

Article

An Approach to Chance Constrained Problems Based on Huge Data Sets Using Weighted Stratified Sampling and Adaptive Differential Evolution [†]

Kiyoharu Tagawa 

Department of Informatics, Kindai University, Osaka 577-8502, Japan; tagawa@info.kindai.ac.jp

[†] This paper is an extended version of our paper published in the 25th International Conference on Information and Software Technologies (ICIST 2019), held in Kaunas University of Technology (Lithuania) on 10–12 October 2019.

Received: 19 March 2020; Accepted: 14 April 2020; Published: 16 April 2020



Abstract: In this paper, a new approach to solve Chance Constrained Problems (CCPs) using huge data sets is proposed. Specifically, instead of the conventional mathematical model, a huge data set is used to formulate CCP. This is because such a large data set is available nowadays due to advanced information technologies. Since the data set is too large to evaluate the probabilistic constraint of CCP, a new data reduction method called Weighted Stratified Sampling (WSS) is proposed to describe a relaxation problem of CCP. An adaptive Differential Evolution combined with a pruning technique is also proposed to solve the relaxation problem of CCP efficiently. The performance of WSS is compared with a well known method, Simple Random Sampling. Then, the proposed approach is applied to a real-world application, namely the flood control planning formulated as CCP.

Keywords: chance constrained problem; data reduction; differential evolution; big data

1. Introduction

In real-world applications, a wide range of uncertainties have to be taken into account. Therefore, optimization problems under uncertainties have been studied for many years. Generally speaking, there are two types of formulations for handling uncertainties in optimization problems. The first one is the deterministic optimization problem [1]. The second one is the stochastic optimization problem [2]. The robust optimization problem is a well-known deterministic formulation [1]. The robust optimization problem always considers the worst-case performance under uncertainties. Therefore, the overestimation of uncertainties may lead to a conservative decision in practice.

The Chance Constrained Problem (CCP), which is also referred to as the probabilistic constrained problem [3], is one of the possible formulations of the stochastic optimization problem. CCP is a risk-averse formulation of problem under uncertainties. Specifically, CCP ensures that the probability of meeting all constraints is above a certain level. Since the balance between optimality and reliability can be designated by CCP, many real-world applications have been formulated as CCPs [3–5].

CCP has been studied in the field of stochastic programming for many years [2]. In stochastic programming, the optimization methods of the nonlinear programming [6] have been used to solve CCP. Recently, Evolutionary Algorithms (EAs) have also been reported for solving CCPs [7–9]. However, in the conventional formulation of CCP, a well known probability distribution such as the normal distribution is used widely as a mathematical model of unknown uncertainties. Then, the pseudo data generated randomly by using the Monte Carlo method based on the mathematical model are used to represent uncertainties [10]. In some cases, the mathematical model is used to derive a deterministic formulation of CCP [2,11]. Otherwise, very few data, or scenarios, observed

actually are used to represent uncertainties. As a drawback of the conventional formulation of CCP, the estimation error of uncertainties is unavoidable in the evaluation of solutions. In other words, if CCP is defined incompletely, the solution is also defective. Then, we cannot enjoy the benefit of CCP.

In recent years, due to advanced information technologies such as Wireless Sensor Networks (WSN) and Internet of Things (IoT) [12], huge data sets called “big data” have been easily obtained in various fields including culture, science, and industry [13]. In many real-world applications, the variance of observed data is caused by some uncertainties. These applications are probably formulated as CCP more accurately by using a large data set instead of the mathematical model.

In this paper, CCP was formulated by using a large data set called a full data set. However, we assumed that the full data set is too large to solve CCP practically. Therefore, in order to evaluate solutions of CCP, we had to reduce the size of the full data set. Clustering is a popular data reduction technique [14]. Clustering divides a data set into some subsets in order to meet two requirements: “Internal cohesion” and “External isolation”. As a drawback of clustering, the result of clustering depends on the structure of data. Moreover, it is not good at dealing with a huge data set [15].

Sampling is another technique of data reduction. In particular, Simple Random Sampling (SRS) is widely used due to its easy execution and simplicity [16]. SRS selects a few samples randomly from a huge data set and discards most of data. As a drawback of SRS, the key information in many data is likely to be lost. Therefore, a new data reduction method called Weighted Stratified Sampling (WSS) has been proposed by authors to use the full data set completely for solving CCP [17].

By using the new data reduction method called WSS, the above CCP based on the full data set is converted into a relaxation problem of CCP. In order to solve the relaxation problem of CCP efficiently, a new optimization method based on Differential Evolution (DE) [18] is also contrived in this paper. In the new optimization method, a pruning technique is introduced into an adaptive DE [19] for reducing the number of candidate solutions to be examined on the process of the optimization.

The proposed approach is applied to a real-world application, namely the flood control planning formulated as CCP [5]. In addition to the conventional reservoir, the water-retaining capacity of the forest is considered in the flood control planning. Incidentally, various reservoir systems have been studied for protecting a downstream area of river from flood damage [20–22]. Even though historical data are used in these studies, many of them have been limited to dealing with problems of deterministic formulation. A stochastic formulation such as CCP is generally a more realistic representation of the flood control planning because stream flows have randomness and are stochastic in nature.

This paper is an extended version of the paper presented in ICIST2019 [17] and differs from the conference paper in the following three points: (1) The necessary sample size for SRS is derived theoretically. Then, it is shown that the theoretical sample size is too large in practice; (2) By using larger data sets, the performance of WSS is examined more intensively by comparison with SRS. Then, it is proven that WSS outperforms SRS in the accuracy of the estimated probability; (3) The ability of the pruning technique to reduce the run time of the adaptive DE is evaluated. Then, it is shown that the effect of the pruning technique increases proportionally to the sample size of WSS.

The remainder of this paper is organized as follows. Section 2 formulates CCP from a full data set. Section 3 explains two data reduction methods, namely the conventional SRS and the proposed WSS. By using a data reduction method, a relaxation problem of CCP is also derived. Section 4 proposes an adaptive DE combined with a pruning technique for solving the relaxation problem of CCP efficiently. Section 5 examines the performance of WSS intensively by comparison with SRS. Section 6 applies the proposed approach to a real-world application, namely the flood control planning formulated as CCP. Section 7 evaluates the ability of the pruning technique to reduce the run time of the adaptive DE on a personal computer. Finally, Section 8 concludes this paper and provides future work.

2. Problem Formulation

In this paper, uncertainties are represented by a vector of random variables $\xi \in \Omega$ with a sample space $\Omega \subseteq \mathfrak{R}^K$. The collection of vectors $\xi \in \Omega \subseteq \mathfrak{R}^K$ is independent and identically distributed (i.i.d.).

It is not necessarily that the collection of the elements $\xi_j \in \mathfrak{R}$ of each vector $\xi \in \mathfrak{R}^K$ is i.i.d. We suppose that both the sample space $\Omega \subseteq \mathfrak{R}^K$ and the distribution of $\xi \in \Omega$ are unknown. However, a full data set $\mathbf{B} \subseteq \Omega$ including a huge number of data $\xi^\ell \in \mathbf{B}$ is available. Incidentally, the full data set can be regarded as a kind of “big data”. Symbols used in this paper are defined as follows:

- A vector of decision variables $x = (x_1, \dots, x_D) \in \mathbf{X} \subseteq \mathfrak{R}^D$
- A vector of random variables $\xi = (\xi_1, \dots, \xi_K) \in \Omega \subseteq \mathfrak{R}^K$
- A huge number of data, or a full data set $\xi^\ell \in \mathbf{B} \subseteq \Omega$
- Measurable function for constraints $g_m : \mathbf{X} \times \Omega \rightarrow \mathfrak{R}, m = 1, \dots, M$
- Objective function to be minimized $f : \mathbf{X} \rightarrow \mathfrak{R}$
- Sufficiency level given by a probability $\alpha \in (0, 1)$.

2.1. Chance Constrained Problem (CCP)

Let $\Pr(\mathcal{A})$ be the probability that an event \mathcal{A} occurs. The joint probability meeting all probabilistic constraints $g_m(x, \xi) \leq 0, m = 1, \dots, M$ for a solution $x \in \mathbf{X}$ is described as

$$p(x, \Omega) = \Pr(\forall \xi \in \Omega : g_m(x, \xi) \leq 0, m = 1, \dots, M). \quad (1)$$

By using the probability in (1), CCP is formulated as

$$\min_{x \in \mathbf{X}} f(x) \quad \text{sub. to} \quad p(x, \Omega) \geq \alpha \quad (2)$$

where a sufficiency level $\alpha \in (0, 1)$ is given by an arbitrary probability.

In real-world applications, both the sample space $\Omega \subseteq \mathfrak{R}^K$ and the distribution of $\xi \in \Omega$ are usually unknown. Therefore, it is impossible to solve CCP in (2) directly.

2.2. Equivalence Problem of CCP

As stated above, due to advanced information technology, we can suppose that a huge data set $\mathbf{B} = \{\xi^\ell\} \subseteq \Omega$ is available for estimating the unknown probability $p(x, \Omega)$ in (1) empirically.

First of all, the indicator function is defined as

$$\mathbb{1}(g_m(x, \xi^\ell) \leq 0) = \begin{cases} 1 & \text{if } g_m(x, \xi^\ell) \leq 0 \\ 0 & \text{otherwise.} \end{cases} \quad (3)$$

By using the indicator function in (3), the unknown probability $p(x, \Omega)$ in (1) can be evaluated empirically from a huge number of data $\xi^\ell \in \mathbf{B}$ for a solution $x \in \mathbf{X}$ as

$$\hat{p}(x, \mathbf{B}) = \frac{1}{|\mathbf{B}|} \sum_{\xi^\ell \in \mathbf{B}} \left(\prod_{m=1}^M \mathbb{1}(g_m(x, \xi^\ell) \leq 0) \right) \quad (4)$$

where $|\mathbf{B}|$ denotes the size of the data set \mathbf{B} , or the total number of data $\xi^\ell \in \mathbf{B}$.

From the law of large numbers [23], we can expect that $p(x, \Omega) \simeq \hat{p}(x, \mathbf{B})$ holds. Therefore, by using the empirical probability in (4), CCP in (2) can be rewritten as

$$\min_{x \in \mathbf{X}} f(x) \quad \text{sub. to} \quad \hat{p}(x, \mathbf{B}) \geq \alpha. \quad (5)$$

It is difficult to solve CCP in (5). In real-world applications, the function value $g_m(x, \xi^\ell)$ in (4) has to be evaluated for each of the data $\xi^\ell \in \mathbf{B}$ through a time-consuming computer simulation. Thus, the full data set \mathbf{B} is too large to evaluate the empirical probability in (4). In order to solve CCP in (5) practically, we need to reduce the number of data $\xi^\ell \in \mathbf{B}$ included in the full data set.

3. Data Reduction Methods

In order to reduce the number of data $\zeta^\ell \in \mathbf{B}$ included in the full data set $\mathbf{B} \subseteq \mathfrak{R}^K$, we explain two data reduction methods. Firstly, Simple Random Sampling (SRS) is the most popular method [24]. SRS is widely used due to its easy execution and simplicity [25]. Secondly, Weighted Stratified Sampling (WSS) is a new data reduction method proposed by authors in ICIST 2019 [17].

