5
SCENARIO GENERATION

5.1
Difficulty in scenario generation

While the number of possible network realizations is computationally tractable, the algorithm presented in Chapter 4 may be used in order to obtain a solution which is within a tolerance level ε from the global optimum of the original problem. However, if one wants to be able to solve large-scale problems, it becomes imperative to have an estimate of the expected value of the second stage cost function which is not based on the complete enumeration of all possible network configurations.

Standard two-stage stochastic programming models usually resort to scenario generation to allow for the evaluation of these multi-dimensional integrals. However, unlike the vast majority of problems studied in the literature, in the humanitarian logistics problem – and, more generally, in the class of problems presented in Section 1.5 – the probability distribution of the random variables is not known before first-stage decisions are determined.

As already pointed out in Section 2.3, this makes it impossible to utilize traditional scenario generation methods such as Monte Carlo sampling, moment matching or minimization of distances between probability measures. In this work, we propose to overcome this obstacle by merging the concepts from importance sampling into a stochastic programming framework, as presented next.

5.2
Importance sampling

In statistics, importance sampling is a technique used to estimate the properties of a certain distribution while only having samples drawn from a different one. In the context of simulation studies, importance sampling is usually employed as a variance reduction technique used in conjunction with the Monte

Carlo method. The basic idea is that certain values of the random variable may have a stronger effect upon the parameter being estimated than others, so it might be interesting to sample these values more frequently than what would otherwise be expected based on the original probability distribution.

As detailed in Rubinstein (1981) [51], the method relies on a simple observation to compute the expected value of a random variable $X \sim F_1(x)$ based on samples from another distribution $F_2(x)$:

$$\mathbb{E}_{f_1}\{x\} = \int_x x f_1(x) dx = \int_x x \frac{f_1(x)}{f_2(x)} f_2(x) dx = \mathbb{E}_{f_2}\left\{x \frac{f_1(x)}{f_2(x)}\right\} \quad (5.1)$$

For a given set of samples x_i $(i = 1, \dots, N)$ drawn according to a probability density function $f_2(X)$, the importance sampling estimator of the mean of distribution $f_1(X)$ is then defined as:

$$\hat{\mu}_X^{IS} = \frac{1}{N} \sum_{i=1}^{N} x_i \cdot \frac{f_1(x_i)}{f_2(x_i)} \quad (5.2)$$

Following expression (5.1), each sample is weighted differently based on the likelihood ratio, i.e. the ratio between the probability of occurrence of that sample under the distribution of interest and the one from which the samples were drawn.

Again according to [51], this estimator is proved to be consistent – it converges to μ_X with probability 1 as the sample size grows to infinity – and unbiased – its expected value is μ_X, whatever the sample size. In the next section, this technique is incorporated into the optimization problem so as to allow for the estimation of the second stage cost function based on scenarios.

5.3 Reformulation

Although the final (*post*-investment) probability distribution of the availability of the edges is not known *a priori*, the initial distribution (i.e., the one which does not consider any reinforcement investments) may be used to generate scenarios of network configuration, for which the probability of occurrence may be easily calculated. This is also the case of the more general class of stochastic programming problems with endogenous uncertainty defined in Chapter 2: the initial probability distribution of the random variables is always known, even though it might change after first-stage decisions are determined.

Additionally, since the linearization technique proposed in Chapter 3 makes it possible to compute the probability of occurrence of any scenario given the first-stage investment decisions (or, at least, an approximation to its value), we may join these pieces of information in order to compute the importance sampling estimator of the expected value of the second stage cost function.

By examining expression (5.2) for the importance sampling estimator, we may identify the corresponding elements of the optimization problem being studied: $f_1(x)$ and $f_2(x)$ are, respectively, the final and initial probability density functions of the scenarios, N is obviously the number of sampled scenarios and the samples x_i represent the values of the scenario-specific second-stage problems which are solved separately, as discussed in Chapter 3. Once again, it is important to stress that the scenarios of network realization are to be sampled according to the initial probability distribution of the edges' availabilities.

This analogy allows us to reformulate problem (3.12) – (3.17) in a way which does not require the full enumeration of all possible network configurations but relies on a smaller subset of randomly generated scenarios, as shown below:

$$(\boldsymbol{P_3}) \quad Min \quad \sum_{e \in E} r_e x_e + \frac{1}{|S|} \sum_{s \in S} g_s \left(\frac{\hat{p}_s}{p_s^{INI}} \right) \tag{5.3}$$

$$subject\ to: \quad Ax \leq b \tag{5.4}$$

$$w_s = \sum_{e \in E} \{\ln(\mathrm{p}_{es}^C) + [\ln(\mathrm{p}_{es}^I) - \ln(\mathrm{p}_{es}^C)] \cdot x_e\} \quad \forall s \in S \tag{5.5}$$

$$\hat{p}_s \geq \alpha_k + \beta_k \cdot w_s \quad \forall s \in S, \forall k \in K \tag{5.6}$$

$$\hat{p} \in \mathbb{R}^+, w \in \mathbb{R} \tag{5.7}$$

$$x \in \{0,1\}^{|E|} \tag{5.8}$$

where:

p_s^{INI} probability of sampled scenario s, calculated based on the initial probability distribution of the availability of each edge, i.e. $p_s^{INI} = \prod_{e \in E} p_{es}^{C}$

Based on a set of scenarios of network realizations, sampled according to the initial probability distribution of the edges' availabilities, a solution to problem (5.3) – (5.8) may be found using the algorithm outlined in Chapter 4.

5.4
Solution robustness

As with any two-stage stochastic program, the solution to these problems depends, essentially, on balancing the trade-off between deterministic first-stage costs and the expected value of probabilistic second-stage costs. It is thus imperative that we have a reasonable estimate of second stage costs in order to be able to have confidence in the quality of the solution obtained.

On the one hand, the larger the set of sampled scenarios, the better the estimate of second-stage expected costs will be. On the other hand, having fewer scenarios makes the problem smaller and solution times are usually faster. Anyhow, once a solution is found for a given set of scenarios, a Monte Carlo simulation – in which the probability distribution of the edges' availabilities takes into account the determined first-stage decisions – may then provide a confidence interval against which the estimate of the expected costs of the second-stage provided at the solution of the problem can be compared in order to assess the need for a larger number of samples. This is discussed in Appendix B, where an algorithm for determining an adequate number of scenarios is described.