

## On a multistage discrete stochastic optimization problem with stochastic constraints and nested sampling

Thuy Anh Ta · Tien Mai · Fabian Bastin ·  
Pierre L'Ecuyer

Received: date / Accepted: date

**Abstract** We consider a multistage stochastic discrete program in which constraints on any stage might 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 at each stage, a number of scenarios are examined and a number of simulation replications are performed for each scenario to estimate the next stage constraints. This approach provides an approximate solution to the multistage problem. To establish the consistency of the SAA approach, we first consider a two-stage problem and 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. These convergence results do not hold uniformly over all possible scenarios for the second stage problem. 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. We also prove exponential convergence of the probability of a large deviation for the optimal value of the SAA, the true value of an optimal solution of the SAA, and the probability that any optimal solution to the SAA is an optimal solution of the true problem. All of these results can be extended to a multistage setting and we explain how to do it. Our framework and SAA results cover

---

Thuy Anh Ta  
ORLab, Faculty of Computer Science, Phenikaa University, Hanoi, Vietnam  
E-mail: anh.tathuy@phenikaa-uni.edu.vn

Tien Mai  
Singapore-MIT Alliance for Research and Techonogies (SMART)  
E-mail: mai.tien@smart.mit.edu

Fabian Bastin  
Department of Computer Science and Operations Research, University of Montreal, and CIRRELT  
E-mail: bastin@iro.umontreal.ca

Pierre L'Ecuyer  
Department of Computer Science and Operations Research, University of Montreal, and CIRRELT  
E-mail: lecuyer@iro.umontreal.ca

a large variety of resource allocation problems for which at each stage after the first one, new information becomes available and the allocation can be readjusted, under constraints that involve expectations estimated by Monte Carlo. 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 later stages.

**Keywords** Sample average approximation · multistage stochastic program · expected value constraints · convergence rate · staffing optimization

## 1 Introduction

### 1.1 Motivation and Problem Formulation

We examine a class of dynamic stochastic optimization problems in which at each stage, a decision must be taken among a finite set of possibilities, under uncertainty, with constraints on expectations that have to be estimated by simulation. These types of problems are common in resource allocation settings with probabilistic constraints on the quality of service, and uncertain demand for which more accurate forecasts become available in later stages of the process. An initial resource allocation must be made in the first stage, but this allocation may have to be modified at some cost in one or more later stage(s) to meet the constraints that correspond to the updated demand forecast.

As an illustration, consider the staffing of a telephone call center in which agents with different skill sets answer different types of calls (Gans et al., 2003; Cezik and L'Ecuyer, 2008; Mehrotra et al., 2010). Typically, an initial staffing and work schedules must be decided and announced to the employees a couple of weeks ahead, based on long-term forecasts of call volumes. Closer to the target period, like in the evening before the target day, or even at lunch time for the afternoon hours, more accurate forecasts of the arrival volumes are often available, and the staffing must be modified (at a cost) to meet the constraints for the new forecasts. Examples of such constraints could be that 80% of the calls must be answered within 20 seconds in the long run, or that with probability at least 0.90, 95% or more of the calls during the day are answered within 6 seconds (e.g., for an emergency call center). These constraints are highly nonlinear in the decision variables and can only be estimated by simulation, due to the complexity of the system. The cost to modify the staffing will generally depend on the current staffing and is not necessarily linear.

Similar resource allocation problems occur for other types of service systems, such as staffing employees in retail stores, servers and cooks in restaurants, nurses and other employees in hospitals, police agents in a city, etc. Our setting covers these types of situations and many more.

We define the model in the general framework of a multistage stochastic program, with  $T$  stages numbered from 1 to  $T$ . For typical practical applications of our results, one can think of two or three stages. At the beginning, a first-stage decision  $x^1$  is

selected from a finite set  $X^1$  and a cost that depends on  $x^1$  is paid. Then at each stage  $t \geq 2$ , new information is revealed and a decision  $x^t$  is selected from the finite set  $X^t$ , by taking into account the previous decision  $x^{t-1}$  and all the information revealed to date, under a set of constraints defined by mathematical expectations. We let  $\xi^t$  denote the cumulative information revealed to date up to stage  $t$ , when decision  $x^t$  is taken, for  $t \geq 2$ . Thus, the sequence  $\xi_2, \xi_3, \dots$  can be seen as a filtration. A cost must also be paid for changing the decision from  $x^{t-1}$  to  $x^t$ . It can be interpreted as the cost for changing the resource allocation.

A first goal is to find an optimal decision for the first stage, to minimize the expected total cost, under the assumption that we will be able to make optimal decisions in the next stages. Then, given a first-stage optimal solution and the information in  $\xi^2$ , the next goals are to select optimal solutions  $x^t$  for stage  $t \geq 2$  given revealed information  $\xi^t$ . We formulate the problem formally as follows:

$$(\mathbf{P0}) \quad \left\{ \begin{array}{l} \min_{x^1 \in X^1} \quad f(x^1) = f^1(x^1) + \mathbb{E}_{\xi^2} [Q^2(x^1, \xi^2)] \\ \text{subject to} \quad g^1(x^1) := \mathbb{E}_{w^1} [G^1(x^1, w^1)] \geq 0, \\ \text{where} \quad Q^t(x^{t-1}, \xi^t) \\ \quad \quad \quad = \min_{x^t \in X^t} f^t(x^{t-1}, x^t, \xi^t) + \mathbb{E}_{\xi^{t+1} | \xi^t} [Q^{t+1}(x^t, \xi^{t+1})] \\ \quad \quad \quad \text{subject to} \quad g^t(x^t, \xi^t) := \mathbb{E}_{w^t} [G^t(x^t, \xi^t, w^t)] \geq 0, \\ \quad \quad \quad \text{for } t = 2, \dots, T, \end{array} \right.$$

and  $Q^{T+1}(\cdot) \equiv 0$ . For  $t = 1, \dots, T$ ,  $x^t \in X^t$  is the vector of decision variables,  $f^t$  is measurable cost function (for changing the decision),  $G^t(\cdot)$  is a random vector,  $w^t$  and  $\xi^{t+1}$  are random vectors whose distributions may depend on  $(x^t, \xi^t)$ , and  $\mathbb{E}_{w^t}$  and  $\mathbb{E}_{\xi^{t+1} | \xi^t}$  denote the mathematical expectations with respect to these variables. We assume that these mathematical expectations cannot be computed exactly and are estimated by (Monte Carlo) simulation. All the random variables are defined on a common probability space.

Our motivation for this problem formulation came from the call center staffing optimization application mentioned earlier. The staffing consists in deciding 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 quality of service (QoS) constraints, under uncertainty in the arrival rate process. Here,  $\xi^t$  represents the accumulated information that can be used to make the forecasts on the demand distributions, and  $x^t$  represents the selected staffing, at stage  $t$ . The function  $f^1$  gives the cost of the initial staffing, while for  $t \geq 2$ ,  $f^t$  gives the cost of changing the staffing from  $x^{t-1}$  to  $x^t$  at stage  $t$ . This change usually means reducing or increasing the number of certain agent types in some periods, or changing the tasks of some agents, etc., to better match the updated forecasts. The new staffing  $x^t \in X^t$  may have to satisfy some QoS constraints expressed as expectations,  $\mathbb{E}_{w^t} [G(x^t, \xi^t, w^t)] \geq 0$ , where  $w^t$  represents all the uncertainty that remains after  $x^t, \xi^t$  are known (e.g., the arrival times and service times of all calls, the abandons, etc.). In practice, these QoS constraints will often appear only for the last stage. But staffing decisions could nevertheless be changed before that,

when it costs less, to avoid higher costs at the last stage. The choice of the chance constraints should reflect the decision maker's risk preferences. They could be constraints on certain probabilities, as suggested earlier in the second paragraph, but could also be constraints on quantiles, or on conditional-value-at-risk, for instance. Instead of representing just the staffing, the decisions  $x^t$  can also contain the work schedules of all the agents, as in Avramidis et al. (2010).

For more details on this call center application, see Cezik and L'Ecuyer (2008); Chan et al. (2014, 2016); Gans et al. (2015); Koole (2013); Mehrotra et al. (2010); Pichitlamken et al. (2003); Ta et al. (2016). Evidence that arrival rates are random and that they can be better predicted with more recent information, based on real call center data and with proper modeling, can be found for example in Avramidis et al. (2004); Ibrahim et al. (2012, 2016); Jaoua et al. (2013); L'Ecuyer et al. (2018); Matteson et al. (2011); Oreshkin et al. (2016).

Several other types of dynamic resource allocation problems fit our framework and can benefit from the theory developed in this paper. Similar staffing and scheduling problems occur for example if we replace the agents in a call center by the employees on the floor and at the pay stations in retail stores or in restaurants, employees in healthcare facilities, number and location of police agents and ambulances, number of trucks and drivers on the road for various types of pick-up and delivery systems, etc. See Defraeye and Van Nieuwenhuysse (2016) for a partial survey. For example, a two-stage stochastic staffing and scheduling problem for nurses are considered in Kim and Mehrotra (2015) and Punnakitikashem et al. (2008, 2013). Two-stage stochastic integer programming applications in energy planning (Haneveld and van der Vlerk, 2001), manufacturing (Dempster et al., 1981), and logistics (Laporte et al., 1992; Pillac et al., 2013), for example, can fit our framework. In all these problems, the demand is stochastic and the forecasts can be improved when we get closer to the target periods, and we want to minimize costs under stochastic constraints. The operational benefit of taking into account the randomness (and distribution) of the forecasts for making the first-stage decision clearly depends on the cost structures and the parameters of the distributions in the target applications.

Note that all of these problems also fit the general framework of stochastic dynamic programming in general state spaces (Bertsekas, 2017) and a solution approach could be based in principle on approximate dynamic programming, in which a value function that represents the expected future cost conditional on the current state, could be approximated in some way at each stage (Bertsekas, 2012). In this paper, we follow a different type of approach, designed for when the number of stages is small, say two or three. The approximate dynamic programming approach would be more appropriate than nested sampling when the number of stages is large.

## 1.2 The Two-Stage Setting

Even though our analysis applies to the multistage setting, to reduce the complexity of the notation and proofs, we will give the detailed proofs in the two-stage setting. After that, we will discuss how to extend the results (and proofs) to the multistage program, by induction. We thus consider the following *two-stage discrete stochastic program*,

in which the notation was modified slightly to avoid carrying too many superscript indices:  $x^1, X^1, x^2, X^2, \xi^2, w^1, w^2, g^1, G^1, g^2, G^2$  are now  $x, X, y, Y, \xi, \omega, w, h, H, g, G$ , respectively.

$$\begin{array}{l}
 \text{(P1)} \quad \left\{ \begin{array}{ll}
 \min_{x \in X} & f(x) = f^1(x) + \mathbb{E}_\xi[Q(x, \xi)] \\
 \text{subject to} & h(x) = \mathbb{E}_\omega[H(x, \omega)] \geq 0, \\
 \text{where} & Q(x, \xi) = \min_{y \in Y} f^2(x, y, \xi) \\
 & \text{subject to } g(y, \xi) = \mathbb{E}_w[G(y, \xi, w)] \geq 0.
 \end{array} \right.
 \end{array}
 \tag{1} \tag{2} \tag{3}$$

In the applications we have in mind,  $\xi$ ,  $w$  and  $\omega$  can be taken as independent. In particular,  $\xi$  and  $w$  and  $\omega$  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. Let  $\mathcal{E}$  and  $\mathcal{W}$  denote the sets in which  $\xi$  and  $w$  take their values. The first-stage decision  $x$  must be taken from the finite set  $X$ , before any of  $\xi$ ,  $w$ , and  $\omega$  can be observed. In the second stage,  $\xi$  is observed and then the *recourse* decision  $y$  must be taken from the finite set  $Y$ . We also define  $Y(\xi)$  as the set of second-stage feasible solutions given  $\xi$ , i.e.,  $Y(\xi) = \{y \in Y \mid g(y, \xi) \geq 0\}$ . The functions  $f^1 : X \rightarrow \mathbb{R}$  and  $f^2 : X \times Y \times \mathcal{E} \rightarrow \mathbb{R}$  are measurable, while  $H(x, \omega) = (H_1(x, \omega), \dots, H_J(x, \omega))$  and  $G(y, \xi, w) = (G_1(y, \xi, w), \dots, G_K(y, \xi, w))$  are random vectors for which  $\mathbb{E}_\omega[|H(x, \omega)|] < \infty$  for all  $x \in X$  and  $\mathbb{E}_w[|G(y, \xi, w)|] < \infty$  for all  $(y, \xi) \in Y \times \mathcal{E}$ . Note that  $w$  is not observed before selecting  $y$  (in fact,  $\omega$  and  $w$  may never be observed).

The stochastic optimization problem (P1) considered here is similar to the two-stage problem discussed in Birge and Louveaux (2011), Section 3.5, except that the first- and second-stage decisions belong to finite sets, and the second-stage constraint itself involves the expectation of a second-stage random variable  $w$  whose realization is never observed. We are interested in the situation in which the expected value functions  $\mathbb{E}_\xi[Q(x, \xi)]$ ,  $\mathbb{E}_\omega[H(x, \omega)]$  and  $\mathbb{E}_w[G(y, \xi, w)]$  cannot be written in a closed form or computed numerically, and are estimated by Monte Carlo. The lack of analytical expression for  $\mathbb{E}_w[G(y, \xi, w)]$  implies that the second-stage problem can only be approximated, in contrast to classical two-stage stochastic problems in which it is assumed to be computable exactly.