3.1. Simple Random Sampling (SRS)

3.1.1. Procedure of SRS

Some samples $\zeta^n \in \mathbf{B}$, $n = 1, \dots, N$ are selected randomly from the full data set for making a sample set $\Xi = \{\zeta^1, \dots, \zeta^N\} \subseteq \mathbf{B}$. The sample size N is far smaller than the data size $|\mathbf{B}|$. By using samples $\zeta^n \in \Xi$, $n = 1, \dots, N$ selected by SRS, an empirical probability is calculated as

$$\hat{p}(\mathbf{x}, \Xi) = \frac{1}{N} \sum_{n=1}^N \left(\prod_{m=1}^M \mathbb{1}(g_m(\mathbf{x}, \zeta^n) \leq 0) \right). \quad (6)$$

The empirical probability $\hat{p}(\mathbf{x}, \Xi)$ in (6) can be used to approximate $\hat{p}(\mathbf{x}, \mathbf{B})$ in (4).

3.1.2. Theoretical Sample Size

Supposing that $\hat{p}(\mathbf{x}, \mathbf{B}) \simeq p(\mathbf{x}, \mathbf{\Omega})$ holds, we can estimate theoretically a necessary sample size for SRS. Let $\mathbf{x} \in \mathbf{X}$ be a solution of CCP in (5). By using random samples $\zeta^n \in \Xi \subseteq \mathbf{B}$, $n = 1, \dots, N$, the empirical probability $\hat{p}(\mathbf{x}, \Xi)$ in (6) is evaluated from $y_n \in \{0, 1\}$ defined as

$$y_n = \prod_{m=1}^M \mathbb{1}(g_m(\mathbf{x}, \zeta^n) \leq 0). \quad (7)$$

The values of $y_n \in \{0, 1\}$ in (7), which depend on random samples $\zeta^n \in \Xi$, $n = 1, \dots, N$, are regarded as random variables obeying a binomial distribution. From the central limit theorem, the binomial distribution can be approximated by a normal distribution [23]. Therefore, for any $\epsilon \in (0, 1)$ and $\delta \in (0, 1)$, the confidence interval of the unknown probability $\hat{p}(\mathbf{x}, \mathbf{B})$ in (4) is described as

$$\Pr(|\hat{p}(\mathbf{x}, \mathbf{B}) - \hat{p}(\mathbf{x}, \Xi)| \leq \epsilon) \geq 1 - \delta. \quad (8)$$

Let $\Phi : \mathfrak{R} \rightarrow [0, 1]$ be the Cumulative Distribution Function (CDF) of the standard normal distribution. From (8), the risk of failure $\delta \in (0, 1)$ is derived as

$$\delta \geq 2\Phi\left(-\epsilon \sqrt{\frac{N}{\hat{p}(\mathbf{x}, \mathbf{B})(1 - \hat{p}(\mathbf{x}, \mathbf{B}))}}\right). \quad (9)$$

From (9), the margin of error $\epsilon \in (0, 1)$ is also derived as

$$\epsilon \geq z_{\delta/2} \sqrt{\frac{\hat{p}(\mathbf{x}, \mathbf{B})(1 - \hat{p}(\mathbf{x}, \mathbf{B}))}{N}} \quad (10)$$

where $z_{\delta/2} \in [0.5, \infty)$ is defined as $z_{\delta/2} = -\Phi^{-1}(\delta/2)$.

From (10), the sample size N necessary for the confidence interval in (8) is derived as

$$N \geq \left(\frac{z_{\delta/2}}{\epsilon}\right)^2 \hat{p}(\mathbf{x}, \mathbf{B})(1 - \hat{p}(\mathbf{x}, \mathbf{B})). \quad (11)$$

Figure 1 shows the margin of error given by (10) for the sample size N and the value $p \in (0, 1)$ of $\hat{p}(\mathbf{x}, \mathbf{B})$. Figure 2 shows the sample size given by (11) for the value of $\hat{p}(\mathbf{x}, \mathbf{B})$ and $\epsilon \in (0, 1)$.

Even if the value of $\hat{p}(x, \mathbf{B})$ in (11) is unknown, the right side of the formula (11) takes its biggest value when $\hat{p}(x, \mathbf{B}) = 0.5$. Therefore, the necessary sample size for SRS is obtained from (11) as

$$N \geq \frac{1}{4} \left(\frac{z_{\delta/2}}{\varepsilon} \right)^2. \quad (12)$$

Unfortunately, the sample size of SRS obtained theoretically by (12) is too large in practice. For example, the sample size in (12) becomes $N = 16,587$ for $\delta = 0.01$ and $\varepsilon = 0.01$.

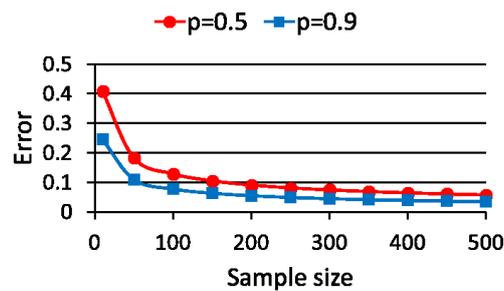


Figure 1. Error for sample size.

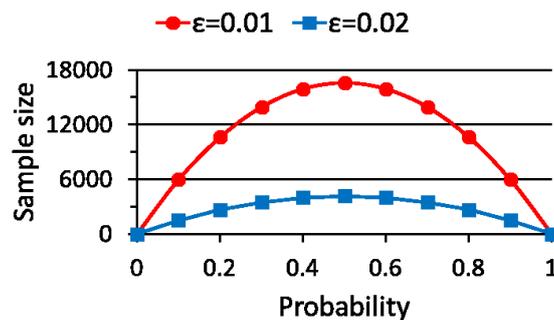


Figure 2. Sample size for probability.

3.2. Weighted Stratified Sampling (WSS)

Since every sample $\xi^n \in \Xi \subseteq \mathbf{B}$ of SRS has the same probability of being chosen from the full data set \mathbf{B} , few samples $\xi^n \in \Xi$ are taken from the sparse part of \mathbf{B} . In addition to the defect of SRS, the majority of data $\xi^l \in \mathbf{B}$ are neglected by SRS even though the full data set \mathbf{B} is very large.

In order to use the full data set \mathbf{B} completely, we proposed a new data reduction method called WSS [17]. In the sampling technique called stratified sampling, which is often used in statistical survey [24], a data set is divided into some strata, or homogeneous subsets. This process is called stratification. Then, a few samples are selected from every stratum. Contrarily to clustering techniques used to divide a data set into some subsets [24], the stratification requires only “Internal cohesion” but not “External isolation”. Therefore, we can choose an arbitrary number of strata for WSS.

3.2.1. Procedure of WSS

A sample set $\Theta \subseteq \mathfrak{R}^K$ is generated from the full data set $\mathbf{B} \subseteq \mathfrak{R}^K$ as follows:

Step 1: By using a K -dimensional histogram, the full data set $\mathbf{B} \subseteq \mathfrak{R}^K$ is divided exclusively into some strata $\mathbf{B}_n \subseteq \mathbf{B}$, $n = 1, \dots, N$ as

$$\mathbf{B} = \mathbf{B}_1 \cup \dots \cup \mathbf{B}_n \cup \dots \cup \mathbf{B}_N \quad (13)$$

where $\mathbf{B}_n \subseteq \mathfrak{R}^K$ and $\mathbf{B}_n \neq \emptyset$, $n = 1, \dots, N$.

Specifically, the K -dimensional histogram is a K -dimensional hypercube that contains the full data set $\mathbf{B} \subseteq \mathfrak{R}^K$. On each side of the K -dimensional hypercube, the entire range of data is divided into a series of intervals. In this paper, the number of intervals is the same on all sides. Moreover, all intervals on each side have equal widths. Therefore, the K -dimensional histogram is an equi-width histogram [26]. Each bin of the K -dimensional histogram is also a K -dimensional hypercube. Then, nonempty bins are used to define strata $\mathbf{B}_n \subseteq \mathbf{B}$, $n = 1, \dots, N$.

Step 2: A new sample point $\theta^n \in \mathfrak{R}^K$ is generated for each stratum $\mathbf{B}_n \subseteq \mathfrak{R}^K$, $n = 1, \dots, N$. Then, the sample set $\Theta \subseteq \mathfrak{R}^K$ is defined as $\Theta = \{\theta^1, \dots, \theta^n, \dots, \theta^N\}$.

Step 3: The weight $w_n \in \mathfrak{R}$ of each sample $\theta^n \in \Theta$ is given by the size of \mathbf{B}_n as $w_n = |\mathbf{B}_n|$.

By using a set of samples $\theta^n \in \Theta$, $n = 1, \dots, N$ and their weights $w_n \in \mathfrak{R}$ obtained by WSS, an empirical probability is calculated to approximate $\hat{p}(x, \mathbf{B})$ in (4) as

$$\hat{p}(x, \Theta) = \frac{1}{W} \sum_{n=1}^N w_n \left(\prod_{m=1}^M \mathbb{1}(g_m(x, \theta^n) \leq 0) \right) \quad (14)$$

where $W = w_1 + \dots + w_n + \dots + w_N$ and $w_n > 0$, $n = 1, \dots, N$.

3.2.2. Sample Generation by WSS

Each of the samples $\theta_n \in \Theta$ of WSS is not necessary to be a sample from the stratum $\mathbf{B}_n \subseteq \mathbf{B}$. For generating samples $\theta_n \in \Theta$, $n = 1, \dots, N$ in Step 2, we think about the optimality of the sample set $\Theta \subseteq \mathfrak{R}^K$. The best sample set $\Theta \subseteq \mathfrak{R}^K$ minimizes the error metric of histogram [27] defined as

$$e_H(\Theta, \mathbf{B}) = \sum_{n=1}^N e_H(\theta^n, \mathbf{B}_n) = \sum_{n=1}^N \sum_{\zeta^\ell \in \mathbf{B}_n} (\theta^n - \zeta^\ell)^2. \quad (15)$$

In order to minimize the error metric of histogram in (15), we will solve the following differential equation about $\theta^n \in \Theta$ because each term $e_H(\theta^n, \mathbf{B}_n)$ in (15) is convex.

$$\frac{\partial e_H(\theta^n, \mathbf{B}_n)}{\partial \theta^n} = 2 \left(|\mathbf{B}_n| \theta^n - \sum_{\zeta^\ell \in \mathbf{B}_n} \zeta^\ell \right) = 0. \quad (16)$$

By solving the differential equation in (16), we can obtain the optimal sample $\theta^n \in \mathfrak{R}^K$ as

$$\theta^n = \frac{1}{|\mathbf{B}_n|} \sum_{\zeta^\ell \in \mathbf{B}_n} \zeta^\ell. \quad (17)$$

Consequently, in Step 2 of WSS, we have only to generate a new sample $\theta^n \in \mathfrak{R}^K$ by using the average point of all data $\zeta^\ell \in \mathbf{B}_n$ included in the stratum $\mathbf{B}_n \subseteq \mathbf{B}$ as shown in (17).

