Methodol Comput Appl Probab (2018) 20:1155–1172
https://doi.org/10.1007/s11009-017-9599-7

Uncertainty Quantification of Stochastic Simulation for Black-box Computer Experiments

Youngjun Choe[1] · Henry Lam[2] · Eunshin Byon[3]

Received: 23 September 2016 / Revised: 15 October 2017 /
Accepted: 18 October 2017 / Published online: 30 October 2017

Abstract Stochastic simulations applied to black-box computer experiments are becoming more widely used to evaluate the reliability of systems. Yet, the reliability evaluation or computer experiments involving many replications of simulations can take significant computational resources as simulators become more realistic. To speed up, importance sampling coupled with near-optimal sampling allocation for these experiments is recently proposed to efficiently estimate the probability associated with the stochastic system output. In this study, we establish the central limit theorem for the probability estimator from such procedure and construct an asymptotically valid confidence interval to quantify estimation uncertainty. We apply the proposed approach to a numerical example and present a case study for evaluating the structural reliability of a wind turbine.

Keywords Central limit theorem · Confidence interval · Importance sampling · Monte Carlo simulation · Variance reduction

Mathematics Subject Classification (2010) 65C05 · 68U20 · 62F12

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11009-017-9599-7) contains supplementary material, which is available to authorized users.

✉ Youngjun Choe
ychoe@uw.edu

[1] Department of Industrial and Systems Engineering, University of Washington, 3900 E Stevens Way NE, Seattle, WA 98195, USA

[2] Department of Industrial Engineering and Operations Research, Columbia University, 500 W. 120th Street, New York, NY 10027, USA

[3] Department of Industrial and Operations Engineering, University of Michigan, 1205 Beal Avenue, Ann Arbor, MI 48109, USA

1 Introduction

This study concerns the reliability evaluation of systems operating under stochastic conditions. An example of such systems is a wind turbine system. Wind turbines experience mechanical and structural loads induced from wind and wave (for offshore turbines), extreme value of which can cause a catastrophic failure. This study is motivated from the need to quantify the uncertainty associated with the wind turbine reliability evaluation.

Computer experiments for reliability evaluation of such stochastic systems typically require two-level simulation procedure. At the first level, the operational condition, $\mathbf{X}$, is generated and in the second level, the system's response, Y, is simulated, given $\mathbf{X}$. In general, the probability distribution of $\mathbf{X}$ is known; it can be empirically estimated from field data (e.g., empirical wind distribution at a site where a turbine is installed) or pre-specified from a design standard.

Unlike $\mathbf{X}$, the conditional distribution of Y given $\mathbf{X}$ is not readily available and Y can be only observed by running a simulator. In particular, this study considers the simulations where the process of getting the response, Y, given $\mathbf{X}$, is unknown. For example, in the wind turbine case, the U.S. Department of Energy (DOE)'s National Renewable Energy Laboratory (NREL) developed aeroelastic simulators such as TurbSim (Jonkman 2009) and FAST (Jonkman and Buhl 2005) to help design processes. Using this set of NREL simulators, we first sample the wind condition, $\mathbf{X}$, from a pre-specified density, f, and feed $\mathbf{X}$ into the simulators to generate the load response, Y. The load response, Y, is random even at a fixed value of $\mathbf{X}$, because the random vector, $\boldsymbol{\epsilon}$, that causes stochastic outputs is embedded inside the simulators and we are only allowed to sample $\mathbf{X}$, but not $\boldsymbol{\epsilon}$ (see Fig. 1). This whole simulation procedure can be viewed as a *two-level stochastic simulation* where the second level uses the stochastic black-box simulator. Such simulators are also called stochastic simulation models in the computer experiment literature (Ankenman et al. 2010; Plumlee and Tuo 2014). Recently, two-level stochastic simulations become increasingly used in other applications in order to study complex real-world systems such as the U.S. power system (Staid et al. 2014).

As simulation models represent real systems more accurately (e.g., denser mesh in finite element analysis), the computational costs of running simulations remain high despite the advance of computing technology. Furthermore, as modern systems are generally expected to meet a high standard of reliability under many uncertainties, simulation studies focus on rare events (e.g., highly reliable system's failure). Estimating the probability of a rare event requires many simulation replications, exacerbating the computational problem.

To improve the computational efficiency of such estimations, two critical questions need to be answered: (1) what is the optimal allocation of computational resources to minimize the estimation uncertainty and (2) how to quantify the estimation uncertainty. In the literature, these two questions have been answered for the simulation model that generates a deterministic output given an input, $\mathbf{X}$ (Kahn and Marshall 1953; Geweke 2005; Koopman et al. 2009). Such simulation models, referred to as deterministic simulation models in this study, use deterministic black-box simulators in the second level, and the stochasticity comes from $\mathbf{X}$ only. Recently, Choe et al. (2015) address the first question and propose

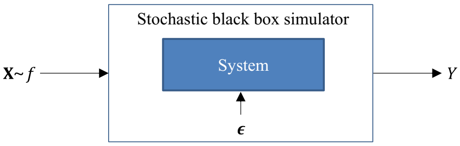

Fig. 1 Two-level stochastic simulation

importance sampling (IS) for two-level stochastic simulations, to efficiently evaluate the system reliability.

