Les Cahiers du GERAD ISSN: 0711–2440

On a two-stage discrete stochastic optimization problem with stochastic constraints and nested sampling

T.A. Ta, T. Mai,
F. Bastin, P. L'Ecuyer

G–2018–59

August 2018

La publication de ces rapports de recherche est rendue possible grâce au soutien de HEC Montréal, Polytechnique Montréal, Université McGill, Université du Québec à Montréal, ainsi que du Fonds de recherche du Québec – Nature et technologies.

Dépôt légal – Bibliothèque et Archives nationales du Québec, 2018
– Bibliothèque et Archives Canada, 2018

The publication of these research reports is made possible thanks to the support of HEC Montréal, Polytechnique Montréal, McGill University, Université du Québec à Montréal, as well as the Fonds de recherche du Québec – Nature et technologies.

Legal deposit – Bibliothèque et Archives nationales du Québec, 2018
– Library and Archives Canada, 2018

GERAD HEC Montréal
3000, chemin de la Côte-Sainte-Catherine
Montréal (Québec) Canada H3T 2A7

Tél. : 514 340-6053
Téléc. : 514 340-5665
info@gerad.ca
www.gerad.ca

On a two-stage discrete stochastic optimization problem with stochastic constraints and nested sampling

Thuy Anh Ta

Tien Mai

Fabian Bastin

Pierre L'Ecuyer

GERAD & DIRO & CIRRELT & IVADO, Pavillon André–Aisenstadt, Montréal (Québec), Canada

tathuyan@iro.umontreal.ca
maianhti@iro.umontreal.ca
bastin@iro.umontreal.ca
lecuyer@iro.umontreal.ca

August 2018
Les Cahiers du GERAD
G–2018–59

Abstract: We consider a two-stage stochastic discrete program in which some of the second stage constraints involve expectations that cannot be computed easily and are approximated by simulation. We study a *sample average approximation* (SAA) approach that uses nested sampling, in which a number of second stage scenarios are examined, and a number of simulation replications are performed for each scenario to estimate the second stage constraints. This approach provides an approximate solution to the two-stage problem. We show that in the second-stage problem, given a scenario, the optimal values and solutions of the SAA converge to those of the true problem with probability one when the sample sizes go to infinity. In the two-stage problem, these convergence results of the second-stage problem do not hold uniformly over all possible scenarios, and this complicates convergence proofs. We are nevertheless able to prove that the optimal values and solutions of the SAA converge to the true ones with probability one when the sample sizes at both stages increase to infinity. As an illustration, we apply this SAA method to a staffing problem in a call center, in which the goal is to optimize the numbers of agents of each type under some constraints on the quality of service (QoS). The staffing allocation has to be decided under an uncertain arrival rate with a prior distribution in the first stage, and can be adjusted at some additional cost when better information on the arrival rate becomes available in the second stage.

Keywords: Sample average approximation, two-stage stochastic program, expected value constraints, convergence rate, staffing optimization

Acknowledgments: This work has been supported by a Canada Research Chair, an Inria International Chair, and a Hydro-Québec research grant to P. L'Ecuyer, by NSERC Discovery Grants to F. Bastin and P. L'Ecuyer, and by scholarships from the CIRRELT, DIRO and Université de Montréal to T.A. Ta.

1 Introduction

We are interested in a class of two-stage stochastic optimization problems in which at each stage, a decision must be taken among a finite set of possibilities, under uncertainty. After making the decision x at the first stage, some information ξ is revealed, then the second-stage decision y is made, under a set of constraints that depend on x and ξ. Some of these constraints at the second stage involve mathematical expectations that cannot be computed exactly and are estimated by Monte Carlo simulation. We pay a cost that depends on x in the first stage, plus a cost that depends on (x, ξ, y) in the second stage. Our first goal is to find an optimal decision x^* for the first stage, to minimize the expected total cost, under the assumption that we will be able to make an optimal decision y in the second stage. Then, given $x = x^*$ and the observation of ξ, our second goal is to select an optimal $y = y^*(x, \xi)$ for that pair (x, ξ).

More formally, the problem can be formulated as follows:

$$
(\mathbf{P1}) \qquad \begin{cases} \min\limits_{x \in X} & f(x) = f_1(x) + \mathbb{E}_\xi[Q(x,\xi)] \\ \text{where} & Q(x,\xi) = \min\limits_{y \in A(x,\xi)} f_2(x,\xi,y) \qquad (1) \\ & \quad \text{subject to} \quad g(x,\xi,y) = \mathbb{E}_w[G(x,\xi,y,w)] \geq 0, \qquad (2) \end{cases}
$$

where $\omega = (\xi, w) \in \Omega = \Xi \times \mathcal{W}$ is distributed according to some probability measure $\mathbb{P}$ over the sample space Ω, and $\mathbb{E}_\xi$ and $\mathbb{E}_w$ denote the expectations with respect to ξ and w. In the applications we have in mind, ξ and w can be taken as independent. In particular, both ξ and w can be viewed as infinite sequences of independent random variables uniformly distributed over $(0, 1)$ and the required randomness is extracted from them (in a Monte Carlo context, these will be the random numbers that drive the simulation), but this interpretation is not essential. The first-stage decision x must be taken from the finite set X. Then ξ is observed and the second-stage *recourse* decision must be taken from the set $A(x, \xi) \subseteq Y$, which may depend on x and ξ, where Y is a finite set. This set $A(x, \xi)$ could be specified by a set of linear inequalities, for example, as will be the case in our illustrations. We also define $Y(x, \xi)$ as the set of second-stage feasible solutions given the pair (x, ξ), i.e., $Y(x, \xi) = \{y \in A(x, \xi) \mid\ g(x, \xi, y) \geq 0\}$. The functions $f_1 : X \to \mathbb{R}$ and $f_2 : X \times \Xi \times Y \to \mathbb{R}$ are measurable, while $G = G(x, \xi, y, w) = (G_1, \ldots, G_K)$ is a random vector for which $\mathbb{E}_w[|G(x, \xi, y, w)|] < \infty$ for all (x, ξ, y) such that $y \in A(x, \xi)$. We are interested in the situation in which the expected value functions $\mathbb{E}_\xi[Q(x, \xi)]$ and $\mathbb{E}_w[G(x, \xi, y, w)]$ cannot be written in a closed form or computed numerically, and are estimated by Monte Carlo.

The stochastic optimization problem considered here occurs in several real-life situations. It was motivated by a staffing optimization problem in telephone call centers, in which one must select a staffing, i.e., decide how many agents of each type to have in the center for each time period of the day, to minimize the operating cost while satisfying some quality of service (QoS) constraints, under uncertainty in the arrival rate process. In the first stage, the manager selects a staffing x for the given day some time in advance, based on an initial forecast of the arrival rate of calls. This staffing has cost $f_1(x)$. Later on, for example in the morning of the given day, an updated (better) forecast of the arrival rate, represented by ξ, becomes available. Based on this new information, the manager can modify the initial staffing by adding or removing some agents to better match the updated forecast by paying some penalty cost $f_2(x, \xi, y)$, where y represents the staffing modification. This y must satisfy a set of linear constraints that generally involve x, ξ, and y, captured here by $y \in A(x, \xi)$, and also some QoS constraints expressed as expectations: $\mathbb{E}_w[G(x, \xi, y, w)] \geq 0$, where w represents all the uncertainty that remains after ξ is known (e.g., the arrival times and service times of calls, abandonments, etc.). For example, one may ask that the expected total waiting time of all calls during the day does not exceed the expected number of calls multiplied by 30 seconds, or that the probability p that at least 95% of calls during the day are answered within 6 seconds is at least 0.90. The choice of these chance constraints reflects the decision maker's risk preferences. We assume that the arrival rate is bounded and that the finite set $A(x, \xi)$ always contains a staffing large enough to satisfy the QoS constraints, uniformly over x and ξ. For more details on this application, see for example Cez̧ik and L'Ecuyer (2008); Chan et al. (2014, 2016); Ta et al. (2016).

In this paper, we study a *sample average approximation* (SAA) approach to solve (P1). The general idea of SAA is to use Monte Carlo sampling to construct sample average functions that approximate the expectations $\mathbb{E}_\xi[Q(x,\xi)]$ and $\mathbb{E}_w[G(x,\xi,y,w)]$ as functions of x and of (x,ξ,y), respectively. In the SAA version of the problem (P1), the expectations are replaced by the sample averages, or equivalently, the exact distributions of ξ and w are approximated by empirical distributions. This permits one to easily compute the expectations as functions of x and y in the SAA problem, and then solve this problem.

The SAA approach itself is not new; see, e.g., Ahmed and Shapiro (2008); Bastin et al. (2006); Robinson (1996); Rubinstein and Shapiro (1993); Shapiro (2003); Shapiro et al. (2014). It is widely used and has been studied at length for solving various types of stochastic optimization problems. A common simple setting is a stochastic programming problem of the form

$$(\mathbf{P2}) \qquad \min_{x\in X} \{f(x) := \mathbb{E}_\omega[F(x,\omega)]\} \tag{3}$$

where $F(x,\omega)$ is a random variable defined over a probability space $(\Omega,\mathcal{F},\mathbb{P})$, the expectation over ω is with respect to the measure $\mathbb{P}$, and X is a set of admissible decisions, often a subset of $\mathbb{R}^n$. The corresponding SAA program is

$$\min_{x\in X} \left\{ \hat{f}_N(x) := \frac{1}{N}\sum_{i=1}^{N} F(x,\omega_i) \right\} \tag{4}$$

where $\omega_1,\dots,\omega_N$ is an independent random sample from $\mathbb{P}$. This independence assumption is relaxed in some papers (not here), e.g., to allow randomized quasi-Monte Carlo sampling (Kim et al., 2015). We refer to (3) and (4) as the *true* and SAA problems, respectively. An optimal solution $\hat{x}_N \in \arg\min_{x\in X} \hat{f}_N(x)$ for (4) and the corresponding optimal value $\hat{v}_N = \hat{f}_N(\hat{x}_N)$ are approximations of an optimal solution x^* and of the optimal value v^* for the true problem (3). Typically, one has $\mathbb{E}[\hat{v}_N] < v^*$; see Shapiro (2003). Another important quantity (perhaps the most relevant) is $f(\hat{x}_N)$, the exact value of a solution $\hat{x}_N$ obtained from the SAA. The difference $f(\hat{x}_N) - v^* \geq 0$ represents the gap between the value of the retained solution and the optimal value. In general there could be multiple optimal solutions x^* and $\hat{x}_N$. We denote by X^* and X_N^* the sets of optimal solutions to (3) and (4), respectively. In the following, x^* and $\hat{x}_N$ denote any of those solutions, in the respective sets. We assume that X^* is not empty and that a finite minimum is attained.

In settings where the space X of solutions is infinite (which is not the case for our problem (P1)), it is typically assumed that X has a norm $\|\cdot\|$ (e.g., the Euclidean norm if X is in the real space), so that the distance between two solutions is well defined, and then one can define the distance from a given solution x to optimality as $\mathrm{dist}(x,X^*) = \inf_{x*\in X^*}\|x-x^*\|$. .

Convergence to zero with probability one (w.p.1) for the three error measures $\mathrm{dist}(\hat{x}_N,X^*)$, $f(\hat{x}_N)-v^*$, and $\hat{v}_N - v^*$ when the sample size $N\to\infty$ has been proved under different sets of (mild) conditions; see Dupacová and Wets (1988); Robinson (1996); Shapiro (2003); Shapiro et al. (2014), for instance. This holds for example if X^* is contained in a compact set $C\subset\mathbb{R}^n$, f is bounded and continuous on C, $\sup_{x\in C}|\hat{f}_N(x)-f(x)|\to 0$ when $N\to\infty$, and $\emptyset \neq X_N^* \subset C$ for N large enough, also w.p.1; see (Shapiro, 2003, Theorem 4). There are also other sets of sufficient conditions.

Knowing that we have convergence w.p.1 is good, but knowing how fast it occurs is better. The speed of convergence of $\hat{x}_N$ to X^* can be measured and studied in various ways. Central limit theorems give estimates of order $O_p(N^{-1/2})$ for the three error measures mentioned above when x^* is unique, $X\subset\mathbb{R}^n$ contains a neighborhood of x^*, and $F(\cdot,\omega)$ is a sufficiently smooth function with bounded variance (Shapiro, 1993).

For $\epsilon\geq 0$, a solution $x\in X$ is said to be ϵ-optimal for the true problem if $f(x)\leq v^*+\epsilon$, and ϵ-optimal for the SAA if $\hat{f}_N(x)\leq v_N^*+\epsilon$. Let X^ϵ and X_N^ϵ denote the sets of ϵ-optimal solutions to the true problem and the SAA problem, respectively. Under appropriate conditions, by using large-deviations theory Dai et al. (2000); Kleywegt et al. (2002); Shapiro and de Mello (2000); Shapiro (2003); Kaniovski et al. (1995), one can prove exponential convergence to zero for the probability of selecting a solution with an optimality gap that exceeds a given value. For example, let $F(x,\omega)$ have a finite moment generating function in a neighborhood of 0, and let $\epsilon > \delta > 0$. If X is finite, or if X is a bounded subset of $\mathbb{R}^n$ and f is Lipschitz-continuous over X with Lipschitz constant L, then there are positive constants K and $\eta = \eta(\delta, \epsilon)$ such that

$$\mathbb{P}[X_N^\delta \subseteq X^\epsilon] \geq 1 - K \exp[-\eta N]. \tag{5}$$

In particular, if the true problem has a unique optimal solution x^* and X is finite, then $\mathbb{P}[\hat{x}_N \neq x^*]$ converges to 0 exponentially fast in N. The constant K can be (at worst) proportional to $|X|$ when X is finite and to L otherwise.

