19 October 2003

Professor RR de Ruyter van Steveninck
Department of Physics
Indiana University
Bloomington, Indiana 47405–7105

Dear Rob,

I am writing to recommend **Ilya Nemenman**, who has applied for a junior faculty position in your department. As you know, I had the good fortune to be Ilya's thesis adviser. He is a young man of considerable breadth, without sacrificing depth. He does this by bringing enormous energy and an impressive intellect to everything that he does. This is a long letter, simply because Ilya has done many things.

Ilya has worked with experimental groups at CERN and at Stanford, making significant contributions to problems at the core of data analysis in complex experiments. At the opposite extreme, he has done formal work on renormalizability in quantum field theory. For his thesis, we worked together on problems that sit at the interface of statistical mechanics with computer science and (eventually) neurobiology.

The central issue in Ilya's thesis is the characterization of complexity. This is an old problem. Most of us share an intuition about what is complex, what is simple, and what is just random, but the challenge is to quantify this intuition. Our bias as physicists is that complexity should have something to do with entropy, but this can't be the whole story since purely random systems have maximal entropy and these are not complex in the intuitive sense. For dynamical systems, Grassberger has argued that the signature of complexity (as opposed to chaos or randomness) is the slow approach of the entropy to its extensive limit. But there is a very different arena where it is important to characterize complexity, and this is in the theory of learning; roughly speaking, it is more difficult to learn a complex model than a simple one, and there are many examples to show that complexity is not just the number of parameters in the model. In fact the complexity of models has many literatures, in computer science, coding theory, statistics, ... Dozens of different measures have been proposed, and it is tempting to view the whole enterprise as a bit of a mess.

Quite dramatically, Ilya has shown that the problem of measuring complexity (at least for one dimensional time series) can be solved in simple physical terms. Generalizing Grassberger's ideas, the subextensive entropy plays the central role. Subextensive components in the entropy are what allow for predictions to be made, and in fact the information that we have about the future of a time series can be written exactly in terms of subextensive entropy. If we think about a time series from which we are supposed to learn a model that describes the data, then prediction or generalization is the essence of success in learning. This suggests a path for unifying the discussion of complexity in dynamical systems and in learning. Ilya has developed this path in detail, and has several major results:

- Time series can be classified based on the asymptotic behavior of the subextensive entropy or predictive information.
- If the data stream is generated by a class of models in which we can learn a finite number of parameters, then the predictive information diverges logarithmically; the coefficient of the divergence counts the number of parameters or more rigorously the phase space dimensionality of the model class.
- Beyond the class of finitely parameterizable models, if we can learn, for example, a (nonparameteric) smooth function to describe the data then the predictive information diverges as a power law.
- Simple models for learning of smooth functions or probability densities can be formulated using ideas from field theory; this leads both to real calculations of the predictive information in these cases and to new and efficient learning algorithms.

This was an impressive body of work.

Although I have emphasized new results—for example, the power law class of problems is outside the traditional domain of learning theory, but is natural from the physics point of view—Ilya also made a brilliant effort to understand how different ideas about complexity are related. The bottom line is that predictive information, and in particular the divergent component of the predictive information, is both the sensible and the unique measure of complexity that satisfies some reasonable constraints, but it was vital to understand how this relates to work (often very formal) in other fields and from other points of view.

Ilya has taken the ideas of his thesis in several different directions, with results just beginning to come out in papers:

Model building. Predictive information measures what is complex or (colloquially) "rich" about a time series. The general problem of providing an efficient representation for this predictive information can be formulated as a variational principle, and simple examples suggest that solving this variational problem can drive the discovery of the correct dynamical model that underlies the time series. Ilya is using these ideas to approach problems such as how we can infer a network of regulatory interactions by observing the dynamics of gene expression.

Network complexity. Having understood the relations among complexity, dimensionality and learning from a unified perspective, Ilya is thinking about adaptation and evolution in biological networks. He has new ideas about how to make precise the intuitive notions of robustness and evolvability, and is trying to relate these to empirical measures of effective dimensionality for these networks.

Adaptation and optimization. The photoreceptor cells in the retina are remarkable physical devices, at one extreme counting single photons but also functioning over an enormous dynamic range. Adaptation is essential, and simple arguments suggest that adaptation allows the cell to make maximally efficient use of its dynamic range in conveying information about fluctuating inputs. But translating this picture into an optimization principle doesn't quite account for the well established phenomenology. Ilya has argued that this "near miss" in fact has the form expected if predictive and not total information were the biologically interesting quantity.

Learning and entropy estimation. Progress in theories of learning should generate better learning algorithms, and Ilya has been working to develop purely information theoretic approaches to regularize the problem of learning discrete distributions. As in the field theoretic approach to continuous distributions, the central role is played by phase space considerations. He has achieved reliable estimates of entropy from surprisingly small numbers of samples, and this approach actually works for the kinds of probability distributions encountered (for example) in the responses of sensory neurons. While entropy estimation is an old problem at the heart of coding and data compression, as well as some strategies for the analysis of dynamical systems, these results seem to advance not just the conceptual framework but also the practical state of the art. This work solves the major technical problem in the application of statistical and information theoretic ideas to the analysis of experiments in neuroscience, and I expect that it will have a similar impact on problems in bioinformatics.

Nemenman combines true theoretical depth, a willingness to roll up his sleeves and analyze real data, and a taste for the conceptual issues at play in biological systems. His ideas about what a theorist can do at the interface of the physical/mathematical sciences and the biological sciences are deep and focused on a search for underlying principles, going far beyond the usual platitudes about large data sets in the genomic era. As an aside, the neuroscience data sets that Ilya has been analyzing—from your experiments—are far larger than many experiments on gene expression dynamics. Even in problems of data analysis, he searches for the compelling theoretical foundations that elude most others. I think his work on learning probability distributions and entropy estimation is a great example where exploring the foundations (which have an independent interest as potential models for learning by organisms) has led to dramatic practical progress.

Ilya is a special fellow, even by the standards of Princeton and Berkeley. Obviously students like this find their own paths. I don't exactly know where Ilya's path will take him, but it will be exciting to be involved, and he deserves every opportunity. It is a pleasure to give him my highest recommendation.

Best wishes,

William Bialek
John Archibald Wheeler/Battelle Professor in Physics
Princeton University