Capitalizing the results in Choe et al. (2015), this study aims to answer the second question by establishing the central limit theorem (CLT) for the IS estimator under near-optimal allocation of the second level simulation effort. This CLT can be readily used to construct confidence intervals (CIs) and thereby quantify the estimation uncertainty. The challenge of analyzing this near-optimally allocated estimator lies in the adaptive nature of the second level replication sizes on all realizations of $\mathbf{X}$ at the first level, which introduces dependencies among all the samples. This source of dependency is quite different from the serial dependencies typically considered in the stochastic literature (e.g., different mixing conditions such as α-mixing (also known as strong mixing) (Rosenblatt 1956), ρ-mixing (Ibragimov 1975), l-mixing (Withers 1981), m-dependence (Hoeffding and Robbins 1948), and positive/negative associations (Roussas 1994)). Rather, it is more closely related to the literature of nested simulation, where analysis of variance (e.g., Sun et al. 2011) and kernel estimator (e.g., Hong et al. 2017) have been suggested. The related problem of estimating conditional density (e.g., Steckley et al. 2016) has also been studied. However, as far as we know, there has been no asymptotic distributional analysis on situations where the second level simulation allocation is adaptive on the first level, which, as Choe et al. (2015) and our subsequent discussion reveal, is capable of significantly reducing the estimation uncertainty. Our main contribution in this paper is to conduct such an analysis rigorously and demonstrate its use in the uncertainty quantification of the considered estimator. We validate the proposed procedure using a numerical study, and demonstrate the utility of the method via a case study on the wind turbine reliability evaluation.

The remainder of this paper is organized as follows. Section 2 reviews the background of this study. Section 3 presents the asymptotic properties of the IS estimator and constructs CIs. Section 4 presents the numerical study and Section 5 details the case study. Section 6 concludes the paper with a summary.

2 Background

To evaluate the system reliability, we consider the failure probability, $p_y \equiv \mathbb{P}(Y > y)$, where $Y \in \mathbb{R}$ is the system output of interest from a simulation run and y is a pre-specified threshold corresponding to the system's resistance level.

The input, $\mathbf{X} \in \mathbb{R}^p$, to a simulator follows a known probability density, f. Due to the random vector, $\boldsymbol{\epsilon}$, hidden inside the black-box simulator, the stochastic simulation model produces the random output, Y, even if the input, $\mathbf{X}$, is fixed. The conditional distribution of Y given $\mathbf{X}$ is unknown. We can estimate the failure probability based on the outputs of multiple simulation runs.

The crude Monte Carlo (CMC) method (Kroese et al. 2011) is one of the most common ways to estimate the failure probability. CMC repeats sampling $\mathbf{X}$ from its known density, f, and passing $\mathbf{X}$ to the simulation model generating the corresponding Y. CMC is, however, inefficient because it does not take into account how likely certain input conditions generate failure events, $\{Y > y\}$.

Alternatively, variance reduction techniques can reduce the number of total simulation replications (or total sample size), n, necessary to achieve a target variance of the probability estimation. IS is regarded as one of the most effective variance reduction techniques (e.g., Glynn and Iglehart 1989, Rubinstein 1999, Botev and Kroese 2008). IS methods have been applied in stochastic simulations in various settings, such as finance (Glasserman and

Li 2005; Kawai 2008), insurance (Asmussen et al. 2000), reliability (Heidelberger 1995; Balesdent et al. 2016), communication networks (Chang et al. 1994), and queueing operations (Sadowsky 1991; Blanchet et al. 2009; Blanchet and Lam 2014). For many of these applications, IS methods are designed by exploiting special structures of the underlying stochastic processes, for instance via large deviations analysis (Juneja and Shahabuddin 2006; Blanchet and Lam 2012; Gatto and Baumgartner 2016).

On the other hand, IS can be also used for stochastic simulations that are not necessarily represented as stochastic processes. Consider two-level simulations where $\mathbf{X}$ is drawn from a known distribution at the first level, but Y is generated from a black-box simulator using $\mathbf{X}$ at the second level. If the black-box simulator is a deterministic simulation model (i.e., Y is deterministic, given $\mathbf{X}$), we can use an IS method, referred to as DIS in this study. This method has been widely used since its development in the '50s (Kahn and Marshall 1953).

If the black-box simulator at the second level is a stochastic simulation model, we can use the recently developed method in Choe et al. (2015), called Stochastic Importance Sampling (SIS). We briefly summarize the method to estimate the failure probability, $p_y = \mathbb{P}(Y > y) = \mathbb{E}[\mathbb{E}[\mathbb{I}(Y > y) \mid \mathbf{X}]]$. Note that to better present our method in this paper, we use slightly different notations from Choe et al. (2015). First, SIS considers that the simulation output, Y, is random at a fixed input, $\mathbf{X}$, and allows multiple simulation replications at the sampled input to capture the randomness in the output. The SIS estimator of the failure probability is

$$\hat{P}_n(y) = \frac{1}{m} \sum_{i=1}^{m} \left(\frac{1}{N_i} \sum_{j=1}^{N_i} \mathbb{I}\left(Y_j^{(i)} > y \right) \right) L_i, \tag{1}$$

where the likelihood ratio, L_i, denotes $f(\mathbf{X}_i)/q(\mathbf{X}_i)$. m is the input sample size, denoting the number of times that the input, $\mathbf{X}$, is sampled independently from a new density, q; N_i is the allocation size, denoting the number of simulation replications alloted to $\mathbf{X}_i$; $Y_j^{(i)}$ is the jth replication output at $\mathbf{X}_i$. In other words, SIS samples m inputs, $\mathbf{X}_1, \ldots, \mathbf{X}_m$, from q, and runs the simulator N_i times at each sampled $\mathbf{X}_i$, $i = 1, \ldots, m$. As a result, we observe the total $n = \sum_{i=1}^{m} N_i$ outputs of $Y_j^{(i)}$ for $i = 1, \ldots, m$, and $j = 1, \ldots, N_i$. Note that the realizations of $\mathbf{X}_1, \ldots, \mathbf{X}_m$ at the first step can inform the choice of N_i, $i = 1, \ldots, m$, at the second step. Thus, SIS allows N_i to be a function of $\mathbf{X}_1, \ldots, \mathbf{X}_m$.

Fixing $N_i = 1$, $i = 1, \ldots, m$, reduces the SIS estimator in Eq. 1 to the SIS2 estimator in Choe et al. (2015), which is therein empirically shown to be inferior to the aforementioned adaptive scheme. This also motivates our analytical focus on the adaptive approach in this paper.