Consider now a two-stage problem like (P1), but without the probabilistic constraints (2), and suppose that the second-stage optimization in (1) is easy to solve for any (x, ξ). It could be a deterministic linear program, for example. Then, since $Q(x, \xi)$ can be computed exactly, by taking $F(x, \xi) = f_1(x) + Q(x, \xi)$ we are back to the setting of (P2) and we can apply the corresponding results. See Shapiro (2003); Shapiro et al. (2014) for further discussion.

Another setting studied earlier (e.g., in Vogel (1994) for a general case and in Atlason et al. (2008); Cez̧ik and L'Ecuyer (2008) in the context of call center staffing) is that of an optimization problem with stochastic constraints:

$$\min_{x \in X} f(x) \qquad \text{subject to} \quad g(x) := \mathbb{E}_\omega[G(x, \omega)] \geq 0, \tag{6}$$

where $f(x)$ is easy to evaluate exactly for all $x \in X$, whereas the expectations in the constraints are estimated by Monte Carlo. In the SAA, one replaces $g(x)$ by $\hat{g}_N(x)$, the Monte Carlo average of N i.i.d. samples of $G(x, \omega)$. Under the assumption that X is finite, that $\hat{g}_N(x) \to g(x)$ w.p.1 when $N \to \infty$, and there is $x^* \in X^*$ such that $g(x^*) > 0$, we have w.p.1 that there is $N_0 > 0$ such that $\hat{x}_N \in X^*$ for all $N \geq N_0$. Under the additional assumption that $G(x, \omega)$ satisfies a large-deviation principle, which implies that $\mathbb{P}[|\hat{g}_N(x) - g(x)| > \epsilon] \to 0$ exponentially fast as a function of N for any $\epsilon > 0$, we also have that $\mathbb{P}[\hat{x}_N \notin X^*] \leq K \exp[-\eta N]$ for some constants K and $\eta > 0$, i.e., the probability of not selecting an optimal decision converges to 0 exponentially fast as a function of N. In Atlason et al. (2008); Cez̧ik and L'Ecuyer (2008), the constraints (6) are on QoS measures which are defined as expectations and x represents a staffing decision (number of agents of each type in each time period). In Avramidis et al. (2010), a similar problem is considered in which x represents the work schedules of all agents.

In this paper we study the convergence of a SAA approximation for the two-stage stochastic program (P1), in which an expectation is estimated by Monte Carlo at each of the two stages. This gives rise to nested (or embedded) Monte Carlo sampling: for each of the N first-stage realizations of ξ (or scenarios), say $\xi_1, \ldots, \xi_N$, we must sample several (say $M_n = M_n(\xi_n)$ for scenario n) second-stage realizations of w to estimate the expectations in the second-stage constraints, because the distribution of G in the second stage depends on ξ. The SAA counterpart of (P1) can be written as

$$(\mathbf{P3}) \qquad \begin{cases} \displaystyle\min_{x \in X} & \hat{f}_N(x) = f_1(x) + \dfrac{1}{N}\displaystyle\sum_{n=1}^{N} \hat{Q}_{M_n}(x, \xi_n) \\ \text{where} & \hat{Q}_{M_n}(x, \xi_n) = \displaystyle\min_{y_n \in A(x, \xi_n)} f_2(x, \xi_n, y_n) \\ & \qquad\qquad \text{subject to} \quad \hat{g}_{M_n}(x, \xi_n, y_n) \geq 0, \end{cases} \tag{7}$$

where $\{\xi_1, \ldots, \xi_N\}$ are i.i.d realizations of ξ and for each n,

$$\hat{g}_{M_n}(x, \xi_n, y_n) := \frac{1}{M_n} \sum_{m=1}^{M_n} G(x, \xi_n, y_n, w_{n,m}),$$

and $\{w_{n,1}, \ldots, w_{n,M_n}\}$ are i.i.d realizations of w. The latter can be independent across values of n, i.e., $\sum_{n=1}^N M_n$ independent realizations of w, or they can be dependent. In particular, one could have $M_n = M$ for all n and $w_{1,m} = \cdots = w_{N,m}$ for all m.

To the best of our knowledge, convergence of the SAA approach has not been studied for this setting. Under appropriate conditions, we prove that w.p.1, the optimal decisions for the SAA converge to the optimal decisions for the true problem when N and the M_n increase toward infinity, in the sense that there are constants N_0 and M_0 such that if $N \geq N_0$ and $\min(M_1, \ldots, M_N) \geq M_0$, the optimal decision at the first stage is the same for the SAA and the true problem. Moreover, for almost all $\xi \in \Xi$, w.p.1 there is an $M_0 = M_0(\xi)$ such that for $M \geq M_0$, the optimal decision at the second stage is the same for the SAA and the true problem. The issue of exponential convergence to 1 of the probability of making an optimal decision is more tricky in our setting than in Problem (P2). We show that this exponential convergence holds at the second stage conditionally on ξ, for almost any fixed ξ, but it does not hold for the unconditional probability. This is related to the fact that the $M_0(\xi)$ in the convergence w.p.1 is not uniformly bounded in ξ in general.

The rest of the paper is organized as follows. In Section 2 we state our results on the consistency of SAA when N and the M_n go to infinity together. In Section 3 we establish the convergence rates of the SAA solutions and optimal values, with respect to N and the M_n. Section 4 illustrates the application of this two-stage SAA approach for solving a staffing optimization application in a call center. Section 5 provides a conclusion.

2 Consistency of the SAA estimators

Let X^* and X_N^* denote the sets of first-stage optimal solutions for the true and SAA problem, respectively. Let v^* and $\hat{v}_N$ be the optimal values for the true and SAA counterpart problems. We also denote by $Y^*(x, \xi)$ the set of optimal solutions for the true second-stage problem given (x, ξ), while $Y_M^*(x, \xi)$ denote its SAA counterparts when using sample size M at the second stage. For $k = 1, \ldots, K$, let $g_k(\cdot)$ and $\hat{g}_{kM}(\cdot)$ denote the k-th elements of $g(\cdot)$ and $\hat{g}_M(\cdot)$ in (7) , respectively.

We first assume that the recourse is relatively complete (see for instance Birge and Louveaux (2011)). Along with the assumption that Y is finite, this implies that the recourse program has at least one optimal solution for every x and $\mathbb{P}$-almost every ξ. Moreover, we assume that the second-stage objective function is almost surely uniformly bounded.

Assumption 1 *X and Y are finite, and for each $x \in X$ and $\mathbb{P}$-almost every $\xi \in \Xi$, $Y(x, \xi) \neq \emptyset$. Moreover, f_2 is bounded uniformly for $\mathbb{P}$-almost every $(x, \xi) \in X \times \Xi$.*

We next assume that for $\mathbb{P}$-almost every scenario ξ, the SAA of the second-stage constraint asymptotically coincide with the true second-stage constraint, and that the true constraint is not active at any true second-stage solution , as otherwise, the SAA constraint could be violated at this solution with a strictly positive probability, for any arbitrary large second-stage sample. Note that in the continuous case, this assumption could be relaxed by assuming that the true and SAA active sets are the same with probability one when the sample size is large enough Bastin et al. (2006); Shapiro (2003).

Assumption 2 *For all $x \in X$ and $\mathbb{P}$-almost all ξ, for all $y \in Y$, $\hat{g}_M(x, \xi, y) \to g(x, \xi, y)$ w.p.1 when $M \to \infty$, and there exists $y \in Y^*(x, \xi)$ such that $g(x, \xi, y) \neq 0$.*

Under Assumption 2, we can apply the known results for the Problem (P2) to the second stage of our problem (P1), to obtain the following proposition, whose proof can be found in Atlason et al. (2004), Atlason et al. (2008).

Proposition 1 *Under Assumptions 1 and 2, and there exists $y \in Y^*(x, \xi)$ such that $g(x, \xi, y) \neq 0$, which occurs for $\mathbb{P}$-almost any ξ, w.p.1 there is a finite $M_0 = M_0(\xi)$ such that for all $M \geq M_0$, $\emptyset \neq Y_M^*(x, \xi) \subseteq Y^*(x, \xi)$ and $\hat{Q}_M(x, \xi) = Q(x, \xi)$. That is, for all $M \geq M_0$, the SAA in the second-stage has at least one optimal solution and any such optimal solution is optimal for the true second-stage problem.*

Moreover, again if there exists $y \in Y^*(x,\xi)$ such that $g(x,\xi,y) \neq 0$, there are positive constants C and $b(\xi)$ such that

$$\mathbb{P}\left[Y_M^*(x,\xi) \subseteq Y^*(x,\xi)\right] \geq 1 - C\exp[-b(\xi)M]. \tag{8}$$

That is, for $\mathbb{P}$-almost any ξ, the probability of missing optimality at the second stage decreases to zero exponentially in M.

It is important to note here is that the sample size M_0 and the constant b in Proposition 1 depend on ξ, and there may be no M_0 and b for which the result holds uniformly in ξ. We give an example of that in the following.

Example 1 Consider the following example of a two-stage program

$$\begin{aligned} \min_{x \in X} \quad & f(x) = x + \mathbb{E}_\xi[Q(x,\xi)] \\ \text{where} \quad & Q(x,\xi) = \min_{y \in Y} \; 2y \\ & \qquad \text{subject to} \;\; \mathbb{E}_w[x + y - 2\xi - w] \geq 0, \end{aligned}$$

where $\xi \sim U(0,1)$ (the uniform distribution), $w \sim \mathcal{N}(0,1)$ (the standard normal distribution), and $X = Y = \{0,1,2\}$. Given $x \in X$, the set of optimal solutions in the second-stage is

$$Y^*(x,\xi) = \arg\min\{2y \mid y \in Y, y \geq 2\xi - x\}.$$

Now, consider the SAA counterpart

$$\begin{aligned} \min_{x \in X} \quad & \hat{f}_N(x) = x + \frac{1}{N}\sum_{n=1}^{N} Q_M(x,\xi_n) \\ \text{where} \quad & Q_M(x,\xi) = \min_{y \in Y} \; 2y \\ & \qquad \text{subject to} \;\; x + y - 2\xi - \hat{w}_M \geq 0, \end{aligned}$$

where $\hat{w}_M$ is a sample average approximation of w by a Monte Carlo method. In this example, for notational simplicity we set $M_1 = \ldots = M_N = M$. Let $x = 1$, we have $Y^*(1,\xi) = \{0\}$ if $\xi \leq 1/2$, and $Y^*(1,\xi) = \{1\}$ if $\xi > 1/2$. So, for a given $\xi \in [0,1/2]$, if we have $\hat{w}_M > 1 - 2\xi$ in the second-stage of the SAA, then the SAA does not return a true second-stage optimal solution, i.e, $Y_M^*(x,\xi) \nsubseteq Y^*(x,\xi)$. Therefore, we have

$$\mathbb{P}\left[Y_M^*(x,\xi) \nsubseteq Y^*(x,\xi)\right] \geq \mathbb{P}\left[\hat{w}_M \geq 1 - 2\xi\right]. \tag{9}$$

Since $\hat{w}_M \sim \mathcal{N}(0,1/M)$, for any $M > 0$ we have

$$\lim_{1-2\xi \to 0} \mathbb{P}\left[\hat{w}_M \geq 1 - 2\xi\right] = \mathbb{P}\left[\hat{w}_M \geq 0\right] = \frac{1}{2}. \tag{10}$$

Hence, if $1 - 2\xi$ can be arbitrarily close to zero, for any given $0 \leq \epsilon < 1/4$, then there is no $M_0 > 0$ such that $\mathbb{P}\left[\hat{w}_M \geq 1 - 2\xi\right] < \epsilon$ for all $M > M_0$ and all $\xi \in [0,1/2)$, and therefore, there is no $M_0 > 0$ such that $\mathbb{P}\left[Y_M^*(x,\xi) \nsubseteq Y^*(x,\xi)\right] < \epsilon$ for all $M > M_0$ and all $\xi \in [0,1/2)$. This also means that there is no M_0 such that, w.p.1, $\hat{Q}_M(x,\xi) = Q(x,\xi)$ for all $M > M_0$ and all $\xi \in [0,1/2)$.

We now show that exponential convergence of the probability of making a wrong decision at the second stage does not hold uniformly in ξ. By contradiction, if there are positive constants C_0, b_0 for which the exponential convergence Proposition 1 holds uniformly in ξ, then for $\mathbb{P}$-almost every $\xi \in \Xi$, we have

$$\ln\left(\mathbb{P}\left[Y_M^*(x,\xi) \nsubseteq Y^*(x,\xi)\right]\right) \leq \ln C_0 - Mb_0, \text{ for all } M > 0. \tag{11}$$

From (9) we have, for $\mathbb{P}$-almost every $\xi \in [0,1/2)$

$$\frac{\ln \mathbb{P}\left[\hat{w}_M \geq 1 - 2\xi\right]}{M} \leq \frac{\ln C_0}{M} - b_0. \tag{12}$$

However, we can always choose M^* large enough such that

$$\frac{\ln(1/4) - \ln C_0}{M^*} > -b_0,$$

and $\xi^* \in [0, 1/2)$ such that $\mathbb{P}[\hat{w}_{M^*} \geq 1 - 2\xi^*] > 1/4$. The latter can be done using (10). Then, we have

$$\frac{\ln \mathbb{P}[\hat{w}_{M^*} \geq 1 - 2\xi^*]}{M^*} - \frac{\ln C_0}{M^*} > \frac{\ln(1/4) - \ln C_0}{M^*} > -b_0,$$

meaning that (12) cannot hold for any $M > 0$ and for almost every $\xi \in [0, 1/2)$.