### 1.3 SAA Approach and Connection with Related Literature

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

One could perhaps think of using *stochastic approximation* (SA) as an alternative to the SAA method. However, since the decision set is discrete, SA does not apply to solve **(P1)**. SA can work well in situations where the decision set is continuous and the objective function is smooth and well-behaved around the optimum. It was originally proposed by Robbins and Monro (1951) to find a zero of a function, and can be applied to find a zero of the gradient of the objective function. For more details on simulation-based optimization via SA, see for example Kushner and Clark (1978); Polyak and Juditsky (1992); L'Ecuyer and Yin (1998). When the objective function is continuous and convex, the approach can be generalized to the use of subgradients Ruszczyński and Syski (1986); Nedic and Bertsekas (2001); Nemirovski et al. (2009). Constraints in expectation are considered in Lan and Zhou (2016). An algorithm that combines SA and SAA is studied in Dussault et al. (1997).

The SAA approach itself is not new; see, e.g., Rubinstein and Shapiro (1993); Robinson (1996); Shapiro (2003); Bastin et al. (2006); Ahmed and Shapiro (2008); 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

$$(P2) \quad \min_{x \in X} \{f(x) := \mathbb{E}_\omega[F(x, \omega)]\} \quad (4)$$

where  $F(x, \omega)$  is a random variable defined over a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ , the expectation over  $\omega$  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

$$\min_{x \in X} \left\{ \hat{f}_N(x) := \frac{1}{N} \sum_{i=1}^N F(x, \omega_i) \right\} \quad (5)$$

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 (4) and (5) as the *true* and SAA problems, respectively. An optimal solution  $\hat{x}_N \in \arg \min_{x \in X} \hat{f}_N(x)$  for (5) 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 (4). 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 (4) and (5), 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  $\text{dist}(x, X^*) = \inf_{x^* \in X^*} \|x - x^*\|$ .

Convergence to zero with probability one (w.p.1) for the three error measures  $\text{dist}(\hat{x}_N, X^*)$ ,  $f(\hat{x}_N) - v^*$ , and  $\hat{v}_N - v^*$  when the sample size  $N \rightarrow \infty$  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  $C$ ,  $\sup_{x \in C} |\hat{f}_N(x) - f(x)| \rightarrow 0$  when  $N \rightarrow \infty$ , 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  $\varepsilon \geq 0$ , a solution  $x \in X$  is said to be  $\varepsilon$ -optimal for the true problem if  $f(x) \leq v^* + \varepsilon$ , and  $\varepsilon$ -optimal for the SAA if  $\hat{f}_N(x) \leq v_N^* + \varepsilon$ . Let  $X^\varepsilon$  and  $X_N^\varepsilon$  denote the sets of  $\varepsilon$ -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  $\varepsilon > \delta > 0$ . 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  $\eta = \eta(\delta, \varepsilon)$  such that

$$\mathbb{P}[X_N^\delta \subseteq X^\varepsilon] \geq 1 - K \exp[-\eta N]. \quad (6)$$

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 (3), and suppose that the second-stage optimization in (2) is easy to solve for any  $(x, \xi)$ . 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) and 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) and Cezik and L'Ecuyer (2008) in the context of call center staffing) is that of an optimization problem with stochastic constraints:

$$\min_{x \in X} f(x) \quad \text{subject to} \quad h(x) := \mathbb{E}_\omega[H(x, \omega)] \geq 0, \quad (7)$$

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  $h(x)$  by  $\hat{h}_N(x)$ , the Monte Carlo average of  $N$  i.i.d. samples of  $H(x, \omega)$ . Under the assumption that  $X$  is finite, that  $\hat{h}_N(x) \rightarrow h(x)$  w.p.1 when  $N \rightarrow \infty$ , and there is  $x^* \in X^*$  such that  $h(x^*) > 0$ , 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  $H(x, \omega)$  satisfies a large-deviation principle, which implies that  $\mathbb{P}[|\hat{h}_N(x) - h(x)| > \varepsilon] \rightarrow 0$  exponentially fast as a function of  $N$

for any  $\varepsilon > 0$ , 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  $N$ . Wang and Ahmed (2008) derived similar results in the more general setting where  $X$  is only assumed to be a nonempty feasible set. In Atlason et al. (2008) and Cezik and L'Ecuyer (2008), the constraints (7) 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.

#### 1.4 Our SAA Formulation and Results

In this paper, we first 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  $\xi$  (or scenarios), say  $\xi_1, \dots, \xi_N$ , 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  $\xi$ . The SAA counterpart of **(P1)** can be written as

$$(\mathbf{P3}) \quad \left\{ \begin{array}{ll} \min_{x \in X} & \hat{f}_{N, M_n}(x) = f^1(x) + \frac{1}{N} \sum_{n=1}^N \hat{Q}_{n, M_n}(x, \xi_n) \\ \text{subject to} & \hat{h}_L(x) \geq 0, \quad (8) \\ \text{where} & \hat{Q}_{n, M_n}(x, \xi_n) = \min_{y_n \in Y} f^2(x, y_n, \xi_n) \quad (9) \\ & \text{subject to } \hat{g}_{n, M_n}(y_n, \xi_n) \geq 0, \quad (10) \end{array} \right.$$

where

$$\hat{h}_L(x) := \frac{1}{L} \sum_{l=1}^L H(x, \omega_l),$$

$(\omega_1, \dots, \omega_L)$  are  $L$  i.i.d. realizations of  $\omega$ ,  $\{\xi_1, \dots, \xi_N\}$  are i.i.d. realizations of  $\xi$ , and for each  $n$ ,

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

where  $\{w_{n, 1}, \dots, w_{n, M_n}\}$  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  $w_{1, m} = \dots = w_{N, m}$  for all  $m$ .

The SAA version of the multistage problem can be formulated in a similar way. That is, at stage  $t = 1, \dots, T$ , given the current state, we generate several realizations of  $w^t$  to estimate  $g^t$  and several realizations of  $\xi^{t+1}$  to estimate the expectation of  $Q^{t+1}$  if  $t < T$ . This can be done recursively until all the expectations are replaced by their SAA versions. Note that for a direct implementation of this, the computational effort increases exponentially with the number of stages, which means that it is practical only for a small number of stages.

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 **(P3)** converge to the optimal decisions for the true problem when  $N$ ,  $L$  and the  $M_n$  increase toward infinity, in the sense that there are constants  $N_0$ ,  $L_0$  and  $M_0$  such that if  $N \geq N_0$ ,  $L \geq L_0$  and  $\min(M_1, \dots, M_N) \geq M_0$ , 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. This proof does not follow from standard results due to the nested nature of the sampling. Similarly, the issue of exponential convergence to 1 of the probability of making an optimal decision is trickier in our setting than in Problem **(P2)**. We show that this exponential convergence holds at the second stage conditionally on  $\xi$ , for almost any fixed  $\xi$ , 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  $\xi$  in general.

We also show that, by using the same techniques used to prove the convergence results for the two-stage case, it is possible to extend the results for the multistage program by induction. More precisely, we show that, w.p.1, the optimal decisions for the SAA converge to the optimal decisions for the true multistage problem when the sample sizes grow to infinity. Moreover, the large deviation probabilities can be bounded from above by functions that converge to 0 exponentially fast when the sample sizes [increase](#). We employ induction techniques to prove such convergence results, noting that we also face the issue that some convergence results may not hold uniformly in the support sets of the corresponding random variables, making the transfer between a stage to its previous stage difficult. We resolve this issue via the same proof techniques as for the two-stage problem.

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 large-deviation probabilities with respect to  $N$  and the  $M_n$ . The extension to a multistage setting is done in Section 4. Section 5 illustrates our convergence results for SAA with a call center staffing optimization example in which arrival rates are uncertain, as discussed earlier. Section 6 provides a conclusion.

## 2 Consistency of the SAA estimators

Let  $X^*$  and  $X_{N, M_n, L}^*$  denote the sets of first-stage optimal solutions for the true and SAA problem, respectively. Let  $v^*$  and  $\hat{v}_{N, M_n, L}$  be the optimal values for the true and SAA counterpart problems. To support the proofs, we also denote by  $\hat{v}_L^*$  the optimal value of the problem

$$\text{(P4)} \quad \min_{x \in X} \{f(x) \mid \hat{h}_L(x) \geq 0\}$$

and note that **(P3)** and **(P4)** have the same set of first-stage feasible solutions, i.e.,  $\{x \in X \mid \hat{h}_L(x) \geq 0\}$ . We also denote by  $Y^*(x, \xi)$  the set of optimal solutions for the true second-stage problem given  $(x, \xi)$ , while  $Y_M^*(x, \xi)$  denotes its SAA counterpart when using sample size  $M$  at the second stage. For  $j = 1, \dots, J$ , let  $h_j(\cdot)$  and  $\hat{h}_{j,L}(\cdot)$

denote the  $j$ -th elements of  $h(\cdot)$  and  $\hat{h}_L(\cdot)$  in (8), respectively. For  $k = 1, \dots, K$ , let  $g_k(\cdot)$  and  $\hat{g}_{k,M}(\cdot)$  denote the  $k$ -th elements of  $g(\cdot)$  and  $\hat{g}_M(\cdot)$  in (10), respectively.

We first assume that the recourse is relatively complete; see for instance Birge and Louveaux (2011) for the definition. 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  $\xi$ . Moreover, we assume that the second-stage objective function is almost surely uniformly bounded.

**Assumption 1**  *$X$  and  $Y$  are finite, and for  $\mathbb{P}$ -almost every  $\xi \in \Xi$ ,  $Y(\xi) \neq \emptyset$ . 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  $\xi$ , the SAA of the second-stage constraint asymptotically coincides 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. 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  $x \in X$ ,  $\hat{h}_L(x) \rightarrow h(x)$  w.p.1 when  $L \rightarrow \infty$ , and for  $\mathbb{P}$ -almost all  $\xi$ , for all  $y \in Y$ ,  $\hat{g}_M(y, \xi) \rightarrow g(y, \xi)$  w.p.1 when  $M \rightarrow \infty$ , there exists  $x \in X^*$  such that  $h(x) \neq 0$ , and there exists  $y \in Y^*(x, \xi)$  such that  $g(y, \xi) \neq 0$ .*

Under Assumption 2, we can apply the known results for 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, 2008).

**Proposition 1** *Under Assumptions 1 and 2, and if there exists  $y \in Y^*(x, \xi)$  such that  $g(y, \xi) \neq 0$ , which occurs for  $\mathbb{P}$ -almost any  $\xi$ , 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  $g(y, \xi) \neq 0$ , there are positive constants  $C$  and  $b(\xi)$  such that*

$$\mathbb{P}[Y_M^*(x, \xi) \subseteq Y^*(x, \xi)] \geq 1 - C \exp[-b(\xi)M]. \quad (11)$$

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

It is important to note that the sample size  $M_0$  and the constant  $b$  in Proposition 1 depend on  $\xi$ , and there may be no  $M_0$  and  $b$  for which the result holds uniformly in  $\xi$ . The next example illustrates this problem. Because of that, we cannot rely on existing results to directly derive exponential bounds for the large-deviation probabilities for the SAA problem (P3).