In contrast to CMC, SIS samples $\mathbf{X}_i$, $i = 1, \ldots, m$, from the new density, q, instead of f. This change of distribution allows sampling efforts to be focused on important input regions where failure events are likely to happen. To compensate for the bias created by the change of distribution, the estimator, $\hat{P}_n(y)$, in Eq. 1 includes the likelihood ratio, $f(\mathbf{X}_i)/q(\mathbf{X}_i)$.

For given n and m, the estimator, $\hat{P}_n(y)$, in Eq. 1 is unbiased and has the minimum variance when we use the optimal SIS density, $q_y^*(\mathbf{x})$, and the optimal allocation size, N_i^*, $i = 1, \ldots, m$, as follows (Choe et al. 2015):

$$q_y^*(\mathbf{x}) = \frac{1}{C_q} f(\mathbf{x}) \sqrt{\frac{1}{n} s_y(\mathbf{x}) \left(1 - s_y(\mathbf{x})\right) + s_y(\mathbf{x})^2} \tag{2}$$

and

$$N_i^* = n \frac{h^*(\mathbf{X}_i)}{\sum_{j=1}^{m} h^*(\mathbf{X}_j)}, \quad i = 1, \ldots, m, \tag{3}$$

where

$$h^*(\mathbf{x}) = \sqrt{\frac{n\left(1 - s_y(\mathbf{x})\right)}{1 + (n-1)\, s_y(\mathbf{x})}}. \tag{4}$$

Here, $s_y(\mathbf{x})$ is $\mathbb{P}(Y > y \mid \mathbf{X} = \mathbf{x})$ and C_q in Eq. 2 is the normalizing constant. Because the conditional probability, $s_y(\mathbf{x})$, is unknown in practice, the optimal solutions in Eqs. 2 and 3 need to be approximated for implementation in practice. For example, our implementation (i.e., the numerical study in Section 4 and the case study in Section 5) approximates $s_y(\mathbf{x})$ by a metamodel (see Section 3 for the implementation guideline). The optimality of SIS is with respect to this approximation. Therefore, the empirical performance of SIS depends on the approximation of $s_y(\mathbf{x})$.

Let the function, $0 \le \hat{s}_y(\mathbf{x}) \le 1$, denote the metamodel (or emulator) of $s_y(\mathbf{X})$. In the implementation, we use the following density that replaces $s_y(\mathbf{x})$ in $q_y^*(\mathbf{x})$ in Eq. 2 by $\hat{s}_y(\mathbf{x})$:

$$q_y(\mathbf{x}) = \frac{1}{\hat{C}_q} f(\mathbf{x}) \sqrt{\frac{1}{n}\hat{s}_y(\mathbf{x})\left(1 - \hat{s}_y(\mathbf{x})\right) + \hat{s}_y(\mathbf{x})^2}, \tag{5}$$

where $\hat{C}_q$ is the normalizing constant of q_y. Similarly, the allocation size, N_i, used in the implementation, is an approximation of the optimal allocation size, N_i^*, in Eq. 3, by substituting $\hat{s}_y(\mathbf{X})$ for $s_y(\mathbf{X})$ and rounding to the nearest positive integer as follows:

$$N_i \equiv \max\left(1, \left\lfloor n\frac{h_n(\mathbf{X}_i)}{\sum_{j=1}^m h_n(\mathbf{X}_j)} + \frac{1}{2}\right\rfloor\right), \quad i = 1, \ldots, m, \tag{6}$$

where

$$h_n(\mathbf{X}) = \sqrt{\frac{n\left(1 - \hat{s}_y(\mathbf{X})\right)}{1 + (n-1)\,\hat{s}_y(\mathbf{X})}}. \tag{7}$$

The floor function, $\lfloor x \rfloor$, in Eq. 6 yields the largest integer not greater than x. Thus, $\lfloor x + 1/2 \rfloor$ is equivalent to rounding x. The sum of N_i, $i = 1, \ldots, m$, in Eq. 6 may deviate slightly from the pre-specified total sample size, n. If we want to ensure $n = \sum_{i=1}^m N_i$ in the implementation, we can adjust either n or some N_i's. For simplicity, we ignore such minor adjustments in the following discussions.

We note that even if the true $s_y(\mathbf{x})$ is used in the implementation to achieve the minimum variance of $\hat{P}_n(y)$, in Eq. 1, the minimum variance is not necessarily zero because we can only optimally allocate the simulation efforts at the first level of simulation and cannot control the randomness within the stochastic simulation model in the second level.

Although the above optimal solutions minimizing the variance of the estimator in Eq. 1 is derived for stochastic simulation models, the distributional property of the SIS estimator is not yet understood well. In particular, quantifying the estimation uncertainty by building a valid CI would be substantially important in practice.

3 Asymptotic Properties of the SIS Estimator

3.1 Confidence Interval for SIS

We first state our main result of the paper, Theorem 1, which constructs the CI for p_y based on the SIS estimator in Eq. 1 with the density q satisfying Assumptions 1–3 as well as both m/n and N_i in Eq. 6 satisfying Assumption 3.

Assumption 1 If $q(\mathbf{x}) = 0$, then $\mathbb{P}(Y > y \mid \mathbf{X} = \mathbf{x})\, f(\mathbf{x}) = 0$ for any $\mathbf{x}$.

Assumption 2 $\mathbb{E}_q\left[\mathbb{I}(Y > y)L^2\right] < \infty$ holds, where the expectation is taken with respect to q.