We now look at the convergence of the optimal value and optimal solution at the first stage of the SAA problem to those of the true problem. We want to show that w.p.1, we have $X_N^* \subseteq X^*$ when $\min(N, M_1, \ldots, M_N)$ is large enough. Since X is finite, there is a fixed $\delta > 0$ such that for every $x \in X \setminus X^*$, $f(x) - v^* \geq \delta$. Then, a sufficient condition for $X_N^* \subseteq X^*$ is that $|\hat{f}_N(x) - f(x)| < \epsilon := \delta/2$ for all $x \in X$. One could think that this last inequality would follow from the observation that since for each ξ_n, $\hat{Q}_{M_n}(x, \xi_n)$ converges to its expectation w.p.1 when $M_n \to \infty$, $|\hat{f}_N(x) - f(x)|$ should converge to 0 w.p.1, so it will eventually be smaller than ϵ. But this simple argument does not really stand (it is not rigorous), because the convergence is not uniform in ξ, so the required M_0 above which $|\hat{f}_N(x) - f(x)| < \epsilon$ when $N > N_0$ and $\min(M_1, \ldots, M_N) > M_0$ may increase without bound when N increases. A more careful argument is needed and this is what we will do now, under our two assumptions. We first introduce some notations, then prove two lemmas which will be used to prove Theorems 1 and 2, which are our main results in this section.

For any $x \in X$ and $\xi \in \Xi$, we define

$$\begin{aligned}
Y_-(x, \xi) &= \{y \in A(x, \xi) \mid \exists\, k \text{ such that } g_k(x, \xi, y) < 0)\}, \\
\bar{\delta}(x, \xi) &= \frac{1}{2} \max_{y \in Y(x,\xi), 1 \leq k \leq K} \{g_k(x, \xi, y) \mid g_k(x, \xi, y) < 0\}, \\
\underline{\delta}(x, \xi) &= \min_{y \in Y^*(x,\xi), 1 \leq k \leq K} \{g_k(x, \xi, y) \mid g_k(x, \xi, y) > 0\}, \\
\delta(x, \xi) &= \min\{-\bar{\delta}(x, \xi), \underline{\delta}(x, \xi)\} > 0, \text{ and} \\
\delta(\xi) &= \min_{x \in X} \delta(x, \xi) > 0.
\end{aligned} \tag{13}$$

By convention, if $Y_-(x, \xi) = \emptyset$ then $\bar{\delta}(x, \xi) = -\infty$, and if $\{(y, k) |\ y \in Y^*(x, \xi),\ g_k(x, \xi, y) > 0\} = \emptyset$, then $\underline{\delta}(x, \xi) = \infty$. Under Assumption 2 we have $\underline{\delta}(x, \xi) < \infty$ for $\mathbb{P}$-almost every $\xi \in \Xi$.

Lemma 1 $\max_{x \in X} |\hat{f}_N(x) - f(x)| \geq |\hat{v}_N - v^*|$.

Proof. Let x^* and x_N^* be optimal solutions to (P1) and (P3), respectively. If $f(x^*) < \hat{f}_N(x_N^*)$, since $\hat{f}_N(x_N^*) \leq \hat{f}_N(x^*)$, we have:

$$|\hat{v}_N - v^*| = |\hat{f}_N(x_N^*) - f(x^*)| \leq |\hat{f}_N(x^*) - f(x^*)| \leq \max_{x \in X} |\hat{f}_N(x) - f(x)|.$$

If $f(x^*) \geq \hat{f}_N(x_N^*)$, since $f(x^*) \leq f(x_N^*)$, we have:

$$|v^* - \hat{v}_N| = |f(x^*) - \hat{f}_N(x_N^*)| \leq |f(x_N^*) - \hat{f}_N(x_N^*)| \leq \max_{x \in X} |\hat{f}_N(x) - f(x)|.$$

In both cases, we have $|\hat{v}_N - v^*| \leq \max_{x \in X} |\hat{f}_N(x) - f(x)|$. □

Lemma 2 *Under Assumptions 1, and 2, for any $x \in X$ and for $\mathbb{P}$-almost every $\xi \in \Xi$, if $|\hat{g}_{kM}(x, \xi, y) - g_k(x, \xi, y)| \leq \delta(x, \xi)$ for all $y \in Y(x, \xi)$ and $k = 1, \ldots, K$, then $\emptyset \neq Y_M^*(x, \xi) \subseteq Y^*(x, \xi)$.*

Proof. Let $Y_M(x,\xi)$ be the set of feasible solutions of the SAA counterpart second-stage problems. Given ξ such that $\underline{\delta}(x,\xi)< \infty$, which holds for $\mathbb{P}$-almost every $\xi \in \Xi$, we have

$$|\hat{g}_{kM}(x,\xi,y) - g_k(x,\xi,y)| \leq \delta(x,\xi) = \min\{-\bar{\delta}(x,\xi), \underline{\delta}(x,\xi)\}.$$

If $y \in Y_-(x,\xi)$, there exists some k such that $g_k(x,\xi,y) < 0$ and

$$\hat{g}_{kM}(x,\xi,y) \leq g_k(x,\xi,y) - \bar{\delta}(x,\xi) < 0.$$

Thus $y \in A(x,\xi)\backslash Y_M(x,\xi)$, and $A(x,\xi)\backslash Y(x,\xi) \subseteq A(x,\xi)\backslash Y_M(x,\xi)$. Since $Y_M(x,\xi) \subseteq A(x,\xi)$, we have $Y_M(x,\xi) \subseteq Y(x,\xi)$. Moreover, w.p.1, there exists $y^* \in Y^*(x,\xi)$ such that $g(x,\xi,y^*) > 0$, we have that for all k,

$$\hat{g}_{kM}(x,\xi,y^*) \geq g_k(x,\xi,y^*) - \underline{\delta}(x,\xi) \geq 0,$$

implying $y^* \in Y_M(x,\xi)$. Moreover, for all $y^*_M \in Y^*_M(x,\xi)$, we have $f_2(y^*) \geq f_2(y^*_M)$. As $Y_M(x,\xi) \subseteq Y(x,\xi)$, we also have $f_2(y^*) \leq f_2(y^*_M)$, and therefore $f_2(y^*) = f_2(y^*_M)$, implying that $y^* \in Y^*_M(x,\xi)$, so $Y^*_M(x,\xi) \neq \emptyset$. This also implies that if $y^*_1 \in Y^*_M(x,\xi)$ and $y^*_2 \in Y^*(x,\xi)$, then $f_2(y^*_1,\xi) = f_2(y^*_2,\xi)$. As $Y^*_M(x,\xi) \subseteq Y_M(x,\xi) \subseteq Y(x,\xi)$, we also have $y^*_1 \in Y(x,\xi)$, and therefore $y^*_1 \in Y^*(x,\xi)$. As a consequence, $\emptyset \neq Y^*_M(x,\xi) \subseteq Y^*(x,\xi)$, which completes the proof. □

Theorem 1 *Under Assumptions 1, and 2, for any $\epsilon > 0$, w.p.1, there are integers $N_0 = N_0(\epsilon)$ and $M_0 = M_0(\epsilon)$ such that for all $N \geq N_0$, and $\min(M_1,\ldots,M_N) \geq M_0$, $|\hat{f}_N(x) - f(x)| \leq \epsilon$ for all $x \in X$, and $|\hat{v}_N - v^*| \leq \epsilon$.*

Proof. We need to prove that for a given $\epsilon > 0$, w.p.1, there are $N_0(\epsilon)$, $M_0(\epsilon) > 0$ such that $|\hat{f}_N(x) - f(x)| \leq \epsilon$ for all $N \geq N_0(\epsilon)$, all $M_1,\ldots,M_N$ such that $\min(M_1,\ldots,M_N) \geq M_0(\epsilon)$, and all $x \in X$. To prove this, we bound $|\hat{f}_N(x) - f(x)|$ using a triangle inequality and then bound each term, as follows.

$$\begin{aligned}\left|\hat{f}_N(x) - f(x)\right| &= \left|\frac{1}{N}\sum_{n=1}^N \hat{Q}_{M_n}(x,\xi_n) - \mathbb{E}_\xi[Q(x,\xi)]\right| \\ &\leq \left|\frac{1}{N}\sum_{n=1}^N Q(x,\xi_n) - \mathbb{E}_\xi[Q(x,\xi)]\right| + \left|\frac{1}{N}\sum_{n=1}^N Q(x,\xi_n) - \frac{1}{N}\sum_{n=1}^N \hat{Q}_{M_n}(x,\xi_n)\right|. \end{aligned} \tag{14}$$

To bound the first term in (14), note that under Assumption 1, $Q(x,\xi)$ is uniformly bounded for $\mathbb{P}$-almost every $\xi \in \Xi$, so the expectation of $Q(x,\xi)$ always exists according to the Lebesgue integration. Thus, this part converges to zero when $N \to \infty$ according to the strong law of large numbers, i.e., w.p.1, there exist $N^1_0(x,\epsilon)$ such that for all $N > N^1_0(x,\epsilon)$,

$$\left|\frac{1}{N}\sum_{n=1}^N Q(x,\xi_n) - \mathbb{E}_\xi[Q(x,\xi)]\right| \leq \frac{\epsilon}{2}. \tag{15}$$

Proving the convergence of the second term is more difficult, because $\hat{Q}_{M_n}(x,\xi)$ may not converge to $Q(x,\xi)$ uniformly in ξ. To prove it, we partition the sample space Ξ into four different subsets as follows. We first define $\bar{\Xi} \subseteq \Xi$ as the set of all scenarios such that Assumptions 1 and 2 hold for every $\xi \in \bar{\Xi}$. Assumptions 1 and 2 imply that $\mathbb{P}(\xi \in \bar{\Xi} | \xi \in \Xi) = 1$. We also choose Ξ_1, Ξ_2 and Ξ_3 as three subsets of $\bar{\Xi}$ such that $\delta(\xi)$ is bounded from below by a positive scalar and the convergence of $\hat{g}_M$ to g holds uniformly on Ξ_3, and for which $\mathbb{P}[\xi \in \Xi_1 \cup \Xi_2]$ can be arbitrarily small. We describe how to choose these sets in the following.

Since $\delta(\xi) > 0$ w.p.1, we have

$$\lim_{\pi\to 0} \mathbb{P}_\xi[\delta(\xi) \leq \pi] = 0.$$

Moreover, from Assumption 2, we can always choose a mapping $M_0 : \Xi \times \mathbb{R} \to \mathbb{N}$ such that given $\xi \in \Xi$ and for any $\epsilon > 0$, w.p.1, we have that

$$|\hat{g}_{kM}(x,\xi,y) - g_k(x,\xi,y)| \leq \epsilon, \tag{16}$$

for all $y \in Y(x,\xi)$, all $M > M_0(\xi,\epsilon)$, and $k \in 1,\dots,K$. Note that $M_0(\xi,\epsilon)$ generally depends on ξ and may be unbounded from above, i.e., we may have $\sup_{\xi\in\Xi} M_0(\xi,\epsilon) = \infty$. However, we have

$$\lim_{M\to\infty} \mathbb{P}_\xi[M_0(\xi,\epsilon) \geq M] = 0.$$

So, there exist $\pi(\epsilon) > 0$ and $M_0^1(\epsilon) > 0$ such that

$$\mathbb{P}[\delta(\xi) \leq \pi(\epsilon)] \leq \frac{\epsilon}{6\alpha} \quad \text{and} \quad \mathbb{P}[M_0(\xi,\pi(\epsilon)) \geq M_0^1(\epsilon)] \leq \frac{\epsilon}{6\alpha},$$

where α is a constant chosen such that $\alpha > \sup_{x\in X, y\in Y,\, \xi\in\Xi\setminus\Xi_0} |2f_2(x,\xi,y)|$. We can simply choose $\alpha = \sup_{x\in X, y\in Y,\, \xi\in\Xi\setminus\Xi_0} |2f_2(x,\xi,y)| + 1$. Hence, we always have $\alpha > |\hat{Q}_{M_n}(x,\xi) - Q(x,\xi)|$ for all $x \in X$, $\xi \in \bar{\Xi}$ and all $n = 1,\dots,N$. This α always exists and is finite because f_2 is bounded uniformly for every $\xi \in \bar{\Xi}$. Let us define

$$\begin{aligned} \Xi_1 &= \{\xi \in \bar{\Xi} \,|\, \delta(\xi) \leq \pi(\epsilon)\}, \\ \Xi_2 &= \{\xi \in \bar{\Xi} \,|\, M_0(\xi,\pi(\epsilon)) \geq M_0^1(\epsilon)\}, \\ \Xi_3 &= \bar{\Xi}\setminus(\Xi_1 \cup \Xi_2). \end{aligned}$$