*Example 1* Consider the two-stage program

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

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

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

Now, consider the SAA counterpart

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

where  $\hat{w}_M$  is a sample average approximation of  $w$  by a Monte Carlo method. In this example, for notational simplicity we set  $M_1 = \dots = M_N = M$ . Regardless of  $x$ , we have  $Y^*(x, \xi) = \{1\}$  if  $\xi \leq 1/2$ , and  $Y^*(x, \xi) = \{2\}$  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.,  $Y_M^*(x, \xi) \not\subseteq Y^*(x, \xi)$ . Therefore, we have

$$\mathbb{P}[Y_M^*(x, \xi) \not\subseteq Y^*(x, \xi)] \geq \mathbb{P}[\hat{w}_M \geq 1 - 2\xi]. \quad (12)$$

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

$$\lim_{1-2\xi \rightarrow 0} \mathbb{P}[\hat{w}_M \geq 1 - 2\xi] = \mathbb{P}[\hat{w}_M \geq 0] = \frac{1}{2}. \quad (13)$$

Hence, if  $1 - 2\xi$  can be arbitrarily close to zero, for any given  $0 \leq \varepsilon < 1/4$ , then there is no  $M_0 > 0$  such that  $\mathbb{P}[\hat{w}_M \geq 1 - 2\xi] < \varepsilon$  for all  $M > M_0$  and all  $\xi \in [0, 1/2)$ , and therefore, there is no  $M_0 > 0$  such that  $\mathbb{P}[Y_M^*(x, \xi) \not\subseteq Y^*(x, \xi)] < \varepsilon$  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  $\xi$ . By contradiction, if there are positive constants  $C_0, b_0$  for which the exponential convergence in Proposition 1 holds uniformly in  $\xi$ , then for  $\mathbb{P}$ -almost every  $\xi \in \mathcal{E}$ , we have

$$\ln(\mathbb{P}[Y_M^*(x, \xi) \not\subseteq Y^*(x, \xi)]) \leq \ln C_0 - Mb_0, \text{ for all } M > 0. \quad (14)$$

From (12) we have, for  $\mathbb{P}$ -almost every  $\xi \in [0, 1/2)$

$$\frac{\ln \mathbb{P}[\hat{w}_M \geq 1 - 2\xi]}{M} \leq \frac{\ln C_0}{M} - b_0. \quad (15)$$

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 (13). 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 (15) 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  $X_{N, M_n, L}^* \subseteq X^*$  when  $\min(N, M_1, \dots, M_N, L)$  is large enough. Since  $X$  is finite, there is a fixed  $\delta > 0$  such that for every  $x \in X \setminus X^*$ ,  $f(x) - v^* \geq \delta$ . Assuming that for a given  $L$  (large enough),  $\{x \in X \mid \hat{h}_L(x) \geq 0\} = \{x \in X \mid h(x) \geq 0\}$ , a sufficient condition for  $X_{N, M_n, L}^* \subseteq X^*$  is that  $|\hat{f}_{N, M_n}(x) - f(x)| < \varepsilon := \delta/2$  for all  $x \in X$ . One could think that this last inequality would follow from the observation that since for each  $\xi_n$ ,  $\hat{Q}_{M_n}(x, \xi_n)$  converges to its expectation w.p.1 when  $M_n \rightarrow \infty$ ,  $|\hat{f}_{N, M_n}(x) - f(x)|$  should converge to 0 w.p.1, so it will eventually be smaller than  $\varepsilon$ . But this simple argument does not really stand (it is not rigorous), because the convergence is not uniform in  $\xi$ , so the required  $M_0$  above which  $|\hat{f}_{N, M_n}(x) - f(x)| < \varepsilon$  when  $N > N_0$  and  $\min(M_1, \dots, M_N) > M_0$  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 three 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

$$\begin{aligned} Y_-(\xi) &= \{y \in Y \mid \exists k \text{ such that } g_k(y, \xi) < 0\}, \\ \bar{\delta}(\xi) &= \frac{1}{2} \max_{y \in Y(\xi), 1 \leq k \leq K} \{g_k(y, \xi) \mid g_k(y, \xi) < 0\}, \\ \underline{\delta}(\xi) &= \min_{y \in Y^*(\xi), 1 \leq k \leq K} \{g_k(y, \xi) \mid g_k(y, \xi) > 0\}, \text{ and} \\ \delta(\xi) &= \min\{-\bar{\delta}(\xi), \underline{\delta}(\xi)\} > 0. \end{aligned} \tag{16}$$

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

**Lemma 1**  $\max_{x \in X} |\hat{f}_{N, M_n}(x) - f(x)| \geq |\hat{v}_{N, M_n, L} - \bar{v}_L^*|$ .

*Proof* Let  $x_{N, M_n, L}^*$  and  $x_L^*$  be optimal solutions to **(P3)** and **(P4)**, respectively. If  $f(x_L^*) < \hat{f}_{N, M_n}(x_{N, M_n, L}^*)$ , since  $\hat{f}_{N, M_n}(x_{N, M_n, L}^*) \leq \hat{f}_{N, M_n}(x_L^*)$ , we have:

$$|\hat{v}_{N, M_n, L} - \bar{v}_L^*| = |\hat{f}_{N, M_n}(x_{N, M_n, L}^*) - f(x_L^*)| \leq |\hat{f}_{N, M_n}(x_L^*) - f(x_L^*)| \leq \max_{x \in X} |\hat{f}_{N, M_n}(x) - f(x)|.$$

If  $f(x_L^*) \geq \hat{f}_{N,M_n}(x_{N,M_n,L}^*)$ , since  $f(x_L^*) \leq f(x_{N,M_n,L}^*)$ , we have  $|\hat{v}_L^* - \hat{v}_{N,M_n,L}| = |f(x_L^*) - \hat{f}_{N,M_n}(x_{N,M_n,L}^*)| \leq |f(x_{N,M_n,L}^*) - \hat{f}_{N,M_n}(x_{N,M_n,L}^*)| \leq \max_{x \in X} |\hat{f}_{N,M_n}(x) - f(x)|$ . In both cases, we have  $|\hat{v}_{N,M_n,L} - \hat{v}_L^*| \leq \max_{x \in X} |\hat{f}_{N,M_n}(x) - f(x)|$ .  $\square$

Also, for almost every scenario, we can recover solutions of the true second-stage problem when the approximate constraints are close enough to the true ones.

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

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

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

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

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

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

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

implying  $y^* \in Y_M(\xi)$ . Moreover, for all  $y_M^* \in Y_M^*(x, \xi)$ , we have  $f^2(x, y^*, \xi) \geq f^2(x, y_M^*, \xi)$ . As  $Y_M(\xi) \subseteq Y$ , we also have  $f^2(x, y^*, \xi) \leq f^2(x, y_M^*, \xi)$ , and therefore  $f^2(x, y^*, \xi) = f^2(x, y_M^*, \xi)$ , 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  $f^2(x, y_1^*, \xi) = f^2(x, y_2^*, \xi)$ . As  $Y_M^*(x, \xi) \subseteq Y_M(\xi) \subseteq Y$ , we also have  $y_1^* \in Y$ , 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.  $\square$

A similar result holds for the first-stage optimal solutions. That is, if we denote

$$\zeta = \min \left\{ \min_{x \in X^*, j} \{h_j(x) \mid h_j(x) > 0\}; \frac{1}{2} \max_{x \in X, j} \{h_j(x) \mid h_j(x) < 0\} \right\}$$

and by  $\tilde{X}_L^*$  the set of optimal solutions to **(P4)**, then the following lemma can be verified in a similar way.

**Lemma 3** *Under Assumptions 1, if  $|\hat{h}_{j,L}(x) - h_j(x)| \leq \zeta$  for all  $x \in X$  and all  $j \in \{1, \dots, J\}$ , then  $\emptyset \neq \tilde{X}_L^* \subseteq X^*$  and  $v^* = \hat{v}_L^*$ .*

**Theorem 1** *Under Assumptions 1, and 2, for any  $\varepsilon > 0$ , w.p.1, there are integers  $N_0 = N_0(\varepsilon)$ ,  $L_0 = L_0(\varepsilon)$  and  $M_0 = M_0(\varepsilon)$  such that for all  $N \geq N_0$ ,  $L \geq L_0$  and  $\min(M_1, \dots, M_N) \geq M_0$ ,  $|\hat{f}_{N,M_n}(x) - f(x)| \leq \varepsilon$  for all  $x \in X$ , and  $|\hat{v}_{N,M_n,L} - v^*| \leq \varepsilon$ .*

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

$$|\hat{f}_{N, M_n}(x) - f(x)| = \left| \frac{1}{N} \sum_{n=1}^N \hat{Q}_{M_n}(x, \xi_n) - \mathbb{E}_\xi [Q(x, \xi)] \right| \quad (17)$$

$$\leq \left| \frac{1}{N} \sum_{n=1}^N Q(x, \xi_n) - \mathbb{E}_\xi [Q(x, \xi)] \right| + \left| \frac{1}{N} \sum_{n=1}^N Q(x, \xi_n) - \frac{1}{N} \sum_{n=1}^N \hat{Q}_{M_n}(x, \xi_n) \right|. \quad (18)$$

To bound the first term in (18), note that under Assumption 1,  $Q(x, \xi)$  is uniformly bounded for  $\mathbb{P}$ -almost every  $\xi \in \mathcal{E}$ , so the expectation of  $Q(x, \xi)$  always exists according to the Lebesgue integration. Thus, this part converges to zero when  $N \rightarrow \infty$  according to the strong law of large numbers, i.e., w.p.1, there exist  $N_0^1(x, \varepsilon)$  such that for all  $N > N_0^1(x, \varepsilon)$ ,

$$\left| \frac{1}{N} \sum_{n=1}^N Q(x, \xi_n) - \mathbb{E}_\xi [Q(x, \xi)] \right| \leq \frac{\varepsilon}{2}. \quad (19)$$

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  $\xi$ . To prove it, we partition the sample space  $\mathcal{E}$  into four different subsets as follows. We first define  $\bar{\mathcal{E}} \subseteq \mathcal{E}$  as the set of all scenarios such that Assumptions 1 and 2 hold for every  $\xi \in \bar{\mathcal{E}}$ . Assumptions 1 and 2 imply that  $\mathbb{P}(\xi \in \bar{\mathcal{E}} | \xi \in \mathcal{E}) = 1$ . We also choose  $\mathcal{E}_1, \mathcal{E}_2$  and  $\mathcal{E}_3$  as three subsets of  $\bar{\mathcal{E}}$  such that  $\delta(\xi)$  is bounded from below by a positive scalar and the convergence of  $\hat{g}_M$  to  $g$  holds uniformly on  $\mathcal{E}_3$ , and for which  $\mathbb{P}[\xi \in \mathcal{E}_1 \cup \mathcal{E}_2]$  can be arbitrarily small. We describe how to choose these sets in the following.

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

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

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

$$|\hat{g}_{k, M}(y, \xi) - g_k(y, \xi)| \leq \varepsilon, \quad (20)$$

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

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

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

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

where  $\alpha$  is a constant chosen such that  $\alpha > \sup_{x \in X, y \in Y, \xi \in \mathcal{E} \setminus \mathcal{E}_0} |2f^2(x, y, \xi)|$ . We can simply choose  $\alpha = \sup_{x \in X, y \in Y, \xi \in \mathcal{E} \setminus \mathcal{E}_0} |2f^2(x, y, \xi)| + 1$ . 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  $\alpha$  always exists and is finite because  $f^2$  is bounded uniformly for every  $\xi \in \bar{\Xi}$ . Let us define

$$\begin{aligned}\mathcal{E}_1 &= \{\xi \in \bar{\Xi} \mid \delta(\xi) \leq \pi(\varepsilon)\}, \\ \mathcal{E}_2 &= \{\xi \in \bar{\Xi} \mid M_0(\xi, \pi(\varepsilon)) \geq M_0^1(\varepsilon)\}, \\ \mathcal{E}_3 &= \bar{\Xi} \setminus (\mathcal{E}_1 \cup \mathcal{E}_2).\end{aligned}$$

Suppose  $\xi_1, \dots, \xi_N \in \bar{\Xi}$ , which happens w.p.1. The second part of (18) can then be written as

$$\begin{aligned}& \left| \frac{1}{N} \sum_{n=1}^N Q(x, \xi_n) - \frac{1}{N} \sum_{n=1}^N \hat{Q}_{M_n}(x, \xi_n) \right| \\ & \leq \frac{1}{N} \sum_{n=1}^N |Q(x, \xi_n) - \hat{Q}_{M_n}(x, \xi_n)| \\ & = \frac{1}{N} \sum_{\xi_n \in \mathcal{E}_1 \cup \mathcal{E}_2} |Q(x, \xi_n) - \hat{Q}_{M_n}(x, \xi_n)| + \frac{1}{N} \sum_{\xi_n \in \mathcal{E}_3} |Q(x, \xi_n) - \hat{Q}_{M_n}(x, \xi_n)| \\ & \leq \frac{1}{N} \sum_{n=1}^N \alpha \mathbb{I}[\xi_n \in \mathcal{E}_1 \cup \mathcal{E}_2] + \frac{1}{N} \sum_{\xi_n \in \mathcal{E}_3} |Q(x, \xi_n) - \hat{Q}_{M_n}(x, \xi_n)|.\end{aligned}\quad (21)$$

The term  $\frac{1}{N} \sum_{n=1}^N \mathbb{I}[\xi_n \in \mathcal{E}_1 \cup \mathcal{E}_2]$  is a sample average of  $\mathbb{P}[\xi_n \in \mathcal{E}_1 \cup \mathcal{E}_2]$ . Therefore, based on the strong law of large numbers, w.p.1, there is  $N_0^2(x, \varepsilon)$  such that, for all  $N \geq N_0^2(x, \varepsilon)$

$$\begin{aligned}\frac{1}{N} \sum_{n=1}^N \mathbb{I}[\xi_n \in \mathcal{E}_1 \cup \mathcal{E}_2] & \leq \mathbb{P}[\xi_n \in \mathcal{E}_1 \cup \mathcal{E}_2] + \frac{\varepsilon}{6\alpha} \\ & \leq \mathbb{P}[\xi_n \in \mathcal{E}_1] + \mathbb{P}[\xi_n \in \mathcal{E}_2] + \frac{\varepsilon}{6\alpha} \\ & \leq \frac{\varepsilon}{6\alpha} + \frac{\varepsilon}{6\alpha} + \frac{\varepsilon}{6\alpha} = \frac{\varepsilon}{2\alpha}.\end{aligned}\quad (22)$$