Assumption 3 For a constant $c_0 \in (0, 1]$, $m/n = c_0 + o(1)$ as $n \to \infty$. The function

$$\tilde{h}(\mathbf{x}) \equiv \sqrt{\frac{1 - \hat{s}_y(\mathbf{x})}{\hat{s}_y(\mathbf{x})}} \tag{8}$$

satisfies the following conditions

$$\mathbb{E}_q\left[\tilde{h}(\mathbf{X})\right] < \infty \tag{9}$$

and

$$\mathbb{P}\left(\frac{\tilde{h}(\mathbf{X})}{c_0 \mathbb{E}_q\left[\tilde{h}(\mathbf{X})\right]} + \frac{1}{2} \in \mathcal{N}\right) = 0 \tag{10}$$

where $\mathcal{N} \equiv \{2, 3, \ldots\}$.

We prove Theorem 1 through this section and defer all the other proofs, except for Theorem 2 which is our main technical result, to the supplementary document. Here, we define $z_{\alpha/2} \equiv \Phi^{-1}(1 - \alpha/2)$ for $\alpha \in (0, 1)$, where $\Phi(\cdot)$ is the cumulative distribution function of $N(0, 1)$.

Theorem 1 ***(CI for SIS)*** *Suppose Assumptions 1–3 hold. Then,*

$$\sqrt{\frac{m}{\hat{\sigma}_y^2}}\left(\hat{P}_n(y) - p_y\right) \xrightarrow{d} N(0, 1) \tag{11}$$

as $m \to \infty$, where

$$\hat{\sigma}_y^2 = \frac{1}{m-1}\sum_{i=1}^{m}\left(\frac{1}{N_i}\sum_{j=1}^{N_i}\mathbb{I}\left(Y_j^{(i)} > y\right)L_i - \hat{P}_n(y)\right)^2. \tag{12}$$

Therefore, $\mathbb{P}\left(p_y \in \left(\hat{P}_n(y) \pm z_{\alpha/2}\hat{\sigma}_y/\sqrt{m}\right)\right) \to 1 - \alpha$ for $\alpha \in (0, 1)$ as $m \to \infty$. That is, $\left(\hat{P}_n(y) \pm z_{\alpha/2}\hat{\sigma}_y/\sqrt{m}\right)$ is a $100(1-\alpha)\%$ asymptotic confidence interval for p_y.

This theorem provides an asymptotically valid way to quantify the uncertainty of estimating p_y based on the SIS estimator in Eq. 1 and the variance estimator in Eq. 12. We below discuss the assumptions used in the theorem and its proof.

Assumption 1 implies that we should use the density, $q(\mathbf{x})$, that makes the estimator, $\hat{P}_n(y)$, in Eq. 1 unbiased. This assumption is satisfied when we use the density, $q_y(\mathbf{x})$, in Eq. 5 with the metamodel, $\hat{s}_y(\mathbf{x})$, satisfying a mild condition (see Proposition 1). In practice, this condition can be readily satisfied by imposing the metamodel, $\hat{s}_y(\mathbf{x})$, to be strictly positive in the support of f.

Proposition 1 *The SIS density, $q_y(\mathbf{x})$, in* Eq. 5 *for $n \geq 1$ satisfies Assumption 1 if $s_y(\mathbf{x}) \neq 0$ implies $\hat{s}_y(\mathbf{x}) \neq 0$ in the support of f.*

Assumption 2 implies that the SIS estimator should have a finite variance. This assumption is also satisfied by the density, $q_y(\mathbf{x})$, in Eq. 5 if the metamodel, $\hat{s}_y(\mathbf{x})$, satisfies another mild condition (see Proposition 2). This condition is slightly stronger than the condition in Proposition 1, as it requires the ratio, $f(\mathbf{X})s_y(\mathbf{X})/\hat{s}_y(\mathbf{X})$, to not explode in the support of f. Analogous to Assumptions 1 and 2, to prove the CLT for DIS, similar or stronger conditions are commonly required in the literature (Koopman et al. 2009).

Proposition 2 *Under Assumption 1, the SIS density, $q_y(\mathbf{x})$, in Eq. 5 for $n \geq 1$ satisfies Assumption 2 if $\mathbb{E}_f\left[s_y(\mathbf{X})/\hat{s}_y(\mathbf{X})\right] < \infty$.*

Assumption 3 generally holds in practical situations. The m/n ratio can be approximately set at a fixed level (e.g., $c_0 = 10\%$ or 30%) according to the empirical finding and implementation guideline suggested in Choe et al. (2015). Proposition 3 proves that $\tilde{h}(\mathbf{x})$ in Eq. 8 satisfies the condition in Eq. 9 for $q = q_y(\mathbf{x})$ in Eq. 5. The condition in Eq. 10 is to address discontinuous points due to the rounding of N_i in Eq. 6, implying that the limit of non-rounded N_i, $\tilde{h}(\mathbf{X}_i)/(c_0\mathbb{E}_q\left[\tilde{h}(\mathbf{X})\right]) + 1/2$, should not belong to a set of integers greater than 1. This condition holds when we impose the continuity on the metamodel, $\hat{s}_y(\cdot)$, for continuous $\mathbf{X}$. In general simulation studies that develop metamodels (or emulators), it is common to model an unknown function as a continuous function (e.g., Plumlee and Tuo 2014; Zhang and Apley 2014, 2016).

Proposition 3 *The function $\tilde{h}(\mathbf{x})$ in* Eq. 8 *satisfies the condition $\mathbb{E}_q\left[\tilde{h}(\mathbf{X})\right] < \infty$ in Eq. 9 for $q = q_y$ in* Eq. 5 *for $n \geq 1$.*

To prove Theorem 1, in the sequel, we prove that the CLT for the SIS estimator in Eq. 1 holds under Assumptions 1–3. We then prove that the variance estimator, $\hat{\sigma}_y^2$, in Eq. 12 converges in probability, and conclude with Slutsky's theorem.

Adding much complexity to DIS, the SIS estimator in Eq. 1 involves the allocation size, N_i, which takes account of all the realizations of the first-level variables. Our main technical result, Theorem 2, builds a technique that handles such complication. As the first step toward proving CLT for SIS, we need to characterize the asymptotic behavior of N_i.

