Optimal Estimation

Jorma Rissanen

Helsinki Institute for Information Technology,
Tampere University of Technology, Finland

1 Modeling Problem

Data $Y = \{y_t : t = 1, 2, \ldots, n\}$, or $Y|X = \{(y_t, x_{1,t}, x_{2,t}, \ldots)\}$, X *explanatory* variables. Want to learn properties in Y expressed by set of distributions as *models*: $f(Y|X_s; \theta, s)$, where $\theta = \theta_1, \ldots, \theta_{k(s)}$ real-valued parameters, s structure parameter: for picking the most important variables in X.

1.1 Models and Estimators

To simplify notations write $y_t, x_{1,t}, x_{2,t}, \ldots$ as x_t; structures determined by number k of real-valued parameters.

Classes of parametric models

$$\mathcal{M}_k = \{f(x^n; \theta, k) : \theta \in \Omega^k \subset R^k\};\ k \leq n$$
$$\mathcal{M} = \{\mathcal{M}_k : k = 1, 2, \ldots, K, K \leq n\}.$$

Sets of *estimator* functions $\bar{\theta}(\cdot), \bar{k}(\cdot)$. Consider the distributions defined by estimators

$$\text{for fixed } k: \quad \bar{f}(x^n; k) = f(x^n; \bar{\theta}(x^n), k)/\bar{C}_{k,n}$$
$$\bar{C}_{k,n} = \int f(y^n; \bar{\theta}(y^n), k)dy^n$$
$$\text{in general}: \quad \bar{f}(x^n) = \bar{f}(x^n; \bar{k}(x^n))/\bar{C}_n$$
$$\bar{C}_n = \sum_k \int_{\bar{k}(y^n)=k} \bar{f}(y^n; k)dy^n$$

Let $\hat{\theta}(\cdot), \hat{k}(\cdot)$ be the estimator that maximizes $\bar{C}_n$:

$$\hat{C}_n = \max_{\bar{\theta}(\cdot), \bar{k}(\cdot)} \bar{C}_n. \tag{1}$$

It also maximizes the probability or density $\hat{f}(x^n)$ on the observed data, which is taken as the single postulate for this theory of estimation. The maximum $\hat{C}_n$ is called the maximum capacity, and it is also the maximum mutual information that any estimator can obtain about the models in the class.

T. Elomaa, J. Hollmén, and H. Mannila (Eds.): DS 2011, LNCS 6926, p. 30, 2011.