https://owos.univie.ac.at

**ONE WORLD OPTIMIZATION SEMINAR**

May 31st 2021 @ 15:30 CEST (Central European Summer Time)

# DANIEL KUHN

## (EPFL)

## A General Framework for Optimal Data-Driven Optimization

**Abstract.** We propose a statistically optimal approach to construct data-driven decisions for stochastic optimization problems. Fundamentally, a data-driven decision is simply a function that maps the available training data to a feasible action. It can always be expressed as the minimizer of a surrogate optimization model constructed from the data. The quality of a data-driven decision is measured by its out-of-sample risk. An additional quality measure is its out-of-sample disappointment, which we define as the probability that the out-of-sample risk exceeds the optimal value of the surrogate optimization model. The crux of data-driven optimization is that the data-generating probability measure is unknown. An ideal data-driven decision should therefore minimize the out-of-sample risk simultaneously with respect to every conceivable probability measure (and thus in particular with respect to the unknown true measure). Unfortunately, such ideal data-driven decisions are generally unavailable. This prompts us to seek data-driven decisions that minimize the out-of-sample risk subject to an upper bound on the out-of-sample disappointment - again simultaneously with respect to every conceivable probability measure. We prove that such Pareto-dominant data-driven decisions exist under conditions that allow for interesting applications: the unknown data-generating probability measure must belong to a parametric ambiguity set, and the corresponding parameters must admit a sufficient statistic that satisfies a large deviation principle. If these conditions hold, we can further prove that the surrogate optimization model generating the optimal data-driven decision must be a distributionally robust optimization problem constructed from the sufficient statistic and the rate function of its large deviation principle. This shows that the optimal method for mapping data to decisions is, in a rigorous statistical sense, to solve a distributionally robust optimization model. Maybe surprisingly, this result holds irrespective of whether the original stochastic optimization problem is convex or not and holds even when the training data is non-i.i.d. As a byproduct, our analysis reveals how the structural properties of the data-generating stochastic process impact the shape of the ambiguity set underlying the optimal distributionally robust optimization model.
This is joint work with Tobias Sutter and Bart Van Parys.

*The link of the zoom-room of the meeting and the corresponding password will be announced the day before the talk on the mailing list of the seminar, to which one can subscribe on https://owos.univie.ac.at.*