

SEARCH STRATEGIES FOR FAILURE CASCADE PATHS IN POWER SYSTEM GRAPHS

JIANFENG ZHANG

Midwest ISO

Carmel, Indiana 46032

E-mail: jzhang@midwestiso.org

JAMES A. BUCKLEW AND FERNANDO L. ALVARADO

Department of Electrical and Computer Engineering

University of Wisconsin

Madison, Wisconsin 53706

E-mail: bucklew@engr.wisc.edu; alvarado@engr.wisc.edu

This paper addresses computer simulation of cascading failures in electric power systems. The paper analyzes the convergence rates of estimator variance in importance sampling and in random search strategies. A uniform search strategy based on the Metropolis algorithm is proposed.

1. INTRODUCTION

In recent years, importance sampling has proven to be a very effective variance reduction technique in rare-event simulation [2–4]. It has been used to study cascading disturbances in large power systems [1]. In this paper we analyze the convergence rates of estimator variance to zero in importance sampling and in random search strategies. We find that the random search strategy may perform better than importance sampling techniques when the state space of the simulation variable is countably finite. Section 2 discusses the motivation for using importance sampling techniques in simulation studies. In Section 3 we analyze the performance of some simple random search strategies. Sections 4–6 restate the problem of the power system security in terms of a network security problem and suggest some graph-theoretic approaches based on the Metropolis algorithm.

2. MOTIVATION FOR IMPORTANCE SAMPLING

Suppose we wish to estimate $\rho = \mathbf{E}\{\phi(Z)\}$, where Z is a random variable describing some observation on a random system. Usually ϕ is the indicator function of some set implying that ρ is the probability of the set. Suppose that the observation random variable Z has probability density function $p(\cdot)$. The direct (Monte Carlo) simulation method would be to generate a sequence of independent and identically distributed (i.i.d.) random numbers Z_1, Z_2, \dots, Z_N from the density $p(\cdot)$ and form the estimate

$$\hat{\rho}_p \doteq \frac{1}{N} \sum_{i=1}^N p(Z_i).$$

By the law of large numbers, $\hat{\rho}_p \rightarrow \rho$ as $N \rightarrow \infty$. Therefore, as the number of observations approaches infinity, we converge to the true value. Suppose instead that we generate a sequence of i.i.d. random numbers Z_1, Z_2, \dots, Z_N using a possibly different density $q(\cdot)$. We then form the estimate

$$\hat{\rho}_q \doteq \frac{1}{N} \sum_{i=1}^N \frac{p(Z_i)}{q(Z_i)} \phi(Z_i).$$

The ratio $p(\cdot)/q(\cdot)$ will be called the *weight function* of the importance sampling estimator. It is simple to verify that the expected value of $\hat{\rho}_q$ under the density $q(\cdot)$ is precisely ρ .

Clearly, the estimate $\hat{\rho}_q$ is unbiased, and as $N \rightarrow \infty$, we also expect it to be converging (by the law of large numbers) to its mean value ρ . The obvious question is: Are there better choices for $q(\cdot)$ than just $p(\cdot)$? The answer is that by making a good choice for $q(\cdot)$, orders of magnitude decrease in the estimator variance can be achieved over a direct Monte Carlo simulation. It is this fact that has spurred most, if not all, of the recent interest in importance sampling techniques.

The important point is that the rate of convergence of the variance to zero as a function of N is precisely $1/N$ regardless of the importance sampling strategy chosen. The next section demonstrates that there are other estimation strategies that can give better rates of convergence in certain types of problems.

3. ANALYSIS OF RANDOM SEARCH STRATEGIES

Suppose we have some countable state space Ω . In the power system security problem [1], an element of Ω would be an ordered sequence of edges deleted from the network due to an initial failure event. Since a power system is a finite graph, Ω would also be finite (although usually very large). Even though the cardinality of Ω may be a very large finite number, on a short-time simulation scale due to the size of these state spaces they may behave as if they were infinite. Therefore, it is instructive to understand the more general countably infinite setting.