Suppose $\xi_1,\dots,\xi_N \in \bar{\Xi}$, which happens w.p.1. The second part of (14) can then be written as

$$\begin{aligned} &\left|\frac{1}{N}\sum_{n=1}^N Q(x,\xi_n) - \frac{1}{N}\sum_{n=1}^N \hat{Q}_{M_n}(x,\xi_n)\right| \\ &\leq \frac{1}{N}\sum_{n=1}^N \left|Q(x,\xi_n) - \hat{Q}_{M_n}(x,\xi_n)\right| \\ &= \frac{1}{N}\sum_{\xi_n\in\Xi_1\cup\Xi_2} \left|Q(x,\xi_n) - \hat{Q}_{M_n}(x,\xi_n)\right| + \frac{1}{N}\sum_{\xi_n\in\Xi_3} \left|Q(x,\xi_n) - \hat{Q}_{M_n}(x,\xi_n)\right| \\ &\leq \frac{1}{N}\sum_{n=1}^N \alpha\mathbb{I}[\xi_n \in \Xi_1\cup\Xi_2] + \frac{1}{N}\sum_{\xi_n\in\Xi_3} \left|Q(x,\xi_n) - \hat{Q}_{M_n}(x,\xi_n)\right|. \end{aligned} \tag{17}$$

The term $\frac{1}{N}\sum_{n=1}^N \mathbb{I}[\xi_n \in \Xi_1\cup\Xi_2]$ is a sample average of $\mathbb{P}[\xi_n \in \Xi_1\cup\Xi_2]$. Therefore, based on the strong law of large numbers, w.p.1, there is $N_0^2(x,\epsilon)$ such that, for all $N \geq N_0^2(x,\epsilon)$

$$\begin{aligned} \frac{1}{N}\sum_{n=1}^N \mathbb{I}[\xi_n \in \Xi_1\cup\Xi_2] &\leq \mathbb{P}[\xi_n \in \Xi_1\cup\Xi_2] + \frac{\epsilon}{6\alpha} \\ &\leq \mathbb{P}[\xi_n \in \Xi_1] + \mathbb{P}[\xi_n \in \Xi_2] + \frac{\epsilon}{6\alpha} \\ &\leq \frac{\epsilon}{6\alpha} + \frac{\epsilon}{6\alpha} + \frac{\epsilon}{6\alpha} = \frac{\epsilon}{2\alpha}. \end{aligned} \tag{18}$$

Moreover, as $\Xi_3 = \{\xi \mid \delta(\xi) > \pi(\epsilon),\, M_0(\xi,\pi(\epsilon)) < M_0^1(\epsilon)\}$, then for any $\xi \in \Xi_3$, w.p.1, we have $|\hat{g}_{kM}(x,\xi,y) - g_k(x,\xi,y)| \leq \pi(\epsilon) < \delta(\xi)$ for all $y \in Y(x,\xi)$, all $M > M_0^1(\epsilon)$, and $k = 1,\dots,K$. So, for any $\xi \in \Xi_3$, w.p.1, $\hat{Q}_M(x,\xi) = Q(x,\xi)$ for all $M > M_0^1(\epsilon)$, or equivalently, w.p.1, for all $M_n > M_0^1(\epsilon)$, $n = 1,\dots,N$, we have

$$\frac{1}{N}\sum_{\{n|\xi_n\in\Xi_3\}} \left|Q(x,\xi_n) - \hat{Q}_{M_n}(x,\xi_n)\right| = 0 \tag{19}$$

Combining (14), (17), (18) and (19) we have, w.p.1, for all $x \in X$, all $N > N_0(\epsilon)$ and $\min\{M_1,\dots,M_N\} > M_0(\epsilon)$,

$$\left|\hat{f}_N(x) - f(x)\right| \leq \epsilon, \tag{20}$$

where $N_0(\epsilon) = \max\{N_0^1(\epsilon), N_0^2(\epsilon)\}$, and $M_0(\epsilon) = M_0^1(\epsilon)$. By combining this with Lemma 1, we obtain that w.p.1, $|\hat{v}_N - v^*| \leq \epsilon$, for all $N > N_0(\epsilon)$ and $\min\{M_1,\dots,M_N\} > M_0(\epsilon)$. □

The next theorem concerns the consistency of the SAA counterpart in terms of first-stage optimal solutions. We show that when the sample sizes are large enough, w.p.1, we can retrieve the true optimal solutions by solving the SAA problem.

Theorem 2 *Under Assumptions 1 and 2, w.p.1, there are integers N_0 and M_0 such that for all $N \geq N_0$, and $\min(M_1, \ldots, M_N) \geq M_0$, $X_N^* \subseteq X^*$.*

Proof. For each $x \in X$ and $x \notin X^*$, we have $f(x) > v^*$, and since X is finite, there exists some $\delta > 0$ such that

$$|f(x) - v^*| > \eta \quad \text{for all } x \in X \backslash X^*.$$

In other words, if $|f(x) - v^*| \leq \eta$, then $x \in X^*$. Now, given $\hat{x}_N \in X_N^*$ we have

$$|f(\hat{x}_N) - v^*| \leq |f(\hat{x}_N) - \hat{f}_N(\hat{x}_N)| + |\hat{f}_N(\hat{x}_N) - v^*|. \tag{21}$$

From Theorem 1, w.p.1, there exist $N_0(\eta)$ and $M_0(\eta) > 0$ such that for all $N \geq N_0(\eta)$, $M_n \geq M_0(\eta)$ for all $n = 1, \ldots, N$,

$$|f(\hat{x}_N) - \hat{f}_N(\hat{x}_N)| \leq \eta/2 \quad \text{and} \quad |\hat{f}_N(\hat{x}_N) - v^*| \leq \eta/2.$$

Thus, w.p.1, there are $N_0, M_0 > 0$ such that for all $N \geq N_0$ and $M_n \geq M_0$, $n = 1, \ldots N$, we have $|f(\hat{x}_N) - v^*| \leq \eta$ and $X_N^* \subseteq X^*$. □

In summary, we have shown that in the first stage, w.p.1, the optimal decision in the SAA becomes equal to that of the true problem when the number of scenarios and the sample size for each SAA second-stage constraint are large enough. Moreover, for any fixed ξ, we can obtain an optimal solution of the corresponding second stage problem by solving its SAA with large enough sample size.

3 Convergence of large-deviation probabilities

In this section, we establish large-deviation principles for the optimal value $\hat{v}_N$ of the SAA, for the true value $f(\hat{x}_N)$ of an optimal solution $\hat{x}_N$ of the SAA, and for the probability that any optimal solution to the SAA is an optimal solution of the true problem. That is, we show that for any $\epsilon > 0$, $\mathbb{P}[|\hat{v}_N - v^*| \leq \epsilon]$, $\mathbb{P}[|f(\hat{x}_N) - v^*| \leq \epsilon]$, and $\mathbb{P}[\emptyset \neq X_N^* \subseteq X^*]$ all converge to 1 exponentially fast when N and the M_n go to ∞. Recall that in Proposition 1 and Example 1, we showed that in the second-stage problem, the probability that a SAA second-stage solution is truly optimal approaches one exponentially fast for any given ξ, but this exponential convergence may not hold uniformly in ξ. For this reason, it is difficult to establish the exponential convergence of $\mathbb{P}[X_N^* \subseteq X^*]$ when N and the M_n go to infinity.

A standard large-deviation result is that if $Z_1, \ldots, Z_M$ are i.i.d replicates of a random variable Z of mean μ and variance $\sigma^2 > 0$ and whose moment generating function is finite in a neighborhood of zero, then for any $\epsilon > 0$ we have (Stroock, 1984; Shapiro, 2003):

$$\mathbb{P}[\hat{Z}_M - \mu > \epsilon] \leq \exp\left(\frac{-M\epsilon^2}{2\sigma^2}\right) \quad \text{and} \quad \mathbb{P}[\hat{Z}_M - \mu < -\epsilon] \leq \exp\left(\frac{-M\epsilon^2}{2\sigma^2}\right). \tag{22}$$

When Z is bounded, as is the case for $Z = Q(x, \xi)$ or Z is given by an indicator function in our setting, its moment generating function is always finite, and we can simply use Hoeffding's equality (Hoeffding, 1963) to establish large-deviation results. We need the following assumption for G.

Assumption 3 *For $\mathbb{P}$-almost every $\xi \in \Xi$, for all $x \in X$ and $y \in Y$, the moment-generating function of $G(x, \xi, y, w)$, i.e. $\mathbb{E}_w\left[\exp\left(tG(x, \xi, y, w)\right)\right]$, is bounded in a neighborhood of $t = 0$.*

The next assumption concerns a finite covering property of the support set Ξ with respect to the function $G_k(x, \xi, y, w)$, given $x \in X$, $y \in Y$ and $w \in \mathcal{W}$. In other words, we require that it is possible to cover the infinite set Ξ by a finite number of subsets such that in each subset, the variation of $G_k(x,\xi,y,w)$, with respect to ξ, is bounded by the size of the subset multiplied by a random variable having a finite moment-generating function. Such an assumption is often made in the stochastic programming literature to establish convergence results with continuous variables (Shapiro et al., 2014; Kim et al., 2015). In our context, the decision variables x and y are discrete, but we need this assumption because the stochastic functions $G(\cdot)$ also depend on ξ whose support may be infinite. In particular, a finite covering property holds if Ξ is compact and $G_k(x,\xi,y,w)$ is Lipschitz continuous in ξ. We introduce the following assumption under a general setting.

Assumption 4 *There is a measurable function $\kappa : \mathcal{W} \to \mathbb{R}^+$ with bounded moment-generating function in a neighborhood of* 0 *such that for any $\upsilon > 0$, there are $H = H(\upsilon) < \infty$ non-empty sets $\Xi^1, \ldots, \Xi^H$ covering Ξ, i.e., $\Xi \subset \bigcup_{h=1}^H \Xi^h$, such that for any $h \in \{1,\ldots,H\}$ and $\mathbb{P}$-almost every $\xi_1,\ \xi_2 \in \Xi^h$, we have*

$$|G_k(x,\xi_2,y,w) - G_k(x,\xi_1,y,w)| \le \kappa(w)\upsilon, \quad \forall x \in X,\ \forall y \in Y,\ k = 1,\ldots,K.$$

It is also convenient in our proofs to assume that the number of distinct values in $\{M_1,\ldots,M_N\}$ is bounded uniformly in N. This is not really restrictive in practice and will permit us to remove the dependence on N when using the finite coverage Assumption 4 to establish an upper bound on the probability

$$\mathbb{P}\left[\left|\frac{1}{N}\sum_{n=1}^N \hat{Q}_{M_n}(x,\xi_n) - Q(x,\xi_n)\right| > \epsilon\right]$$

for large N. Without Assumptions 4 and 5, we are still able to establish "weaker" large-deviation results; see Theorem 4.

Assumption 5 *The number of distinct values in $\{M_1,\ldots,M_N\}$ is bounded uniformly in N.*

We are now in a position to provide large-deviation bounds for the optimal value of the SAA problem and for the true value of an optimal solution to the SAA.

Theorem 3 *Suppose Assumptions 1 to 5 hold. Then for any $\epsilon > 0$, there exist positive constants C_1, C_2, $b_1(\epsilon)$, and $b_2(\epsilon)$ that do not depend on N and the M_n, $n = 1,\ldots,N$, such that*

$$\begin{aligned}\mathbb{P}\left[|\hat{v}_N - v^*| > \epsilon\right] &\le C_1\exp[-b_1(\epsilon)N] + C_2\exp[-b_2(\epsilon)\overline{M}] \quad \text{and}\\ \mathbb{P}\left[|f(\hat{x}_N) - v^*| > \epsilon\right] &\le C_1\exp[-b_1(\epsilon)N] + C_2\exp[-b_2(\epsilon)\overline{M}],\end{aligned}$$

where $\hat{x}_N$ is an arbitrary optimal solution to the SAA problem and $\overline{M} = \min_{n=1,\ldots,N} M_n$.