The allocation size, N_i, in Eq. 6 depends not only on $\mathbf{X}_i$ but also on all $\mathbf{X}_j$, $j = 1, \ldots, m$. Accordingly, N_i is not independent of N_j for $j \neq i$. We address this dependency issue in Lemma 1 by showing that under certain regularity conditions, the allocation size with a fixed index, say N_k, becomes asymptotically independent of N_j, $j \neq k$, as m increases.

Lemma 1 ***(Asymptotic independence between the allocation sizes)*** *Suppose that Assumption 3 holds. Then, for any fixed index k,*

$$N_k \xrightarrow{P} \tilde{N}_k \tag{13}$$

$$\equiv \max\left(1, \left\lfloor \frac{\tilde{h}(\mathbf{X}_k)}{c_0\mathbb{E}_q\left[\tilde{h}(\mathbf{X})\right]} + \frac{1}{2} \right\rfloor\right), \tag{14}$$

as $m \to \infty$.

Note that the convergence in Eq. 13 is not uniform over k. Our derivation for Theorem 2 shows that, despite the lack of uniform convergence, the CLT for the SIS estimator based on two-level stochastic simulations still holds.

Building upon Lemma 1 that characterizes the asymptotic independence of the allocation sizes, we derive the CLT for SIS in Theorem 2. Its proof is lengthy, so we defer it to Appendix.

Theorem 2 ***(CLT for SIS estimator)*** *Suppose Assumptions 1–3 hold. Then,*

$$\sqrt{\frac{m}{\sigma_y^2}}\left(\hat{P}_n(y) - p_y\right) \xrightarrow{d} N(0,1) \tag{15}$$

as $m \to \infty$, where

$$\sigma_y^2 = \mathbb{E}_q\left[\frac{1}{\tilde{N}} s_y(\mathbf{X})\left(1 - s_y(\mathbf{X})\right) L^2\right] + \mathbb{E}_q\left[s_y(\mathbf{X})^2 L^2\right] - p_y^2 \tag{16}$$

with

$$\tilde{N} = \max\left(1, \left\lfloor \frac{\tilde{h}(\mathbf{X})}{c_0 \mathbb{E}_q\left[\tilde{h}(\mathbf{X})\right]} + \frac{1}{2} \right\rfloor\right).$$

Theorem 2 describes the asymptotic normality of the SIS estimator, $\hat{P}_n(y)$, in Eq. 1. As m increases, the SIS estimator becomes close to a normal random variable with the mean of p_y and the variance of σ_y^2/m. We note that '$m \to \infty$' is equivalent to '$n \to \infty$' under Assumption 3. $\tilde{N}$ is a random variable depending on $\mathbf{X}$ that follows the density, q.

Theorem 2 provides the information on the distributional properties of SIS estimator in the asymptotic regime. Yet, the asymptotic variance is unknown, because σ_y^2 in Eq. 16 involves $s_y(\mathbf{X}) = \mathbb{P}(Y > y \mid \mathbf{X})$ and p_y. Lemma 2 states that $\hat{\sigma}_y^2$ in Eq. 12 is a consistent estimator of σ_y^2.

Lemma 2 ***(Consistency of a variance estimator)*** *Suppose Assumptions 1–3 hold. Then,*

$$\hat{\sigma}_y^2 \xrightarrow{P} \sigma_y^2 \tag{17}$$

as $m \to \infty$.

The statement in Eq. 11 in Theorem 1 follows from Theorem 2 and Lemma 2 by Slutsky's theorem. This completes the proof of Theorem 1.

3.2 Confidence Intervals With Different Thresholds

The optimal SIS method depends on the failure threshold, y, leading to the sampling and simulation results optimized for the particular y. Suppose we obtain the simulation outputs with y. We can still use the obtained simulation outputs to estimate the failure probability at a different threshold, $\tilde{y}$, for $\tilde{y} > y$ without conducting experiments again.

Suppose that given y, we sample $\mathbf{X}_i, i = 1, \ldots, m$, from the density, q, satisfying Assumption 1 and obtain the simulation outputs, $Y_j^{(i)}$ for $i = 1, \ldots, m$ and $j = 1, \ldots, N_i$. Then, to estimate the failure probability, $p_{\tilde{y}} = \mathbb{P}(Y > \tilde{y})$, we can replace y with $\tilde{y}$ and use the SIS estimator, $\hat{P}_n(\tilde{y})$, in Eq. 1. The estimator, $\hat{P}_n(\tilde{y})$, is an unbiased estimator of $p_{\tilde{y}}$ for $\tilde{y} > y$ because if $q(\mathbf{x}) = 0$, then $\mathbb{P}(Y > \tilde{y} \mid \mathbf{X} = \mathbf{x}) f(\mathbf{x}) = 0$ for any $\mathbf{x}$, due to the fact that $\mathbb{P}(Y > y \mid \mathbf{X} = \mathbf{x}) f(\mathbf{x}) = 0$ under Assumption 1 and $\mathbb{P}(Y > \tilde{y} \mid \mathbf{X} = \mathbf{x}) f(\mathbf{x}) \leq \mathbb{P}(Y > y \mid \mathbf{X} = \mathbf{x}) f(\mathbf{x})$ for $\tilde{y} > y$ (Choe et al. 2016). On the other hand, for $\tilde{y} < y$, the unbiasedness of the estimator, $\hat{P}_n(\tilde{y})$, does not necessarily hold because $\mathbb{P}(Y > \tilde{y} \mid \mathbf{X} = \mathbf{x}) f(\mathbf{x})$

is not necessarily zero for some $\mathbf{x}$ even if $q(\mathbf{x}) = 0$. Corollary 1 below constructs the pointwise CI for $p_{\tilde{y}}$ for $\tilde{y} > y$ using the simulation outputs optimized for estimating p_y.