Suppose we have some probability measure P over the subsets of Ω and we are interested in estimating $P(A)$ for some particular subset of Ω . For example, A could

be the set of all ordered sequences that cause the graph to be disconnected. To estimate $P(A)$, we choose a “search” probability measure F over the (Borel) subsets of Ω . In many situations, this measure F is not known explicitly. In general, it would be very hard to write down explicitly what this measure is for some arbitrary random path search on a graph.

We choose N independent samples $\{x_i \in \Omega : i = 1, \dots, N\}$ from the F distribution. We suppose that we can evaluate $P(x_i)$ for each of these chosen values. We keep track of the total number of unique points that we find in set A . We suppose that in N trials, we find K such points.

Define

$$\begin{aligned}
 j_1 &= \left\{ \min_i: x_i \in A \right\}, \\
 j_2 &= \left\{ \min_{i:i>j_1}: x_i \in A \text{ and } x_i \neq x_{j_1} \right\}, \\
 &\vdots \\
 j_i &= \left\{ \min_{i:i>j_{i-1}}: x_i \in A \text{ and } x_i \neq x_{j_1}, x_{j_2}, \dots, x_{j_{i-1}} \right\}, \\
 &\vdots \\
 j_K &= \left\{ \min_{i:i>j_{K-1}}: x_i \in A \text{ and } x_i \neq x_{j_1}, x_{j_2}, \dots, x_{j_{K-1}} \right\}.
 \end{aligned}$$

We can then write our estimator as

$$\hat{P}(A) = \sum_{i=1}^K P(x_{j_i}).$$

Of course, this estimator is biased. In fact (as pointed out in [1]), the estimator will always be a lower bound (a useful fact in itself) to the true value of $P(A)$. To get a better notion of its performance, let us compute the mean squared error of the estimate:

$$\begin{aligned}
 &E[(P(A) - \hat{P}(A))^2] \\
 &= E \left[\left(\sum_{\substack{x_i \in A: x_i \text{ not chosen in} \\ N \text{ independent} \\ \text{samples from } F}} P(x_i) \right)^2 \right] \\
 &= E \left[\left(\sum_{x_i \in A} P(x_i) 1_{\{x_i \text{ not in } N \text{ samples}\}} \right)^2 \right] \\
 &= E \left[\sum_{x_i \in A} \sum_{x_j \in A} P(x_i) P(x_j) 1_{\{x_i \text{ not in } N \text{ samples}\}} 1_{\{x_j \text{ not in } N \text{ samples}\}} \right] \\
 &= \sum_{x_i \in A} \sum_{x_j \in A \text{ and } j \neq i} P(x_i) P(x_j) [1 - F(x_i) - F(x_j)]^N + \sum_{x_i \in A} P(x_i)^2 [1 - F(x_i)]^N,
 \end{aligned}$$

where $1_{\{B\}} = 1$ if B is true, and zero otherwise. Without loss of generality, we can assume $P(x_i) > 0$ for all $x_i \in A$ (if not, just delete those points from A). From the above equation, we can immediately draw some important conclusions. If the state space Ω is finite, then we obtain that the mean squared error decreases exponentially (or geometrically) fast in N . This is an important point and we will return to it later.

When Ω is finite, let us define $f_1 = \min_{x_i \in A} F(x_i)$ with minimum occurring at the value x_{m_1} [i.e., $F(x_{m_1}) = f_1$]. Of course, this minimum could occur at several points and so x_{m_1} is not necessarily unique. For any $x_i \in A$ and $x_j \in A$, we have

$$0 < 1 - F(x_i) - F(x_j) < 1 - F(x_i) \leq 1 - f_1 < 1.$$

Suppose $|\Omega| < \infty$; it is obvious that

$$\begin{aligned} & \lim_{N \rightarrow \infty} \frac{1}{N} \log E[(P(A) - \hat{P}(A))^2] \\ &= \lim_{N \rightarrow \infty} \frac{1}{N} \log \left([1 - f_1]^N \left[\sum_{x_i \in A, F(x_i)=f_1} P(x_i)^2 + \varepsilon(N) \right] \right) \\ &= \lim_{N \rightarrow \infty} \frac{1}{N} \left(\log [1 - f_1]^N + \log \left[\sum_{x_i \in A, F(x_i)=f_1} P(x_i)^2 + \varepsilon(N) \right] \right) \\ &= \log [1 - f_1], \end{aligned}$$

where $\lim_{N \rightarrow \infty} \varepsilon(N) = 0$. Then we have the following theorem.