Proof. We use again the triangle inequality in (14). For any $\epsilon > 0$, we have

$$\begin{aligned}
&\mathbb{P}\left[\max_{x\in X}\left|\hat{f}_N(x) - f(x)\right| > \epsilon\right] = P\left[\max_{x\in X}\left|\frac{1}{N}\sum_{n=1}^N \hat{Q}_{M_n}(x,\xi_n) - \mathbb{E}_\xi[Q(x,\xi)]\right| > \epsilon\right]\\
&\le \mathbb{P}\left[\max_{x\in X}\left|\frac{1}{N}\sum_{n=1}^N \hat{Q}_{M_n}(x,\xi_n) - \frac{1}{N}\sum_{n=1}^N Q(x,\xi_n)\right| + \max_{x\in X}\left|\frac{1}{N}\sum_{n=1}^N Q(x,\xi_n) - \mathbb{E}_\xi[Q(x,\xi)]\right| > \epsilon\right]\\
&\le \mathbb{P}\left[\left(\max_{x\in X}\left|\frac{1}{N}\sum_{n=1}^N\left(\hat{Q}_{M_n}(x,\xi_n) - Q(x,\xi_n)\right)\right| > \frac{\epsilon}{2}\right)\bigcup\left(\max_{x\in X}\left|\frac{1}{N}\sum_{n=1}^N Q(x,\xi_n) - \mathbb{E}_\xi[Q(x,\xi)]\right| > \frac{\epsilon}{2}\right)\right]\\
&\le \mathbb{P}\left[\max_{x\in X}\left|\frac{1}{N}\sum_{n=1}^N\left(\hat{Q}_{M_n}(x,\xi_n) - Q(x,\xi_n)\right)\right| > \frac{\epsilon}{2}\right] + \mathbb{P}\left[\max_{x\in X}\left|\frac{1}{N}\sum_{n=1}^N Q(x,\xi_n) - \mathbb{E}_\xi[Q(x,\xi)]\right| > \frac{\epsilon}{2}\right]\\
&\le \sum_{x\in X}\left(\mathbb{P}\left[\left|\frac{1}{N}\sum_{n=1}^N\left(\hat{Q}_{M_n}(x,\xi_n) - Q(x,\xi_n)\right)\right| > \frac{\epsilon}{2}\right] + \mathbb{P}\left[\left|\frac{1}{N}\sum_{n=1}^N Q(x,\xi_n) - \mathbb{E}_\xi[Q(x,\xi)]\right| > \frac{\epsilon}{2}\right]\right). \qquad (23)
\end{aligned}$$

Considering the second part of (23) and given the fact that $Q(x,\xi)$ is bounded by the interval $[-\alpha,\alpha]$ for $\mathbb{P}$-almost every ξ, where α is defined as in the proof of Theorem 1, we obtain the following from Hoeffding's inequality (Hoeffding, 1963):

$$\mathbb{P}\left[\left|\frac{1}{N}\sum_{n=1}^{N} Q(x,\xi_n) - \mathbb{E}_\xi[Q(x,\xi)]\right| > \frac{\epsilon}{2}\right] \le 2\exp\left(\frac{-N\epsilon^2}{8\alpha^2}\right). \tag{24}$$

As discussed earlier, the convergence in probability of $\hat{Q}_M(x,\xi) \to Q(x,\xi)$ does not hold uniformly on Ξ. To deal with this issue, similar to the proof of Theorem 1, we divide the support set Ξ into smaller sub-sets. First, we define $\bar{\Xi} \subseteq \Xi$ as the set of all scenarios $\xi \in \Xi$ for which Assumptions 1, 2 and 3 hold. Note that $\mathbb{P}[\xi \in \bar{\Xi}] = 1$. We select $\pi(\epsilon) > 0$ such that

$$\mathbb{P}_\xi[\delta(\xi) \le \pi(\epsilon)] \le \frac{\epsilon}{6\alpha},$$

where $\delta(\xi)$ is defined in (13). Let also define $\Xi_1 = \{\xi \in \bar{\Xi} | \delta(\xi) \le \pi(\epsilon)\}$, and $\Xi_2 = \bar{\Xi}\backslash\Xi_1)$. We write the first part of (23) as

$$\begin{aligned}
&\mathbb{P}\left[\left|\frac{1}{N}\sum_{n=1}^{N} \hat{Q}_{M_n}(x,\xi_n) - \frac{1}{N}\sum_{n=1}^{N} Q(x,\xi_n)\right| > \frac{\epsilon}{2}\right] \\
&\le \mathbb{P}\left[\frac{1}{N}\sum_{\xi_n \in \Xi_1 \cup \Xi_2} \left|\hat{Q}_{M_n}(x,\xi_n) - Q(x,\xi_n)\right| > \frac{\epsilon}{2}\right] \\
&\le \mathbb{P}\left[\frac{1}{N}\sum_{\xi_n \in \Xi_1} \left|\hat{Q}_{M_n}(x,\xi_n) - Q(x,\xi_n)\right| > \frac{\epsilon}{4}\right] + \mathbb{P}\left[\frac{1}{N}\sum_{\xi_n \in \Xi_2} \left|\hat{Q}_{M_n}(x,\xi_n) - Q(x,\xi_n)\right| > \frac{\epsilon}{4}\right] \\
&\le \mathbb{P}\left[\frac{1}{N}\sum_{n=1}^{N} \alpha\mathbb{I}[\xi_n \in \Xi_1] > \frac{\epsilon}{4}\right] + \mathbb{P}\left[\frac{1}{N}\sum_{\xi_n \in \Xi_2} \left|\hat{Q}_{M_n}(x,\xi_n) - Q(x,\xi_n)\right| > \frac{\epsilon}{4}\right].
\end{aligned} \tag{25}$$

The first term in (25) concerns a sample average approximation of $\alpha\mathbb{P}[\xi \in \Xi_1]$, and we have $\alpha\mathbb{P}[\xi \in \Xi_1] \le \epsilon/6 < \epsilon/4$. Moreover, $\mathbb{I}[\xi \in \Xi_1]$ only takes values in $\{0,1\}$, so by Hoeffding's inequality we have

$$\mathbb{P}\left[\frac{1}{N}\sum_{n=1}^{N} \mathbb{I}[\xi_n \in \Xi_1] > \frac{\epsilon}{4\alpha}\right] \le \exp\left(\frac{-N\epsilon^2}{72\alpha^2}\right). \tag{26}$$

For the second term of (25), we have

$$\begin{aligned}
&\mathbb{P}\left[\frac{1}{N}\sum_{\xi_n \in \Xi_2} \left|\hat{Q}_{M_n}(x,\xi_n) - Q(x,\xi_n)\right| \le \frac{\epsilon}{4}\right] \\
&\ge \mathbb{P}\left[\left|\hat{Q}_{M_n}(x,\xi_n) - Q(x,\xi_n)\right| = 0,\ \forall \xi_n \in \Xi_2,\ n = 1,\dots,N\right] \\
&\ge \mathbb{P}\left[\left|\hat{g}_{kM(\xi)}(x,\xi,y) - g_k(x,\xi,y)\right| \le \delta(\xi),\ \forall \xi \in \Xi_2,\ \forall y \in Y,\ k = 1,\dots,K\right] \\
&\ge \mathbb{P}\left[\left|\hat{g}_{kM(\xi)}(x,\xi,y) - g_k(x,\xi,y)\right| \le \pi(\epsilon),\ \forall \xi \in \Xi_2,\ \forall y \in Y,\ k = 1,\dots,K\right],
\end{aligned}$$

where $M(\xi)$ is a mapping from Ξ to $\mathbb{N}^+$ such that $M(\xi_n) = M_n$, $n = 1,\dots,N$, and we assume that $M(\xi) = \overline{M}$ for all $\xi \ne \xi_n$, $n = 1,\dots,N$. Moreover, as the number of distinct values in $\{M_1,\dots,M_N\}$ is bounded uniformly, there exists $T \in \mathbb{N}^+$ that is independent of N and T values $\{\mathcal{M}_1,\dots,\mathcal{M}_T\}$ such that $M(\xi) \in \{\mathcal{M}_1,\dots,\mathcal{M}_T\}$ for all $\xi \in \Xi$. Hence, we have

$$
\begin{aligned}
&\mathbb{P}\left[\frac{1}{N}\sum_{\xi_n\in\Xi_2}\left|\hat{Q}_{M_n}(x,\xi_n)-Q(x,\xi_n)\right|>\frac{\epsilon}{4}\right]\\
&\le \mathbb{P}\left[\exists(\xi,y,k)\middle|\ \xi\in\Xi_2,\ y\in Y,\ k\in\{1,\ldots,K\},\ \left|\hat{g}_{kM(\xi)}(x,\xi,y)-g_k(x,\xi,y)\right|>\pi(\epsilon)\right]\\
&\le \sum_{y\in Y}\sum_{k=1}^{K}\mathbb{P}\left[\sup_{\xi\in\Xi_2}\left|\hat{g}_{kM(\xi)}(x,\xi,y)-g_k(x,\xi,y)\right|>\pi(\epsilon)\right]\\
&\le \sum_{y\in Y}\sum_{k=1}^{K}\sum_{t=1}^{T}\mathbb{P}\left[\sup_{\xi\in\Xi_2}\left|\hat{g}_{k\mathcal{M}_t}(x,\xi,y)-g_k(x,\xi,y)\right|>\pi(\epsilon)\right].
\end{aligned}
\tag{27}
$$

Basically, given a scenario $\xi\in\Xi_2$, we bound the probability $\mathbb{P}[|\hat{g}_{k\mathcal{M}_t}(x,\xi,y)-g_k(x,\xi,y)|>\pi(\epsilon)]$ using LD theory. So, the probability $\mathbb{P}[\sup_{\xi\in\Xi_2}|\hat{g}_{kM(\xi)}(x,\xi,y)-g_k(x,\xi,y)|>\pi(\epsilon)]$ can be bounded using LD theory if $|\Xi_2|$ is finite. If $|\Xi_2|$ is infinite, we use a discretization technique over set Ξ_2 as in the following.

Under Assumption 4, if we define $\Xi_2^h=\Xi_2\cap\Xi^h$, $h=1,\ldots,H$, then for $\mathbb{P}$-almost every $\xi,\xi_1\in\Xi_2^h$ and for all $x\in X$, $y\in Y$, $k=1\ldots,K$, we have

$$|G_k(x,\xi,y,w)-G_k(x,\xi_1,y,w)|\le\kappa(w)\upsilon.$$

For each set Ξ_2^h, $h=1,\ldots,H$, we choose a representative point $\bar{\xi}_h\in\Xi_2^h$ such that for $\mathbb{P}$-almost every $\xi\in\Xi_2^h$ and for all $x\in X$, $y\in Y$, $k=1\ldots,K$, we have

$$|G_k(x,\xi,y,w)-G_k(x,\bar{\xi}_h,y,w)|\le\kappa(w)\upsilon.$$

We also define the corresponding mapping $h(\xi)=\bar{\xi}_h$ if $\xi\in\Xi_2^h$. We have the following inequality

$$
\begin{aligned}
|\hat{g}_{kM}(x,\xi,y)-g_k(x,\xi,y)|&\le|\hat{g}_{kM}(x,\xi,y)-\hat{g}_{kM}(x,h(\xi),y)|\\
&\quad+|\hat{g}_{kM}(x,h(\xi),y)-g_k(x,h(\xi),y)|+|g_k(x,h(\xi),y)-g_k(x,\xi,y)|.
\end{aligned}
\tag{28}
$$

Here, we assume that $\hat{g}_{kM}(x,\xi,y)$ and $\hat{g}_{kM}(x,h(\xi),y)$ are computed by the same set of realizations of w. We also have $\hat{g}_{kM}(x,\xi,y)-\hat{g}_{kM}(x,h(\xi),y)$ is a SAA of $g_k(x,\xi,y)-g_k(x,h(\xi),y)$, therefore, for $\mathbb{P}$-almost every $\xi\in\Xi_2^h$ we can write

$$
\begin{aligned}
|\hat{g}_{kM}(x,\xi,y)-\hat{g}_{kM}(x,h(\xi),y)|&=\frac{1}{M}\left|\sum_{m=1}^{M}\left(G_k(x,\xi,y,w_m)-G_k(x,h(\xi),y,w_m)\right)\right|\\
&\le\frac{1}{M}\sum_{m=1}^{M}|G_k(x,\xi,y,w_m)-G_k(x,h(\xi),y,w_m)|\\
&\le\frac{1}{M}\sum_{m=1}^{M}\kappa(w_m)\upsilon.
\end{aligned}
$$

So, for $\mathbb{P}$-almost every $\xi\in\Xi_2^h$,

$$|\hat{g}_{kM}(x,\xi,y)-\hat{g}_{kM}(x,h(\xi),y)|\le\hat{\kappa}_M\upsilon,\tag{29}$$

where $\hat{\kappa}_M=M^{-1}\sum_{m=1}^{M}\kappa(w_m)$ is a sample average version of $\mathbb{E}_w[\kappa(w)]$. We also have that, for $\mathbb{P}$-almost every $\xi\in\Xi_2^h$,

$$|g_k(x,\xi,y)-g_k(x,h(\xi),y)|\le\mathbb{E}_w[\kappa(w)]\upsilon.\tag{30}$$

From the assumption that the moment-generating function of $\kappa(w)$ is finite valued in a neighborhood of 0, we have $\mathbb{E}_w[\kappa(w)]$ is finite. We define $L_\kappa=\mathbb{E}_w[\kappa(w)]$. From (30) we have $|g_k(x,\xi,y)-g_k(x,h(\xi),y)|\le L_\kappa\upsilon$

for $\mathbb{P}$-almost every $\xi \in \Xi_2^h$. Thus, for $\mathbb{P}$-almost every $\xi \in \Xi_2^h$, we have

$$\begin{aligned}
&|\hat{g}_{kM}(x,\xi,y) - g_k(x,\xi,y)| \\
&\leq |\hat{g}_{kM}(x,\xi,y) - \hat{g}_{kM}(x,h(\xi),y)| + |\hat{g}_{kM}(x,h(\xi),y) - g_k(x,h(\xi),y)| \\
&\quad + |g_k(x,h(\xi),y) - g_k(x,\xi,y)| \\
&\leq \hat{\kappa}_M \upsilon + |\hat{g}_{kM}(x,h(\xi),y) - g_k(x,h(\xi),y)| + L_k \upsilon.
\end{aligned}$$

Let us return to the evaluation of (27). If we set $\upsilon = \pi(\epsilon)/(4L_\kappa)$, then from (28), (29) and (30), we have

$$\begin{aligned}
&\mathbb{P}\left[\sup_{\xi\in\Xi_2} |\hat{g}_{k\mathcal{M}_t}(x,\xi,y) - \mathbb{E}[G_k(x,\xi,y)]| > \pi(\epsilon)\right] \\
&\leq \mathbb{P}\left[\max_{h=1,\ldots,H} \left|\hat{g}_{k\mathcal{M}_t}(x,\bar{\xi}_h,y) - g_k(x,\bar{\xi}_h,y)\right| > \frac{\pi(\epsilon)}{3}\right] \\
&\qquad + \mathbb{P}\left[\max_{h=1,\ldots,H} \hat{\kappa}_{\mathcal{M}_t} > \frac{\pi(\epsilon)}{3\upsilon}\right] + \mathbb{P}\left[L_\kappa \upsilon > \frac{\pi(\epsilon)}{3}\right] \\
&\leq \sum_{h=1}^{H} \left(\mathbb{P}\left[\left|\hat{g}_{k\mathcal{M}_t}(x,\bar{\xi}_h,y) - g_k(x,\bar{\xi}_h,y)\right| > \frac{\pi(\epsilon)}{3}\right] + \mathbb{P}\left[\hat{\kappa}_{\mathcal{M}_t} > \frac{4L_\kappa}{3}\right]\right).
\end{aligned} \tag{31}$$