3.3. Relaxation Problems of CCP

As stated above, the equivalence problem of CCP in (5) is hard to solve. That is because the full data set \mathbf{B} is too large. Therefore, a data reduction method, namely SRS or WSS, is employed to formulate a relaxation problem of CCP. By using $\hat{p}(x, \Xi)$ in (6) or $\hat{p}(x, \Theta)$ in (14) to approximate $\hat{p}(x, \mathbf{B})$ in (4), the relaxation problem of CCP is formulated as

$$\min_{x \in X} f(x) \quad \text{sub. to} \quad \hat{p}(x, \mathbf{S}) \geq \beta \quad (18)$$

where the sample set $\mathbf{S} \subseteq \mathfrak{R}^K$ denotes either $\Xi \subseteq \mathfrak{R}^K$ or $\Theta \subseteq \mathfrak{R}^K$. The correction level $\beta \in (0, 1]$ is chosen as $\beta \geq \alpha$ to compensate the margin of error $\varepsilon \in (0, 1)$ caused by SRS or WSS.

4. Adaptive Differential Evolution with Pruning Technique

Differential Evolution (DE) has been proven to be one of the most powerful global optimization algorithms [28,29]. Unfortunately, cardinal DE [18] is only applicable to unconstrained problems. Moreover, the performance of DE is significantly influenced by its control parameter settings. Therefore, in order to solve the relaxation problem of CCP shown in (18) efficiently, a new optimization method called Adaptive DE with Pruning technique (ADEP) is proposed. In the proposed ADEP, three techniques are introduced into cardinal DE [18]: (1) Adaptive control of parameters [19]; (2) Constraint handling based on feasibility rule [30]; and (3) Pruning technique in selection [17].

4.1. Strategy of DE

As well as cardinal DE [18], ADEP has a set of candidate solutions $x_i \in P_t, i = 1, \dots, NP$ called population in each generation t . Each candidate solution $x_i \in X$ is a vector of decision variables. The initial population $x_i \in P_0 \subseteq X$ is randomly generated according to a uniform distribution.

At each generation t , every $x_i \in P_t$ is assigned to a target vector in turn. By using the basic strategy of DE called "DE/rand/1/bin" [18], a trial vector $z_i \in X$ is generated from the target vector $x_i \in P_t$. Specifically, except for the current target vector $x_i \in P_t$, three other distinct vectors, say x_{r1}, x_{r2} , and $x_{r3}, i \neq r1 \neq r2 \neq r3$, are selected randomly from the population P_t . By using the three vectors, the differential mutation generates a new real vector $v_i \in \mathbb{R}^D$ called mutated vector as

$$v_i = x_{r1} + F(x_{r2} - x_{r3}) \quad (19)$$

where $F \in [0.1, 1]$ is a control parameter called scale factor.

The binomial crossover between the mutated vector $v_i \in \mathbb{R}^D$ and the target vector $x_i \in P_t$ generates another real vector $z_i = (z_{1,i}, \dots, z_{D,i}) \in X$ called trial vector. Specifically, each component $z_{j,i} \in \mathbb{R}, j = 1, \dots, D$ of the trial vector $z_i \in X$ is inherited from either $v_i \in X$ or $x_i \in P_t$ as

$$z_{j,i} = \begin{cases} v_{j,i} & \text{if } \text{rand}_j \leq CR \vee j = j_r \\ x_{j,i} & \text{otherwise} \end{cases} \quad (20)$$

where $CR \in [0, 1]$ is a control parameter called crossover rate. $\text{rand}_j \in [0, 1]$ denotes a uniformly distributed random value. The subscript $j_r \in [1, D]$ is selected randomly every time, which ensures that the newborn vector $z_i \in X$ differs from the existing one $x_i \in P_t$ at least one element.

4.2. Adaptive Control of Parameters

The performance of DE depends on control parameters, namely the scale factor $F \in [0.1, 1]$ in (19) and the crossover rate $CR \in [0, 1]$ in (20). Therefore, various parameter adaptation mechanisms have been reported [28,29,31]. ADEP employs an adaptive parameter control mechanism in which feedback from the evolutionary search is used to dynamically change the control parameters [19].

According to the adaptive parameter control mechanism [19], all vectors $x_i \in P_t, i = 1, \dots, NP$ have their own control parameters, namely $F_{i,t} \in [0.1, 1]$ and $CR_{i,t} \in [0, 1]$, at each generation t . Then, these control parameters are initialized as $F_{i,0} = 0.5$ and $CR_{i,0} = 0.9, i = 1, \dots, NP$.

For generating the mutated vector $v_i \in \mathbb{R}^D$ as shown in (19), the scale factor $F \in \mathbb{R}$ is decided by using the control parameter $F_{i,t} \in [0.1, 1]$ associated with the target vector $x_i \in P_t$ as

$$F = \begin{cases} 0.1 + \text{rand}_1 \cdot 0.9 & \text{if } \text{rand}_2 < 0.1 \\ F_{i,t} & \text{otherwise} \end{cases} \quad (21)$$

where $\text{rand}_j \in [0, 1], j = 1, 2$ are uniformly distributed values.

Similarly, for generating the trial vector $u_i \in \mathbb{R}^D$ as shown in (20), the crossover rate $CR \in \mathbb{R}$ is decided by using the control parameter $CR_{i,t} \in [0, 1]$ associated with the target vector $x_i \in P_t$ as

$$CR = \begin{cases} \text{rand}_3 & \text{if } \text{rand}_4 < 0.1 \\ CR_{i,t} & \text{otherwise} \end{cases} \quad (22)$$

where $\text{rand}_j \in [0, 1]$, $j = 3, 4$ are uniformly distributed values.

The trial vector $z_i \in X$ generated by using $F \in [0.1, 1]$ in (21) and $CR \in [0, 1]$ in (22) is compared with the target vector $x_i \in P_t$ as described below. If $z_i \in X$ is better than $x_i \in P_t$, $z_i \in X$ is selected for a new vector $x_i \in P_{t+1}$ of the next generation, and the control parameters of $x_i \in P_{t+1}$ are decided as $F_{i,t+1} = F$ and $CR_{i,t+1} = CR$. Otherwise, the target vector $x_i \in P_t$ survives to the next generation, and the control parameters of $x_i \in P_{t+1}$ are inherited from $x_i \in P_t$ as $F_{i,t+1} = F_{i,t}$ and $CR_{i,t+1} = CR_{i,t}$.

4.3. Constraint Handling and Pruning Technique in Selection

Evolutionary Algorithms (EAs) including DE are typically applied to problems in which bounds are the only constraints. Therefore, a number of Constraint Handling Techniques (CHTs) have been proposed in order to apply EAs to constrained optimization problems [32]. Among those CHTs, the feasibility rule [30] is one of the most widely used CHTs because of its simplicity and efficiency. Thus, ADEP uses a feasibility rule with the amount of constraint violation defined from (18) as

$$h(x_i) = \max\{\beta - \hat{p}(x_i, S), 0\} \quad (23)$$

where the candidate solution $x_i \in P_t$ of CCP in (18) is feasible if $h(x_i) = 0$ holds.

At each generation t , the trial vector $z_i \in X$, $i = 1, \dots, NP$ is compared with the corresponding target vector $x_i \in P_t$. Then, either the trial vector $z_i \in X$ or the target vector $x_i \in P_t$ is selected for a vector $x_i \in P_{t+1}$ of the next generation. First of all, if the following condition is satisfied,

$$(h(x_i) = 0) \wedge (f(x_i) < f(z_i)) \quad (24)$$

the trial vector $z_i \in X$ is discarded immediately because $x_i \in P_t$ is better than $z_i \in X$. Then, the trial vector $x_i \in P_t$ survives to the next generation. Since the pruning technique based on the condition in (24) does not require the value of $h(z_i)$, it is very effective to reduce the run time of ADEP.

Only when the condition in (24) is not satisfied, the probability $\hat{p}(z_i, S)$ is evaluated by using the sample set $S \subseteq \mathbb{R}^K$ to get the value of $h(z_i)$. If either of the following conditions is satisfied,

$$\begin{cases} h(z_i) < h(x_i) \\ (h(z_i) = h(x_i)) \wedge (f(z_i) \leq f(x_i)) \end{cases} \quad (25)$$

the trial vector $z_i \in X$ is selected for a new vector $x_i \in P_{t+1}$ of the next generation. Otherwise, the current target vector $x_i \in P_t$ survives to the next generation and becomes a vector $x_i \in P_{t+1}$.

4.4. Proposed Algorithm of ADEP

The algorithm of ADEP is described as follows. The maximum number of generations NT is given as the termination condition. The population size NP is chosen as $NP = 5D$ [18].

Step 1: Randomly generate the initial population $x_i \in P_0 \subseteq X$, $i = 1, \dots, NP$. $t = 0$.

Step 2: For $i = 1$ to NP , evaluate $f(x_i)$ and $h(x_i)$ for each vector $x_i \in P_0$.

Step 3: If $t = NT$ holds, output the best solution $x_b \in P_t$ and terminate ADEP.

Step 4: For $i = 1$ to NP , generate the trial vector $z_i \in X$ from the target vector $x_i \in P_t$.

Step 5: For $i = 1$ to NP , evaluate $f(z_i)$ for the trial vector $z_i \in X$.

Step 6: For $i = 1$ to NP , evaluate $h(z_i)$ for $z_i \in X$ only if the condition in (24) is not satisfied.

Step 7: For $i = 1$ to NP , select either $z_i \in X$ or $x_i \in P_t$ for $x_i \in P_{t+1}$. $t = t + 1$.

Step 8: Go back to Step 3.

5. Performance Evaluation of WSS

We evaluate the performance of the proposed WSS by comparison with the conventional SRS. Specifically, by using WSS and SRS, we estimate the probability meeting $g(\mathbf{x}, \boldsymbol{\zeta}^\ell) \leq \gamma$ as

$$\hat{p}(\mathbf{x}, \mathbf{B}) = \frac{1}{|\mathbf{B}|} \sum_{\boldsymbol{\zeta}^\ell \in \mathbf{B}} \mathbb{1}(g(\mathbf{x}, \boldsymbol{\zeta}^\ell) \leq \gamma) \quad (26)$$

where $g : \mathbf{X} \times \boldsymbol{\Omega} \rightarrow \mathfrak{R}$ is a measurable function.

From (9) to (11), the performance of SRS depends on the value of the probability $\hat{p}(\mathbf{x}, \mathbf{B})$ to be estimated. Therefore, by changing the value of $\gamma \in \mathfrak{R}$ in (26), we change the value of the probability to be estimated by SRS and WSS. From a full data set $\mathbf{B} \subseteq \mathfrak{R}^K$, the sample sets $\Xi \subseteq \mathfrak{R}^K$ and $\Theta \subseteq \mathfrak{R}^K$ are generated, respectively, by using SRS and WSS. Then, the estimation error is defined as

$$\varepsilon_p(\mathbf{S}, \mathbf{B}) = |\hat{p}(\mathbf{x}, \mathbf{S}) - \hat{p}(\mathbf{x}, \mathbf{B})| \quad (27)$$

where the sample set $\mathbf{S} \subseteq \mathfrak{R}^K$ denotes either $\Xi \subseteq \mathfrak{R}^K$ or $\Theta \subseteq \mathfrak{R}^K$.

5.1. Case Study 1

Each value $\xi_i \in \mathfrak{R}$ of practical data $\boldsymbol{\zeta} \in \boldsymbol{\Omega}$ in (2) usually has a range. Therefore, a full data set $\boldsymbol{\zeta}^\ell = (\zeta_1, \zeta_2) \in \mathbf{B} \subseteq \mathfrak{R}^2$ is generated randomly by using a truncated normal distribution as

$$\zeta_j \sim \mathcal{N}(\mu_j, \sigma_j^2, \mu_j - 3\sigma_j, \mu_j + 3\sigma_j), \quad j = 1, 2 \quad (28)$$

where the mean is $\boldsymbol{\mu} = (\mu_1, \mu_2) = (1, 2)$ and the variance is $\boldsymbol{\sigma}^2 = (\sigma_1^2, \sigma_2^2) = (0.1^2, 0.2^2)$.