Moreover, as  $\mathcal{E}_3 = \{\xi \mid \delta(\xi) > \pi(\varepsilon), M_0(\xi, \pi(\varepsilon)) < M_0^1(\varepsilon)\}$ , then for any  $\xi \in \mathcal{E}_3$ , w.p.1, we have  $|\hat{g}_{k,M}(y, \xi) - g_k(y, \xi)| \leq \pi(\varepsilon) < \delta(\xi)$  for all  $y \in Y(x, \xi)$ , all  $M > M_0^1(\varepsilon)$ , and  $k = 1, \dots, K$ . So, for any  $\xi \in \mathcal{E}_3$ , w.p.1,  $\hat{Q}_M(x, \xi) = Q(x, \xi)$  for all  $M > M_0^1(\varepsilon)$ , or equivalently, w.p.1, for all  $M_n > M_0^1(\varepsilon)$ ,  $n = 1, \dots, N$ , we have

$$\frac{1}{N} \sum_{\{n \mid \xi_n \in \mathcal{E}_3\}} |Q(x, \xi_n) - \hat{Q}_{M_n}(x, \xi_n)| = 0 \quad (23)$$

Combining (18), (21), (22) and (23) we have, w.p.1, for all  $x \in X$ , all  $N > N_0(\varepsilon)$  and  $\min\{M_1, \dots, M_N\} > M_0(\varepsilon)$ ,

$$|\hat{f}_{N, M_n}(x) - f(x)| \leq \varepsilon, \quad (24)$$

where  $N_0(\varepsilon) = \max\{N_0^1(\varepsilon), N_0^2(\varepsilon)\}$ , and  $M_0(\varepsilon) = M_0^1(\varepsilon)$ . Furthermore, we have  $|\hat{v}_N - v^*| \leq |\hat{v}_{N, M_n, L} - \hat{v}_L^*| + |\hat{v}_L^* - v^*|$ . By combining this with Lemmas 1 and 3, we obtain that w.p.1, there are  $N_0(\varepsilon)$ ,  $L_0(\varepsilon)$  and  $M_0(\varepsilon)$  such that  $|\hat{v}_{N, M_n, L} - v^*| \leq \varepsilon$  for all  $N > N_0$ ,  $L > L_0$  and  $\min\{M_1, \dots, M_N\} > M_0$ .  $\square$

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 a true optimal solution by solving the SAA problem.

**Theorem 2** *Under Assumptions 1 and 2, w.p.1, there are integers  $N_0$ ,  $L_0$  and  $M_0$  such that for all  $N \geq N_0$ ,  $L \geq L_0$  and  $\min(M_1, \dots, M_N) \geq M_0$ ,  $X_{N, M_n, L}^* \subseteq X^*$ .*

*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

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

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

$$|f(\hat{x}) - v^*| \leq |f(\hat{x}) - \hat{f}_{N, M_n}(\hat{x})| + |\hat{f}_{N, M_n}(\hat{x}) - v^*|. \quad (25)$$

From Theorem 1, w.p.1, there exist  $N_0(\eta)$ ,  $L_0(\eta)$  and  $M_0(\eta) > 0$  such that for all  $N \geq N_0(\eta)$ ,  $L \geq L_0(\eta)$ ,  $M_n \geq M_0(\eta)$  for all  $n = 1, \dots, N$ ,

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

Thus, w.p.1, there are  $N_0, L_0, M_0 > 0$  such that for all  $N \geq N_0$ ,  $L \geq L_0$  and  $M_n \geq M_0$ ,  $n = 1, \dots, N$ , we have  $|f(\hat{x}) - v^*| \leq \eta$  and  $X_{N, M_n, L}^* \subseteq X^*$ .  $\square$

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  $\xi$ , 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  $f(\hat{x}_{N, M_n, L})$  of an optimal solution  $\hat{x}_{N, M_n, L}$  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  $\varepsilon > 0$ ,  $\mathbb{P}[|\hat{v}_{N, M_n, L} - v^*| \leq \varepsilon]$ ,  $\mathbb{P}[|f(\hat{x}_{N, M_n, L}) - v^*| \leq \varepsilon]$ , and  $\mathbb{P}[\emptyset \neq X_{N, M_n, L}^* \subseteq X^*]$  all converge to 1 exponentially fast when  $L, N$  and the  $M_n$  go to  $\infty$ . 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  $\xi$ , but this exponential convergence may not hold uniformly in  $\xi$ . For this reason, it is difficult to establish the exponential convergence of  $\mathbb{P}[X_{N, M_n, L}^* \subseteq X^*]$  when  $L, N$  and the  $M_n$  go to infinity.

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

$$\mathbb{P}[\hat{Z}_M - \mu > \varepsilon] \leq \exp\left(\frac{-M\varepsilon^2}{2\sigma^2}\right) \quad \text{and} \quad \mathbb{P}[\hat{Z}_M - \mu < -\varepsilon] \leq \exp\left(\frac{-M\varepsilon^2}{2\sigma^2}\right). \quad (26)$$

When  $Z$  is bounded, as is the case for  $Z = Q(x, \xi)$  or if  $Z$  is given by an indicator function as we will have later on in **(P8)**, 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 all  $x \in X$ , the moment-generating function of  $H(x, \omega)$ , defined as  $\mathbb{E}_\omega[\exp(tH(x, \omega))]$ , is bounded in a neighborhood of  $t = 0$ , and for  $\mathbb{P}$ -almost every  $\xi \in \Xi$ , for all  $x \in X$  and  $y \in Y$ , the moment-generating function of  $G(y, \xi, w)$ , i.e.  $\mathbb{E}_w[\exp(tG(y, \xi, w))]$ , is bounded in a neighborhood of  $t = 0$ .*

The next assumption concerns a finite covering property of the support set  $\Xi$  with respect to the function  $G_k(y, \xi, w)$ , 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  $\Xi$  by a finite number of subsets such that in each subset, the variation of  $G_k(y, \xi, w)$ , with respect to  $\xi$ , 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  $\xi$  whose support may be infinite. In particular, a finite covering property holds if  $\Xi$  is compact and  $G_k(y, \xi, w)$  is Lipschitz continuous in  $\xi$ . We introduce the following assumption under a general setting.

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

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

It is also convenient in our proofs to assume that the number of distinct values in  $\{M_1, \dots, M_N\}$  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

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

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  $\{M_1, \dots, M_N\}$  is bounded uniformly in  $N$ .*

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

**Theorem 3** Suppose Assumptions 1 to 5 hold. Then for any  $\varepsilon > 0$ , there exist positive constants  $C_0, C_1, C_2, b_0, b_1(\varepsilon)$ , and  $b_2(\varepsilon)$  that do not depend on  $L, N$  and the  $M_n, n = 1, \dots, N$ , such that

$$\mathbb{P} [|\hat{v}_{N, M_n, L} - v^*| > \varepsilon] \leq C_0 \exp[-b_0 L] + C_1 \exp[-b_1(\varepsilon)N] + C_2 \exp[-b_2(\varepsilon)\bar{M}],$$

where  $\bar{M} = \min_{n=1, \dots, N} M_n$ . There also exist positive constants  $C'_0, C'_1, C'_2, b'_0, b'_1(\varepsilon)$ , and  $b'_2(\varepsilon)$  such that

$$\mathbb{P} [|\hat{f}(\hat{x}_N) - v^*| > \varepsilon] \leq C'_0 \exp[-b'_0 L] + C'_1 \exp[-b'_1(\varepsilon)N] + C'_2 \exp[-b'_2(\varepsilon)\bar{M}],$$

where  $\hat{x}_{N, M_n, L}$  is an arbitrary optimal solution to the SAA problem.

*Proof* We use again the triangle inequality in (18). For any  $\varepsilon > 0$ , we have

$$\begin{aligned} \mathbb{P} \left[ \max_{x \in X} |\hat{f}_{N, M_n}(x) - f(x)| > \varepsilon \right] &= \mathbb{P} \left[ \max_{x \in X} \left| \frac{1}{N} \sum_{n=1}^N \hat{Q}_{M_n}(x, \xi_n) - \mathbb{E}_\xi [Q(x, \xi)] \right| > \varepsilon \right] \quad (27) \\ &\leq \mathbb{P} \left[ \max_{x \in X} \left| \frac{1}{N} \sum_{n=1}^N \hat{Q}_{M_n}(x, \xi_n) - \frac{1}{N} \sum_{n=1}^N Q(x, \xi_n) \right| + \max_{x \in X} \left| \frac{1}{N} \sum_{n=1}^N Q(x, \xi_n) - \mathbb{E}_\xi [Q(x, \xi)] \right| > \varepsilon \right] \\ &\leq \mathbb{P} \left[ \left( \max_{x \in X} \left| \frac{1}{N} \sum_{n=1}^N (\hat{Q}_{M_n}(x, \xi_n) - Q(x, \xi_n)) \right| > \frac{\varepsilon}{2} \right) \cup \left( \max_{x \in X} \left| \frac{1}{N} \sum_{n=1}^N Q(x, \xi_n) - \mathbb{E}_\xi [Q(x, \xi)] \right| > \frac{\varepsilon}{2} \right) \right] \\ &\leq \mathbb{P} \left[ \max_{x \in X} \left| \frac{1}{N} \sum_{n=1}^N (\hat{Q}_{M_n}(x, \xi_n) - Q(x, \xi_n)) \right| > \frac{\varepsilon}{2} \right] + \mathbb{P} \left[ \max_{x \in X} \left| \frac{1}{N} \sum_{n=1}^N Q(x, \xi_n) - \mathbb{E}_\xi [Q(x, \xi)] \right| > \frac{\varepsilon}{2} \right] \\ &\leq \sum_{x \in X} \left( \mathbb{P} \left[ \left| \frac{1}{N} \sum_{n=1}^N (\hat{Q}_{M_n}(x, \xi_n) - Q(x, \xi_n)) \right| > \frac{\varepsilon}{2} \right] + \mathbb{P} \left[ \left| \frac{1}{N} \sum_{n=1}^N Q(x, \xi_n) - \mathbb{E}_\xi [Q(x, \xi)] \right| > \frac{\varepsilon}{2} \right] \right). \quad (28) \end{aligned}$$

Considering the second part of (28) and given the fact that  $Q(x, \xi)$  is bounded by the interval  $[-\alpha, \alpha]$  for  $\mathbb{P}$ -almost every  $\xi$ , where  $\alpha$  is defined as in the proof of Theorem 1, we obtain the following from Hoeffding's inequality (Hoeffding, 1963):

$$\mathbb{P} \left[ \left| \frac{1}{N} \sum_{n=1}^N Q(x, \xi_n) - \mathbb{E}_\xi [Q(x, \xi)] \right| > \frac{\varepsilon}{2} \right] \leq 2 \exp \left( \frac{-N\varepsilon^2}{8\alpha^2} \right). \quad (29)$$

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

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

where  $\delta(\xi)$  is defined in (16). Let also define  $\Xi_1 = \{\xi \in \bar{\Xi} \mid \delta(\xi) \leq \pi(\varepsilon)\}$ , and  $\Xi_2 = \bar{\Xi} \setminus \Xi_1$ . We write the first part of (28) as

$$\begin{aligned} & \mathbb{P} \left[ \left| \frac{1}{N} \sum_{n=1}^N \hat{Q}_{M_n}(x, \xi_n) - \frac{1}{N} \sum_{n=1}^N Q(x, \xi_n) \right| > \frac{\varepsilon}{2} \right] \\ & \leq \mathbb{P} \left[ \frac{1}{N} \sum_{\xi_n \in \Xi_1 \cup \Xi_2} |\hat{Q}_{M_n}(x, \xi_n) - Q(x, \xi_n)| > \frac{\varepsilon}{2} \right] \\ & \leq \mathbb{P} \left[ \frac{1}{N} \sum_{\xi_n \in \Xi_1} |\hat{Q}_{M_n}(x, \xi_n) - Q(x, \xi_n)| > \frac{\varepsilon}{4} \right] + \mathbb{P} \left[ \frac{1}{N} \sum_{\xi_n \in \Xi_2} |\hat{Q}_{M_n}(x, \xi_n) - Q(x, \xi_n)| > \frac{\varepsilon}{4} \right] \\ & \leq \mathbb{P} \left[ \frac{1}{N} \sum_{n=1}^N \alpha \mathbb{I}[\xi_n \in \Xi_1] > \frac{\varepsilon}{4} \right] + \mathbb{P} \left[ \frac{1}{N} \sum_{\xi_n \in \Xi_2} |\hat{Q}_{M_n}(x, \xi_n) - Q(x, \xi_n)| > \frac{\varepsilon}{4} \right]. \quad (30) \end{aligned}$$

The first term in (30) concerns a sample average approximation of  $\alpha \mathbb{P}[\xi \in \Xi_1]$ , and we have  $\alpha \mathbb{P}[\xi \in \Xi_1] \leq \varepsilon/6 < \varepsilon/4$ . Moreover,  $\mathbb{I}[\xi \in \Xi_1]$  only takes values in  $\{0, 1\}$ , so by Hoeffding's inequality we have

$$\mathbb{P} \left[ \frac{1}{N} \sum_{n=1}^N \mathbb{I}[\xi_n \in \Xi_1] > \frac{\varepsilon}{4\alpha} \right] \leq \exp \left( \frac{-N\varepsilon^2}{72\alpha^2} \right). \quad (31)$$

For the second term of (30), we have

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

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

$$\begin{aligned} & \mathbb{P} \left[ \frac{1}{N} \sum_{\xi_n \in \Xi_2} |\hat{Q}_{M_n}(x, \xi_n) - Q(x, \xi_n)| > \frac{\varepsilon}{4} \right] \\ & \leq \mathbb{P} \left[ \exists (\xi, y, k) \mid \xi \in \Xi_2, y \in Y, k \in \{1, \dots, K\}, |\hat{g}_{k, M(\xi)}(y, \xi) - g_k(y, \xi)| > \pi(\varepsilon) \right] \\ & \leq \sum_{y \in Y} \sum_{k=1}^K \mathbb{P} \left[ \sup_{\xi \in \Xi_2} |\hat{g}_{k, M(\xi)}(y, \xi) - g_k(y, \xi)| > \pi(\varepsilon) \right] \\ & \leq \sum_{y \in Y} \sum_{k=1}^K \sum_{t=1}^{\mathcal{T}} \mathbb{P} \left[ \sup_{\xi \in \Xi_2} |\hat{g}_{k, \mathcal{M}_t}(y, \xi) - g_k(y, \xi)| > \pi(\varepsilon) \right]. \quad (32) \end{aligned}$$

Basically, given a scenario  $\xi \in \mathcal{E}_2$ , we bound the probability  $\mathbb{P}[|\hat{g}_{k,M}(y, \xi) - g_k(y, \xi)| > \pi(\varepsilon)]$  using LD theory. So, the probability  $\mathbb{P}[\sup_{\xi \in \mathcal{E}_2} |\hat{g}_{k,M}(\xi)(y, \xi) - g_k(y, \xi)| > \pi(\varepsilon)]$  can be bounded using LD theory if  $|\mathcal{E}_2|$  is finite. If  $|\mathcal{E}_2|$  is infinite, we use a discretization technique over set  $\mathcal{E}_2$  as in the following.

