Finding $(\alpha,\vartheta)$-solutions via Sampled SCSPs

Roberto Rossi,² Brahim Hnich,² S. Armagan Tarım⁴ and Steven Prestwich⁴

¹Logistics, Decision and Information Sciences, Wageningen University, the Netherlands, roberto.rossi@wur.nl
²Department of Computer Engineering, Izmir University of Economics, Turkey, brahim.hnich@ieu.edu.tr
³Department of Management, Hacettepe University, Ankara, Turkey, armagan.tarim@hacettepe.edu.tr
⁴Cork Constraint Computation Centre, University College Cork, Ireland, s.prestwich@uc.ie

Abstract. We propose a novel approach for dealing with single-stage stochastic constraint satisfaction problems (SCSPs) that include random variables over a continuous or large discrete support. Our approach is based on two novel tools: sampled SCSPs and $(\alpha,\vartheta)$-solutions. Instead of explicitly enumerating a very large or infinite set of future scenarios, we employ statistical estimation to determine if a given assignment is consistent for a SCSP. As in statistical estimation, the quality of our estimate is determined via confidence interval analysis. In contrast to existing approaches based on sampling, we provide likelihood guarantees for the quality of the solutions found. Our approach can be used in concert with existing strategies for solving SCSPs.

Stochastic Constraint Satisfaction Problems (SCSPs)

An n-stage SCSP [Walsh, 2002; Tarım et al., 2006; Hnich et al., 2000] is defined as a 7-tuple $(V, S, D, P, F, C, \omega)$ where:

- $V$ is a set of decision variables;
- $S$ is a set of random variables;
- $D$ is a function mapping each element of $V$ (respectively, $S$) to a domain (respectively, support) of potential values. In classical SCSPs both decision variable domains and random variable supports are assumed to be finite.
- $P$ is a function mapping each element of $S$ to a probability distribution for its associated support.
- $C$ is a set of chance-constraints over a non-empty subset of decision variables and a subset of random variables. $\beta$ is a function mapping each chance-constraint $h$ to the threshold value in the interval $[0, 1]$.
- $\{h_1, h_2, \ldots, h_k\}$ is a list of decision stages such that each $h_i: V_{i-1} \times V_i \times S_i \rightarrow \mathbb{R}$, $V_i$ form a partition of $V$, and $S_i$ form a partition of $S$.

Example 1. Let us consider a two-stage SCSP in which $V_1 = \{a, b\}$ and $V_2 = \{c, d\}$, $S_1 = \{|\}\}$ and $S_2 = \{|\}\}$. Random variable $s_2$ may take two possible values, 5 and 4, each with probability 0.5. Random variable $s_1$ may also take two possible values, 3 and 4, each with probability 0.5. The domain of $s_1$ is $\{3, 4\}$, the domain of $s_2$ is $\{5, 6\}$. There are two chance-constraints in $C$: $p_1(s_1, s_2) = 0.75$ and $p_2(s_1, s_2) = 12.5$.

Policy tree

To solve an n-stage SCSP on assignment to the variables in $V$, we must find such that, given random values for $S$, assignments can be found for $V_i$ such that, given random values for $S_{i-1}$, assignments can be found for $V_{i-1}$ etc. in the order of the stages. Under the assumption that random variable supports are finite, the solution of an n-stage SCSP is, in general, represented by means of a policy tree (Furno et al., 2000). The arcs in such a policy tree represent values observed for random variables whereas nodes at each level represent the decisions associated with the different stages. We call the policy tree of an n-stage SCSP that is a solution a satisfying policy tree.

A possible solution to the SCSP in Example 1 is the satisfying policy tree shown in Fig. 1.

Sampled SCSPs

Consider a SCSP over a set $S$ of random variables. Assume that random variables are defined on continuous or large discrete supports. Solving the original SCSP clearly poses a hard combinatorial challenge, in fact the policy tree comprises a number of scenarios that depend on the number of random variable domains. Furthermore, if the random variable support is continuous, the policy tree comprises an infinite number of scenarios.

We propose to sample a more compact SCSP, which comprises at most $N$ scenarios, out of the original problem and we shall call this new problem a “sampled” SCSP over $N$ scenarios. Intuitively, a sampled SCSP is a random version of the original problem, the solution of which is a policy tree that comprises a bounded number of paths sampled out of the original policy tree. In Fig. 2 we show a sampled SCSP comprising 3 sampled scenarios for the problem discussed in Example 1.

Confidence interval analysis for Binomial proportions

Confidence interval analysis is a well established technique in statistics. Informally, confidence intervals are a useful tool for computing, from a given set of experimental results, a range of values that, with a certain confidence level (or confidence probability), will contain the true value of a parameter. We will focus on the binomial case, where the parameter is the probability of success.