Corollary 1 ***(Pointwise CI for*** $\tilde{y} > y$***)*** *Suppose the conditions in Theorem 1 hold. Then, for* $\tilde{y} > y$*, the CI for* $p_{\tilde{y}}$, $\left(\hat{P}_n(\tilde{y}) \pm z_{\alpha/2}\hat{\sigma}_{\tilde{y}}/\sqrt{m}\right)$ *is asymptotically valid, i.e.,* $\mathbb{P}\left(p_{\tilde{y}} \in \left(\hat{P}_n(\tilde{y}) \pm z_{\alpha/2}\hat{\sigma}_{\tilde{y}}/\sqrt{m}\right)\right) \to 1 - \alpha$ *for* $\alpha \in (0, 1)$ *as* $m \to \infty$.

We believe that the result in Corollary 1, which justifies the pointwise CI for $\tilde{y} > y$, is practically desirable. At the system design stage, designers want to estimate the failure probability and quantify the estimation uncertainties at multiple design parameters, $\tilde{y}$'s, rather than at a single value of y. In particular, system designers are interested in a large resistance level, $\tilde{y}$, which corresponds to a small failure probability, $p_{\tilde{y}}$, to ensure a high level of system reliability. Corollary 1 suggests that we can construct the CI for $p_{\tilde{y}}$ using the results optimized for p_y, without rerunning the simulation with each $\tilde{y}$.

We note that the result in Corollary 1 may not hold when $\tilde{y} < y$. Also, we remark that our practical interest in this study lies in the pointwise uncertainty of estimating $p_{\tilde{y}}$ at certain $\tilde{y}$'s. Under stronger assumptions, simultaneous CI may be established if one wants to evaluate the simultaneous uncertainty of estimating $p_{\tilde{y}}$ over a range of $\tilde{y}$, which is beyond the scope of this study.

3.3 Implementation Guideline

We summarize the implementation procedure to obtain the CI for SIS estimator.

Implementation procedure

1. Given y, sample $\mathbf{X}_i$, $i = 1, \ldots, m$, from the SIS density, q_y in Eq. 5.
2. For each $\mathbf{X}_i$, run the simulator N_i (in Eq. 6) times to obtain $Y_j^{(i)}$ for $i = 1, \ldots, m$ and $j = 1, \ldots, N_i$.
3. Estimate the failure probability for $\tilde{y}$ by $\hat{P}_n(\tilde{y})$ in Eq. 1 for $\tilde{y} \geq y$.
4. Obtain $\hat{\sigma}_{\tilde{y}}$ in Eq. 12.
5. Construct the $100(1-\alpha)\%$ pointwise CI, $\left(\hat{P}_n(\tilde{y}) \pm z_{\alpha/2}\hat{\sigma}_{\tilde{y}}/\sqrt{m}\right)$, for $p_{\tilde{y}}$.

In Steps 1 and 2, as noted in Section 2, the SIS density and allocation size need the metamodel, $\hat{s}_y(\mathbf{x})$, of the conditional probability, $s_y(\mathbf{x}) = \mathbb{P}(Y > y \mid \mathbf{X} = \mathbf{x})$. Depending on applications, different methods can be used for constructing the metamodel, $\hat{s}_y(\mathbf{x})$. For example, Choe et al. (2015) provide a guideline on how to build the metamodel for wind turbine reliability estimation, which we employ in our implementation. Specifically, we build the metamodel based on a small pilot sample, using the generalized additive model for location, scale and shape (GAMLSS) (Rigby and Stasinopoulos 2005) that allows us to model the conditional distribution of Y at $\mathbf{X} = \mathbf{x}$ with a parametric distribution (e.g., we use the normal distribution in Section 4 and the generalized extreme value (GEV) distribution (Coles 2001) in Section 5). In GAMLSS, the conditional distribution parameters can be modeled using cubic spline functions of $\mathbf{x}$. The model fitting is done by the backfitting algorithm commonly used in generalized additive modeling (Rigby and Stasinopoulos 2005). The smoothing parameters in the model are determined by minimizing the Bayesian information criterion (Schwarz 1978), as suggested by Rigby and Stasinopoulos (2005). The metamodel's goodness-of-fit can be tested using the Kolmogorov-Smirnov test (Choe et al.

2015). Once we have the metamodel, we can draw **X** from the SIS density in Eq. 5 using the acceptance-rejection algorithm (Kroese et al. 2011).

4 Numerical Study

This section presents a numerical example to show that the empirical coverage level of the proposed CI agrees with the target coverage probability, $1-\alpha$, under various settings.

Cannamela et al. (2008) originally develop a deterministic simulation model example, which is later modified by Choe et al. (2015) into the stochastic simulation model example. We use the latter data generating model as follows:

$$X \sim N(0,1)\,, \qquad Y|X \sim N\Big(\mu(X)\,, \sigma^2(X)\Big)\,, \tag{18}$$

with

$$\begin{aligned} \mu(X) &= 0.95\delta X^2\,(1+0.5\cos(5X)+0.5\cos(10X))\,, \\ \sigma(X) &= 1+0.7\,|X|+0.4\cos(X)+0.3\cos(14X). \end{aligned} \tag{19}$$

The metamodel of the conditional distribution of $Y|X$ is set as the normal distribution with its mean and standard deviation modeled as cubic spline functions of X in the GAMLSS framework described in Section 3. The model is fitted to a pilot sample of size 600 that includes X's uniformly sampled between $(-4, 4)$ and their corresponding Y's from the true conditional distribution.

In this example, all model and experiment parameters are set as in Choe et al. (2015). The parameter, δ, in Eq. 19 determines the similarity of the optimal SIS density, q_y^*, in Eq. 2 to the original input density, f. For $\delta = 1$ (-1), the important regions are far from (close to) $X = 0$, which is the mode of f, the density of $N(0, 1)$. Consequently, the SIS densities that focus on the important regions differ significantly for different δ's. We use the failure threshold, y, that corresponds to the true failure probability, $p_y = 0.01$. The ratio of m/n is set as 30%.