The first part of (31) can be handled using LD theory, i.e., under Assumption 3 and using (22), we obtain

$$\mathbb{P}\left[\left|\hat{g}_{k\mathcal{M}_t}(x,\bar{\xi}_h,y) - g_k(x,\bar{\xi}_h,y)\right| > \frac{\pi(\epsilon)}{3}\right] \leq 2\exp\left(\frac{-\mathcal{M}_t \pi^2(\epsilon)}{18\sigma_g^2}\right) \leq 2\exp\left(\frac{-\overline{M}\pi^2(\epsilon)}{18\sigma_g^2}\right), \tag{32}$$

where $\sigma_g^2 = \sup_{x,y,k,\xi} \mathrm{Var}_w[G_k(x,\xi,y,w)]$. For the second part of (31), using again LD theory we obtain

$$\mathbb{P}\left[\hat{\kappa}_{h\mathcal{M}_t} > \frac{4L_\kappa}{3}\right] \leq \exp\left(\frac{-\overline{M}L_\kappa^2}{18\sigma_\kappa^2}\right), \tag{33}$$

where $\sigma_\kappa^2 = \mathrm{Var}_w[\kappa(w)]$. Combining (32) and (33), we have

$$\mathbb{P}\left[\sup_{\xi\in\Xi_2} |\hat{g}_{k\mathcal{M}_t}(x,\xi,y) - g_k(x,\xi,y)| > \pi(\epsilon)\right] \leq H\left(2\exp\left(\frac{-\overline{M}\pi^2(\epsilon)}{18\sigma_g^2}\right) + \exp\left(\frac{-\overline{M}L_\kappa^2}{18\sigma_\kappa^2}\right)\right),$$

and, from (27),

$$\mathbb{P}\left[\frac{1}{N}\sum_{\xi_n\in\Xi_2} \left|\hat{Q}_{M_n}(x,\xi_n) - Q(x,\xi_n)\right| > \frac{\epsilon}{4}\right] \leq K|Y|HT\left(2\exp\left(\frac{-\overline{M}\pi^2(\epsilon)}{18\sigma_g^2}\right) + \exp\left(\frac{-\overline{M}L_\kappa^2}{18\sigma_\kappa^2}\right)\right). \tag{34}$$

Combining (25), (26) and (34), we have

$$\begin{aligned}
&\mathbb{P}\left[\left|\frac{1}{N}\sum_{n=1}^{N} \hat{Q}_{M_n}(x,\xi_n) - \frac{1}{N}\sum_{n=1}^{N} Q(x,\xi_n)\right| > \frac{\epsilon}{2}\right] \\
&\leq \exp\left(\frac{-N\epsilon^2}{72\alpha^2}\right) + |Y|KHT\left(2\exp\left(\frac{-\overline{M}\pi^2(\epsilon)}{18\sigma_g^2}\right) + \exp\left(\frac{-\overline{M}L_\kappa^2}{18\sigma_\kappa^2}\right)\right).
\end{aligned}$$

Along with (23) and (24), this leads to

$$\begin{aligned}
&\mathbb{P}\left[\max_{x\in X} \left|\hat{f}_N(x) - f(x)\right| > \epsilon\right] \\
&\leq 2|X|\exp\left(\frac{-N\epsilon^2}{8\alpha^2}\right) + |X|\exp\left(\frac{-N\epsilon^2}{72\alpha^2}\right) \\
&\quad + |X||Y|KHT\left(2\exp\left(\frac{-\overline{M}\pi^2(\epsilon)}{18\sigma_g^2}\right) + \exp\left(\frac{-\overline{M}L_\kappa^2}{18\sigma_\kappa^2}\right)\right).
\end{aligned}$$

So, in summary, there exist positive constants C_1, C_2, $b_1(\epsilon)$, $b_2(\epsilon)$, where b_1, b_2 depend on ϵ, and C_1, C_2 depend on $|X|$, $|Y|$, K, H and T such that

$$P\left[\max_{x\in X}\left|\hat{f}_N(x)-f(x)\right|\leq\epsilon\right]\geq 1-C_1\exp(-Nb_1(\epsilon))-C_2\exp(-\overline{M}b_2(\epsilon)). \tag{35}$$

Combining this result with Lemma 1, we also have

$$P\left[|\hat{v}_N-v^*|\leq\epsilon\right]\geq 1-C_1\exp(-Nb_1(\epsilon))-C_2\exp(-\overline{M}b_2(\epsilon))$$

and this completes the proof. □

In the next theorem we relax assumptions Assumptions 4 and 5 (finite coverage and bounded number of distinct values for the M_n), and prove a weaker results under the remaining assumptions. Note that there is now an extra $\ln N$ in the exponent of the second exponential.

Theorem 4 *Suppose that Assumptions 1, 2, and 3 hold. Given $\epsilon>0$, there are positive constants C_1, $b_1(\epsilon)$, C_2, $b_2(\epsilon)$ such that*

$$\begin{aligned}\mathbb{P}\left[|\hat{v}_N-v^*|>\epsilon\right] &\leq C_1\exp(-b_1(\epsilon)N)+C_2\exp(-b_2(\epsilon)\overline{M}+\ln N) \qquad \textit{and}\\ \mathbb{P}\left[|f(\hat{x}_N)-v^*|>\epsilon\right] &\leq C_1\exp(-b_1(\epsilon)N)+C_2\exp(-b_2(\epsilon)\overline{M}+\ln N)\end{aligned}$$

where $\overline{M}=\min_{n=1,\ldots,N}M_n$, and $\hat{x}_N$ is an optimal solution to the SAA problem.

Proof. We use the same notation and definitions as in the proof of Theorem 3. However, instead of using a discretization technique for the support set Ξ_2, we just consider (25) and derive the following inequalities

$$\begin{aligned}&\mathbb{P}\left[\frac{1}{N}\sum_{\xi_n\in\Xi_2}\left|\hat{Q}_{M_n}(x,\xi_n)-Q(x,\xi_n)\right|>\frac{\epsilon}{4}\right]\\
&\leq\mathbb{P}\left[\exists\,\xi_n\in\Xi_2\ \Big|\ \left|\hat{Q}_{M_n}(x,\xi_n)-Q(x,\xi_n)\right|>\frac{\epsilon}{4}\right]\\
&\leq\sum_{\substack{\xi_n\in\Xi_2\\ n=1,\ldots,N}}\mathbb{P}\left[\left|\hat{Q}_{M_n}(x,\xi_n)-Q(x,\xi_n)\right|>\frac{\epsilon}{4}\right]\\
&\leq\sum_{\substack{\xi_n\in\Xi_2\\ n=1,\ldots,N}}\mathbb{P}\left[\exists y,k\Big|\ |\hat{g}_{kM_n}(x,\xi_n,y)-g_k(x,\xi_n,y)|>\delta(\xi_n)\right]\\
&\leq\sum_{\substack{\xi_n\in\Xi_2\\ n=1,\ldots,N}}\sum_{y\in Y}\sum_{k=1}^{K}\mathbb{P}\left[\ |\hat{g}_{kM_n}(x,\xi_n,y)-g_k(x,\xi_n,y)|>\pi(\epsilon)\right]\\
&\leq 2NK|Y|\exp\left(\frac{-\overline{M}\pi^2(\epsilon)}{2\sigma_g^2}\right)=2K|Y|\exp\left(\frac{-\overline{M}\pi^2(\epsilon)}{2\sigma_g^2}+\ln N\right).\end{aligned}$$

And similarly to the proof of Theorem 3 we also have

$$\begin{aligned}&\mathbb{P}\left[\max_{x\in X}\left|\hat{f}_N(x)-f(x)\right|>\epsilon\right]\\
&\leq 2|X|\exp\left(\frac{-N\epsilon^2}{8\alpha^2}\right)+|X|\exp\left(\frac{-N\epsilon^2}{72\alpha^2}\right)\\
&+2|X||Y|K\exp\left(\frac{-\overline{M}\pi^2(\epsilon)}{2\sigma_g^2}+\ln N\right).\end{aligned} \tag{36}$$

We complete the proof by selecting $C_1=3|X|$, $b_1(\epsilon)=\epsilon^2/(72\alpha^2)$, $C_2=2|X||Y|K$, and $b_2(\epsilon)=\pi^2(\epsilon)/(2\sigma_g^2)$, and using Lemma 1. □

Although Theorem 4 is *"weaker"* than Theorem 3 due to the term $\ln N$, if $\overline{M}$ increases at least as fast as N, for instance if $\overline{M} \geq N$, we have that $(\ln N)/\overline{M} \to 0$ when $N \to \infty$, meaning that we can neglect the term $\ln N$ when N and $\overline{M}$ are large enough. Formally speaking, there are $N_0 > 0$ and $b_2' < b_2$ such that for all $\overline{M} > N > N_0$, we have that $-\overline{M}b_2 + \ln N < -\overline{M}b_2'$. This means that, without Assumption 4 and 5, we still obtain bounds that converge at the same (asymptotic) rates as in Theorem 3 when $\overline{M}$ and N are large enough.

The next theorem tells us that with a probability that converges to 1 exponentially fast in N and $\overline{M}$, the SAA has a non-empty set of optimal solutions and each one is also an optimal (feasible) solution for the true problem. The proof is based on the the results of Theorems 3 and 4, and uses the fact that the set of first-stage feasible solutions is finite.

Theorem 5 *If Assumptions 1 to 5 hold, there exist positive constants C_1, b_1, C_2, and b_2, such that*

$$\mathbb{P}\left[\emptyset \neq X_N^* \subseteq X^*\right] \geq 1 - C_1 \exp(-b_1 N) - C_2 \exp(-b_2 \overline{M}),$$

where $\overline{M} = \min_{n=1,\ldots,N} M_n$. If Assumptions 1 to 3 hold, there exist positive constants C_1, b_1, C_2, and b_2, such that

$$\mathbb{P}\left[\emptyset \neq X_N^* \subseteq X^*\right] \geq 1 - C_1 \exp(-b_1 N) - C_2 \exp(-b_2 \overline{M} + \ln N).$$

Proof. Under the Assumption 1, X_N^* is not empty, and since $|X|$ is finite, there always exits $\rho > 0$ such that

$$|f(x) - v^*| > \rho, \text{ for all } x \in X \backslash X^*, \tag{37}$$

where ρ can be chosen such that $0 < \rho < \min_{x \in X \backslash X^*} |f(x) - v^*|$. In other words, if $x \in X$ such that $|f(x) - v^*| \leq \rho$ then $x \in X^*$. Now, using the inequality in (21) and Lemma 1 we have

$$\begin{aligned} \mathbb{P}\left[X_N^* \subseteq X^*\right] &\geq \mathbb{P}\left[|f(\hat{x}_N) - v^*| \leq \rho \text{ for all } \hat{x}_N \in X_N^*\right] \\ &\geq \mathbb{P}\left[\max_{x \in X} \left|\hat{f}_N(x) - f(x)\right| \leq \rho/2\right]. \end{aligned}$$

Moreover, under Assumptions 1 to 5, using (35) we have that there are positive constants C_1, C_2, b_1, and b_2 such that

$$\mathbb{P}\left[\max_{x \in X} \left|\hat{f}_N(x) - f(x)\right| > \rho/2\right] \leq C_1 \exp(-N b_1) + C_2 \exp(-\overline{M} b_2).$$

If only Assumptions 1, 2, and 3 hold, we use (36) to obtain that there exist positive constants C_1, C_2, b_1, and b_2 such that

$$\mathbb{P}\left[\max_{x \in X} \left|\hat{f}_N(x) - f(x)\right| > \frac{\rho}{2}\right] \leq C_1 \exp(-N b_1) + C_2 \exp(-\overline{M} b_2 + \ln N).$$

This completes the proof. □

Theorems 3, 4, and 5 do not tell us explicitly how large N and M_n must be for the probability of getting an exact optimal solution to exceed a given target value. The next result provides such explicit sufficient conditions.

Corollary 1 ***(Sample size estimates)***
Suppose Assumptions 1 to 5 hold. We have that $P\left[X_N^* \subseteq X^*\right] \geq 1 - \beta$ if

$$\begin{aligned} N &\geq \left(\frac{288\alpha^2}{\rho^2}\right) \ln\left(\frac{6|X|}{\beta}\right) \quad \textit{and} \\ M_n &\geq \max\left\{\frac{18\sigma_g^2}{\pi^2(\rho/2)}, \frac{18\sigma_\kappa^2}{L_\kappa^2}\right\} \ln\left(\frac{6|X||Y|KHT}{\beta}\right), \ n = 1, \ldots, N. \end{aligned}$$

If only Assumptions 1 to 3 hold, we have the following sufficient values:

$$N \geq \left(\frac{288\alpha^2}{\rho^2}\right) \ln\left(\frac{6|X|}{\beta}\right) \quad \textit{and}$$
$$M_n \geq \frac{2\sigma_g^2}{\pi^2(\rho/2)} \ln\left(\frac{4|X||Y|KN}{\beta}\right), \; n = 1, \ldots, N.$$

These sufficient conditions on N and the M_n are probably too conservative and difficult to compute to provide practical concrete numbers, but they provide insight by showing that N depends logarithmically on the size of the feasible set X and on the tolerance probability β, while M depends logarithmically on the sizes of the feasible sets X and Y as well as the tolerance β.

