain and described by a triangular distribution with support $[x_*, x^*]$ and mode $\widetilde{x}$. Write a function that generates focal intervals of the predictive belief function on Y. Plot the lower and upper cdfs, as well as the contour function of this belief function for different values of x_*, x^* and $\widetilde{x}$.

(Hint: cf. Section 5.3 in the paper "O. Kanjanatarakul, T. Denoeux and S. Sriboonchitta. Prediction of future observations using belief functions: a likelihood-based approach. International Journal of Approximate Reasoning, Vol. 72, pages 71-94, 2016.".)