**Populærvitenskapelig framstilling, 2009**

Statistics for Innovation $(sfi)^2$ is one of the 14 Norwegian Centres for Research-based Innovation - Sentre for forskningsdrevet innovasjon (SFI), with the aim to enhance the capability of the Norwegian economy to innovate by focusing on long-term research, forging alliances between research-intensive enterprises and prominent research groups. Statistics for Innovation develops core statistical methodologies, strategically necessary to achieve innovation goals in four key sectors: petroleum, finance & insurance, marine and health. The acronym $(sfi)^2$ comes from the identical abbreviations of Statistics for Innovation and Sentre for Forskningsdrevet Innovasjon, both as sfi, with the square power indicating that we work for a more than linear growth and investment return! $(sfi)^2$ operates from 2007 and is part of Norsk Regnesentral (Norwegian Computing Center - NR) in Oslo, in partnership with the University of Oslo (UiO), the Norwegian University of Science and Technology (NTNU) and 11 partners: Biomolex, DnBNOR, Gjensidige, Hydro, The Institute of Marine Research, Pubgene, Oslo University Hospital as, Sencel, Smerud Medical Research, Spermatech and Statoil. (sfi)2 is funded by the Research Council of Norway and by the partners, with an annual budget of about 40 mill NOK, including in-kind contribution as research. About 100 researchers actively join the research projects of the centre. Statistics for Innovation participates to the UiO training programs and funds many PhD students, researchers and postdocs.

Science, industry and business possess the technology to collect, store, distribute and process huge amounts of data. Fantastically precious data, yet complex and incomplete. Statistics provides the methods for understanding data and learning about the world from data. Because knowledge is fragmented and data are incomplete, discoveries and decisions are uncertain. Statistics is the science to quantify how confident we are in our knowledge, to quantify risk and to make decisions under uncertainty. Statistics for Innovation is built on the fact that discovery and innovation in the four sectors (petroleum, finance & insurance, marine and health) pose common challenges to modern statistics. There is a common need for new statistical methods, theories and tools: much data from different sources is collected, but it is often hard to detect the few critical and informative ones; data is extremely complex, because we measure more in depth; the systems to be understood or predicted are *per se* more complex; the easy part of the analysis has been done, and it is necessary to discover "second order effects" which can give competitive advantages; and finally, evidence-based decision making is more and more required in society at large. Moving from this perspective, we work towards the identification and solution of specific and ambitious innovation aims, in all four areas and with our partners, finding the right combination of strong methodology, competitive knowledge, unique data, and cross-disciplinary research team.

Success for $(sfi)^2$ is measured in many ways. First of all, *valuable innovation*: we develop new statistical methodology which is necessary to achieve valuable innovation aims in the four target areas. Second, *culture*: we are building a new research culture, at the interface between statistical science and long term innovation, in industry, finance, business, science. Third, *people*: we train a new generation of statisticians, experts in the modern statistics, ready to enter the job market in key positions for innovation. In this global and highly competitive world, in order to be protagonists in producing outstanding innovation, to be *best and first*, we need to build on our competitive edges. Our research project portfolio is rich of highly ambitions projects, grand

challenges, where we take significant risk, but the stake is high. We are very happy to see, already now, very promising results in many projects, and clear successes, in some.

We develop new methods that can help Statoil to increase their efficiency in discovering new oil and gas. For each area, there, geology and geophysics allow the estimation of a probability of discovery. The problem when planning a drilling program is that there is no good model for updating these probabilities when data from newly drilled wells become available. We model dependency between Un-successful wells, and make the updating of the probability of discovery coherent and automatic. How many well should be drilled before we give up the prospect? In what order should locations be inspected? We use Bayesian networks to model dependencies. Bayesian networks is a class of stochastic models, represented as a network, where the elements are possible locations to drill, and the links between locations represent geophysics and geology. As with many other statistical powerful tools, they appear central in many of our $(sfi)^2$ is projects: Versions of bayesian networks are used to understand how genes collaborate in tumor growth, or how electricity prices in Germany depend on electricity prices in the Nordic countries, together with another dozen global economical indicators. The challenge is to capture important dependencies between locations, genes or time series. At Statoil we are currently evaluating our new approach, applied on an area where there are many wells, but still interesting exploration prospects left as well. The Bayesian networks will be tested by feeding it with historical data, where we now know the truth, and compare exploration decisions based on them with the decisions that were historically taken. Results are very encouraging, and we believe that this approach will be used to plan further drilling.