Under Assumption 4, if we define  $\mathcal{E}_2^s = \mathcal{E}_2 \cap \mathcal{E}^s$ ,  $s = 1, \dots, S$ , then for  $\mathbb{P}$ -almost every  $\xi, \xi_1 \in \mathcal{E}_2^s$  and for all  $x \in X, y \in Y, k = 1 \dots, K$ , we have

$$|G_k(y, \xi, w) - G_k(y, \xi_1, w)| \leq \kappa(w)v.$$

For each set  $\mathcal{E}_2^s$ ,  $s = 1, \dots, S$ , we choose a representative point  $\bar{\xi}_s \in \mathcal{E}_2^s$  such that for  $\mathbb{P}$ -almost every  $\xi \in \mathcal{E}_2^s$  and for all  $x \in X, y \in Y, k = 1 \dots, K$ , we have

$$|G_k(y, \xi, w) - G_k(y, \bar{\xi}_s, w)| \leq \kappa(w)v.$$

We also define the corresponding mapping  $s(\xi) = \bar{\xi}_s$  if  $\xi \in \mathcal{E}_2^s$ . We have the following inequality

$$\begin{aligned} |\hat{g}_{k,M}(y, \xi) - g_k(y, \xi)| &\leq |\hat{g}_{k,M}(y, \xi) - \hat{g}_{k,M}(y, s(\xi))| \\ &\quad + |\hat{g}_{k,M}(y, s(\xi)) - g_k(y, s(\xi))| + |g_k(y, s(\xi)) - g_k(y, \xi)|. \end{aligned} \quad (33)$$

Here, we assume that  $\hat{g}_{k,M}(y, \xi)$  and  $\hat{g}_{k,M}(y, s(\xi))$  are computed by the same set of realizations of  $w$ . We also have  $\hat{g}_{k,M}(y, \xi) - \hat{g}_{k,M}(y, s(\xi))$  is a SAA of  $g_k(y, \xi) - g_k(y, s(\xi))$ , therefore, for  $\mathbb{P}$ -almost every  $\xi \in \mathcal{E}_2^s$  we can write

$$\begin{aligned} |\hat{g}_{k,M}(y, \xi) - \hat{g}_{k,M}(y, s(\xi))| &= \frac{1}{M} \left| \sum_{m=1}^M (G_k(y, \xi, w_m) - G_k(y, s(\xi), w_m)) \right| \\ &\leq \frac{1}{M} \sum_{m=1}^M |G_k(y, \xi, w_m) - G_k(y, s(\xi), w_m)| \\ &\leq \frac{1}{M} \sum_{m=1}^M \kappa(w_m)v. \end{aligned}$$

So, for  $\mathbb{P}$ -almost every  $\xi \in \mathcal{E}_2^s$ ,

$$|\hat{g}_{k,M}(y, \xi) - \hat{g}_{k,M}(y, s(\xi))| \leq \hat{\kappa}_M v, \quad (34)$$

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  $\xi \in \mathcal{E}_2^s$ ,

$$|g_k(y, \xi) - g_k(y, s(\xi))| \leq \mathbb{E}_w[\kappa(w)]v. \quad (35)$$

From the assumption that the moment-generating function of  $\kappa(w)$  is finite in a neighborhood of 0,  $\mathbb{E}_w[\kappa(w)]$  is finite. We define  $\chi_\kappa = \mathbb{E}_w[\kappa(w)]$ . From (35) we have  $|g_k(y, \xi) - g_k(y, s(\xi))| \leq \chi_\kappa v$  for  $\mathbb{P}$ -almost every  $\xi \in \mathcal{E}_2^s$ . Thus, for  $\mathbb{P}$ -almost every  $\xi \in \mathcal{E}_2^s$ , we have

$$\begin{aligned} &|\hat{g}_{k,M}(y, \xi) - g_k(y, \xi)| \\ &\leq |\hat{g}_{k,M}(y, \xi) - \hat{g}_{k,M}(y, s(\xi))| + |\hat{g}_{k,M}(y, s(\xi)) - g_k(y, s(\xi))| \\ &\quad + |g_k(y, s(\xi)) - g_k(y, \xi)| \\ &\leq \hat{\kappa}_M v + |\hat{g}_{k,M}(y, s(\xi)) - g_k(y, s(\xi))| + \chi_\kappa v. \end{aligned}$$

Let us return to the evaluation of (32). If we set  $v = \pi(\varepsilon)/(4\chi_\kappa)$ , then from (33), (34) and (35), we have

$$\begin{aligned} & \mathbb{P} \left[ \sup_{\xi \in \bar{\mathcal{E}}_2} |\hat{g}_{k, \mathcal{M}_t}(y, \xi) - \mathbb{E}[G_k(y, \xi, w)]| > \pi(\varepsilon) \right] \\ & \leq \mathbb{P} \left[ \max_{s=1, \dots, S} |\hat{g}_{k, \mathcal{M}_t}(x, \bar{\xi}_s, y) - g_k(x, \bar{\xi}_s, y)| > \frac{\pi(\varepsilon)}{3} \right] \\ & \quad + \mathbb{P} \left[ \max_{s=1, \dots, S} \hat{\kappa}_{\mathcal{M}_t} > \frac{\pi(\varepsilon)}{3v} \right] + \mathbb{P} \left[ \chi_\kappa v > \frac{\pi(\varepsilon)}{3} \right] \\ & \leq \sum_{s=1}^S \left( \mathbb{P} \left[ |\hat{g}_{k, \mathcal{M}_t}(y, \bar{\xi}_s) - g_k(y, \bar{\xi}_s)| > \frac{\pi(\varepsilon)}{3} \right] + \mathbb{P} \left[ \hat{\kappa}_{\mathcal{M}_t} > \frac{4\chi_\kappa}{3} \right] \right). \end{aligned} \quad (36)$$

The first part of (36) can be handled using LD theory, i.e., under Assumption 3 and using (26), we obtain

$$\mathbb{P} \left[ |\hat{g}_{k, \mathcal{M}_t}(y, \bar{\xi}_s) - g_k(y, \bar{\xi}_s)| > \frac{\pi(\varepsilon)}{3} \right] \leq 2 \exp \left( \frac{-\mathcal{M}_t \pi^2(\varepsilon)}{18\sigma_g^2} \right) \leq 2 \exp \left( \frac{-\bar{M} \pi^2(\varepsilon)}{18\sigma_g^2} \right), \quad (37)$$

where  $\sigma_g^2 = \sup_{y, k, \xi} \text{Var}_w[G_k(y, \xi, w)]$ . For the second part of (36), using again LD theory we obtain

$$\mathbb{P} \left[ \hat{\kappa}_{h, \mathcal{M}_t} > \frac{4\chi_\kappa}{3} \right] \leq \exp \left( \frac{-\bar{M} \chi_\kappa^2}{18\sigma_\kappa^2} \right), \quad (38)$$

where  $\sigma_\kappa^2 = \text{Var}_w[\kappa(w)]$ . Combining (37) and (38), we have

$$\mathbb{P} \left[ \sup_{\xi \in \bar{\mathcal{E}}_2} |\hat{g}_{k, \mathcal{M}_t}(y, \xi) - g_k(y, \xi)| > \pi(\varepsilon) \right] \leq S \left( 2 \exp \left( \frac{-\bar{M} \pi^2(\varepsilon)}{18\sigma_g^2} \right) + \exp \left( \frac{-\bar{M} \chi_\kappa^2}{18\sigma_\kappa^2} \right) \right),$$

and, from (32),

$$\mathbb{P} \left[ \frac{1}{N} \sum_{\xi_n \in \bar{\mathcal{E}}_2} |\hat{Q}_{M_n}(x, \xi_n) - Q(x, \xi_n)| > \frac{\varepsilon}{4} \right] \leq K|Y|S\mathcal{T} \left( 2 \exp \left( \frac{-\bar{M} \pi^2(\varepsilon)}{18\sigma_g^2} \right) + \exp \left( \frac{-\bar{M} \chi_\kappa^2}{18\sigma_\kappa^2} \right) \right). \quad (39)$$

Combining (30), (31) and (39), we have

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

Along with (28) and (29), this leads to

$$\begin{aligned}
& \mathbb{P} \left[ \max_{x \in X} |\hat{f}_{N, M_n}(x) - f(x)| > \varepsilon \right] \\
& \leq 2|X| \exp\left(\frac{-N\varepsilon^2}{8\alpha^2}\right) + |X| \exp\left(\frac{-N\varepsilon^2}{72\alpha^2}\right) \\
& + |X||Y|KS\mathcal{T} \left( 2 \exp\left(\frac{-\bar{M}\pi^2(\varepsilon)}{18\sigma_g^2}\right) + \exp\left(\frac{-\bar{M}\chi_k^2}{18\sigma_k^2}\right) \right). \\
& < 3|X| \exp\left(\frac{-N\varepsilon^2}{72\alpha^2}\right) + |X||Y|KS\mathcal{T} \left( 2 \exp\left(\frac{-\bar{M}\pi^2(\varepsilon)}{18\sigma_g^2}\right) + \exp\left(\frac{-\bar{M}\chi_k^2}{18\sigma_k^2}\right) \right)
\end{aligned} \tag{40}$$

Now, combining this result with Lemmas 1 and 3, we also have

$$\begin{aligned}
& \mathbb{P} [|\hat{v}_{N, M_n, L} - v^*| > 2\varepsilon] \leq \mathbb{P} [|\hat{v}_{N, M_n, L} - \hat{v}_L^*| > \varepsilon] + \mathbb{P} [|\hat{v}_L^* - v^*| > \varepsilon] \\
& \leq \mathbb{P} \left[ \max_{x \in X} |\hat{f}_{N, M_n}(x) - f(x)| > \varepsilon \right] + \mathbb{P} \left[ \exists(x, j) \left| \hat{h}_{j, L}(x) - u_j(x) \right| > \zeta \right] \\
& \leq \mathbb{P} \left[ \max_{x \in X} |\hat{f}_{N, M_n}(x) - f(x)| > \varepsilon \right] + \sum_{x \in X} \sum_{j=1}^J \mathbb{P} \left[ |\hat{h}_{j, L}(x) - u_j(x)| > \zeta \right] \\
& \leq \mathbb{P} \left[ \max_{x \in X} |\hat{f}_{N, M_n}(x) - f(x)| > \varepsilon \right] + |X|J \exp\left(\frac{-L\zeta^2}{2\sigma_u^2}\right),
\end{aligned} \tag{41}$$

where  $\sigma_u = \max_x \text{Var}_\omega[H_j(x, \omega)]$ . Combining (40) and (41), we have that there exist positive constants  $C_0, C_1, C_2, b_0, b_1(\varepsilon), b_2(\varepsilon)$ , where  $b_1, b_2$  depend on  $\varepsilon$ , and  $C_0, C_1, C_2$  depend on  $|X|, |Y|, K, S, J$  and  $\mathcal{T}$  such that

$$\mathbb{P} [|\hat{v}_{N, M_n, L} - v^*| > \varepsilon] \leq C_0 \exp(-Lb_0) + C_1 \exp(-Nb_1(\varepsilon)) + C_2 \exp(-\bar{M}b_2(\varepsilon)), \tag{42}$$