The truncated normal distribution in (28) is the probability distribution derived from that of a normally distributed random variable by bounding the random variable as $\zeta_j \in [\mu_j - 3\sigma_j, \mu_j + 3\sigma_j]$. The correlation matrix of the random variables $\zeta_j \in \mathfrak{R}$, $j = 1, 2$ is also given as

$$\mathbf{R} = \begin{pmatrix} 1.0 & \rho_{12} \\ \rho_{21} & 1.0 \end{pmatrix} = \begin{pmatrix} 1.0 & -0.8 \\ -0.8 & 1.0 \end{pmatrix}. \quad (29)$$

Please notice that the full data set $\boldsymbol{\zeta}^\ell \in \mathbf{B}$ is i.i.d. even if the elements of $\boldsymbol{\zeta}^\ell \in \mathbf{B}$ have a correlation. The size of the full data set $\mathbf{B} \subseteq \mathfrak{R}^2$ generated randomly is $|\mathbf{B}| = 10^7$. Figure 3 shows the spatial patterns of the full data $\boldsymbol{\zeta}^\ell \in \mathbf{B}$ and the random samples $\boldsymbol{\zeta}^n \in \Xi \subseteq \mathbf{B}$, $n = 1, \dots, N$ selected by SRS. Figure 4 also shows the full data $\boldsymbol{\zeta}^\ell \in \mathbf{B}$ and the weighted samples $\boldsymbol{\theta}^n \in \Theta$ generated by WSS.

From Figures 3 and 4, we can see that the weighted samples $\boldsymbol{\theta}^n \in \Theta$, $n = 1, \dots, N$ of WSS are scattered more widely as compared to the random samples $\boldsymbol{\zeta}^n \in \Xi$ of SRS. Especially, SRS has not taken any samples $\boldsymbol{\zeta}^n \in \Xi$ from the sparse part of the full data set $\mathbf{B} \subseteq \mathfrak{R}^2$ in Figure 3.

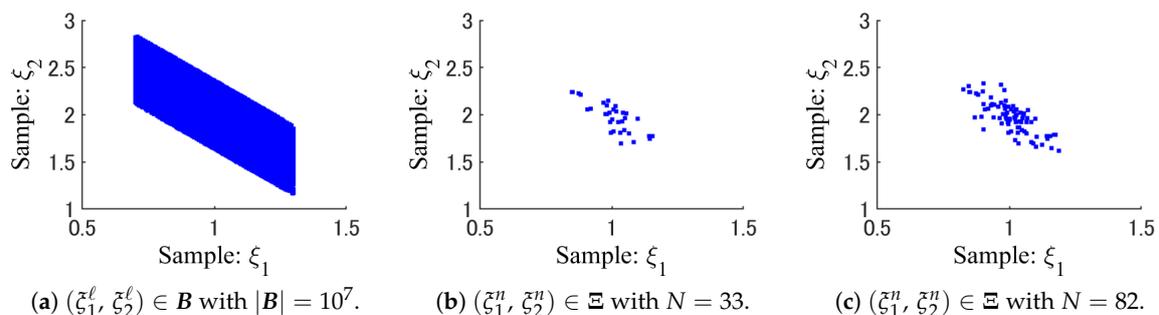


Figure 3. Patterns of the full data $\boldsymbol{\zeta}^\ell \in \mathbf{B} \subseteq \mathfrak{R}^2$ and the random samples $\boldsymbol{\zeta}^n \in \Xi$ selected by SRS.

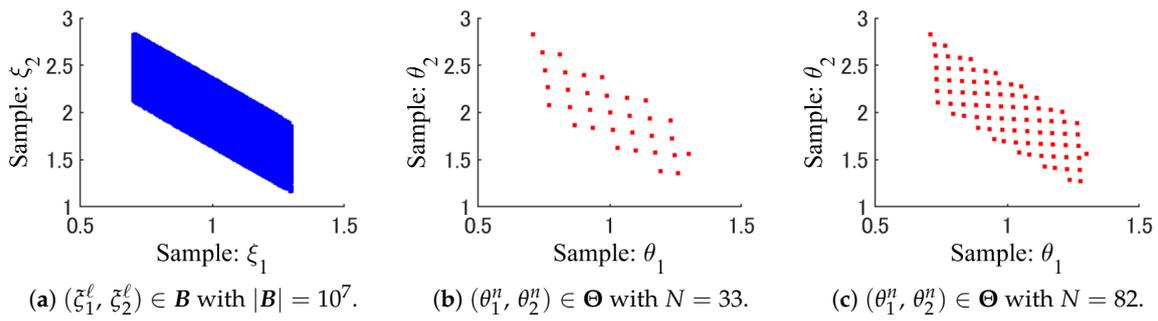


Figure 4. Patterns of the full data $\zeta^\ell \in \mathbf{B} \subseteq \mathfrak{R}^2$ and the weighted samples $\theta^n \in \Theta$ generated by WSS.

For the probability $\hat{p}(x, \mathbf{B})$ in (26), a function $g(x, \zeta^\ell)$ is defined as

$$g(x, \zeta^\ell) = \frac{x_1 \zeta_1 + x_2 \zeta_2}{2} = \frac{\zeta_1 + \zeta_2}{2} \tag{30}$$

where the vector of decision variables is given as $x = (x_1, x_2) = (1, 1)$.

The probability in (26) becomes $\hat{p}(x, \mathbf{B}) \simeq 0.933$ when $\gamma = 1.6$ is chosen. Then, the estimation errors in (27), namely $\varepsilon_p(\Theta, \mathbf{B})$ and $\varepsilon_p(\Xi, \mathbf{B})$, are evaluated 100 times for each sample size by using different full data sets $\mathbf{B} \subseteq \mathfrak{R}^2$ and summarized in Figure 5. From Figure 5, the average value and the standard deviation of $\varepsilon_p(\Theta, \mathbf{B})$ are smaller than those of $\varepsilon_p(\Xi, \mathbf{B})$ for any sample sizes. Furthermore, $\varepsilon_p(\Theta, \mathbf{B})$ converges to almost zero faster than $\varepsilon_p(\Xi, \mathbf{B})$ on average. Consequently, we can say that the proposed WSS outperforms the conventional SRS in the accuracy of the estimated probability.

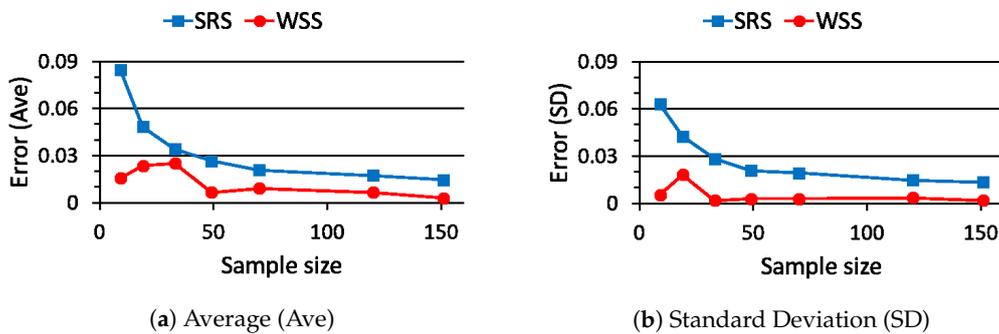


Figure 5. Estimation error for sample size when $\hat{p}(x, \mathbf{B}) \simeq 0.933$ ($K = 2$).

The probability in (26) becomes $\hat{p}(x, \mathbf{B}) \simeq 0.773$ when $\gamma = 1.55$ is chosen. The estimation errors in (27) are evaluated for WSS and SRS with the probability $\hat{p}(x, \mathbf{B}) \simeq 0.773$ as stated above and summarized in Figure 6. The probability in (26) becomes $\hat{p}(x, \mathbf{B}) \simeq 0.500$ when $\gamma = 1.5$ is chosen. The estimation errors in (27) are also evaluated for WSS and SRS with the probability $\hat{p}(x, \mathbf{B}) \simeq 0.500$ and summarized in Figure 7. From Figures 6 and 7, the standard deviation of $\varepsilon_p(\Theta, \mathbf{B})$ is always smaller than that of $\varepsilon_p(\Xi, \mathbf{B})$. On the other hand, in the average value of the estimation errors, WSS seems to lose the advantage over SRS when the probability value gets closer to $\hat{p}(x, \mathbf{B}) = 0.5$.

From Figures 5–7, the performance of WSS depends not only on the sample size but also on the probability value to be estimated. Then, WSS outperforms SRS when the probability value is large. A large value is usually chosen for the probability meeting all constraints in CCP. Consequently, we can say that WSS is more suitable for formulating the relaxation problem of CCP in (18).

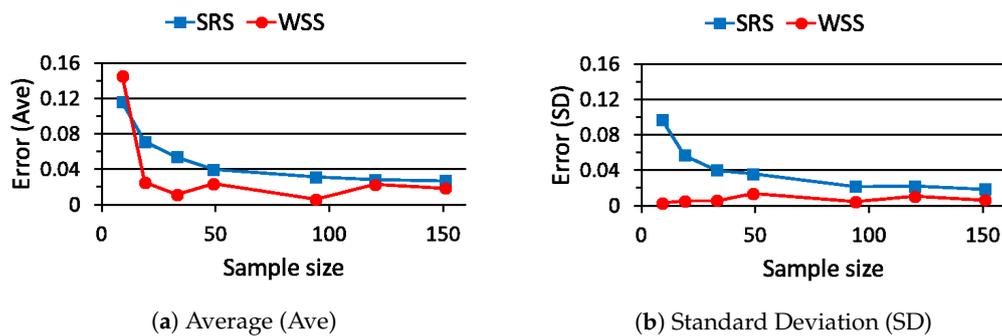


Figure 6. Estimation error for sample size when $\hat{p}(x, B) \simeq 0.773$ ($K = 2$).

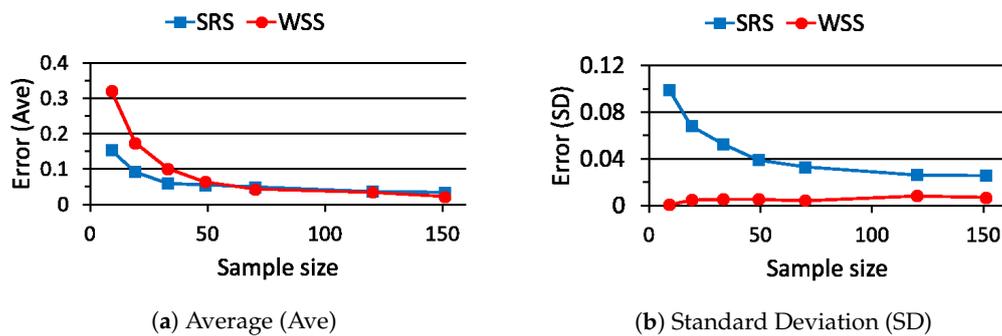


Figure 7. Estimation error for sample size when $\hat{p}(x, B) \simeq 0.500$ ($K = 2$).