To compute the empirical coverage level, we repeatedly construct the $100(1-\alpha)\%$ CI, $\left(\hat{P}_n(y) \pm z_{\alpha/2}\hat{\sigma}_y/\sqrt{m}\right)$, 10,000 times and calculate the proportion of the CIs covering the true failure probability, p_y. We consider the target coverage probability, $1-\alpha$, of 0.90 and 0.95. Table 1 shows the experiment results. We summarize the key observations as follows:

Table 1 Empirical coverage level vs. the total sample size, n, in the numerical study

	$\delta = 1$		$\delta = -1$	
$1-\alpha$	0.90	0.95	0.90	0.95
n	Empirical coverage level			
1000	0.89	0.94	0.87	0.92
10000	0.89	0.94	0.89	0.95
100000	0.90	0.95	0.90	0.95

NOTE: The empirical coverage level is the proportion of CIs (out of 10,000 experiments) that include the true failure probability, $p_y = 0.01$.

- With the moderate size of n of 1000 (note that $p_y = 0.01$), the corresponding empirical coverages are close to the target coverage probabilities, $1-\alpha$.
- As n increases, the empirical coverage level reaches the target coverage probability, $1-\alpha$. This result agrees with the asymptotic result stated in Theorem 1.
- The parameters, α and δ, do not appear to significantly affect the behavior of CI coverage.

5 Case Study

Wind turbines experience stochastic weather conditions (Byon et al. 2010). To reflect the randomness into the reliability evaluation of the turbine design, the international standard, IEC 61400-1 (International Electrotechnical Commission 2005), requires the turbine designer to use stochastic simulations. The wind industry commonly uses NREL aeroelastic simulators that simulate stochastic loads imposed on a turbine (Moriarty 2008; Manuel et al. 2013; Choe et al. 2016). We focus our analysis on a load response type, namely, the bending moment at a blade root. The blade bending moment represents the structural response of a blade due to an external force or moment causing the blade to bend (Soleimanzadeh et al. 2012; Byon et al. 2016). The extreme bending moment at a blade root can lead to the structural failure of the blade and is extensively studied in the wind industry to ensure the turbine reliability. Therefore, our simulation output of interest, Y, is the blade bending moment. We estimate a small failure probability associated with an extreme load level, which can be observed rarely with the probability less than, or equal to, 0.01.

5.1 Details of the Simulations

We use the benchmark simulation setup required in the IEC 61400-1 standard (International Electrotechnical Commission 2005) with the same setting used in the studies by Moriarty (2008) and Choe et al. (2015), and obtain the blade bending moment, following the procedure in Moriarty (2008). The simulation input, $\mathbf{X}$, is the wind speed (unit: m/s) sampled from a truncated Rayleigh distribution (with the support of $[3, 25]$ and the scale parameter of $10\sqrt{2/\pi}$). The supplementary document includes the implementation details on the NREL simulators.

In the implementation, one simulation replication represents sampling one $\mathbf{X}$ from the truncated Rayleigh distribution and running the NREL simulators to simulate 10-min turbine operations using the input, $\mathbf{X}$. Each simulation replication takes about 1-min wall-clock time by a regular computer available nowadays. To sample $\mathbf{X}$ from the SIS density, we use the same metamodel and sampling technique used in Choe et al. (2015). For the metamodel to approximate $s_y(\mathbf{x}) = \mathbb{P}(Y > y \mid \mathbf{X} = \mathbf{x})$, the GEV distribution is used with its location and scale parameters varying with $\mathbf{x}$ while keeping the shape parameter constant in the GAMLSS framework described in Section 3. The model is fitted to a pilot sample of size 600 including $\mathbf{X}$'s uniformly sampled from $[3, 25]$ and the corresponding Y's generated from the NREL simulators. The model's goodness-of-fit is tested and confirmed using the Kolmogorov-Smirnov test (Choe et al. 2015).

5.2 Implementation Results

We first test whether the empirical coverage level of CI is similar to the target coverage probability. Unlike the numerical studies in Section 4 where we repeat the experiment 10,000

Table 2 $100(1-\alpha)\%$ CI's empirical coverage level based on 200 experiments in the case study

Target coverage, $1-\alpha$	0.70	0.80	0.90	0.95	0.99
Empirical coverage	0.755	0.85	0.93	0.95	0.99
(95% CI)	(0.695, 0.815)	(0.80, 0.90)	(0.89, 0.97)	(0.92, 0.98)	(0.98, 1.00)

NOTE: The empirical coverage level denotes the proportion of experiments whose CIs include the estimated p_y in the total 200 experiments, e.g., 151/200 = 0.755 for $1-\alpha = 0.70$. Because the empirical coverage level is a binomial proportion estimator, we compute and present the 95% normal approximate CI of the coverage probability in the parentheses.

times, we limit the repetition to 200 times in this case study because of the high computational cost (high-performance computing with 28 cores takes about 6 weeks of wall-clock time). We use $y = 14,300$ *kNm*, $n = 9,000$, and $m/n = 30\%$. Because p_y is unknown in this case study, we estimate it with the sample average of the 200 failure probability estimates , namely, 0.0105. We compute the empirical coverage level by obtaining the proportion of CIs that cover the estimated p_y.

Table 2 shows the empirical coverage levels for different target coverage probabilities ($1-\alpha = 0.70$, 0.80, 0.90, 0.95, and 0.99). The observed coverage level appears to generally match the target level, as the 95% CIs of the coverage probabilities all cover the target coverage levels and are reasonably narrow. We note that 200 repetitions result in the coverage probability estimates accurate enough for our validation purpose.