Consider a discrete random variable $X$ that follows a Bernoulli distribution. Accordingly, each variable may produce only two outcomes, i.e., “yes” and “no”, with probability $p$ and $1-p$, respectively. Let us assume that the value of $p$ — the “yes” probability — is unknown. Obviously, if we observe the outcome of a Bernoulli trial once, the data collected will not reveal much about the value of $p$. Nevertheless, in practice, we may be interested in “estimating” $p$, by repeatedly observing the behavior of the random variable in a sequence of Bernoulli trials. This problem is well known in statistics and both exact and approximate techniques are available for performing this estimation (Clopper and Pearson, 1934; Agresti and Coull, 1998). The estimation produced by the binomial cumulative distribution function (CDF) is in order to build the interval from the data observed. The Clopper-Pearson interval can be written as $(\hat{p}, \hat{p}_{UB})$, where

$$\hat{p}_{UB} = \frac{X + \frac{1}{2} \alpha_{LB}}{N + \frac{3}{2} \alpha_{LB}}$$

The number of successes ($\text{"yes"}$ events) observed in the sample, $X$, is a binomial random variable with $N$ trials and probability of success $p$ and is in the confidence probability. Note that we assume $D(x) = 0$ when $X = N$ and $D(x) = 1$ when $N = X$.

Properties of SCSPs solutions

We will now characterize the probability that the solution of a sampled SCSP over $N$ scenarios is a solution to the original single-stage SCSP. We firstly discuss what the minimum value for $N$ is, in order to achieve a predefined probability $\alpha$ that a policy tree that satisfies a chance-constraint of the SCSP sampled solutions also satisfies the same chance-constraint in the original SCSP. Since a policy tree for the sampled SCSP is a solution for a subset of all the paths that constitute a policy tree for the original SCSP, this policy tree in order to satisfy $h$ in the original SCSP, this policy tree must clearly provide a sufficient satisfaction probability regardless of the scenarios that have been ignored by the sampling process.

Consider a confidence probability and a margin of error $\epsilon$. The number of scenarios $N$ for the sampled SCSP depends on $\alpha$ and $\epsilon$, which we recall is the satisfaction probability we aim for our chance-constraint $h$.

Definition 2: $N$ is computed as the minimum value for which $\text{Pr} \{D(D) = (0, 200), P(D) = \text{uniform}(0, 200), C = \{c : 0.5 \leq c \leq 0.6\} \} \geq (1 - \alpha)/2$. In order to achieve a predefined probability $\alpha$ that a given policy tree satisfies a chance-constraint $h$ in the sampled SCSP also satisfies the same chance-constraint in the original SCSP. Since a policy tree for the sampled SCSP is a solution for a subset of all the paths that constitute a policy tree for the original SCSP, this policy tree in order to satisfy $h$ in the original SCSP, this policy tree must clearly provide a sufficient satisfaction probability regardless of the scenarios that have been ignored by the sampling process.

Definition 4: An $(\alpha, \vartheta)$-solution to a SCSP $P$ is a policy tree that at least with probability $\vartheta$ for every chance-constraint $h$, with satisfaction threshold $\vartheta$, a satisfaction probability greater or equal to $\vartheta$ in the original SCSP.

It is apparent that $\alpha$ may be interpreted as a parameter that the user can set in order to define a “region of indifference” for the satisfaction probability. In actual region, we assume that the minimum is $N$ and that guarantees $D(D) = (0, 200), P(D) = \text{uniform}(0, 200), C = \{c : 0.5 \leq c \leq 0.6\}$, and that satisfaction probabilities remain in an acceptable range.

It can be proved that any solution to a sampled SCSP over $N$ scenarios, where $N$ is computed according to Definition 2, is an $(\alpha, \vartheta)$-solution to the original SCSP.

Example 2 (Rehearsal problem).

Consider the following SCSP $P = \{V, S, D, F, C, \omega\}$ where $V = \{a, b, c, d\}$, $\omega = \{0, 1\}$, and $D(a) = \{0, 1\}$, $D(b) = \{0, 1\}$, $D(c) = \{0, 1\}$, and $D(d) = \{0, 1\}$, a [1/2] analytical solution for the above problem is $\{0, 1\}$. Consider $\alpha = 0.9$ and $\vartheta = 0.95$, this directly implies that $N = 36$, and since this is the minimum $N$ that guarantees $D(D) = (0, 200), P(D) = \text{uniform}(0, 200), C = \{c : 0.5 \leq c \leq 0.6\}$, and that satisfaction probabilities remain in an acceptable range. In order to achieve a predefined probability $\alpha$ that a given policy tree satisfies a chance-constraint $h$ in the sampled SCSP also satisfies the same chance-constraint in the original SCSP.

The number of random variables in the domain of $D$ can be proved to satisfy $\vartheta$ or not satisfy the chance-constraint with the prescribed confidence probability $\vartheta$. More specifically, only values in $(0.75, 0.3)$ and $(0.3, 0.05)$ can be proved to not satisfy and satisfy the chance-constraint (see Fig. 3). Moreover, every value smaller than 0.75 will be correctly classified with confidence level greater or equal to $\alpha$.

In other words, assignments that provide a satisfaction probability less or equal to $\vartheta$ in the original SCSP, will be correctly classified as infeasible with probability greater or equal to $\alpha$. Thus, the solution of the sampled SCSP is an $(\alpha, \vartheta)$-solution to the original SCSP.

Our approach exploits sampling in order to keep under control the number of scenarios that must be analyzed in order to find a solution. Intuitively, our approach “estimates” if a given assignment is consistent or not with respect to a given set of chance-constraints. As in statistical estimation, the quality of this estimator is determined by confidence interval analysis. In contrast to Furno et al. (2000), we provide likelihood guarantees for the solutions found. In fact, we explicitly exploit a confidence probability that bounds the actual probability of making a wrong estimation.