5.2. Case Study 2

A full data set $\zeta^\ell = (\zeta_1, \zeta_2, \zeta_3) \in B \subseteq \mathfrak{R}^3$ is generated randomly as

$$\zeta_j \sim \mathcal{N}(\mu_j, \sigma_j^2, \mu_j - 3\sigma_j, \mu_j + 3\sigma_j), j = 1, 2, 3 \quad (31)$$

where $\mu = (\mu_1, \mu_2, \mu_3) = (1.5, 2.0, 1.0)$ and $\sigma^2 = (\sigma_1^2, \sigma_2^2, \sigma_3^2) = (0.2^2, 0.1^2, 0.1^2)$.

The correlation matrix of the random variables $\zeta_j \in \mathfrak{R}, j = 1, 2, 3$ is also given as

$$\mathbf{R} = \begin{pmatrix} 1.0 & \rho_{12} & \rho_{13} \\ \rho_{21} & 1.0 & \rho_{23} \\ \rho_{31} & \rho_{32} & 1.0 \end{pmatrix} = \begin{pmatrix} 1.0 & 0.6 & 0.0 \\ 0.6 & 1.0 & -0.4 \\ 0.0 & -0.4 & 1.0 \end{pmatrix}. \quad (32)$$

The size of the full data set $B \subseteq \mathfrak{R}^3$ generated randomly is $|B| = 10^7$. Figure 8 shows the spatial patterns of the full data $\zeta^\ell \in B$ and the random samples $\zeta^n \in \Xi \subseteq B, n = 1, \dots, N$ selected by SRS. Figure 9 also shows the full data $\zeta^\ell \in B$ and the weighted samples $\theta^n \in \Theta$ generated by WSS.

From Figures 8 and 9, we can see that the weighted samples $\theta^n \in \Theta, n = 1, \dots, N$ of WSS are scattered more widely than the random samples $\zeta^n \in \Xi$ of SRS. Some weighted samples $\theta^n \in \Theta$ shown in Figure 9 seem to be overlapped each other due to the high dimensionality $K = 3$ of them. However, we can recognizable the uniformity in the pattern of the weighted samples $\theta^n \in \Theta$.

For the probability $\hat{p}(x, B)$ in (26), a linear function $g(x, \zeta^\ell)$ is defined as

$$g(x, \zeta^\ell) = \frac{x_1 \zeta_1 + x_2 \zeta_2 + x_3 \zeta_3}{3} = \frac{\zeta_1 + \zeta_2 + \zeta_3}{3} \quad (33)$$

where the vector of decision variables is given as $x = (x_1, x_2, x_3) = (1, 1, 1)$.

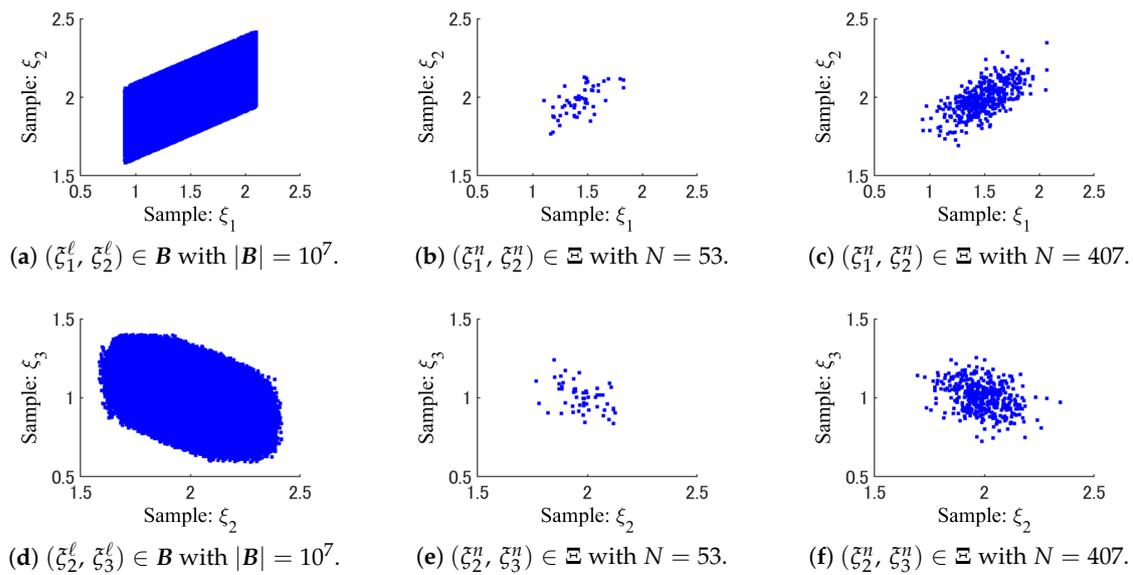


Figure 8. Patterns of the full data $\xi^\ell \in \mathbf{B} \subseteq \mathfrak{R}^3$ and the random samples $\xi^n \in \Xi$ selected by SRS.

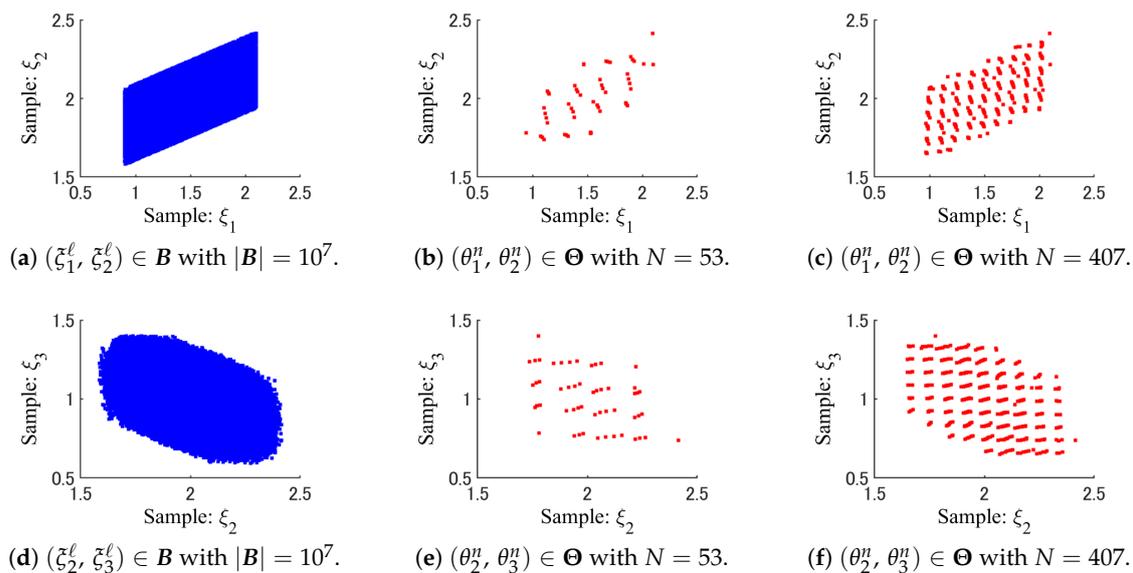


Figure 9. Patterns of the full data $\xi^\ell \in \mathbf{B} \subseteq \mathfrak{R}^3$ and the weighted samples $\theta^n \in \Theta$ generated by WSS.

The probability in (26) becomes $\hat{p}(x, \mathbf{B}) \simeq 0.986$ when $\gamma = 1.7$ is chosen. Then, the estimation errors in (27), namely $\varepsilon_p(\Theta, \mathbf{B})$ and $\varepsilon_p(\Xi, \mathbf{B})$, are evaluated 100 times for each sample size by using different full data sets $\mathbf{B} \subseteq \mathfrak{R}^3$ and summarized in Figure 10. From Figure 10, the average value and the standard deviation of $\varepsilon_p(\Theta, \mathbf{B})$ are smaller than those of $\varepsilon_p(\Xi, \mathbf{B})$ for any sample sizes. From these results, we confirm that the proposed WSS outperforms the conventional SRS in this case.

5.3. Case Study 3

A full data set $\mathbf{B} \subseteq \mathfrak{R}^3$, $|\mathbf{B}| = 10^7$ is generated randomly as shown in (31) and (32). For the probability $\hat{p}(x, \mathbf{B})$ in (26), a non-linear function $g(x, \xi^\ell)$ is defined as

$$g(x, \xi^\ell) = \frac{x_1 \xi_1^2 + x_2 \xi_2^2 + x_3 \xi_3^2}{3} = \frac{\xi_1^2 + \xi_2^2 + \xi_3^2}{3} \tag{34}$$

where the vector of decision variables is given as $x = (x_1, x_2, x_3) = (1, 1, 1)$.

The probability in (26) becomes $\hat{p}(x, \mathbf{B}) \simeq 0.817$ when $\gamma = 1.7$ is chosen. The estimation errors in (27) are evaluated for WSS and SRS with the probability $\hat{p}(x, \mathbf{B}) \simeq 0.817$ as stated above and summarized in Figure 11. From Figure 11, WSS is also better than SRS because the average value and the standard deviation of $\varepsilon_p(\Theta, \mathbf{B})$ are smaller than those of $\varepsilon_p(\Xi, \mathbf{B})$ for any sample sizes.

From Figures 10 and 11, we can see that the margin of error provided by WSS converges fast, and then it does not decrease even if the sample size is increased. Therefore, WSS can estimate the probability accurately with a small sample size if we choose the sample size properly.

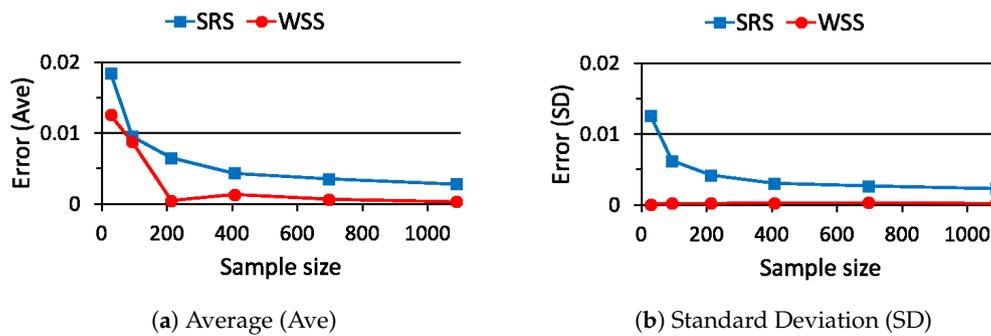


Figure 10. Estimation error for sample size with the linear function in (33).

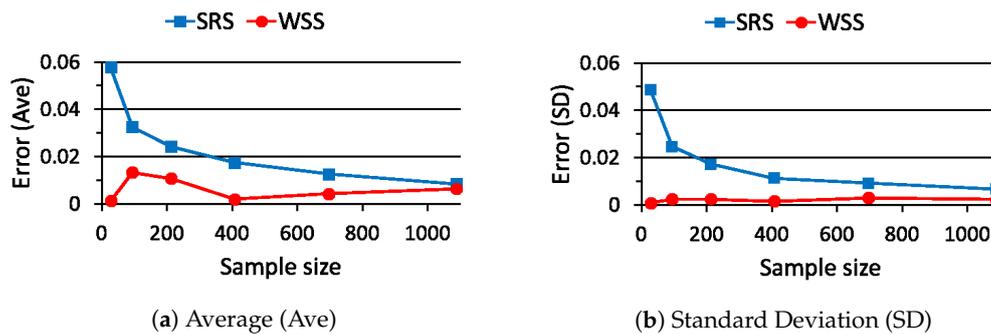


Figure 11. Estimation error for sample size with the non-linear function in (34).