THEOREM 3.1: *Suppose $|\Omega| < \infty$. Then*

$$\lim_{N \rightarrow \infty} \frac{1}{N} \log E[(P(A) - \hat{P}(A))^2] = \log [1 - f_1].$$

The limit $\log [1 - f_1]$ indicates how fast the mean squared error of the estimate converges to zero. The larger the value of f_1 , the faster the mean squared error converges to zero. When Ω is finite, the distribution that has maximal f_1 is the uniform distribution. An immediate consequence of the theorem is the following.

LEMMA 3.1: *Suppose $|\Omega| < \infty$. The optimal search distribution is the uniform distribution.*

When the cardinality of the Ω space is countably infinite, determining the behavior of the random search estimator is quite a bit more complicated. Let us first consider a thought experiment. For simplicity (and without loss of generality), let us suppose that $\Omega = \mathbb{Z}^+$, the nonnegative integers. Define $P(j) = p_j$ and $F(j) = f_j$ for all $j \in \mathbb{Z}^+$. Suppose $p_i = C/i^{1+\gamma}$ (where C is a constant). Suppose we are interested in $A = \{i : i \geq T\}$. Further, suppose that we decide to sample deterministically from Ω (instead of choosing random samples from some F). The best possible choice (which will not usually be possible) would be to choose the values $T, T + 1,$

$T + 2, \dots, T + N - 1$ and form the estimate $\hat{P}(A) = \sum_{i=T}^{T+N-1} p_i$. Thus, for the mean squared error,

$$\begin{aligned} E[(P(A) - \hat{P}(A))^2] &= \left| \sum_{i=T+N}^{\infty} p_i \right|^2 \\ &\approx \frac{C_1}{(T+N)^{2\gamma}}. \end{aligned}$$

This, of course, is the best possible rate for the mean squared error. Therefore, when $\gamma < \frac{1}{2}$, the possible rate is worse than a relative frequency-type estimator, which always has a rate of $1/N$.

One may ask: How much does one have to give up to use a random search strategy? Let us consider the $\gamma = 1$ case in a little more detail. We get the following useful upper bound:

$$\begin{aligned} E[(P(A) - \hat{P}(A))^2] &= \sum_{x_i \in A} \sum_{x_j \in A \text{ and } j \neq i} P(x_i)P(x_j)[1 - F(x_i) - F(x_j)]^N + \sum_{x_i \in A} P(x_i)^2[1 - F(x_i)]^N \\ &\leq \sum_{x_i \in A} \sum_{x_j \in A} P(x_i)P(x_j)[1 - F(x_i)]^N. \end{aligned}$$

Suppose that we choose $f_i = C'/i^{1+\varepsilon}$ for some $\varepsilon > 0$. We can approximate the above upper bounding sum with the integral

$$\begin{aligned} &\int_T^\infty \int_T^\infty \frac{1}{x^2} \frac{1}{y^2} \left[1 - \frac{1}{x^{1+\varepsilon}}\right]^N dx dy \\ &= \int_0^{1/T} \int_0^{1/T} [1 - u^{1+\varepsilon}]^N du dv \\ &= \int_0^{1/N^{(1-\delta)/(1+\varepsilon)}} \int_0^{1/N^{(1-\delta)/(1+\varepsilon)}} [1 - u^{1+\varepsilon}]^N du dv \\ &\quad + \int_{1/N^{(1-\delta)/(1+\varepsilon)}}^{1/T} \int_{1/N^{(1-\delta)/(1+\varepsilon)}}^{1/T} [1 - u^{1+\varepsilon}]^N du dv \\ &\leq \left(\frac{1}{N^{(1-\delta)/(1+\varepsilon)}}\right)^2 + \left(\frac{1}{T} - \frac{1}{N^{(1-\delta)/(1+\varepsilon)}}\right)^2 \left[1 - \frac{1}{N^{1-\delta}}\right]^N \\ &\leq \frac{1}{N^{2(1-\delta)/(1+\varepsilon)}} + C''e^{-N^\delta}. \end{aligned}$$

Thus, by choosing δ (and ε) small enough, we can get arbitrarily close to the best possible rate $1/N^2$. (Interestingly enough, when $\varepsilon = 1$ we can show that the integral in question has exactly rate $1/N$. Therefore, a finer analysis than the above may be possible.)