The Genomic Hyperbrowser, is a major project, involving a large research team, with computer scientists, molecular biologists from the Oslo University Hospital, software engineers and statisticians, who are creating one of the newest bioinformatics discovery tool in the world. With this web-based tool, it will be possible to compare in real time different genomic traces, representing precious molecular, biological, clinical information along the human genome. We are building a high level, abstract description of many generic biological questions, which represent a large portion of the typical questions which molecular biologists would ask. We translate biology into mathematical statistics, in order to gain analytical power from such generality, use rigorous quantitative investigations to establish relations that describe the complexity of elements of life. This abstract mathematical language is matched with powerful statistical tests, performing local analysis along the genome. The system is sufficiently flexible to handle disparate data types, and with the power and intelligence to handle many types of biological inquiries in real time. It is open source and integrated with a well established framework. A first beta version is soon ready.

Adaptation to climate change is a necessity. While every effort should be made to produce global mitigation strategies, mankind has to cope with the effects of climate change in the decades to come. We have launched a major effort to quantify the effect of climate change for the insurance industry and help design innovative coverage strategies. Equipped with years of experience in studying the effect of weather and climate on insurance losses in Norway, we have launched a four year European project, together with Gjensidige, Lloyds and the London School of Economics, and funded by the European Union, CCII-Climate Change for the Insurance Industry, with the aim to develop tools for risk management.

It is more important than ever for financial institutions to quantify and evaluate their risk exposures. The recent financial crises, and the evolution of increasingly complex financial products, substantiate the need for improved knowledge and practice of risk. The current global financial crisis calls for better quantitative methodology as an important component of financial risk management. Statistical methods play the central role. We have been working in the last years, with DnB NOR, at the interface between operational quantitative risk analysis and innovative statistical methodologies. In this position we hold a competitive advantage that produces innovation ideas and products for our partners and beyond. We have been developing a new statistical methodology, based on a graphical representation of multivariate distributions. We have shown how multivariate data can be modelled using a cascade of simple building blocks called pair-copulae. This probabilistic construction represents a radically new way of constructing complex multivariate highly dependent models. Our results have attracted attention internationally. Our approach will be important also in other areas, beyond finance, whenever the aim is to model how many quantities behave simultaneously, and in particular how they depend on each other when they are very large. We look to hydrology, meteorology and climate sciences.

We have been constructing powerful statistical models which assist fishery managers in setting quotas in an international setting for Havforskningsinstituttet. At $(sfi)^2$ we introducing important new aspects into the model, merging different sources of data. We look to new ways of monitoring the Barents Sea for juvenile fish, important in shrimp fishing, where we expect to be able to quantify the advantage of our approach in terms of reduced sampling needed to reach the same precision, or a much higher precision when operating optimally.

Integrating different sources of data is also the core of our projects in clinical trials. With Smerud Medical Research we developed new statistical tools that will allow to analysis together different and independent clinical studies of similar drugs: This allows a much more powerful analysis, given the same costs. Bayesian methods are used, like in many of the other projects, and simulation based estimation is performed. This approach is now considered from the point of view of the market, with particular reference to the merger of pharmaceutical companies.

For further information, please contact the director of Statistics for Innovation, Professor Arnoldo Frigessi (frigessi@medisin.uio.no) or the host institution, Norsk regnesentral, with its CEO Lars Holden (lars.holden@nr.no).

1.12.2009
Arnoldo Frigessi