delivering the first inequality. Moreover, from the triangular inequality (25),

$$\mathbb{P} [|\hat{f}_{N, M_n, L} - v^*| > 2\varepsilon] \tag{43}$$

$$\leq \mathbb{P} [|\hat{f}_{N, M_n, L} - \hat{f}_{N, M_n}(\hat{x}_{N, M_n, L})| > \varepsilon] + \mathbb{P} [|\hat{f}_{N, M_n}(\hat{x}_{N, M_n, L}) - v^*| > \varepsilon],$$

$$\leq \mathbb{P} \left[ \max_{x \in X} |f(x) - \hat{f}_{N, M_n}(x)| > \varepsilon \right] + \mathbb{P} [|\hat{v}_{N, M_n, L} - v^*| > \varepsilon], \tag{44}$$

and combining again (40) and (41), we obtain the second inequality.  $\square$

In the next theorem we relax 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  $\varepsilon > 0$ , there are positive constants  $C_0, b_0, C_1, b_1(\varepsilon), C_2, b_2(\varepsilon)$  such that*

$$\mathbb{P} [|\hat{v}_{N, M_n, L} - v^*| > \varepsilon] \leq C_0 \exp(-b_0L) + C_1 \exp(-b_1(\varepsilon)N) + C_2 \exp(-b_2(\varepsilon)\bar{M} + \ln N),$$

where  $\bar{M} = \min_{n=1, \dots, N} M_n$ . There also exist positive constants  $C'_0, C'_1, C'_2, b'_0, b'_1(\varepsilon)$ , and  $b'_2(\varepsilon)$  such that

$$\mathbb{P}[|f(\hat{x}_{N, M_n, L}) - v^*| > \varepsilon] \leq C'_0 \exp[-b'_0 L] + C'_1 \exp[-b'_1(\varepsilon)N] + C'_2 \exp[-b'_2(\varepsilon)\bar{M} + \ln N],$$

where  $\hat{x}_{N, M_n, L}$  is any 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  $\mathcal{E}_2$ , we just consider (30) and derive the following inequalities

$$\begin{aligned} & \mathbb{P}\left[\frac{1}{N} \sum_{\xi_n \in \mathcal{E}_2} |\hat{Q}_{M_n}(x, \xi_n) - Q(x, \xi_n)| > \frac{\varepsilon}{4}\right] \\ & \leq \mathbb{P}\left[\exists \xi_n \in \mathcal{E}_2 \mid |\hat{Q}_{M_n}(x, \xi_n) - Q(x, \xi_n)| > \frac{\varepsilon}{4}\right] \\ & \leq \sum_{\substack{\xi_n \in \mathcal{E}_2 \\ n=1, \dots, N}} \mathbb{P}\left[|\hat{Q}_{M_n}(x, \xi_n) - Q(x, \xi_n)| > \frac{\varepsilon}{4}\right] \\ & \leq \sum_{\substack{\xi_n \in \mathcal{E}_2 \\ n=1, \dots, N}} \mathbb{P}\left[\exists y, k \mid |\hat{g}_{k, M_n}(y, \xi_n) - g_k(y, \xi_n)| > \delta(\xi_n)\right] \\ & \leq \sum_{\substack{\xi_n \in \mathcal{E}_2 \\ n=1, \dots, N}} \sum_{y \in Y} \sum_{k=1}^K \mathbb{P}\left[|\hat{g}_{k, M_n}(y, \xi_n) - g_k(y, \xi_n)| > \pi(\varepsilon)\right] \\ & \leq 2NK|Y| \exp\left(\frac{-\bar{M}\pi^2(\varepsilon)}{2\sigma_g^2}\right) = 2K|Y| \exp\left(\frac{-\bar{M}\pi^2(\varepsilon)}{2\sigma_g^2} + \ln N\right). \end{aligned}$$

And similarly to the proof of Theorem 3 we also have

$$\begin{aligned} & \mathbb{P}\left[\max_{x \in X} |\hat{f}_{N, M_n}(x) - f(x)| > \varepsilon\right] \\ & \leq 2|X| \exp\left(\frac{-N\varepsilon^2}{8\alpha^2}\right) + |X| \exp\left(\frac{-N\varepsilon^2}{72\alpha^2}\right) \\ & \quad + 2|X||Y|K \exp\left(\frac{-\bar{M}\pi^2(\varepsilon)}{2\sigma_g^2} + \ln N\right). \end{aligned} \tag{45}$$

Then we can use (41), (44), Lemmas 1 and 2 to obtain the desired inequalities and complete the proof.  $\square$

Although Theorem 4 is “weaker” than Theorem 3 due to the term  $\ln N$ , if  $\bar{M}$  increases at least as fast as  $N$ , for instance if  $\bar{M} \geq N$ , we have that  $(\ln N)/\bar{M} \rightarrow 0$  when  $N \rightarrow \infty$ , meaning that we can neglect the term  $\ln N$  when  $N$  and  $\bar{M}$  are large enough. Formally speaking, there are  $N_0 > 0$  and  $b'_2 < b_2$  such that for all  $\bar{M} > N > N_0$ , we have that  $-\bar{M}b_2 + \ln N < -\bar{M}b'_2$ . This means that, without Assumption 4 and 5, we still obtain bounds that converge at the same (asymptotic) rates as in Theorem 3 when  $\bar{M}$  and  $N$  are large enough.

The next theorem tells us that with a probability that converges to 1 exponentially fast in  $N$  and  $\bar{M}$ , 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 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  $C_0$ ,  $b_0$ ,  $C_1$ ,  $b_1$ ,  $C_2$ , and  $b_2$ , such that*

$$\mathbb{P}[\emptyset \neq X_{N,M_n,L}^* \subseteq X^*] \geq 1 - C_0 \exp(-b_0 L) - C_1 \exp(-b_1 N) - C_2 \exp(-b_2 \bar{M}),$$

where  $\bar{M} = \min_{n=1,\dots,N} M_n$ . If Assumptions 1 to 3 hold, there exist positive constants  $C_0$ ,  $b_0$ ,  $C_1$ ,  $b_1$ ,  $C_2$ , and  $b_2$ , such that

$$\mathbb{P}[\emptyset \neq X_{N,M_n,L}^* \subseteq X^*] \geq 1 - C_0 \exp(-b_0 L) - C_1 \exp(-b_1 N) - C_2 \exp(-b_2 \bar{M} + \ln N).$$

*Proof* Under Assumption 1,  $X_{N,M_n,L}^*$  is not empty, and since  $|X|$  is finite, there always exists  $\rho > 0$  such that

$$|f(x) - v^*| > \rho, \text{ for all } x \in X \setminus X^*, \quad (46)$$

where  $\rho$  can be chosen such that  $0 < \rho < \min_{x \in X \setminus X^*} |f(x) - v^*|$ . In other words, if  $x \in X$  such that  $|f(x) - v^*| \leq \rho$  then  $x \in X^*$ . Now, we have

$$\begin{aligned} \mathbb{P}[\emptyset \neq X_{N,M_n,L}^* \subseteq X^*] &\geq \mathbb{P}[\forall \hat{x} \in X_{N,M_n,L}^* |f(\hat{x}) - v^*| \leq \rho] \\ &= 1 - \mathbb{P}[\exists \hat{x} \in X_{N,M_n,L}^* |f(\hat{x}) - v^*| > \rho] \\ &\geq 1 - \sum_{\hat{x} \in X_{N,M_n,L}^*} \mathbb{P}[|f(\hat{x}) - v^*| > \rho]. \end{aligned} \quad (47)$$

Using the second inequalities of Theorem 3 and 4 and under the assumption that  $X$  is finite, we obtain the desired results.  $\square$

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  $P[X_{N,M_n,L}^* \subseteq X^*] \geq 1 - \beta$  if

$$\begin{aligned} N &\geq \left( \frac{1152\alpha^2}{\rho^2} \right) \ln \left( \frac{18|X|^2}{\beta} \right), \\ L &\geq \left( \frac{2\sigma_u^2}{\zeta^2} \right) \ln \left( \frac{3|X|^2 J}{\beta} \right) \quad \text{and} \\ M_n &\geq \max \left\{ \frac{18\sigma_g^2}{\pi^2(\rho/4)}, \frac{18\sigma_k^2}{\chi_k^2} \right\} \ln \left( \frac{18|X|^2 |Y| KS \mathcal{T}}{\beta} \right), \quad n = 1, \dots, N. \end{aligned}$$

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

$$\begin{aligned} N &\geq \left( \frac{1152\alpha^2}{\rho^2} \right) \ln \left( \frac{18|X|^2}{\beta} \right) \\ L &\geq \left( \frac{2\sigma_u^2}{\zeta^2} \right) \ln \left( \frac{3|X|^2 J}{\beta} \right) \quad \text{and} \\ M_n &\geq \frac{2\sigma_g^2}{\pi^2(\rho/4)} \ln \left( \frac{12|X|^2 |Y| K N}{\beta} \right), \quad n = 1, \dots, N. \end{aligned}$$

*Proof* We first show the estimates for the case that Assumptions 1 to 5 hold. The estimates then can be obtained using the proof of Theorem 5 and 3 as follows. Using (44), (47) and (41), we have the chain of inequalities

$$\begin{aligned} 1 - P[X_{N,M_n,L}^* \subseteq X^*] &\leq \sum_{\hat{x} \in X_{N,M_n,L}^*} \mathbb{P}[|f(\hat{x}) - v^*| > \rho] \\ &\leq |X| \mathbb{P} \left[ \max_{x \in X} |f(x) - \hat{f}_{N,M_n}(x)| > \frac{\rho}{2} \right] + |X| \mathbb{P} \left[ |\hat{v}_{N,M_n,L} - v^*| > \frac{\rho}{2} \right] \\ &\leq 2|X| \mathbb{P} \left[ \max_{x \in X} |f(x) - \hat{f}_{N,M_n}(x)| > \frac{\rho}{4} \right] + |X|^2 J \exp \left( \frac{-L\zeta^2}{2\sigma_u^2} \right) \end{aligned} \quad (48)$$

We can bound the first part of (48) using (40) and obtain

$$\begin{aligned} 1 - P[X_{N,M_n,L}^* \subseteq X^*] &\leq |X|^2 J \exp \left( \frac{-L\zeta^2}{2\sigma_u^2} \right) + 6|X|^2 \exp \left( \frac{-N\rho^2}{1152\alpha^2} \right) + \\ &\quad + 2|X|^2 |Y| K \mathcal{S} \left( 2 \exp \left( \frac{-\bar{M}\pi^2(\rho/4)}{18\sigma_g^2} \right) + \exp \left( \frac{-\bar{M}\chi_k^2}{18\sigma_k^2} \right) \right). \end{aligned}$$

So, if we choose  $N, L, M_n$  as in the corollary, then  $1 - P[X_{N,M_n,L}^* \subseteq X^*] \leq \beta$  or  $P[X_{N,M_n,L}^* \subseteq X^*] \geq 1 - \beta$  as desired.

If only Assumptions 1 to 3 hold, we make use of the proof of Theorem 4 to have

$$\begin{aligned} 1 - P[X_{N,M_n,L}^* \subseteq X^*] &\leq 2|X| \mathbb{P} \left[ \max_{x \in X} |f(x) - \hat{f}_{N,M_n}(x)| > \frac{\rho}{4} \right] + |X|^2 J \exp \left( \frac{-L\zeta^2}{2\sigma_u^2} \right) \\ &< |X|^2 J \exp \left( \frac{-L\zeta^2}{2\sigma_u^2} \right) + 6|X|^2 \exp \left( \frac{-N\rho^2}{1152\alpha^2} \right) \\ &\quad + 4|X|^2 |Y| K \exp \left( \frac{-\bar{M}\pi^2(\rho/4)}{2\sigma_g^2} + \ln N \right) \end{aligned} \quad (49)$$

So the selection of  $N, L, M_n, n = 1, \dots, N$ , in the corollary is sufficient to guarantee that  $P[X_{N,M_n,L}^* \subseteq X^*] \geq 1 - \beta$ .