The lesson to be gained from this analysis is that the sampling distribution should be in some sense as “heavy tailed” or as maximally spread out as we can get it. Of course, in the finite state space case, this is the uniform distribution.

In the power systems problem [1], the random search strategy works because the state space Ω is finite and fairly small. As a result, the guaranteed asymptotic convergence rate is observed early. The exponential rate promise (in the finite state space setting) is really conditioned upon visiting all the points in Ω with a high probability. Many graph problems have their complexity; hence the number of points in Ω increases exponentially fast as the size increases. Consequently, random search strategies are doomed to perform more poorly as the size of Ω increases. Obviously, the random search strategy would fail completely for an uncountably infinite state space where each point in the space has probability zero of occurrence. Another major drawback is the necessity to store all previously chosen samples to make sure that a given point in Ω is not used twice.

4. A UNIFORM SEARCH STRATEGY

As we saw in the previous section, the optimal strategy for a random search is to search uniformly over the state space. In a path search over a large graph, it is not at all clear how one goes about searching uniformly over the set of all paths leading to some failure set. Indeed, in most applications we do not even know the size of the failure set or the total number of paths. In this section, we propose a search technique based on the Metropolis Algorithm of Markov chain Monte Carlo, which allows us to easily specify a uniform search distribution.

Suppose we have an undirected graph $G = (V, E)$. We are interested in the phenomenological effects of cascading link (link = edge) failures. We suppose that an initial event triggers a failure on one of the edges, say $j \in E$. Define the neighborhood edge set $N(j) = \{e \in E : e \text{ and } j \text{ share a common vertex}\}$. (Each element of $N(j)$ represents a possible failure following j in practical power system cascading failure problems.) With the failure of j , we suppose that each of the elements of $N(j)$ is at “risk” to fail also with given a priori probabilities $p_e, e \in N(j)$. In the model we may allow these probabilities to change with the “state” of the graph after a chain of previous failures have occurred. We are interested in sequences of failures that disconnect the graph, cause large load losses, violate security criteria, have large numbers of lines lost—any or all of these. We denote the set of subgraphs of G that have these properties of large load lost, massive disconnection, and so forth that result from these unfortunate chains of events as B . Quite simply then, we wish to estimate the probability that the original graph G moves into one of the states B through a sequence of link failures. We suppose (although it is not completely necessary) that the disconnected graphs are always at least a subset of B .

Typical random searches suffer from the problem that they tend to find the short-path-length elements of B much more readily than the longer paths associated perhaps with greater calamities. We propose an algorithm based on techniques from Markov chain Monte Carlo (MCMC) to allow us to sample uniformly (or with any other distribution!) from the set B regardless of the individual element’s path length.

We can think of the sequence of line outages as forming a tree. The original lost link j is the first node of the tree. The neighborhood set of that link then gives rise to the second level of the tree, which will have $|N(j)|$ nodes. Each of these nodes will have a neighborhood set that will give rise to the third level of the tree and so on. Eventually, a sequence of link outages will terminate with a graph in B .

We first propose a Markov chain with state space Z , the nodes of the tree. We wish this Markov chain to have stationary distribution π . Usually we choose π to be uniform over Z . Let Z denote the nodes of the tree. We specify the transition structure of the Markov chain as

$$p_{ij} = q_{ij}\alpha_{ij}, \quad i, j \in Z, j \neq i.$$

$Q = (q_{ij})$ is called the ‘‘candidate generating’’ transition matrix and is arbitrary as long as it is chosen to be irreducible. The idea is simple: When the present state is i , the next ‘‘tentative’’ state j is chosen with probability q_{ij} . When $j \neq i$, this new state is ‘‘accepted’’ with probability α_{ij} . Hence, the probability of moving from i to j when $j \neq i$ is given as above.

