Semiparametric Estimation for Kernel Families

Abram Kagan

Dept. of Mathematics, University of Maryland, College Park

Abstract

Suppose that the statistician has a sample of (large) size n from a population P_θ where not only a (finite-dimensional) parameter of interest θ is unknown, but the available information on the family$\{P_\theta,\ \theta \in \Theta\}$ is in the form of a sample of (large) size m from P_{θ_0} where $\theta_0 \in \Theta$ is known. How to estimate θ in this setup?
We consider a class of these problems that may, on one side, be useful for applications (at least, in the speaker's opinion) and, on the other, allows a rigorous mathematical statistical analysis.
A family $\mathcal{P} = \{P_\theta,\ \theta \in \Theta\}$ of probability distributions on a measurable space $(\mathcal{X},\ \mathcal{A})$ is called a *kernel* family with kernel $h(x;\theta) \geq 0$, generator P (a probability distribution on $(\mathcal{X},\ \mathcal{A})$), and parameter θ if

$$dP_\theta(x) = C(\theta)h(x;\theta)dP(x)$$

with a normalizer $C(\theta)$.
A classical example of a kernel family is a natural exponential family (NEF) of distribution functions $F(x;\theta)$ on $(\mathbb{R},\ \mathcal{B})$,

$$dF(x;\theta) = \exp\{\theta x - \psi(\theta)\}dF(x)$$

depending on a scalar parameter θ.
A general setup

$$dF(x;\theta) = C(\theta)h(x;\theta)dF(x)$$

is considered when both θ and F are unknown and the available data are two independent samples, $(x'_1, \ldots, x'_m)$ from $F(x)$ and $(x_1, \ldots, x_n)$ from $F(x;\theta)$. The setup can be interpreted as a case-control study when F is the distribution of a characteristic of interest in the control group and the parameter θ measures the effect of a treatment.
The asymptotic behavior as $m, n \to \infty$ of the method of moments estimator (MME) and the maximum likelihood estimator (MLE) of θ is described in cases of (i) $m = cn(1 + o(1))$, $c > 0$, (ii) $n = o(m)$ and (iii) $m = o(n)$. An interesting phenomenon is that the asymptotic variance of the MLE is not always less than that of the MME.