6. Flood Control Planning

6.1. Formulation of CCP

Reservoirs are constructed to protect an urban area at the lower part of river from the flood damage caused by torrential rain. The flood control reservoir system design has been formulated as CCP [33]. In addition to the reservoir, the water-retaining capacity of forest is counted to prevent the flood caused by heavy rainfall. Thereby, the flood control planning is formulated as CCP [5].

Figure 12 shows a topological river model. Symbol \bigcirc denotes a forest considered in the flood control planning. There are three forests in watersheds. The gross area of each forest $a_j, j = 1, 2, 3$ is a constant. The amount of rainfall $\xi_j \in \mathfrak{R}$ per unit area in each of the forests is a random variable. The water-retaining capacity of forest $x_j \in \mathfrak{R}, j = 1, 2, 3$ per unit area is regarded as a decision variable because it can be controlled through the forest maintenance such as afforestation. According to the model of the forest mechanism [34], the inflow of water $q_j \in \mathfrak{R}$ from each forest to the river is

$$q_j(x_j, \xi_j) = a_j (\xi_j - x_j (1 - \exp(-\xi_j/x_j))) \quad (35)$$

where the effect of past rainfall is not considered in the model [34].

Symbol Δ in Figure 12 denotes a reservoir. Three reservoirs are constructed in the river. The capacity of each reservoir $x_j \in \mathfrak{R}$, $j = 4, 5, 6$ is also a decision variable. From $q_j(x_j, \zeta_j) \in \mathfrak{R}$ in (35), the inflow of water from the river to the town located at the lower part of the river is calculated as

$$g(x, \zeta) = \max\{\max\{\max\{q_3 - x_6, 0\} + q_2 - x_5, 0\} + q_1 - x_4, 0\} \quad (36)$$

where $q_j \in \mathfrak{R}$, $j = 1, 2, 3$ denotes $q_j(x_j, \zeta_j) \in \mathfrak{R}$ defined by (35).

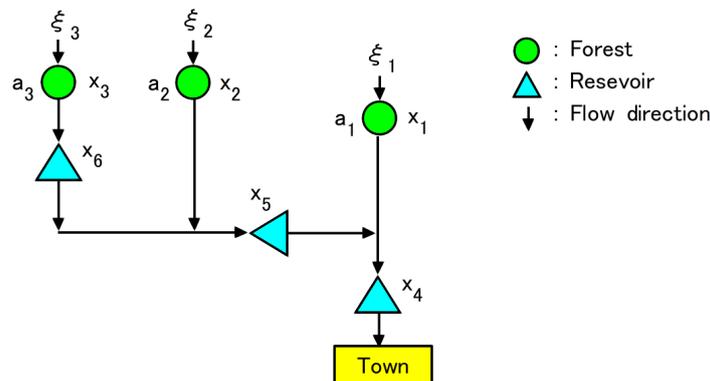


Figure 12. Topological river model.

The probability of meeting $g(x, \zeta) \leq 0$ has to be greater than $\alpha \in (0, 1)$. The maintenance cost of a forest is proportional to its capacity. The construction cost of a reservoir is proportional to the square of its capacity. Then, the flood control planning to minimize the total cost is formulated as

$$\left[\begin{array}{l} \min_{x \in X} \quad f(x) = \sum_{j=1}^3 a_j x_j + \sum_{j=4}^6 x_j^2 \\ \text{sub. to} \quad \Pr(\forall \zeta \in \Omega : g_m(x, \zeta) \leq 0, m = 1, 2, 3) \geq \alpha \\ \quad \quad \quad 0.5 \leq x_1 \leq 1.5, 0.5 \leq x_2 \leq 1.5, 0.5 \leq x_3 \leq 1.5, \\ \quad \quad \quad 0 \leq x_4 \leq 3, 0 \leq x_5 \leq 3, 0 \leq x_6 \leq 4 \end{array} \right. \quad (37)$$

where $a_j = 2$, $j = 1, 2, 3$. From (36), functions $g_m(x, \zeta)$, $m = 1, 2, 3$ are derived as

$$\left[\begin{array}{l} g_1(x, \zeta) = q_1(x_1, \zeta_1) - x_4 \\ g_2(x, \zeta) = q_1(x_1, \zeta_1) + q_2(x_2, \zeta_2) - x_4 - x_5 \\ g_3(x, \zeta) = q_1(x_1, \zeta_1) + q_2(x_2, \zeta_2) + q_3(x_3, \zeta_3) - x_4 - x_5 - x_6. \end{array} \right. \quad (38)$$

6.2. Comparison of SRS and WSS

We suppose that the amount of rainfall $\zeta \in \Omega$ in (37) is given as “big data”. For convenience, the full data set $\zeta^\ell = (\zeta_1, \zeta_2, \zeta_3) \in \mathbf{B} \subseteq \Omega$ is generated randomly by using the truncated normal distribution shown in (31). Please notice that $\zeta^\ell \in \mathbf{B}$ denotes the amount of rainfall in one period, but the river flow depending on an actual time. The inflow of water $q_j \in \mathfrak{R}$ is derived from $\zeta_j \in \mathfrak{R}$ as shown in (35). Therefore, the full data set $\zeta^\ell \in \mathbf{B}$ is i.i.d. On the other hand, a set of $\zeta_j \in \mathfrak{R}$, $j = 1, 2, 3$ is not i.i.d. The correlation matrix of the amounts of rainfalls $\zeta_j \in \mathfrak{R}$, $j = 1, 2, 3$ is given as

$$\mathbf{R} = \begin{pmatrix} 1.0 & \rho_{12} & \rho_{13} \\ \rho_{21} & 1.0 & \rho_{23} \\ \rho_{31} & \rho_{32} & 1.0 \end{pmatrix} = \begin{pmatrix} 1.0 & 0.5 & 0.0 \\ 0.5 & 1.0 & 0.3 \\ 0.0 & 0.3 & 1.0 \end{pmatrix}. \quad (39)$$

The size of the full data set $\mathbf{B} \subseteq \mathbb{R}^3$ generated randomly is $|\mathbf{B}| = 10^7$. Figure 13 shows the spatial patterns of the full data $\zeta^\ell \in \mathbf{B}$, the random samples $\zeta^n \in \Xi \subseteq \mathbf{B}, n = 1, \dots, N$ selected by SRS, and the weighted samples $\theta^n \in \Theta, n = 1, \dots, N$ generated by WSS.

From Figure 13, we can see that the weighted samples $\theta^n \in \Theta, n = 1, \dots, N$ of WSS are scattered more widely than the random samples $\zeta^n \in \Xi, n = 1, \dots, N$ of SRS. Some weighted samples $\theta^n \in \Theta$ of WSS seem to be overlapped each other in Figure 13 due to the high dimensionality $K = 3$ of them. However, we can recognize the uniformity in the pattern of the weighted samples $\theta^n \in \Theta$.

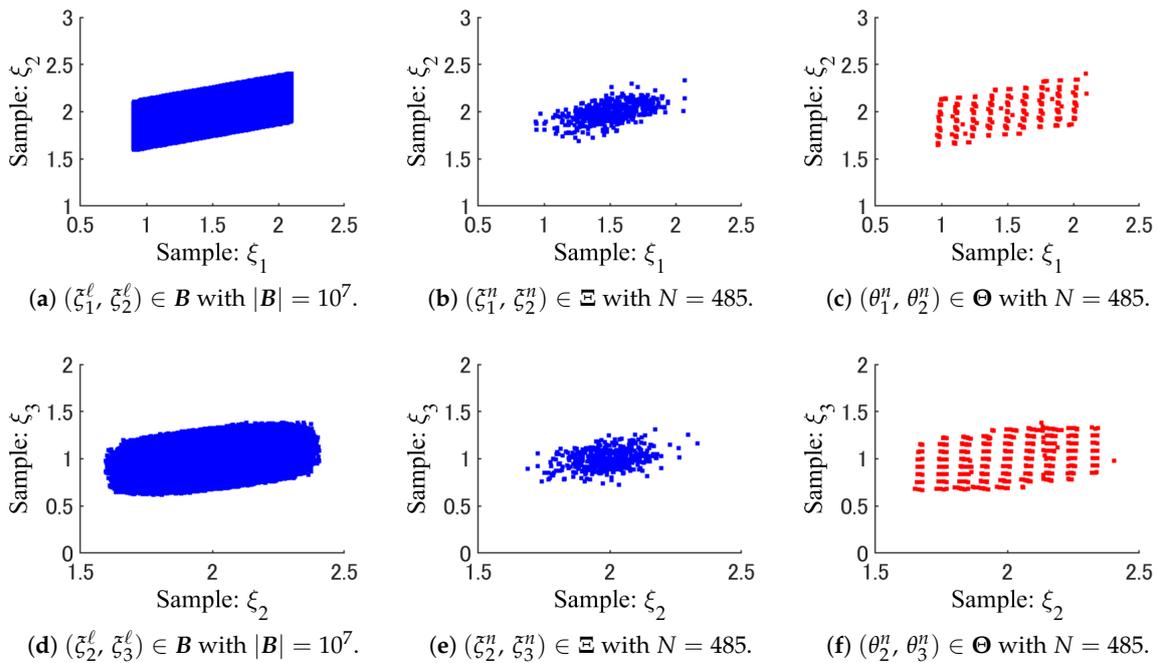


Figure 13. Patterns of the full data set, the random samples of SRS, and the weighted samples of WSS.

The flood control planning formulated as CCP in (37) is transformed into an equivalence problem of CCP as shown in (5) by using the above full data set $\mathbf{B} \subseteq \mathbb{R}^3, |\mathbf{B}| = 10^7$. For a solution $x \in X$ of the equivalence problem of CCP, the joint probability defined by (4) and (37) becomes $\hat{p}(x, \mathbf{B}) \simeq 0.906$. By using SRS and WSS respectively, we estimate the value of the joint probability $\hat{p}(x, \mathbf{B})$.

The estimation errors in (27), namely $\varepsilon_p(\Theta, \mathbf{B})$ and $\varepsilon_p(\Xi, \mathbf{B})$, are evaluated 100 times for each sample size by using different full data sets $\mathbf{B} \subseteq \mathbb{R}^3$ and summarized in Figure 14. From Figure 14, the average value and the standard deviation of $\varepsilon_p(\Theta, \mathbf{B})$ are smaller than those of $\varepsilon_p(\Xi, \mathbf{B})$ for any sample sizes. From these results, we confirm that WSS outperforms SRS in this case too.

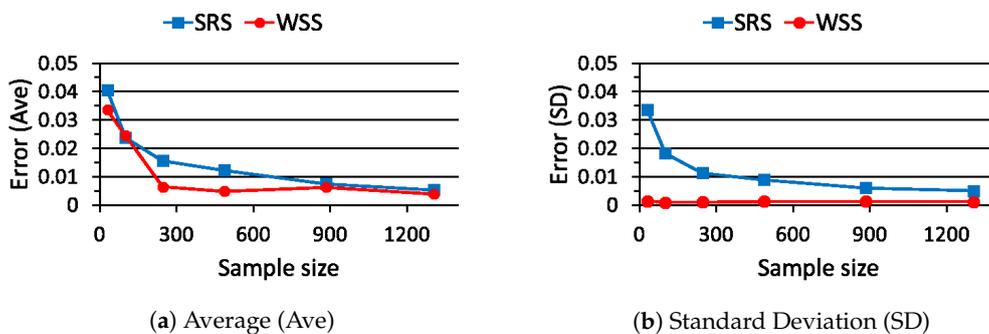


Figure 14. Estimation error for sample size in the flood control planning.