We need to specify Q for our tree search. At node $k \in Z$, there are m_k nodes directly below it. We denote these as $k_{d,1}, k_{d,2}, \dots, k_{d,m_k}$. There is, of course, only one node directly above node k , which we denote as k_u . We will only allow transitions between k and these neighboring nodes. Let $0 < p < 1$ be chosen. We can then choose

$$q_{kk_{d,i}} = \frac{1-p}{m_k}, \quad i = 1, 2, \dots, m_k,$$

and

$$q_{kk_u} = p.$$

When k is a leaf, there are no ‘‘down’’ nodes, and so we may choose

$$q_{kk_u} = 1 \quad \text{when } k \text{ is a leaf.}$$

These choices, of course, give rise to an irreducible Q . Furthermore, we use the Metropolis algorithm choice for α_{ij} as

$$\alpha_{ij} = \min\left(1, \frac{q_{ji}\pi(j)}{q_{ij}\pi(i)}\right).$$

When π is uniform, this simplifies to

$$\alpha_{kk_{d,i}} = \min\left(1, \frac{m_k p}{1-p}\right), \quad i = 1, 2, \dots, m_k$$

and

$$\alpha_{k k_u} = \min\left(1, \frac{1-p}{m_{k_u} p}\right), \quad i = 1, 2, \dots, m_k,$$

$$\alpha_{k k_u} = \frac{1-p}{m_{k_u}} \quad \text{when } k \text{ is a leaf.}$$

Thus, we have constructed a Markov chain with a uniform stationary distribution on the nodes of the tree. The leaf nodes are visited uniformly, and because every subgraph in B is associated with a unique leaf, we are also visiting these subgraphs uniformly.

5. A MODIFIED UNIFORM SEARCH STRATEGY

The strategy described in Section 4 constructs a Markov chain with a uniform stationary distribution on the nodes of the tree. Since we only need to produce a uniform sampling distribution on the leaf nodes, we can modify the algorithm to reduce the number of visits of nonleaf nodes while keeping a uniform stationary distribution on the leaf nodes. We denote L as the state space of all leaf nodes in Z . We start by modifying the transition matrix $Q = (q_{ij})$. In the previous section, q_{ij} is time invariant (i.e., it is not time dependent). Now we let q_{ij} also depend on the last state s . When the last state is s and the present state is i , the next “tentative” state j is chosen with probability $q_{ij}(s)$. As in Section 4, we only allow transitions between neighboring nodes. At node $k \in Z$, let $0 < p_k < 1$ be chosen. p_k can be either constant or k dependent, but it must be time invariant. If k is the root, let $p_k \equiv 0$. Using the same denotation as in Section 4, we then choose

$$q_{k k_d, i}(k_u) = \frac{1}{m_k}, \quad i = 1, 2, \dots, m_k,$$

$$q_{k k_d, i}(k_{d, l}) = \frac{1-p_k}{m_k - 1}, \quad i = 1, 2, \dots, l-1, l+1, \dots, m_k,$$

$$q_{k k_u}(k_{d, l}) = p_k, \quad l = 1, 2, \dots, m_k.$$

If $m_k = 1$, we choose $p_k = 1$. If k is a leaf, then we choose $q_{k k_u} = 1$.

Suppose we generate a random sequence $\{x_m | x_m \in Z\}$ by applying transition matrix $Q = (q_{ij}(s))$ and assuming $\alpha_{ij} \equiv 1$. Because Q is not time invariant, the sequence $\{x_m\}$ is not a Markov chain. Let $\{y_n\}$ be the subsequence of $\{x_m\}$ that contains all of the leaf nodes in $\{x_m\}$ (i.e., $y_n \in L$). Let

$$t_{ij} = \text{Prob}\{y_{n+1} = j | y_n = i\}.$$

It is easy to prove that the matrix $T = (t_{ij})$ is time invariant and irreducible [7].

