

# DISTRIBUTIONALLY ROBUST OPTIMIZATION IN CONTEXT OF DATA-DRIVEN PROBLEMS

Erick Hans Delage, Ph.D.  
Stanford University, 2009

Reading Committee Members:  
Yinyu Ye, Stephen P. Boyd, and Shie Mannor

A decision maker is often confronted with situations where he has incomplete knowledge about some of the parameters of the problem that he is addressing: *e.g.*, market demand, future cost of merchandise, or quality of the available resources. While a simple approach that lets these parameters take their most likely values might lead to good solutions, it can easily lead to unexpected results in general. To avoid unnecessary deceptions, one should instead modify his decision model so that it accounts for the uncertainty that is present.

At a conceptual level, stochastic programming is an effective approach since it leads to decisions that directly trades off risk and performance. This is done by requiring that the decision maker formulate a distribution describing the probability that the parameters take any given value. The framework then offers an exhaustive array of risk measures to choose from for evaluating the performance of a candidate solution. Unfortunately, while this decision problem can generally be solved efficiently, the resulting “optimal risk-sensitive solution” can be misleading in applications where there is ambiguity in the choice of a distribution for representing the uncertain parameters. This is for instance the case in data-driven problems, where information about the distribution of parameters is mostly derived from the observation of historical realizations. After a limited amount of observations, a decision maker is often unable to fully determine the underlying distribution. This leads to three important questions for making a stochastic programming approach practical:

1. How accurate is a distribution estimate that is based only on a finite set of samples?
2. Can we account for distribution uncertainty when solving a stochastic program?
3. What is the computational cost associated with taking this uncertainty into account?

This thesis develops new theoretical foundations for a quantitative methodology that provides answers to these questions and can be used in a wide range of applications.

We first address these questions from a frequentist point of view. In doing so, we initially derive a new confidence region for the covariance matrix of a random vector given a finite set of independent and identically distributed samples. Unlike typical confidence regions, this new one constrains the terms of the matrix jointly with respect to their effect on the variance of any one dimensional projection of the random vector. The result allows us to define a “well structured” set of distributions that is guaranteed with high probability to contain the distribution from which the samples were drawn.

With the objective of accounting for such distribution uncertainty, we analyze the computational difficulty related to solving the distributionally robust form of the stochastic program. This model suggests making decisions according to the worst distribution in the uncertainty set. Although the model has been widely studied since Scarf introduced it in 1958, we are the first to use duality theory to show that for cases where the objective is convex in the decision variables and “piecewise concave” in the parameters, there exists a polynomial time algorithm that can solve the problem to any level of accuracy. In fact, we show that, for some of these models, the distributionally robust form of the stochastic program is much easier to solve than its original form. This analysis leads to a framework that provides reliable solutions to data-driven problems. Many applications can benefit from this framework, including both a fleet mix optimization problem and a portfolio selection problem. In particular, experiments with real stock market data confirm that this approach accounts more effectively for uncertainty in the future value of financial assets.

In the final part of this work, we choose to adopt a Bayesian approach. In this case, it is well known that one can choose to represent distribution uncertainty with a hyper-distribution (*i.e.*, a distribution over distributions). After formulating the distributionally robust problem in this context, we contrast the robust approach to a percentile criterion (*a.k.a.*, value-at-risk) which is more natural to consider. A study of Markov decision processes leads to interesting new results on the computational difficulties related to solving this percentile optimization problem. More specifically, we show that while this problem can be solved efficiently for a class of cost uncertainty, it can also lead to NP-hard problems. After providing an approximation algorithm for the intractable form, we perform a comparative study of how the robust and percentile based methods exploit the distribution information present in a machine replacement problem.