4 Illustration with a staffing optimization problem

In this section we illustrate consistency on of the SAA approach on the call center staffing application mentioned in the introduction. In the first stage, the arrival rate is assumed uncertain with some prior continuous distribution, then in the second stage some additional information is revealed that changes this distribution. We first formulate the problem and show how it fits our framework. Then we give numerical illustrations.

4.1 A two-stage staffing problem with chance constraints

We consider a multi-skill call center with K call types (numbered from 1 to K), and I agent groups (numbered from 1 to I). Agents within each group i are assumed to be homogeneous and can answer the same set of call types. Each group can handle a specific set of call types, which are not disjoint. The calls are assigned to agents by a router. The staffing vector is $z = (z_1, \ldots, z_I)^{\mathsf{T}}$, where z_i is the number of agents in group i. To keep the present example simpler, we consider a single time period, which we call a "day."

For a "random" day, the arrival process for call type k is assumed to be time-homogeneous Poisson with rate Λ^k for the entire day, for each k, where $\Lambda = (\Lambda^1, \ldots, \Lambda^K)$ is a random vector, and we assume that these K Poisson processes are independent. We also suppose that several days in advance, in the first stage, Λ has a prior distribution which corresponds to some initial distributional forecast. At a later time (the second stage), the distributional forecast is updated, which means that Λ has a new distribution, typically with less uncertainty (smaller variance) but not necessarily. To fit our setting, we assume that ξ is a parameter of the distribution of Λ. Before stage 1, ξ is unknown but we know its probability distribution. At stage 2, we know ξ, but we may not know yet Λ.

Given the staffing vector z, let $\mathcal{S}_k(z) = \mathcal{S}_k(z, w)$ be the service level (SL) of call type k during the day, defined as the proportion of all calls that are answered within τ_k seconds, and let $\mathcal{S}_0(z) = \mathcal{S}_0(z, w)$ be the aggregate SL of the day over all calls, which is the proportion of all calls answered within τ_0 seconds. All of these are random variables whose distributions depend on the staffing z and are also functions of the random element w, which represents the randomness that remains after z and ξ are known. Our stochastic constraints at the second stage will be the following chance constraints on the SLs:

$$\mathbb{P}[\mathcal{S}_k(z) \geq l_k] \geq 1 - \pi_k, \qquad 0 \leq k \leq K, \tag{38}$$

where the probability is with respect to w, and for each k, l_k is a given SL target and π_k is a risk threshold which represents the maximum acceptable value for the probability of missing the SL target for call type k. Note that each constraint in (38) can be rewritten in the form (2) as $\mathbb{E}[\mathbb{I}[\mathcal{S}_k(z) \geq l_k]] + \pi_k - 1 \geq 0$, where $\mathbb{I}[\cdot]$ is the indicator function.

In the first stage, the manager must select an initial staffing $x = (x_1, \ldots, x_I)^{\mathsf{T}}$, at the corresponding cost per agent of $c = (c_1, \ldots, c_I)^{\mathsf{T}}$, based on an initial forecast that gives a prior distribution for ξ. In the second stage, the realization of ξ becomes available, which provides an updated distributional forecast of the arrival rate, and the manager can modify the initial staffing x by adding or removing agents at some penalty costs. More specifically, given ξ, the manager can add $r_i^+(\xi)$ extra agents to group i at cost $c_i^+ > c_i$ per agent, or remove $r_i^-(\xi) \le x_i$ agents in group i and save c_i^- per agent, where $0 \le c_i^- < c_i$. After this recourse, the new number of agents in group i is $z_i(\xi) = x_i + r_i^+(\xi) - r_i^-(\xi)$. Let c, c^+, c^-, and $z(\xi)$ be the vectors with components c_i, c_i^+, c_i^-, and $z_i(\xi)$, respectively. We define the recourse vectors as $r^+(\xi) = (r_1^+(\xi), \dots, r_I^+(\xi))^{\mathrm{T}}$, and $r^-(\xi) = (r_1^-(\xi), \dots, r_I^-(\xi))^{\mathrm{T}}$. The cost of the recourse $y = (r^+(\xi), r^-(\xi))$ is $f_2(x, \xi, y) = (c^+)^{\mathrm{T}} r^+(\xi) - (c^-)^{\mathrm{T}} r^-(\xi)$. The realized staffing used for the day is $z = z(\xi)$. The corresponding two-stage staffing problem can be written as

$$
(\mathbf{P4}) \qquad \begin{cases} \displaystyle\min_{x \in X} \quad c^{\mathrm{T}} x + \mathbb{E}_\xi \left[Q(x, \xi) \right], & \\ \text{where} \quad Q(x, \xi) = \min \quad \left\{ (c^+)^{\mathrm{T}} r^+(\xi) - (c^-)^{\mathrm{T}} r^-(\xi) \right\} & \\ \qquad\qquad \text{subject to} \quad x + r^+(\xi) - r^-(\xi) = z(\xi), & \\ \qquad\qquad\qquad\qquad \mathbb{P}[\mathcal{S}_k(z(\xi)) \ge l_k] \ge 1 - \pi_k, \quad k = 0, \dots, K, & \\ \qquad\qquad\qquad\qquad 0 \le r_i^-(\xi) \le x_i, \qquad i = 1, \dots, I, & \\ \qquad\qquad\qquad\qquad r^+(\xi), r^-(\xi) \ge 0 \text{ and integer.} & \end{cases}
$$

In (P4), X is the set of initial staffing vectors that the manager can select at the first stage, and Y is a set of possible corrections at the second stage. Some assumptions must be made here to make sure that Assumptions 1 and 2 are satisfied. First, we assume that the arrival rate vector Λ has a continuous distribution and an upper bound vector $\bar{\Lambda} = (\bar{\Lambda}^1, \dots, \bar{\Lambda}^K)$, i.e., $\sup_{\xi \in \Xi} \Lambda^k(\xi) \le \bar{\Lambda}^k$, and that there is at least one solution $x \in X$ large enough to satisfy all the SL constraints whenever $\Lambda \le \bar{\Lambda}$. Moreover, as the arrival rates are bounded, there exists $\bar{x} \in \mathbb{N}^I$ such that $\mathbb{P}[\mathcal{S}_k(z) \ge l_k] \ge 1 - \pi_k$, $\forall z \ge \bar{x}$, $k = 1, \dots, K$. Then, it is sufficient to choose $X = \{x \in \mathbb{N}^I \mid 0 \le x \le \bar{x}\}$, and $Y = \{y = (r^+, r^-) \in \mathbb{N}^{2I} \mid \min\{r^+, r^-\} = 0$ and $\max\{r^+, r^-\} \le \bar{x}\}$. We also choose $A(x, \xi) = \{(r^+, r^-) \in Y \mid x + r^+ \le \bar{x} \text{ and } x - r^- \ge 0\}$. Indeed, X and Y are finite. Furthermore, the objective at the first stage is $f_1(x) = c^{\mathrm{T}} x$ and at the second stage is $f_2(x, \xi, y) = (c^+)^{\mathrm{T}} r^+ - (c^-)^{\mathrm{T}} r^-$. Since X and Y are finite, $f_1(.)$ and $f_2(.)$ are also bounded.

For Assumption 2, here we have $g(x, \xi, y) = \mathbb{P}[\mathcal{S}_k(z) \ge l_k] + \pi_k - 1$. Note that for any fixed Λ, the SL $\mathcal{S}_k(z)$ has a discrete distribution over the rational numbers (the SL is always a ratio of integers). Given that the arrival processes are time-homogeneous Poisson with rate Λ, one can write the probability $\mathbb{P}[\mathcal{S}_k(z) \ge l_k \mid \Lambda]$ as an infinite sum of continuous functions of Λ, and from this one can prove that $\mathbb{P}[\mathcal{S}_k(z) \ge l_k \mid \Lambda]$ is also continuous in Λ (see the appendix for a detailed proof). Then, under the assumption that the prior distribution of Λ is continuous, the a priori probability that $g(x, \xi, y) = 0$ is zero.

Thus, our example satisfies all the assumptions for the consistency of the SAA. Assumption 4 is harder to verify and may not always hold in our call center example, as the SL $\mathcal{S}_k(z)$ is a ratio of two integers and can take an infinite number of rational values. However, even without Assumption 4, we still have the weaker LD result of Theorem 4.

For the SAA problem, let $r_n^+ = r^+(\xi_n)$, $r_n^- = r^-(\xi_n)$ and $z_n = z(\xi_n)$ denote the recourse and final staffing vectors for scenario n, we can formulate the SAA problem as

$$
(\mathbf{P5}) \qquad \begin{cases} \displaystyle\min\, c^{\mathrm{T}} x + \frac{1}{N} \sum_{n=1}^{N} \left[(c^+)^{\mathrm{T}} r_n^+ - (c^-)^{\mathrm{T}} r_n^- \right] \\ \text{subject to} \begin{cases} x + r_n^+ - r_n^- = z_n, \quad n = 1, \dots, N, \\ \displaystyle\frac{1}{M_n} \sum_{m=1}^{M_n} \mathbb{I}[\hat{\mathcal{S}}_k^m(z_n) \ge l_k] \ge 1 - \pi_k, \quad k = 0, \dots, K,\ n = 1, \dots, N \\ 0 \le r_n^- \le x, \quad n = 1, \dots, N \\ x, r_n^+, r_n^- \ge 0 \text{ and integer}, \quad n = 1, \dots, N, \end{cases} \end{cases}
$$

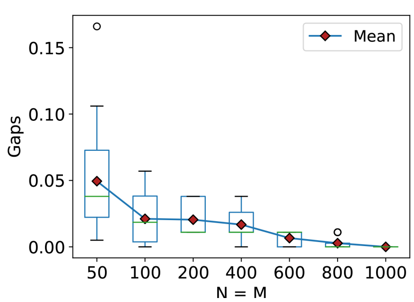
where $\hat{\mathcal{S}}_k^m(z_n)$ is the SL of call type k (the aggregated SL if $k = 0$) in the m-th second-stage simulation for scenario n. The SAA problem above can be solved by a simulation-based cutting plane method proposed in Chan et al. (2016). The main idea of this algorithm is to replace the chance constraints by linear cuts and solve the resulting mixed integer linear programming by a linear solver such as CPLEX.

4.2 Numerical experiments

Here we report a numerical experiment to illustrate the consistency of the SAA estimator, with a small example. Numerical experiments with larger examples are presented in Ta et al. (2018). We consider a call center with $K = 2$ call types and $I = 2$ agent groups, with $\mathcal{S}_1 = \{1\}$ and $\mathcal{S}_2 = \{1, 2\}$. The cost per agent in Stage 1 is $c_1 = 1$ and $c_2 = 1.1$. The recourse costs are $c_i^+ = 2c_i$ and $c_i^- = 0.5c_i$, for $i = 1, 2$. We assume that for the two call types, (i) each caller abandons with probability 0.02 if it has to wait, (ii) patience times (for those who do not abandon immediately on arrival) are exponential with means 10 and 6 minutes, (iii) the service times are exponential with means 10 and 7.5 minutes. The arrival rate for call type k is $\Lambda^k = \xi^k\beta^k$, where β^k is a random busyness factor for the day, which follows a symmetric triangular distribution with mean and mode 1, minimum 0.8, and maximum 1.2, while ξ^k is an independent random factor having a truncated normal distribution with means 70 and 100, standard deviations 10.5 and 15, and truncated to the intervals $[50, 90]$ and $[80, 120]$, for the two call types. These random variables are assumed independent across the two call types. We take $\tau_k = \tau_0 = 120$ (seconds), $l_k = 0.8$ for $k = 1, \ldots, K$, and $l_0 = 0.85$, $\pi_k = 0.2$ for $k = 1, \ldots, K$, and $\pi_0 = 0.15$.

The simulations were performed using the ContactCenter simulation software (Buist and L'Ecuyer, 2005, 2012), developed with the SSJ simulation library (L'Ecuyer et al., 2002). The SAA problems were solved with MATLAB linked to IBM-ILOG CPLEX version 12.6, using the cutting plane method described in Chan et al. (2016).

In the experiment, we aim at evaluating the quality of SAA optimal solutions given by different pairs of M, N, where $M_1 = M_2 = \ldots = M_N = M$. To do so we increase M and N simultaneously. We take $M = N =$ 50, 100, 200, 400, 600, 800, and 1000. For each pair (M, N), we generate 20 sets of scenarios, and for each set of scenarios we approximate the chance constraints by independent realizations of w across scenarios. Each set of scenarios gives a SAA optimal solution $\hat{x}_N$ whose quality can be measured by the gap $f(\hat{x}_N) - v^*$ between the true value of $\hat{x}_N$ and the optimal value v^*. We cannot compute $f(\hat{x}_N)$ and v^* exactly in general, but we can estimate the gaps out of sample. For this, we consider a SAA with $M = N = 1000$ as a validation problem, in which the set of scenarios is independent of those used to obtain $\hat{x}_N$. We then compute the gaps between the costs given by these SAA solutions and the optimal costs given by the validation problem. Let $\bar{f}$ and $\bar{f}^*$ denote the first-stage cost function and the optimal cost given by the SAA validation problem. We estimate the gap by $\bar{f}(\hat{x}_N) - \bar{f}^*$. In Figure 1, on the left side we show box plots of the estimated gaps and on the right side we report the number of zero gaps, for the selected values of $N = M$. We see that when $M = N$ increase above 400, the number of SAA solutions that are also optimal for the validation problem increases quickly with N. When $M = N = 1000$, the corresponding SAA solutions are all the same, and identical to the optimal solution of the validation problem.