Now, we can construct a Markov chain with state space L , the leaf nodes of the tree. We specify the transition structure of the Markov chain as follows:

$$p_{ij} = t_{ij}\beta_{ij}, \quad i, j \in L, j \neq i,$$

where t_{ij} is as defined earlier and β_{ij} is chosen as

$$\beta_{ij} = \min\left(1, \frac{t_{ji}}{t_{ij}}\right).$$

The implementation is simple. Start from a leaf y_n , walk in the tree by the rule of Q until another leaf \tilde{y} is reached. Let $y_{n+1} = y_n$ or $y_{n+1} = \tilde{y}$ by the rule of β . Then start from y_{n+1} to generate y_{n+2} by the same rules. By the theory of the Metropolis algorithm, the generated sequence is the Markov chain with a uniform stationary distribution on the leaf nodes of the tree.

6. IMPORTANCE SAMPLING IS NOT DEAD

Denote the sequence of subgraphs from B generated by the above Metropolis algorithm as S_1, S_2, \dots, S_N with associated probabilities of occurrence $p(S_1), p(S_2), \dots, p(S_N)$. Of course, this sequence of subgraphs is *not* independent, but the variance reduction technique of importance sampling does not require that in order to give an unbiased estimate.

We propose the importance sampling estimator

$$\hat{\rho}_{IS} = |B| \frac{1}{N} \sum_{j=1}^N p(S_j).$$

An exhaustive search algorithm would find the $K \ll N$ unique graphs from the set $\{S_1, S_2, \dots, S_N\}$ denoted as $S^{(1)}, S^{(2)}, \dots, S^{(K)}$ and form the estimate

$$\hat{\rho}_{ES} = \sum_{j=1}^K p(S^{(j)}).$$

The exhaustive search requires the saving of all the search paths and it becomes impractical when the size of the system is very large. The importance sampling estimator is unbiased but, unfortunately, requires the computation of $|B|$, which can be difficult. In applications applied to power systems, the computation of $|B|$ is often unnecessary, especially when the study is to compare different dispatch strategies [7]. Reference 7 also gives a method to estimate $|B|$ through simulation.

We simulated on an artificial tree to compare different important sampling strategies. The artificial tree is constructed to mimic the propagation of disturbances in large power systems. The details of the construction of the tree can be found in [7]. Figure 1 shows simulation results on a tree, which has about 2×10^{11} leaf nodes. One of the reasons we choose to use an artificial tree is that we can construct the tree in a way that all of the statistical properties of the tree are computable; therefore we can evaluate the performance of our simulations conveniently. Simulation