6.3. Solution of CCP

By using the weighted samples $\theta^n \in \Theta$, $n = 1, \dots, N$ of WSS, the flood control planning formulated as CCP in (37) is transformed into a relaxation problem of CCP as shown in (18). That is because the result in Figure 14 shows that WSS estimates the value of $\hat{p}(x, \mathbf{B})$ using fewer samples than SRS. From the average value of $\varepsilon_p(\Theta, \mathbf{B})$ in Figure 14, the sample size is chosen as $N \simeq 482$.

The proposed ADEP is coded in MATLAB [35]. The parameters of ADEP and WSS are chosen as shown in Table 1. As stated above, the sample size is chosen as $N \simeq 482$. For a sufficiency level $\alpha \in (0, 1)$ in (37), the correction level in (18) is chosen as $\beta = \alpha + 0.03$. The population size is decided as $NP = 5D$ by recommendation of the literature [18]. The maximum number of generations is decided as $NT = 80$ through a preliminary experiment. Figure 15 shows the convergence graph of ADEP when the sufficiency level is chosen as $\alpha = 0.9$ or $\alpha = 0.6$. The horizontal axis of Figure 15 is the number of generations. The vertical axis is the best objective function value achieved at each generation. From Figure 15, we can confirm that $NT = 80$ is a sufficiently large number of generations.

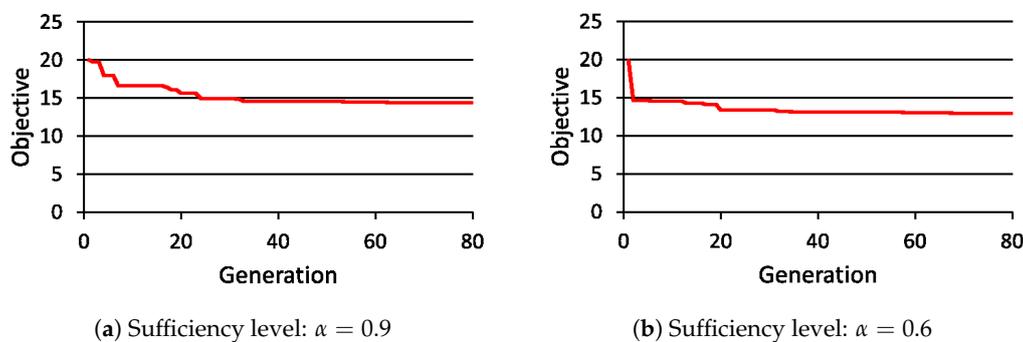


Figure 15. Convergence graph of ADEP.

Table 1. Parameters of ADEP and WSS for the relaxation problem of CCP.

Generation: NT	Population Size: NP	Sample Size: N	Correction Level: β
80	30	482	$\alpha + 0.03$

ADEP is applied to the relaxation problem of CCP 50 times. For the respective runs of ADEP, different full data sets $\mathbf{B} \subseteq \mathcal{R}^3$ and initial populations $P_0 \subseteq \mathbf{X}$ are generated randomly. Every solution $x_b \in \mathbf{X}$ obtained by ADEP for the relaxation problem of CCP is checked whether it also satisfies the constraint of the equivalence problem of CCP. From the ratio of the infeasible solutions $x_b \in \mathbf{X}$ that do not meet the constraint $\hat{p}(x_b, \mathbf{B}) \geq \alpha$, the risk of failure $\delta \in (0, 1)$ in (8) is evaluated empirically.

Table 2 shows the results of the experiments conducted for several sufficiency levels $\alpha \in (0, 1)$ in which $f(x_b)$ denotes the objective function value of the best solution $x_b \in \mathbf{X}$ obtained by ADEP for a given $\alpha \in (0, 1)$; $\hat{p}(x_b, \Theta)$ and $\hat{p}(x_b, \mathbf{B})$ are empirical probabilities provided by $x_b \in \mathbf{X}$; $\varepsilon_p(\Theta, \mathbf{B})$ is the estimation error defined by (27); $\hat{\delta} \in [0, 1]$ is the risk of failure evaluated empirically as stated above. Except the risk of failure $\hat{\delta} \in [0, 1]$, the results in Table 2 are averaged over 50 runs.

Table 2. Solutions obtained by ADEP for CCP in (37).

α	β	$f(x_b)$	$\hat{p}(x_b, \Theta)$	$\hat{p}(x_b, \mathbf{B})$	$\varepsilon_p(\Theta, \mathbf{B})$	$\hat{\delta}$
0.9	0.93	14.438	0.934	0.921	0.012	0.00
0.8	0.83	13.725	0.837	0.809	0.028	0.00
0.7	0.73	13.355	0.737	0.725	0.012	0.00
0.6	0.63	12.949	0.636	0.609	0.027	0.16

From Table 2, we can confirm the usefulness of the proposed method. Even though the sample size N of WSS is small, the value of $\hat{p}(x_b, \Theta)$ is very close to $\hat{p}(x_b, \mathbf{B})$, and $\varepsilon_p(\Theta, \mathbf{B}) < 0.03$ holds for every $\alpha \in (0, 1)$. Moreover, the majority of the solutions $x_b \in \mathbf{X}$ satisfy the constraint $\hat{p}(x_b, \mathbf{B}) \geq \alpha$. Therefore, if we suppose that $\hat{p}(x_b, \mathbf{B}) \simeq p(x_b, \Omega)$ holds, the solution $x_b \in \mathbf{X}$ is regarded as a feasible solution of CCP in (37). We also see the trade-off between the optimality of the solution $x_b \in \mathbf{X}$ evaluated by $f(x_b)$ and the reliability specified by $\alpha \in (0, 1)$. From the value of $\delta \in [0, 1]$, it seems to be hard to obtain feasible solutions of CCP in (37) for a small sufficiency level: $\alpha = 0.6$. That is because WSS is suitable for estimating a large value of $\hat{p}(x_b, \mathbf{B})$ as shown in Figures 5–7.

7. Performance Evaluation of ADEP

For solving the relaxation problem of CCP efficiently, the pruning technique shown in (24) is introduced into an Adaptive DE (ADE) and ADEP is proposed. By comparing ADE with ADEP, the ability of the pruning technique to reduce the run time of ADE is evaluated. The flood control planning formulated as CCP in (37) is used to draw a comparison between ADE and ADEP. Therefore, the control parameters of them are given by Table 1 except the sample size N . Thereby, ADE and ADEP are executed on a personal computer (CPU: Intel(R) Core(TM) i7-3770@3.40GHz, Memory: 16.0GB).

By changing the value of $\beta \in (0, 1)$ with a sample size $N \simeq 482$, ADE and ADEP are applied to the relaxation problem of CCP, respectively, 50 times. Table 3 shows the results of the experiments average over 50 runs in which $f(x_b)$ is the objective function value of the best solution $x_b \in \mathbf{X}$; $\hat{p}(x_b, \Theta)$ is the empirical probability provided by $x_b \in \mathbf{X}$. The run time of each algorithm except the generation of the full data set $\mathbf{B} \subseteq \mathbb{R}^3$, $|\mathbf{B}| = 10^7$ is also shown in Table 3. Rate in Table 3 means the percentage of the trial vectors $z_i \in \mathbf{X}$ which are discarded by the pruning technique used in ADEP. Furthermore, the numbers in parenthesis indicate the standard deviations of the respective values in Table 3.

From Table 3, we confirm that the pruning technique works well for reducing the run time of ADEP. Besides, the high rate in Table 3 shows that more than half of the trial vectors $z_i \in \mathbf{X}$ are eliminated by the pruning technique without evaluating the value of $h(z_i)$ in (23). From the values of $f(x_b)$ and $\hat{p}(x_b, \Theta)$ in Table 3, we can also see that ADE and ADEP find the same solution $x_b \in \mathbf{X}$. Therefore, the pruning technique does not harm the quality of the solution obtained by ADEP.

Table 3. Comparison between ADE and ADEP with sample size $N \simeq 482$.

β	ADE			ADEP			
	$f(x_b)$	$\hat{p}(x_b, \Theta)$	Time [s]	$f(x_b)$	$\hat{p}(x_b, \Theta)$	Time [s]	Rate
0.93	14.438 (0.035)	0.934 (0.001)	0.263 (0.023)	14.438 (0.035)	0.934 (0.001)	0.189 (0.013)	0.509 (0.015)
0.83	13.725 (0.034)	0.837 (0.002)	0.267 (0.011)	13.725 (0.034)	0.837 (0.002)	0.190 (0.010)	0.514 (0.013)
0.73	13.355 (0.032)	0.737 (0.009)	0.263 (0.015)	13.355 (0.032)	0.737 (0.009)	0.186 (0.012)	0.521 (0.015)
0.63	12.949 (0.032)	0.636 (0.009)	0.265 (0.018)	12.949 (0.032)	0.636 (0.009)	0.185 (0.009)	0.519 (0.013)

By using a larger sample size $N \simeq 1304$, ADE and ADEP are applied to the relaxation problem of CCP again 50 times. Table 4 shows the results of the experiments in the same way with Table 3. From Table 4, we can also confirm the effectiveness of the pruning technique used in ADEP.

From Tables 3 and 4, the run times of ADE and ADEP depend on the sample size N of WSS. The pruning technique of ADEP is more effective when a large sample size is required. We can also see that the sample size $N = 482$ is large enough for solving the flood control planning because there is not much difference between the qualities of the solutions shown in Tables 3 and 4.

The advantage of the pruning technique might not be demonstrated well enough due to the short run times of ADE shown in Tables 3 and 4. The short run time of ADE is attributable to the

simple forest mechanism model given by (35). If the inflow of water is estimated through a complex mathematical computation taking hours [36] or the amount of rainfall is predicted from a huge weather data set [37], we must realize the advantage of the pruning technique that surely reduces the run time of ADE without harming the quality of the obtained solution. In any case, we can confirm the expected performance of the pruning technique from the high rates shown in Tables 3 and 4.

Table 4. Comparison between ADE and ADEP with sample size $N \simeq 1304$.

β	ADE			ADEP			Rate
	$f(x_b)$	$\hat{p}(x_b, \Theta)$	Time [s]	$f(x_b)$	$\hat{p}(x_b, \Theta)$	Time [s]	
0.93	14.480 (0.036)	0.932 (0.001)	0.335 (0.012)	14.480 (0.036)	0.932 (0.001)	0.223 (0.011)	0.512 (0.014)
0.83	13.790 (0.026)	0.834 (0.003)	0.335 (0.011)	13.790 (0.026)	0.834 (0.003)	0.223 (0.011)	0.513 (0.014)
0.73	13.341 (0.025)	0.732 (0.002)	0.337 (0.016)	13.341 (0.025)	0.732 (0.002)	0.222 (0.010)	0.516 (0.015)
0.63	13.002 (0.023)	0.633 (0.002)	0.334 (0.009)	13.002 (0.023)	0.633 (0.002)	0.222 (0.013)	0.520 (0.014)

8. Conclusions

For solving CCP formulated from a huge data set, or a full data set, a new approach is proposed. By using the full data set instead of the mathematical model simulating uncertainties, the estimation error of uncertainties caused by the mathematical model can be eliminated. However, the full data set is usually too large to solve CCP practically. Therefore, a relaxation problem of CCP is derived by using a data reduction method. As a new data reduction method based on the stratified sampling, WSS is proposed and evaluated in this paper. Contrary to the well-known SRS, WSS can use the information of the full data set completely. Besides, it is shown that WSS outperforms SRS in the accuracy of the estimated probability. In order to solve the relaxation problem of CCP efficiently, an Adaptive DE combined with a Pruning technique (ADEP) is also proposed. The proposed approach is demonstrated through a real-world application, namely the flood control planning formulated as CCP.

