Probab. Theory Relat. Fields **113**, 415–419 (1999)

Probability Theory and Related Fields

On rate and sharp optimal estimation*

Sam Efromovich

Department of Mathematics and Statistics, University of New Mexico, Albuquerque, NM 87131, USA

Andrew, Lucien and Pascal are to be congratulated for describing and studying an important problem of model selection via penalization. I also offer my congratulations for the solution of many well known subproblems, such as rate-optimal adaptation over the range of all ellipsoids, adaptive estimation for rich sieves and estimating the support of a distribution. I believe that your interpretation of the results of Talagrand will be employed by many statisticians.

I would like to comment on some of the obtained results (mainly for the case of nested family of sieves) via some known and recent results on sharp minimax estimation where the squared integrated risk converges with optimal rate and the best constant.

Projection estimation versus sharp minimax estimation for Ellipsoids

Let us, following Section 2.4.2, consider the case of the Ellipsoid $\mathscr{E}(\alpha)$ and the Fourier basis. Then any projection estimator with the cutoff m proportional to $n^{1/(2\alpha+1)}$ is rate minimax. On the other hand, to get the best constant of the minimax mean integrated squared error convergence, the cutoff is to be equal to $m^*(\alpha, n) = c^*(\alpha)n^{1/(2\alpha+1)}$ with some specific $c^*(\alpha)$ and then estimated Fourier coefficients are to be smoothed by the weights $(1 - (j/m^*(\alpha, n))^\alpha)$, see Efromovich and Pinsker (1982) and Efromovich (1996).

* This paper discusses some aspects of the preceding paper by Barron et al. in this same issue [Barron A., Birgé L., Massart P.: Risk bounds for model selection via penalization. Probab. Theory Relat. Fields **113**, 301–413 (1999)].

Thus, unfortunately, for this particular class of Sobolev functions there is no chance that the projection estimator will give us the best constant. However, there is a nice class of functions where the projection estimator is sharp.

Projection estimation is sharp for analytic functions

Consider classes of functions on the torus or on an interval whose Fourier coefficients belong to the Ellipsoid with $|a_j| = Ce^{-\gamma j}$ or to the similar hyperrectangle. Analytic functions are the familiar examples of such classes, see Section 7.8 in DeVore and Lorentz (1993).

For this case a projection estimator is sharp minimax. Moreover, we can show that: (i) this projection estimator is sharp for pointwise minimax approach; (ii) derivatives of this estimator are sharp minimax estimators of the corresponding derivatives; (iii) for the case of density, integration of the minimax projection estimator leads to the second order efficient estimation of the cumulative distribution function.

Thus, the accuracy index is an almost perfect oracle whenever the underlying function is sufficiently smooth.

I have been able to find a data-driven projection estimator (where the cutoff is based only on data) that matches all these properties of the oracle except for the best constant of the pointwise risk where one must lose it. On the other hand, it is not clear from the paper that the penalized estimator matches these properties. Is it possible to clarify this particular example?

To match a minimax estimator or the accuracy index?

I believe that the minimaxity is stressed too much in the paper. The main result is that the penalized estimator matches (up to a constant and sometimes up to a logarithmic factor) the accuracy index that is based on the underlying function. This is the main result and the fact that this accuracy index for some settings gives us minimax rates is secondary. There are many estimators that try to match a minimax estimator, for instance, minimax plug-in estimators. However, a minimax estimator is not the best for a majority of functions from a considered class and therefore matching the accuracy index is more promising. The following example sheds light on the issue.

Simultaneous sharp estimation of a function and its derivatives

This is a curious example that shows that an adaptive estimator may perform better than a minimax oracle. Consider the problem of sharp minimax estimation of functions and their derivatives for the case of an Ellipsoid. It is known that derivatives of the minimax oracle are not the minimax estimators of these derivatives, see Tsybakov (1997). On the other hand, derivatives of the Efromovich-Pinsker's adaptive estimator (this estimator matches the sharp accuracy index) are minimax estimators of the corresponding derivatives. Thus, an adaptive estimator may be both minimax and may have properties that the corresponding minimax oracle does not have. I am sure that the authors could suggest similar examples as well.