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

#### 4 Extension to a Multistage Setting

We now discuss how to extend our results to the multistage program. We construct a SAA of Problem **(P0)** by replacing all the expectations by Monte Carlo estimates obtained via nested sampling. These estimates can be obtained with different sample sizes at the different nodes of the scenario tree, but to keep the discussion and notation simpler here, we assume here that the sample sizes are the same in all branches: at stage  $t$ , we use  $N^t$  realizations for  $\xi^t$  and  $M^t$  realizations for  $w^t$ . This assumption is not essential for the proofs. It must be understood that the distribution of  $\xi^t$  generally depends on  $\xi^{t-1}$ , which can be different between all tree nodes at stage  $t$ , although this dependence is not indicated explicitly in the formulation below, to avoid overly complicated notation. The SAA of **(P0)** can be stated as follows:

$$(\text{P6}) \quad \left\{ \begin{array}{l} \min_{x^1 \in X^1} \quad \widehat{f}(x^1) = f^1(x^1) + \widehat{Q}_{N^2}^2(x^1) = f^1(x^1) + \frac{1}{N^2} \sum_{n=1}^{N^2} \widehat{Q}^2(x^1, \xi_n^2) \\ \text{subject to} \quad \widehat{g}_{M^1}^1(x^1) := \frac{1}{M^1} \sum_{m=1}^{M^1} G(x^1, w_m^1) \geq 0, \\ \text{where} \quad \widehat{Q}^t(x^{t-1}, \xi_n^t) \\ \quad = \min_{x^t \in X^t} f^t(x^{t-1}, x^t, \xi_n^t) + \widehat{Q}_{N^{t+1}}^{t+1}(x^t, \xi_n^t) \\ \quad \quad \text{subject to} \quad \widehat{g}_{M^t}^t(x^t, \xi_n^t) := \frac{1}{M^t} \sum_{m=1}^{M^t} G^t(x^t, \xi_n^t, w_m^t) \geq 0, \\ \quad \quad \text{with} \quad \widehat{Q}_{N^t}^t(x^{t-1}, \xi_n^{t-1}) := \frac{1}{N^t} \sum_{n'=1}^{N^t} \widehat{Q}^t(x^{t-1}, \xi_{n'}^t) \\ \quad \quad \text{for } t = 2, \dots, T, \end{array} \right.$$

and  $\widehat{Q}_{N^{T+1}}^{T+1}(\cdot) \equiv 0$ . In this formulation,  $\widehat{Q}_{N^t}^t(x^{t-1}, \xi_n^{t-1})$  is the SAA counterpart of  $\mathbb{E}_{\xi^t | \xi^{t-1}}[\widehat{Q}^t(x^{t-1}, \xi^t)]$  and  $\widehat{g}_{M^t}^t(x^t, \xi_n^t)$  is the SAA of  $g^t(x^t, \xi^t)$ . We recall that the  $\xi_n^t$ 's and the  $w_m^t$  are generally different across the different nodes of the scenario tree at level  $t$ . We rely on some basic assumptions to establish the convergence results, e.g., for all  $t = 1, \dots, T$ ,  $f^t$  is measurable and bounded uniformly,  $X^t$  is finite, and the SAA  $\widehat{g}_{M^t}^t$  converges to  $g^t$  w.p.1 when the sample size  $M^t$  goes to infinity.

#### 4.1 Consistency of the SAA

The strong consistency results proved for the two-stage case in Section 2 can be extended to the multistage case by induction on  $T$ . For  $T = 2$  strong consistency is proved in Theorems 1 and 2. The next theorem extends it to  $T > 2$ . We use the shorthand notation  $X_{N^t, M^t}^*$  for the set of optimal solutions to the SAA **(P6)**, keeping in mind that the subscript represents *all* the values of  $N^t$  and  $M^t$  at the different nodes of the scenario tree.

**Theorem 6** *Under the obvious multistage extension of Assumptions 1 and 2, for any  $\varepsilon > 0$ , w.p.1, there are integers  $N_0 = N_0(\varepsilon)$  and  $M_0 = M_0(\varepsilon)$  such that if  $N^t \geq N_0$  and  $M^t \geq M_0$  for all  $t$  and at all tree nodes, then  $|\hat{f}(x^1) - f(x^1)| \leq \varepsilon$  for all  $x^1 \in X^1$ ,  $|\hat{v} - v^*| \leq \varepsilon$ , and  $X_{N^t, M^t}^* \subseteq X^*$ .*

*Proof* For  $T = 2$  the result follows from Theorems 1 and 2, so it remains to do the induction step, and we give a sketch of how this can be done. Suppose the result holds for  $T - 1$ . Then the  $T$ -stage problem **(P0)** can be rewritten as a “two-stage” program by regrouping the time periods  $t = 2, \dots, T$  into a “second stage”. The strong consistency results for the second stage of this two-stage problem follow from the induction hypothesis. Then, given strong consistency of the second stage, we can complete the proof by using the same proof arguments as for the two-stage problem in Section 2.

#### 4.2 Large-Deviation Properties

It is also possible to establish exponential bounds on the large-deviation probabilities for the multistage case. The idea is to find an upper bound for  $\mathbb{P} \left[ \max_{x^1 \in X^1} |\hat{f}(x^1) - f(x^1)| \geq \varepsilon \right]$ , given any  $\varepsilon > 0$ . Under assumptions that are similar to Assumption 1–3, but for all the stages, we decompose  $|\hat{f}(x^1) - f(x^1)|$  as

$$\begin{aligned} & \mathbb{P} \left[ \max_{x^1 \in X^1} |\hat{f}(x^1) - f(x^1)| \geq \varepsilon \right] \\ &= \mathbb{P} \left[ \left| \frac{1}{N^2} \sum_{n=1}^{N^2} \hat{Q}^2(x^1, \xi_n^2) - \mathbb{E}_{\xi^2} [Q^2(x^1, \xi^2)] \right| \geq \varepsilon \right] \\ &\leq \sum_{x^1 \in X^1} \mathbb{P} \left[ \left| \frac{1}{N^2} \sum_{n=1}^{N^2} Q^2(x^1, \xi_n^2) - \mathbb{E}_{\xi^2} [Q^2(x^1, \xi^2)] \right| \geq \frac{\varepsilon}{2} \right] \\ &\quad + \sum_{x^1 \in X^1} \mathbb{P} \left[ \left| \frac{1}{N^2} \sum_{n=1}^{N^2} (\hat{Q}^2(x^1, \xi_n^2) - Q^2(x^1, \xi_n^2)) \right| \geq \frac{\varepsilon}{2} \right]. \quad (50) \end{aligned}$$

Using Hoeffding’s inequality, the first part of (50) can be bounded by an exponential function of the form  $C_1 \exp(-N^2 b_1(\varepsilon))$ . For the second part of (50), we again note that  $Q^2(x^1, \xi^2)$  is the value of a  $(T - 1)$ -stage program, and we can use induction on  $T$  again as in Section 4.1. Here we also face the issue that the convergence in

probability of  $\widehat{Q}^2(x^1, \xi^2) \rightarrow Q^2(x^1, \xi^2)$  might not hold uniformly in the support set of  $\xi^2$ , i.e., the exponents of the exponential bounds may depend on  $\xi^2$ . This issue, again, can be resolved by decomposing the support set of  $\xi^2$  into two parts, where the results hold uniformly in first path and the second part has a probability of less than  $\varepsilon/(6\tau)$ , where  $\tau$  is an upper bound on  $|\widehat{Q}^2(x^1, \xi_n^2) - Q^2(x^1, \xi_n^2)|$  (assuming that this bound always exists). This way of solving the “non-uniform” issue is similar to steps from Equation (28) to (30) in the proof of Theorem 3. This allows to recursively apply Hoeffding’s inequality at each node of the scenario tree to derive a bound of the form

$$\sum_{t=2}^T C_t \exp\left(-b_t(\varepsilon)N^t + \sum_{h=2}^{t-1} \ln N^h\right) + \sum_{t=1}^T D_t \exp\left(-e_t(\varepsilon)M^t + \sum_{h=2}^t \ln N^h\right) \quad (51)$$

where  $C_2, \dots, C_T, b_2(\varepsilon), \dots, b_T(\varepsilon), D_1, \dots, D_T$  and  $e_1(\varepsilon), \dots, e_T(\varepsilon)$  are positive constants which do not depend on any  $N^t$  and  $M^t$ .

In summary, we can bound large-deviation probabilities and the probability of getting wrong solutions when solving the SAA by a function of the form (51), which converges to 0 exponentially fast when  $\min_t \min(N^t, M^t) - \max_t \log N^t \rightarrow \infty$ . Here we note that we are unable to use “finite covering” assumptions (e.g. Assumption 4) to remove the terms containing  $\ln N^h$  in the exponents of the bounds due to the complexity of the proofs in the multistage case.

## 5 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.

### 5.1 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  $y = (y_1, \dots, y_I)^T$ , where  $y_i$  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.”

The exact arrival rates are usually unknown, and several authors consider stochastic optimization to capture this uncertainty; see for instance Harrison and Zeevi (2005); Bassamboo et al. (2006); Liao et al. (2012, 2013); Gurvich et al. (2010); Helber and Henken (2010); Robbins and Harrison (2010), and Gans et al. (2015). Here we assume that for a “random” day, the arrival process for call type  $k$  is time-homogeneous

Poisson with rate  $\Lambda^k$  for the entire day, for each  $k$ , where  $\Lambda = (\Lambda^1, \dots, \Lambda^K)$  is a random vector, and we assume that these  $K$  Poisson processes are independent.

We suppose that several days in advance, in the first stage,  $\Lambda$  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  $\Lambda$  has a new distribution, typically with less uncertainty (smaller variance) but not necessarily. To fit our setting, we assume that  $\xi$  is a parameter of the distribution of  $\Lambda$ . Before stage 1,  $\xi$  is unknown but we know its probability distribution. At stage 2, we know  $\xi$ , but we may not know yet  $\Lambda$ .

The manager of the call of center typically has to ensure sufficient *quality of service* (QoS) in order to meet client satisfaction criteria, to be competitive, or to comply with regulation. The QoS is often measured by the *service level* (SL), defined as the fraction of calls answered within a given time limit, called the *acceptable wait threshold* (AWT). Given the staffing vector  $y$ , the service level of call type  $k$  during the day is defined by

$$\mathcal{S}_k(y) = \mathcal{S}_k(y, w) = \frac{A_k(y)}{T_k - L_k(y)}$$

where  $T_k$  is the total number of calls of type  $k$  that arrived in the period,  $A_k(y)$  is the number of those calls served after waiting at most  $\tau_k$  seconds, and  $L_k(y)$  is the number of them that abandoned after waiting more than  $\tau_k$  seconds. We also denote by  $\mathcal{S}_0(y) = \mathcal{S}_0(y, w)$  the aggregate SL of the day over all calls, which is the proportion of all calls answered within  $\tau_0$  seconds. All of these are random variables whose distributions depend on the staffing  $y$  and are also functions of the random element  $w$ , which represents the randomness that remains after  $y$  and  $\xi$  are known. For other definitions of service level, see Jouini et al. (2013).

Our stochastic constraints at the second stage will be the following chance constraints on the SLs:

$$\mathbb{P}[\mathcal{S}_k(y) \geq l_k] \geq 1 - \pi_k, \quad 0 \leq k \leq K, \quad (52)$$

where the probability is with respect to  $w$ , and for each  $k$ ,  $l_k$  is a given SL target and  $\pi_k$  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 (52) can be rewritten in the form (3) as  $\mathbb{E}[\mathbb{I}[\mathcal{S}_k(y) \geq l_k]] + \pi_k - 1 \geq 0$ , where  $\mathbb{I}[\cdot]$  is the indicator function.

In the first stage, the manager selects an initial staffing  $x = (x_1, \dots, x_I)^\top$ , at the corresponding cost per agent of  $c = (c_1, \dots, c_I)^\top$ , based on an initial forecast that gives a prior distribution for  $\xi$ . In the second stage, the realization of  $\xi$  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  $\xi$ , the manager can choose a new number of agents  $y_i$  for group  $i$  at additional cost  $c_i^+(y_i - x_i)$  if  $y_i > x_i$  and at cost  $c_i^-(y_i - x_i)$  if  $y_i < x_i$ , where  $0 \leq c_i^- < c_i$ . The latter cost is negative, which implies that the manager can make a profit by removing agents from the initial staffing.

Let  $c$ ,  $c^+$ ,  $c^-$ , and  $y(\xi)$  be the vectors with components  $c_i$ ,  $c_i^+$ ,  $c_i^-$ , and  $y_i(\xi)$ , respectively. The first-stage cost is  $f^1(x) = c^\top x$  and the cost of the new staffing  $y$  is

$f^2(x, y, \xi) = \sum_{i=1}^I c_i^+(y_i(\xi) - x_i) \mathbb{I}[y_i(\xi) > x_i] + c_i^-(y_i(\xi) - x_i) \mathbb{I}[y_i(\xi) < x_i]$ . The corresponding two-stage staffing problem can be written as

$$(\mathbf{P7}) \quad \begin{cases} \min_{x \in X} c^\top x + \mathbb{E}_\xi [Q(x, \xi)], \\ \text{where } Q(x, \xi) = \min \left\{ \sum_{i=1}^I c_i^+(y_i(\xi) - x_i) \mathbb{I}[y_i(\xi) > x_i] + c_i^-(y_i(\xi) - x_i) \mathbb{I}[y_i(\xi) < x_i] \right\} \\ \text{subject to } \mathbb{P}[\mathcal{S}_k(y(\xi)) \geq l_k] \geq 1 - \pi_k, \quad k = 0, \dots, K, \\ y(\xi) \geq 0 \text{ and integer.} \end{cases}$$

In **(P7)**,  $X$  is the set of initial staffing vectors that the manager can select at the first stage, and  $Y$  is a set of possible staffing vectors 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  $\Lambda$  has a continuous distribution and an upper bound vector  $\bar{\Lambda} = (\bar{\Lambda}^1, \dots, \bar{\Lambda}^K)$ , i.e.,  $\sup_{\xi \in \Xi} \Lambda^k(\xi) \leq \bar{\Lambda}^k$ , and that there is at least one solution  $x \in X$  large enough to satisfy all the SL constraints whenever  $\Lambda \leq \bar{\Lambda}$ . Moreover, as the arrival rates are bounded, there exists  $\bar{x} \in \mathbb{N}^I$  such that  $\mathbb{P}[\mathcal{S}_k(y) \geq l_k] \geq 1 - \pi_k$ ,  $\forall y \geq \bar{x}$ ,  $k = 1, \dots, K$ . Then, it is sufficient to choose  $X = \{x \in \mathbb{N}^I \mid 0 \leq x \leq \bar{x}\}$ , and  $Y = \{y \in \mathbb{N}^I \mid 0 \leq y \leq \bar{x}\}$ . Indeed,  $X$  and  $Y$  are finite. Since  $X$  and  $Y$  are finite,  $f^1(\cdot)$  and  $f^2(\cdot)$  are also bounded. Here,  $f^2(\cdot)$  is piecewise linear, but we could have more general recourse functions, where the marginal costs to add or remove agents would depend on the agents already affected in the first-stage.