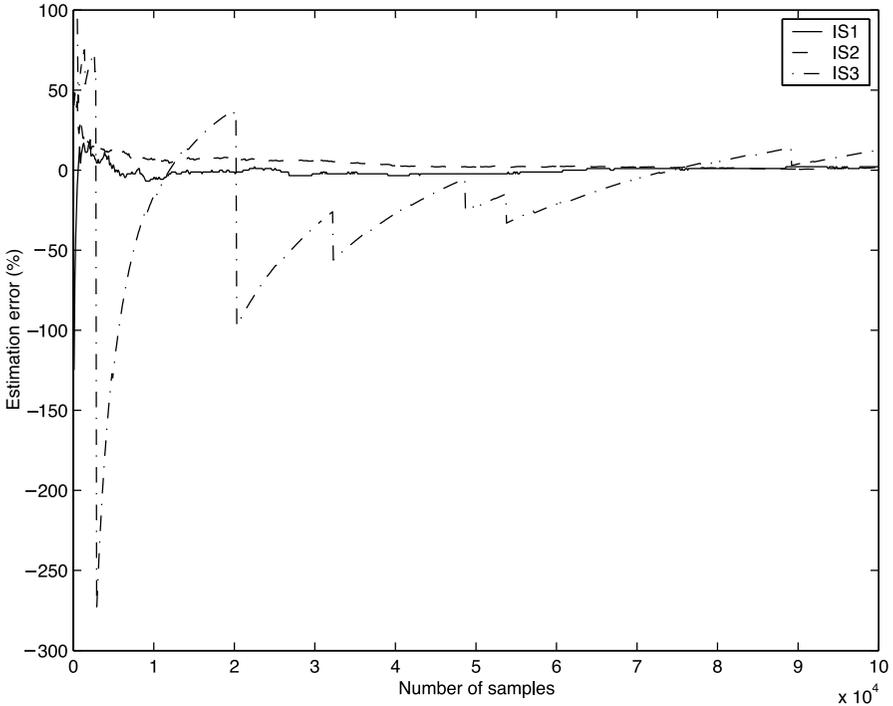


FIGURE 1. Simulation results.

results from three sampling strategies are shown in Figure 1. IS1 (the solid line) represents the Metropolis algorithm described in Section 5. IS2 (the dashed line) represents a simple importance sampling strategy. In IS2, each leaf node sample is reached by starting a search from the root. At each node in the search path, a topology study is performed and a child node is chosen randomly. In this strategy, each child node has equal probability to be chosen. IS3 (the dot dashed line) differs from IS2 in that some child nodes have a higher priority to be chosen in the simulation so that some search paths have significant higher probabilities than other paths. Figure 1 shows that IS3 does not perform as well as the other two. This illustrates the importance of the proper choice of sample strategy in importance sampling techniques. The simulation results are consistent with our theory that the sampling distribution should be in some sense as “heavy tailed” or as maximally spread out as we can get it.

In our simulation, IS1 and IS2 showed similar performance. This is because the structure of the artificial tree makes IS2 behave similarly to uniform sampling. We expect IS1 and IS2 could perform differently when the structure of the tree becomes more complex. We also want to point out that IS1 has less average computation cost per sample than IS2 does. For power system simulations, this means

that, on average, IS1 requires fewer power flow calculations to get the same number of samples. However, the tuning of a Metropolis algorithm can be tricky [5].

7. CONCLUSION

Cascading equipment trips occur fairly frequently on electric transmission systems. Occasionally, cascading disturbances lead to widespread blackouts, which have caused tremendous economic losses to society and severe damage to equipment. Concern for security always come first in power system operation. Because of the complexity of a large power system, computer simulation is essential in system security studies. However, as pointed out in [6], “lack of computational resources and of efficient algorithms have been major obstacles in studying large blackouts.” Importance sampling and random search strategies are two approaches applied in recent studies [1,6,7]. In this paper, we analyzed the convergence rates of estimator variance in importance sampling and in random search strategies. We find that the optimal sampling distribution should be in some sense as “heavy tailed” or as maximally spread out as we can get it. In the finite state space case, this is the uniform distribution. This result provides a theoretical foundation for the choice of better sampling techniques.

In practice, it is not clear how one goes about searching uniformly over the set of all paths leading to a failure. In this paper, we described a search strategy that achieves uniform distribution approximately. This provides an alternative choice of sampling strategy for future studies.

References

1. Bae, K. & Thorp, J.S. (1997). An importance sampling application: 179 bus WSCC system under voltage based hidden failures and relay misoperations. *Proceedings of the Hawaii International Conference on System Science*.
2. Bucklew, J.A. (2004). *Introduction to rare event simulation*. New York: Springer-Verlag.
3. Heidelberger, P. (1995). Fast simulation of rare events in queuing and reliability models. *ACM Transactions on Modeling and Computer Simulation* 5(1): 43–85.
4. Lieber, D., Rubinstein, R.Y., & Elmakis, D. (1997). Quick simulation of rare events in stochastic networks. *IEEE Transactions on Reliability* 46(2): 254–265.
5. Robert, C.P. & Casella, G. (1999). *Monte Carlo statistical methods*. New York: Springer-Verlag.
6. Thorp, J.S. & Casella, G. (2001). Computer simulation of cascading disturbances in electric power systems—Impact of protection systems on transmission system reliability. Final Project Report, Power Systems Engineering Research Center Publication 01-01.
7. Zhang, J. (2003). Modeling and simulation of cascading contingencies. Ph.D. dissertation, University of Wisconsin, Madison, WI.