

**Dissertation release**

**09.02.2016**

## **New methods developed for improving the reliability of scientific discoveries**

<b>Title of the dissertation</b>	Sampling from scarcely defined distributions: methods and applications in data mining Otanta niukasti määritellyistä jakaumista: menetelmät ja sovellukset tiedonlouhinnassa
<b>Contents of the dissertation</b>	<p>Reliability and reproducibility of discoveries is essential for scientific progress. M.Sc. Aleksi Kallio studied difficult cases of scientific data analytics and developed new methods and approaches to assess the statistical significance of discoveries. Improved methods are needed due to rapidly growing volumes of data and more complex analytical questions that are faced in modern research.</p> <p>The dissertation introduces the term scarcely defined distributions to describe difficult statistical distributions that are common in modern data analytics. The dissertation discusses methods and applications of data mining, in which scarcely defined distributions emerge. Several strategies are put forth that allow to analyze complex datasets. Applications are reviewed from several fields, including bioinformatics, paleontology and ecology. A common factor for the application areas is the complexity of the underlying processes and error sources.</p> <p>The work concludes that development of new and flexible analytical methods is crucial for all fields that desire to use data to support decision making and prediction. If testing for significance and reliability is not on par with the rest of the data processing machinery then the future of data driven discovery will be plagued with false interpretations. The applicability of the research extends beyond the fields that were discussed. The generic methods and approaches can be adopted to many use cases where complex data sources are relevant, including major social questions related to medicine, climate and social networks.</p>
<b>Field of the dissertation</b>	Computer and information science, Data mining
<b>Doctoral candidate</b>	Aleksi Kallio, M.Sc. Born in Ristijärvi 1981
<b>Time of the defence</b>	19.2.2016 at 12:00
<b>Place of the defence</b>	Aalto University School of Science, lecture hall T2, Konemiehentie 2, Espoo
<b>Opponent</b>	Dr. Pauli Miettinen, Max-Planck-Institut für Informatik, Germany
<b>Custos</b>	Professor Aristides Gionis, Aalto University School of Science, Department of Computer Science
<b>Electronic dissertation</b>	<a href="http://urn.fi/URN:ISBN:978-952-60-6654-7">http://urn.fi/URN:ISBN:978-952-60-6654-7</a>
<b>School of Science electronic dissertations</b>	<a href="https://aaltodoc.aalto.fi/handle/123456789/52">https://aaltodoc.aalto.fi/handle/123456789/52</a>
<b>Doctoral candidate's contact information</b>	Aleksi Kallio, CSC – IT Center for Science Ltd., <a href="mailto:aleksi.kallio@csc.fi">aleksi.kallio@csc.fi</a>