For Assumption 2, here we have  $g(y, \xi) = \mathbb{P}[\mathcal{S}_k(y) \geq l_k] + \pi_k - 1$ . Note that for any fixed  $\Lambda$ , the SL  $\mathcal{S}_k(y)$  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  $\Lambda$ , one can write the probability  $\mathbb{P}[\mathcal{S}_k(y) \geq l_k \mid \Lambda]$  as an infinite sum of continuous functions of  $\Lambda$ , and from this one can prove that  $\mathbb{P}[\mathcal{S}_k(y) \geq l_k \mid \Lambda]$  is also continuous in  $\Lambda$ . 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  $\mathcal{S}_k(y)$  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  $y_n = y(\xi_n)$  denote the second-stage staffing vectors for scenario  $n$ , we can formulate the SAA problem as

$$(\mathbf{P8}) \quad \begin{cases} \min c^\top x + \frac{1}{N} \sum_{n=1}^N \left[ \sum_{i=1}^I c_i^+(y_{n,i} - x_i) \mathbb{I}[y_{n,i} > x_i] + c_i^-(y_{n,i} - x_i) \mathbb{I}[y_{n,i} < x_i] \right] \\ \text{subject to } \begin{cases} \frac{1}{M_n} \sum_{m=1}^{M_n} \mathbb{I}[\hat{\mathcal{S}}_k^m(y_n) \geq l_k] \geq 1 - \pi_k, \quad k = 0, \dots, K, \quad n = 1, \dots, N \\ x, y_n \geq 0 \text{ and integer,} \quad n = 1, \dots, N, \end{cases} \end{cases}$$

where  $\hat{\mathcal{S}}_k^m(y_n)$  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.

A numerically more efficient decomposition technique, developed in Ta et al. (2019), can also be used. We also refer to those papers for more details about our application.

## 5.2 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. (2019). 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  $\beta^k$  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  $\xi^k$  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  $k = 1, \dots, K$ , and  $l_0 = 0.85$ ,  $\pi_k = 0.2$  for  $k = 1, \dots, K$ , 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).

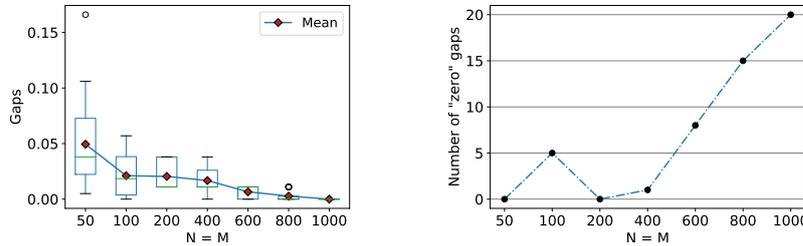


Fig. 1: Gaps between the costs given by SAA solutions with  $M = N = 50, 100, 200, 400, 600, 800, 1000$  and the optimal cost given by the validation problem.

In the experiment, we aim at evaluating the quality of the SAA optimal solutions given by different pairs of  $M, N$ , where  $M_1 = M_2 = \dots = M_N = M$ . 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  $\hat{x}_{N,M}$  whose

quality can be measured by the gap  $f(\hat{x}_{N,M}) - v^*$  between the true value of  $\hat{x}_{N,M}$  and the optimal value  $v^*$ . We cannot compute  $f(\hat{x}_{N,M})$  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  $\hat{x}_{N,M}$ . 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  $\bar{f}(\hat{x}_{N,M}) - \bar{f}^*$ . 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.

## 6 Conclusion

We have considered a multistage stochastic discrete programming problem with constraints in expectation on any stage. We studied the consistency of the SAA method with nested sampling to solve this problem. We have considered a multistage stochastic discrete programming problem with constraints in expectation on any stage. We studied the consistency of the SAA method with nested sampling to solve this problem, and we also proved exponential convergence of the probability of the probability of a large deviation for the optimal value of the SAA with respect to true optimal value, and of making incorrect decisions. We used a call center staffing problem under arrival rate uncertainty to illustrate our theoretical findings in a two-stage context. 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 critical scenarios. Another important aspect is to develop effective methods for solving the SAA in large-scale settings.

**Acknowledgements** 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.

## 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, Deslauriers A, L'Ecuyer P (2004) Modeling daily arrivals to a telephone call center. *Management Science* 50(7):896–908
- 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
- Bassamboo A, Harrison JM, Zeevi A (2006) Design and control of a large call center: Asymptotic analysis of an LP-based method. *Operations Research* 54(3):419–435
- 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
- Bertsekas DP (2012) *Dynamic Programming and Optimal Control, Volume II*, 4th edn. Athena Scientific, Belmont, Mass.
- Bertsekas DP (2017) *Dynamic Programming and Optimal Control, Volume I*, 4th edn. Athena Scientific, Belmont, Mass.
- Birge JR, Louveaux F (2011) *Introduction to Stochastic Programming*, 2nd edn. 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>
- Cezik 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 CH, Birge JR (2000) Convergence properties of two-stage stochastic programming. *Journal of Optimization Theory and Applications* 106(3):489–509
- Defraeye M, Van Nieuwenhuysse I (2016) Staffing and scheduling under nonstationary demand for service: A literature review. *Omega* 58:4–25, DOI <https://doi.org/10.1016/j.omega.2015.04.002>, URL <http://www.sciencedirect.com/science/article/pii/S0305048315000754>
- Dempster MAH, Fisher ML, Jansen L, Lageweg BJ, Lenstra JK, Rinnooy Kan AHG (1981) Analytical evaluation of hierarchical planning systems. *Operations Research* 29(4):707–716
- Dupacova J, Wets R (1988) Asymptotic behavior of statistical estimators and of optimal solutions of stochastic optimization problems. *The Annals of Statistics* pp 1517–1549
- Dussault JP, Labrecque D, L'Ecuyer P, Rubinstein R (1997) Combining the stochastic counterpart and stochastic approximation methods. *Discrete Event Dynamic Systems: Theory and Applications* 7(1):5–28

- Gans N, Koole G, Mandelbaum A (2003) Telephone call centers: Tutorial, review, and research prospects. *Manufacturing and Service Operations Management* 5:79–141
- Gans N, Shen H, Zhou YP, Korolev N, McCord A, Ristock H (2015) Parametric forecasting and stochastic programming models for call-center workforce scheduling. *Manufacturing and Service Operations Management* 17(4):571–588
- Gurvich I, Luedtke J, Tezcan T (2010) Staffing call centers with uncertain demand forecasts: A chance-constrained optimization approach. *Management Science* 56(7):1093–1115
- Haneveld WKK, van der Vlerk MH (2001) Optimizing electricity distribution using two-stage integer recourse models. In: *Stochastic optimization: Algorithms and applications*, Springer, pp 137–154
- Harrison JM, Zeevi A (2005) A method for staffing large call centers based on stochastic fluid models. *Manufacturing and Service Operations Management* 7(1):20–36
- Helber S, Henken K (2010) Profit-oriented shift scheduling of inbound contact centers with skills-based routing, impatient customers, and retrials. *OR Spectrum* 32:109–134
- Hoeffding W (1963) Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association* 58:13–29
- Ibrahim R, L'Ecuyer P, Régnard N, Shen H (2012) On the modeling and forecasting of call center arrivals. In: Laroque C, Himmelspach J, Pasupathy R, Rose O, Uhrmacher AM (eds) *Proceedings of the 2012 Winter Simulation Conference*, IEEE Press, pp 256–267
- Ibrahim R, Ye H, L'Ecuyer P, Shen H (2016) Modeling and forecasting call center arrivals: A literature study and a case study. *International Journal of Forecasting* 32(3):865–874
- Jaoua A, L'Ecuyer P, Delorme L (2013) Call type dependence in multiskill call centers. *Simulation* 89(6):722–734
- Jouini O, Koole G, Roubos A (2013) Performance indicators for call centers with impatient customers. *IIE Transactions* 45(3):341–354
- Kaniovski YM, King AJ, Wets RJB (1995) Probabilistic bounds (via large deviations) for the solutions of stochastic programming problems. *Annals of Operations Research* 56(1):189–208
- Kim K, Mehrotra S (2015) A two-stage stochastic integer programming approach to integrated staffing and scheduling with application to nurse management. *Operations Research* 63(6):1431–1451
- 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
- Koole G (2013) *Call Center Optimization*. MG books, Amsterdam
- Kushner HJ, Clark DS (1978) *Stochastic Approximation Methods for Constrained and Unconstrained Systems*, Applied Mathematical Sciences, vol 26. Springer-Verlag

- Lan G, Zhou Z (2016) Algorithms for stochastic optimization with functional of expectation constraints, arXiv preprint 1604.03887
- Laporte G, Louveaux F, Mercure H (1992) The vehicle routing problem with stochastic travel times. *Transportation science* 26(3):161–170
- L'Ecuyer P, Yin G (1998) Budget-dependent convergence rate for stochastic approximation. *SIAM Journal on Optimization* 8(1):217–247
- 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
- L'Ecuyer P, Gustavsson K, Olsson L (2018) Modeling bursts in the arrival process to an emergency call center. In: *Proceedings of the 2018 Winter Simulation Conference*, IEEE Press, pp 525–536
- Liao S, van Delft C, Koole G, Jouini O (2012) Staffing a call center with uncertain non-stationary arrival rate and flexibility. *OR Spectrum* 34:691–721
- Liao S, van Delft C, Vial JP (2013) Distributionally robust workforce scheduling in call centers with uncertain arrival rates. *Optimization Methods and Software* 28(3):501–522
- Matteson DS, McLean MW, Woodard DB, Henderson SG (2011) Forecasting emergency medical service call arrival rates. *Annals of Applied Statistics* 5(2B):1379–1406
- Mehrotra V, Ozlük O, Saltzman R (2010) Intelligent procedures for intra-day updating of call center agent schedules. *Production and Operations Management* 19(3):353–367
- Nedic A, Bertsekas DP (2001) Incremental subgradient methods for nondifferentiable optimization. *SIAM Journal on Optimization* 12(1):109–138
- Nemirovski A, Juditsky A, Lan G, Shapiro A (2009) Robust stochastic approximation approach to stochastic programming. *SIAM Journal on Optimization* 19(4):1574–1609
- Oreshkin B, Régnard N, L'Ecuyer P (2016) Rate-based daily arrival process models with application to call centers. *Operations Research* 64(2):510–527
- Pichitlamken J, Deslauriers A, L'Ecuyer P, Avramidis AN (2003) Modeling and simulation of a telephone call center. In: *Proceedings of the 2003 Winter Simulation Conference*, IEEE Press, pp 1805–1812
- Pillac V, Gendreau M, Guéret C, Medaglia AL (2013) A review of dynamic vehicle routing problems. *European Journal of Operational Research* 225(1):1–11
- Polyak BT, Juditsky AB (1992) Acceleration of stochastic approximation by averaging. *SIAM Journal on Control and Optimization* 30(4):838–855
- Punnakitikashem P, Rosenberger JM, Buckley-Behan D (2008) Stochastic programming for nurse assignment. *Computational Optimization and Applications* 40(3):321–349
- Punnakitikashem P, Rosenberger JM, Buckley-Behan DF (2013) A stochastic programming approach for integrated nurse staffing and assignment. *IIE Transactions* 45(10):1059–1076
- Robbins H, Monro S (1951) A stochastic approximation method. *Annals of Mathematical Statistics* 22:400–407

- Robbins TR, Harrison TP (2010) A stochastic programming model for scheduling call centers with global service level agreements. *European Journal of Operational Research* 207(3):1608–1619
- 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
- Ruszczyński A, Syski W (1986) On convergence of the stochastic subgradient method with on-line stepsize rules. *Journal of Mathematical Analysis and Applications* 114(2):512–527
- 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. *Handbooks in Operations Research and Management Science*, 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, Bastin F, L'Ecuyer P (2019) A simulation-based decomposition approach for two-stage staffing optimization in call centers under arrival rate uncertainty. Submitted for publication
- Vogel S (1994) A stochastic approach to stability in stochastic programming. *Journal of Computational and Applied Mathematics* 56:65–96
- Wang W, Ahmed S (2008) Sample average approximation of expected value constrained stochastic programs. *Operations Research Letters* 36:515–519