Aggregation for road traffic prediction

Amélie DETAIS
Laboratoire de Statistique et Probabilités
Université Paul Sabatier, 118 route de Narbonne
31062 Toulouse cedex 9
detais@cict.fr

Forecasting of road traffic travel time has recently become an increasingly studied topic in statistics. Various methods have been proposed in the statistical literature, such as mixture models, hierarchical classification, constant speed method,... (see Loubes *et al.*, *Road trafficking description and short term travel time forecasting, with a classification method*, to appear in The Canadian Journal of Statistics). Past studies have shown that none of these methods is better than the others.

Therefore, a natural idea is to combine different procedures, in order to retain the advantages of each one. This has motivated us to study an aggregation method to forecast road traffic travel time.

For each of several speed measurement stations located along some road axe of interest, a model for road traffic can be written as :

$$Y_j = \sum_{p=1}^{P} \mathbb{I}_p(X_j) f_p + \varepsilon_j, \text{ for all } j \in \{1, ..., J\},$$

where :

- Y_j is the vector of observed speeds at discrete times during the j-th day,
- f_p is the p-th vector of speed archetype,
- $(X_j)_{j\in\{1,...,J\}}$ are i.i.d. unobserved random variables taking values in $\{1, ..., P\}$, indicating the speed archetype of the current day,
- $(\varepsilon_j)_{j\in\{1,...,J\}}$ are independent centered gaussian vectors, with covariance matrix Γ.

We aim at establishing the properties of an aggregated estimator of the day archetype, obtained from M estimation methods. Given a vector of observed speeds until time n, each of these M methods gives us an estimator $\hat{f_m}$ $(m = 1, \dots, M)$ of the day archetype. For $\theta \in \mathbb{R}^M$, we define an aggregated estimator as

$$\hat{f}_\theta = \sum_{m=1}^{M} \theta_m \hat{f_m}.$$

We first estimate the vector θ by using a least-square penalized method, leading to the estimator $\hat{\theta}$. We then evaluate the qualities of the aggregated estimator $\widehat{f_{\hat{\theta}}}$ by providing an upper bound for its quadratic error. This is obtained by using a Log-Sobolev inequality.