The multivariate case

The examples in Sections 2.3 and 3.1.1 about polynomials on a sphere are mainly given to support the claim that the suggested method can be applied to a wide variety of problems including the multivariate ones. It looks to me that aside from overcoming some technical difficulties, this case does not shed any new light on the rate-optimal estimation.

On the other hand, this example has enlighten the problem of sharp adaptive estimation via Efromovich-Pinsker estimator. The underlying idea of this estimator is to divide the frequency domain into coronas and then to mimic the best linear estimator that uses the same shrinking within each corona. For the case of the torus, merits and demerits of this estimator are simplicity of calculation of the estimate (because no optimization problem is to be solved) and the necessity to combine different zonal eigenspaces into the same corona, respectively. For the multivariate case, the merit is the same but the demerit disappears because here we can use the zonal eigenspaces as the coronas.

Example when ordering of sieves does matter

I believe that exploring "enriched" families of sieves is very promising. However, for the discussed rate-optimal approach ordering of the sieves is not so crucial. Here I would like to give an example where this ordering is important for sharp-optimal estimation.

Consider the well known problem of estimating aperiodic differentiable functions from an Ellipsoid. It is well known that the trigonometric basis is not reach enough to get even rate-optimal estimation and therefore one has to enrich it, for instance, by a linear function.

To get rate-optimal estimation, it is sufficient to apply the authors' method to this enriched family of sieves. Note that it is absolutely irrelevant what index or series of indexes is assigned to the linear sieve whenever it is under the consideration of the penalized procedure. On the other hand, this index is to be a very special one to get sharp minimax estimation.

I believe that similar curious phenomena exist for rate-optimal estimation as well. Do the authors know some of them?

Small samples

The reader who is interested in the numerical analysis of the authors' accuracy index, optimal truncation m and the corresponding adaptive projection estimators is referred to Efromovich (1996) where the case of 18 specific probability densities and sample sizes from 50 to 1000 is studied. While the authors give some precise constants in their bounds, I simply computed the discussed characteristics for this set of functions and the sample sizes. Interestingly, this study has showed that the asymptotic research sheds light even on the case of these small sample sizes.

Do we always need an adaptation?

On first glance, the question looks very strange. Adaptation is the core topic of the paper. However, I would like to give two related examples where no adaptation is necessary at all for sharp and rate-optimal estimation.

The first problem is the probability density estimation where the random variable of interest is observed with additive normal (or any so-called supersmooth) measurement error. In this case we observe realizations of the random variable $Y = X + \epsilon$ and we are to estimate the probability density of X. It is assumed that this density belongs to an Ellipsoid and the distribution of the measurement error is known. Then, there exists a special projection minimax estimator whose cutoff m depends only on n, that is, there is no need in adaptation to unknown smoothness of the estimated density.

Similar phenomenon occurs for the problem of demixing in Poisson Mixture models where a special projection estimator is minimax, see Hengartner (1997).

Is everything rosy for adaptation?

The reader, after looking through the paper that manifests the complete victory of adaptation, might suppose that adaptation always leads us to matching (apart of at most a logarithmic factor) minimax oracles. Unfortunately, this is not the case. Consider the well known problem of sequential estimation of a curve with a given assigned risk where the quality of sequential estimation is defined via the minimax mean sample size. Efromovich (1995) shows that a minimax oracle beats rates of any adaptive sequential estimator (except of some particular cases). Thus, you “inverse” the setting and then matching rates of minimax oracles is impossible.

Question

In Birgé and Massart (1994a) the authors note that it is unclear whether the extra $\ln(n)$ factor is necessary or it is due to some weakness of the approach. Is there any progress in this direction?

References

Efromovich, S.: Adaptive orthogonal series density estimation for small samples, Computational Statistics and Data Analysis **22**, 599–617 (1996)

Efromovich, S.: Sequential nonparametric estimation with assigned risk, Annals of Statistics **23**, 1376–1392 (1995)

Efromovich, S.: On nonparametric regression for iid observations in general setting, Annals of Statistics **24**, 682–686 (1996)

Hengartner, N.: Adaptive demixing in Poisson mixture models, Annals of Statistics **25**, 917–928 (1996)

Tsybakov, A.: Asymptotically efficient signal estimation in L_2 under general loss functions, Problems of Inform. Transmit. **33**, 94–106 (1997)