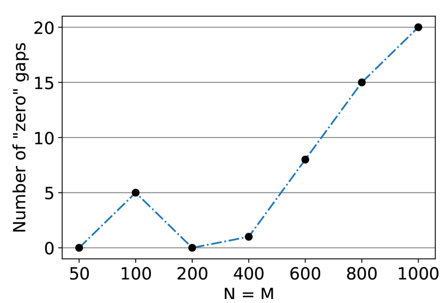


Figure 1: Gaps between the costs given by SAA solutions with $M = N =$ 50, 100, 200, 400, 600, 800, 1000 and the optimal cost given by the validation problem.

5 Conclusion

We have considered a two-stage stochastic programming problem with stochastic constraints in the second stages. We have studied the consistency of the SAA method with nested sampling to solve this problem, and we also proved exponential convergence of the probability of making incorrect decisions. We used a call center staffing problem under arrival rate uncertainty to illustrate our theoretical findings. For future work, it would be interesting to investigate methods for choosing the sample size at the second stage adaptively, e.g., with larger sample sizes for the more important scenarios. Another important aspect is to develop effective methods for solving the SAA in large-scale settings.

Appendix

Proposition 2 *Given a vector of staffing z, the function $h_k(\Lambda) = \mathbb{P}[\mathcal{S}_k(z) \geq l_k \mid \Lambda]$ is a continuous function of Λ.*

Proof. Let denote the number of calls as the vector $C = (C_1, \ldots, C_K)$ where C_k is the number of arrival calls of call type k. As the arrival process for call type k is time-homogeneous Poisson with rate Λ^k, we can write the probability that the service level is at least some values as

$$\begin{aligned} h_k(\Lambda) &= \mathbb{P}[\mathcal{S}_k \geq l_k \mid \Lambda] \\ &= \sum_{r=0}^{\infty} \sum_{\substack{c \in \mathbb{N}^K \\ ||c||_1 = r}} \mathbb{P}[\mathcal{S}_k \geq l_k \mid C = c]\mathbb{P}[C = c|\Lambda] \\ &= \sum_{r=0}^{\infty} \sum_{\substack{c \in \mathbb{N}^K \\ ||c||_1 = r}} \alpha_c \mathbb{P}[C = c \,|\, \Lambda] \\ &= \sum_{r=0}^{\infty} \sum_{\substack{c \in \mathbb{N}^K \\ ||c||_1 = r}} \alpha_c \prod_{k=1}^{K} \mathbb{P}\left[C_k = c_k \,|\, \Lambda^k\right], \end{aligned} \tag{39}$$

where $c = (c_1, \ldots, c_K)$, $||c||_1 = \sum_{k=1}^{K} |c_k|$, and $\alpha_c = \mathbb{P}[\mathcal{S}_k \geq l_k \,|\, C = c] \leq 1$. Moreover, each term $\mathbb{P}[C_k = c_k \,|\, \Lambda^k]$ is a continuous function with respect to Λ^k. So, $\mathbb{P}[C = c \,|\Lambda]$ is also a continuous function with respect to Λ, and $h_k(\Lambda)$ can be written as an infinite sum of continuous functions. From the definition of continuity, $h_k(\Lambda)$ is continuous if for any Λ_0, and for any $\delta > 0$, there exists $\epsilon_1 > 0$ such that for all Λ satisfies $\|\Lambda - \Lambda_0\| \leq \epsilon_1$, we always have

$$|h_k(\Lambda) - h_k(\Lambda_0)| \leq \delta, \tag{40}$$

where $\| \cdot \|$ is the Euclidean norm.

To prove the continuity of $h_k(\Lambda)$, as $\lim_{t\to\infty} \mathbb{P}[C_k > t] = 0$, we first have that, given any $\delta > 0$, there always exists $t_1 > 0$ large enough such that

$$\sum_{\substack{c \in \mathbb{N}^K \\ c_k > t_1, k=1,\ldots,K}} \alpha_c \mathbb{P}[C = c \,|\, \bar{\Lambda}] \leq \sum_{\substack{c \in \mathbb{N}^K \\ c_k > t_1, k=1,\ldots,K}} \mathbb{P}[C = c \,|\, \bar{\Lambda}] = \prod_{k=1}^{K} \mathbb{P}\left[C_k > t_1 | \bar{\Lambda}\right] \leq \frac{\delta}{4}. \tag{41}$$

Moreover, one can show that there exists $t_2 > 0$ such that for all $c_k > t_2$, $k = 1, \ldots, K$, the function $\mathbb{P}[C_k = c_k \,|\, \Lambda^k]$ is monotonically increasing with respect to Λ^k. This can be verified by considering the first-order derivative of $\mathbb{P}[C_k = c_k \,|\, \Lambda^k]$ with respect to Λ^k

$$\frac{\partial \mathbb{P}[C_k = c_k \,|\, \Lambda^k]}{\partial \Lambda^k} = \frac{(\Lambda^k)^{c_k - 1}}{(c_k - 1)!} e^{-\Lambda^k} - \frac{(\Lambda^k)^{c_k}}{(c_k)!} e^{-\Lambda^k} = \frac{(\Lambda^k)^{c_k - 1}}{(c_k - 1)!} e^{-\Lambda^k} \left(1 - \frac{\Lambda^k}{c_k}\right), \tag{42}$$

which is positive if $1 - \Lambda^k/c_k > 0$. Since $1 - \Lambda^k/c_k \geq 1 - \bar{\Lambda}/c_k$, it suffices to take $t_2 \geq \bar{\Lambda}$. Combine (41) and (42), and by choosing $t_0 = \max\{t_1, t_2\}$ we obtain

$$\sum_{\substack{c \in \mathbb{N}^K \\ c_k > t_0, k=1,\ldots,K}} \alpha_c \mathbb{P}[C = c \,|\, \Lambda] \leq \sum_{\substack{c \in \mathbb{N}^K \\ c_k > t_0, k=1,\ldots,K}} \alpha_c \mathbb{P}[C = c \,|\, \bar{\Lambda}] \leq \frac{\delta}{4}, \text{ for all } \Lambda \leq \bar{\Lambda}. \tag{43}$$

Define

$$\mathcal{T}_k(\Lambda) = \sum_{\substack{c \in \mathbb{N}^K \\ 0 \leq c_k \leq t_0, k=1,\ldots,K}} \alpha_c \mathbb{P}[C = c \,|\, \Lambda] \text{ and } \mathcal{H}_k(\Lambda) = \sum_{\substack{c \in \mathbb{N}^K \\ c_k > t_0, k=1,\ldots,K}} \alpha_c \mathbb{P}[C = c \,|\, \Lambda].$$

We then can write $h_k(\Lambda) = \mathcal{T}_k(\Lambda) + \mathcal{H}_k(\Lambda)$, noting that $\mathcal{T}_k(\Lambda)$ is a finite sum of continuous functions, so $\mathcal{T}_k(\Lambda)$ is continuous. We are now ready to prove (40). Consider the following triangle inequality

$$|h_k(\Lambda) - h_k(\Lambda_0)| \leq |\mathcal{T}_k(\Lambda) - \mathcal{T}_k(\Lambda_0)| + |\mathcal{H}_k(\Lambda) - \mathcal{H}_k(\Lambda_0)|\,. \tag{44}$$

As $\mathcal{T}_k(\Lambda)$ is a continuous function, for any $\delta > 0$, there exists ϵ_2 such that $|\mathcal{T}_k(\Lambda) - \mathcal{T}_k(\Lambda_0)| \leq \frac{\delta}{2}$, for all Λ satisfies $||\Lambda - \Lambda_0|| \leq \epsilon_2$. Let $\epsilon = \max\{\epsilon_1, \epsilon_2\}$, then from (43) and (44), we obtain

$$|h_k(\Lambda) - h_k(\Lambda_0)| \leq \frac{\delta}{2} + |\mathcal{H}_k(\Lambda)| + |\mathcal{H}_k(\Lambda_0)| \leq \delta,$$

proving (40). □

References

Ahmed S, Shapiro A (2008) Solving chance-constrained stochastic programs via sampling and integer programming. In: Tutorials in Operations Research, INFORMS, pp 261–269

Atlason J, Epelman MA, Henderson SG (2004) Call center staffing with simulation and cutting plane methods. Annals of Operations Research 127:333–358

Atlason J, Epelman MA, Henderson SG (2008) Optimizing call center staffing using simulation and analytic center cutting plane methods. Management Science 54(2):295–309

Avramidis AN, Chan W, Gendreau M, L'Ecuyer P, Pisacane O (2010) Optimizing daily agent scheduling in a multiskill call centers. European Journal of Operational Research 200(3):822–832

Bastin F, Cirillo C, Toint PhL (2006) Convergence theory for nonconvex stochastic programming with an application to mixed logit. Mathematical Programming, Series B 108(2–3):207–234

Birge JR, Louveaux F (2011) Introduction to Stochastic Programming, 2nd ed. Springer-Verlag, New York, NY, USA

Buist E, L'Ecuyer P (2005) A Java library for simulating contact centers. In: Kuhl ME, Steiger NM, Armstrong FB, Joines JA (eds) Proceedings of the 2005 Winter Simulation Conference, IEEE Press, pp 556–565

Buist E, L'Ecuyer P (2012) ContactCenters: A Java Library for Simulating Contact Centers. Software user's guide, available at http://www.simul.umontreal.ca/contactcenters

Cez̧ik MT, L'Ecuyer P (2008) Staffing multiskill call centers via linear programming and simulation. Management Science 54(2):310–323

Chan W, Ta TA, L'Ecuyer P, Bastin F (2014) Chance-constrained staffing with recourse for multi-skill call centers with arrival-rate uncertainty. In: Proceedings of the 2014 Winter Simulation Conference, IEEE Press, pp 4103–4104

Chan W, Ta TA, L'Ecuyer P, Bastin F (2016) Two-stage chance-constrained staffing with agent recourse for multi-skill call centers. In: Proceedings of the 2016 Winter Simulation Conference, IEEE Press, Piscataway, NJ, USA, pp 3189–3200

Dai L, Chen C, Birge J (2000) Convergence properties of two-stage stochastic programming. Journal of Optimization Theory and Applications 106(3):489–509

Dupacová J, Wets R (1988) Asymptotic behavior of statistical estimators and of optimal solutions of stochastic optimization problems. The Annals of Statistics pp 1517–1549

Hoeffding W (1963) Probability inequalities for sums of bounded random variables. Journal of the American statistical association 58(301):13–30

Kaniovski YM, King AJ, Wets RJ (1995) Probabilistic bounds (via large deviations) for the solutions of stochastic programming problems. Annals of Operations Research 56(1):189–208

Kim S, Pasupathy R, Henderson SG (2015) A guide to sample average approximation. In: Fu MC (ed) Handbook of Simulation Optimization, Springer, New York, NY, USA, pp 207–243

Kleywegt AJ, Shapiro A, Homem-de Mello T (2002) The sample average approximation method for stochastic discrete optimization. SIAM Journal on Optimization 12(2):479–502

L'Ecuyer P, Meliani L, Vaucher J (2002) SSJ: A framework for stochastic simulation in Java. In: Yücesan E, Chen CH, Snowdon JL, Charnes JM (eds) Proceedings of the 2002 Winter Simulation Conference, IEEE Press, pp 234–242

Robinson SM (1996) Analysis of sample path optimization. Mathematics of Operations Research 21:513–528

Rubinstein RY, Shapiro A (1993) Discrete Event Systems: Sensitivity Analysis and Stochastic Optimization by the Score Function Method. Wiley, New York

Shapiro A (1993) Asymptotic behavior of optimal solutions in stochastic programming. Mathematics of Operations Research 18(4):829–845

Shapiro A (2003) Monte Carlo sampling methods. In: Ruszczyński A, Shapiro A (eds) Stochastic Programming, Handbooks in Operations Research and Management Science, Elsevier, Amsterdam, The Netherlands, pp 353–425, chapter 6

Shapiro A, de Mello TH (2000) On rate of convergence of Monte Carlo approximations of stochastic programs. SIAM Journal on Optimization 11(1):70–86

Shapiro A, Dentcheva D, Ruszczyński A (2014) Lecture Notes on Stochastic Programming: Modeling and Theory, 2nd edn. SIAM, Philadelphia

Stroock DW (1984) An introduction to the theory of large deviations. Springer-Verlag New York

Ta TA, L'Ecuyer P, Bastin F (2016) Staffing optimization with chance constraints for emergency call centers. In: MOSIM 2016–11th International Conference on Modeling, Optimization and Simulation, see http://www.iro.umontreal.ca/~lecuyer/myftp/papers/mosim16emergency.pdf

Ta TA, Chan W, L'Ecuyer P, Bastin F (2018) A simulation-based Benders decomposition method for the staffing problem in multiskill call centers under arrival rate uncertainty. Working Paper

Vogel S (1994) A stochastic approach to stability in stochastic programming. Journal of Computational and Applied Mathematics 56:65–96