Since huge data sets are available in various fields nowadays, many real-world applications can be formulated as CCPs without making mathematical models. Therefore, the combination of ADEP and WSS seems to be a promising approach to CCP formulated by using a huge data set. Especially, ADEP is applicable to any CCP in which the probabilistic constraint has to be evaluated empirically from a set of samples. On the other hand, there are the following open problems about WSS.

- How to properly make the strata from a full data set for WSS: The performance of WSS depends on the stratification method such as the number of strata and the shape of each stratum. By improving the stratification method, the optimal sample size of WSS will also be found.
- How to feedback the values of functions $g_m(x, \theta^n)$ to generate samples $\theta^n \in \Theta$: If we can use the function values effectively, we must be able to make the strata for WSS adaptively.
- How to cope with high-dimensional data sets: Since the similarity of data $\zeta^\ell \in B_n$ which are assigned in the same stratum $B_n \subseteq B$ is reduced in proportion to the dimensionality of the full data set, it may be hard to represent all data $\zeta^\ell \in B_n$ only by one sample $\theta^n \in \Theta$.

In our future work, we will tackle the above open problems about WSS. Moreover, we would like to demonstrate the usefulness of the proposed approach through the various real-world applications which are formulated as CCPs by using huge data sets. In particular, it is necessary that the proposed approach to CCP be evaluated by using real data sets [36]. We also need to compare ADEP with state-of-the-art optimization methods such as Ant Colony Optimization (ACO) algorithm [38].

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ACO	Ant Colony Optimization
ADE	Adaptive Differential Evolution
ADEP	Adaptive Differential Evolution with Pruning technique
CCP	Chance Constrained Problem
CHT	Constraint Handling Technique
DE	Differential Evolution
EA	Evolutionary Algorithm
SRS	Simple Random Sampling
WSS	Weighted Stratified Sampling

References

1. Ben-Tal, A.; Ghaoui, L.E.; Nemirovski, A. *Robust Optimization*; Princeton University Press: Princeton, NJ, USA, 2009.
2. Prékopa, A. *Stochastic Programming*; Kluwer Academic Publishers: Berlin, Germany, 1995.
3. Uryasev, S.P. *Probabilistic Constrained Optimization: Methodology and Applications*; Kluwer Academic Publishers: Berlin, Germany, 2001.
4. Lubin, M.; Dvorkin, Y.; Backhaus, S. A robust approach to chance constrained optimal power flow with renewable generation. *IEEE Trans. Power Syst.* **2016**, *31*, 3840–3849. [[CrossRef](#)]
5. Tagawa, K.; Miyanaga, S. An approach to chance constrained problems using weighted empirical distribution and differential evolution with application to flood control planning. *Electron. Commun. Jpn.* **2019**, *102*, 45–55. [[CrossRef](#)]
6. Bazaraa, M.S.; Sherali, H.D.; Shetty, C.M. *Nonlinear Programming: Theory and Algorithm*; John Wiley & Sons: Hoboken, NJ, USA, 2006.
7. Poojari, C.A.; Varghese, B. Genetic algorithm based technique for solving chance constrained problems. *Eur. J. Oper. Res.* **2008**, *185*, 1128–1154. [[CrossRef](#)]
8. Liu, B.; Zhang, Q.; Fernández, F.V.; Gielen, G.G.E. An efficient evolutionary algorithm for chance-constrained bi-objective stochastic optimization. *IEEE Trans. Evol. Comput.* **2013**, *17*, 786–796. [[CrossRef](#)]
9. Tagawa, K.; Miyanaga, S. Weighted empirical distribution based approach to chance constrained optimization problems using differential evolution. In Proceedings of the IEEE Congress on Evolutionary Computation (CEC2017), San Sebastian, Spain, 5–8 June 2017; pp. 97–104.
10. Kroese, D.P.; Taimre, T.; Botev, Z.I. *Handbook of Monte Carlo Methods*; Wiley: Hoboken, NJ, USA, 2011.
11. Tagawa, K. Group-based adaptive differential evolution for chance constrained portfolio optimization using bank deposit and bank loan. In Proceedings of the IEEE Congress on Evolutionary Computation (CEC2019), Wellington, New Zealand, 10–13 June 2019; pp. 1556–1563.
12. Xu, L.D.; He, W.; Li, S. Internet of things in industries: A survey. *IEEE Trans. Ind. Inform.* **2014**, *10*, 2233–2243. [[CrossRef](#)]
13. Rossi, E.; Rubattion, C.; Viscusi, G. Big data use and challenges: Insights from two internet-mediated surveys. *Computers* **2019**, *8*, 73. [[CrossRef](#)]
14. Kile, H.; Uhlen, K. Data reduction via clustering and averaging for contingency and reliability analysis. *Electr. Power Energy Syst.* **2012**, *43*, 1435–1442. [[CrossRef](#)]
15. Fahad, A.; Alshatri, N.; Tari, Z.; Alamri, A.; Khalil, I.; Zomaya, A.Y.; Fofou, S.; Bouras, A. A survey of clustering algorithms for big data: Taxonomy and empirical analysis. *IEEE Trans. Emerg. Top. Comput.* **2014**, *2*, 267–279. [[CrossRef](#)]
16. Jayaram, N.; Baker, J.W. Efficient sampling and data reduction techniques for probabilistic seismic lifeline risk assessment. *Earthq. Eng. Struct. Dyn.* **2010**, *39*, 1109–1131. [[CrossRef](#)]
17. Tagawa, K. Data reduction via stratified sampling for chance constrained optimization with application to flood control planning. In Proceedings of the ICIST 2019, CCIS 1078, Vilnius, Lithuania, 10–12 October 2019; Springer: Cham, Switzerland, 2019; pp. 485–497.

18. Price, K.; Storn, R.M.; Lampinen, J.A. *Differential Evolution: A Practical Approach to Global Optimization*; Springer: Cham, Switzerland, 2005.
19. Brest, J.; Greiner, S.; Boskovic, B.; Merink, M.; Zumer, V. Self-adapting control parameters in differential evolution: A comparative study on numerical benchmark problems. *IEEE Trans. Evol. Comput.* **2006**, *10*, 646–657. [[CrossRef](#)]
20. Yazdi, J.; Neyshabouri, S.A.A.S. Optimal design of flood-control multi-reservoir system on a watershed scale. *Nat. Hazards* **2012**, *63*, 629–646. [[CrossRef](#)]
21. Zhang, W.; Liu, P.; Chen, X.; Wang, L.; Ai, X.; Feng, M.; Liu, D.; Liu, Y. Optimal operation of multi-reservoir systems considering time-lags of flood routing. *Water Resour. Manag.* **2016**, *30*, 523–540. [[CrossRef](#)]
22. Zhou, C.; Sun, N.; Chen, L.; Ding, Y.; Zhou, J.; Zha, G.; Luo, G.; Dai, L.; Yang, X. Optimal operation of cascade reservoirs for flood control of multiple areas downstream: A case study in the upper Yangtze river basin. *Water* **2018**, *10*, 1250. [[CrossRef](#)]
23. Ash, R.B. *Basic Probability Theory*; Dover: Downers Grove, IL, USA, 2008.
24. Han, J.; Kamber, M.; Pei, J. *Data Mining—Concepts and Techniques*; Morgan Kaufmann: Burlington, MA, USA, 2012.
25. Tempo, R.; Calafiore, G.; Dabbene, F. *Randomized Algorithms for Analysis and Control of Uncertain Systems: With Applications*; Springer: Cham, Switzerland, 2012.
26. Poosala, V.; Ioannidis, Y.E.; Haas, P.J.; Shekita, E.J. Improved histograms for selectivity estimation of range predicates. In Proceedings of the ACM SIGMOD International Conference on Management of Data, Montreal, QC, Canada, 4–6 June 1996, pp. 294–305.
27. Cormode, G.; Garofalakis, M. Histograms and wavelets on probabilistic data. *IEEE Trans. Knowl. Data Eng.* **2010**, *22*, 1142–1157. [[CrossRef](#)]
28. Das, S.; Suganthan, P.N. Differential evolution: A survey of the state-of-the-art. *IEEE Trans. Evol. Comput.* **2011**, *15*, 4–31. [[CrossRef](#)]
29. Eltaeib, T.; Mahmood, A. Differential evolution: A survey and analysis. *Appl. Sci.* **2018**, *8*, 1945. [[CrossRef](#)]
30. Deb, K. An efficient constraint handling method for genetic algorithms. *Comput. Methods Appl. Mech. Eng.* **2000**, *186*, 311–338. [[CrossRef](#)]
31. Tanabe, R.; Fukunaga, A. Reviewing and benchmarking parameter control methods in differential evolution. *IEEE Trans. Evol. Comput.* **2020**, *50*, 1170–1184. [[CrossRef](#)]
32. Montes, E.E.; Coello, C.A.C. Constraint-handling in nature inspired numerical optimization: Past, present and future. *Swarm Evol. Comput.* **2011**, *1*, 173–194. [[CrossRef](#)]
33. Prékopa, A.; Szántai, T. Flood control reservoir system design using stochastic programming. *Math. Progr. Study* **1978**, *9*, 138–151.
34. Maita, E.; Suzuki, M. Quantitative analysis of direct runoff in a forested mountainous, small watershed. *J. Jpn. Soc. Hydrol. Water Resour.* **2009**, *22*, 342–355. [[CrossRef](#)]
35. Martinez, A.R.; Martinez, W.L. *Computational Statistics Handbook with MATLAB*® , 2nd ed.; Chapman & Hall/CRC: London, UK, 2008.
36. Monrat, A.A.; Islam, R.U.; Hossain, M.S.; Andersson, K. Challenges and opportunities of using big data for assessing flood risks. In *Applications of Big Data Analytics*; Alani, M.M., Tawfik, H., Saeed, M., Anya, O., Eds.; Springer: Cham, Switzerland, 2018; Chapter 2, pp. 31–42.
37. Reddy, P.C.; Babu, A.S. Survey on weather prediction using big data analytics. In Proceedings of the IEEE 2nd International Conference on Electrical, Computer and Communication Technologies, Coimbatore, India, 22–24 February 2017; pp. 1–6.
38. Brociek, R.; Słota, D. Application of real ant colony optimization algorithm to solve space and time fractional heat conduction inverse problem. In Proceedings of the 22nd International Conference on Information and Software Technologies, ICIST2016, Communications in Computer and Information Science, Druskininkai, Lithuania, 13–15 October 2016; Springer: Cham, Switzerland, 2016; Volume 639, pp. 369–379.