Next, to illustrate how the CIs can help a design process, we estimate the failure probability of 10^{-2} or less because such a small failure probability is desired in the wind industry (Lee et al. 2013). To do so, we pool all the results from the 200 experiments. The pooled estimator of the failure probability, p_y, is $\hat{P}_{200n}(y)$ in Eq. 1 with m replaced by $200m$. We also use the result in Theorem 1 with $200n$ and $200m$ in place of n and m, respectively, and construct the CIs of $p_{\tilde{y}}$ for some $\tilde{y}$'s greater than y, based on Corollary 1.

Table 3 shows the point estimates and CIs for the failure probabilities corresponding to three $\tilde{y}$'s greater than $y = 14{,}300$ *kNm*. We note that the ratio of CI width to the corresponding point estimate increases as $\tilde{y}$ increases, reflecting the increasing uncertainty in the distribution tail. This is because the experiments were optimized to estimate p_y for $y = 14{,}300$ *kNm*. As the threshold, $\tilde{y}$, increases, a smaller number of simulation outputs, which were generated from the original experiments with $y = 14{,}300$ *kNm*, contribute to $\hat{P}_{200n}(\tilde{y})$ in Eq. 1 and the corresponding CI in Corollary 1, because a larger number of outputs result in $\mathbb{I}\left(Y_j^{(i)} > \tilde{y}\right) = 0$ in Eq. 1. Accordingly, as $\tilde{y}$ becomes greater than y, the estimation uncertainty gets larger.

Table 3 Failure probability point estimates and 95% CIs (in parentheses) for blade bending moments using the simulation outputs from 200 experiments with $y = 14{,}300$ *kNm*

	Failure threshold, $\tilde{y}$ (*kNm*)		
	14,500	15,000	15,500
Point Est.	0.00515	0.000733	0.0000928
95% CI	(0.00508, 0.00522)	(0.000704, 0.000761)	(0.0000819, 0.0001038)

6 Summary

SIS can significantly save computational resources in estimating the probability associated with the output of two-level stochastic simulation (Choe et al. 2015). This paper studies the asymptotic properties of the SIS estimator with a focus on measuring the estimation uncertainty. We establish the CLT for the SIS estimator and construct the asymptotically valid CI that uses a consistent variance estimator. Our numerical study shows that the asymptotic CI's empirical coverage level indeed converges to the target coverage probability. In our case study, we use the CI to quantify the uncertainty of the failure probability estimation for wind turbine reliability evaluation.

Acknowledgements The authors greatly appreciate an editorial board member and an anonymous reviewer for their thorough review and comments that helped improve the manuscript greatly. This work was partially supported by the National Science Foundation (Grant No. CMMI-1362513, CMMI-1523453, CMMI-1542020, CAREER CMMI-1653339, and IIS-1741166) and the University of Michigan MCubed Grant.

Appendix: Proof of Theorem 2

To prove the CLT in Eq. 15,

$$\sqrt{\frac{m}{\sigma_y^2}}\left(\hat{P}_n(y) - p_y\right) \xrightarrow{d} N(0,1),$$

we introduce the following estimator:

$$\tilde{P}_n(y) = \frac{1}{m}\sum_{i=1}^{m}\left(\frac{1}{\tilde{N}_i}\sum_{j=1}^{\tilde{N}_i}\mathbb{I}\left(Y_j^{(i)} > y\right)\right)\frac{f(\mathbf{X}_i)}{q(\mathbf{X}_i)}, \tag{20}$$

where $\tilde{N}_i$ is defined in Eq. 14. Then, we express the left-hand side of Eq. 15 as

$$\begin{aligned}\sqrt{\frac{m}{\sigma_y^2}}\left(\hat{P}_n(y) - p_y\right) &= \sqrt{\frac{m}{\sigma_y^2}}\left(\hat{P}_n(y) - \tilde{P}_n(y) + \tilde{P}_n(y) - p_y\right)\\ &= \sqrt{\frac{m}{\sigma_y^2}}\left(\hat{P}_n(y) - \tilde{P}_n(y)\right) + \sqrt{\frac{m}{\sigma_y^2}}\left(\tilde{P}_n(y) - p_y\right)\end{aligned} \tag{21}$$

Our proof for Eq. 15 *consists of three main steps*:

1. Proof for the first term in Eq. 21 converging to zero in probability:

$$\sqrt{\frac{m}{\sigma_y^2}}\left(\hat{P}_n(y) - \tilde{P}_n(y)\right) \xrightarrow{P} 0. \tag{22}$$

2. Proof for the second term in Eq. 21 converging to $N(0,1)$ in distribution:

$$\sqrt{\frac{m}{\sigma_y^2}}\left(\tilde{P}_n(y) - p_y\right) \xrightarrow{d} N(0,1). \tag{23}$$

3. Application of Slutsky's theorem to Eq. 21.

To prove the first main step's result in Eq. 22, we show

$$\mathbb{P}\left(\left|\sqrt{m}\left(\hat{P}_n(y) - \tilde{P}_n(y)\right)\right| > \epsilon\right) \to 0 \tag{24}$$

for any $\epsilon > 0$ as $m \to \infty$. Both estimators, $\hat{P}_n(y)$ and $\tilde{P}_n(y)$, are unbiased estimators of p_y by Assumption 1, making

$$\begin{aligned}\mathbb{E}_q\left[\hat{P}_n(y) - \tilde{P}_n(y)\right] &= \mathbb{E}_q\left[\hat{P}_n(y)\right] - \mathbb{E}_q\left[\tilde{P}_n(y)\right] \\ &= p_y - p_y \\ &= 0.\end{aligned}$$

By Chebyshev's inequality, the left-hand side of Eq. 24 is bounded from above as follows:

$$\mathbb{P}\left(\left|\sqrt{m}\left(\hat{P}_n(y) - \tilde{P}_n(y)\right)\right| > \epsilon\right) \leq \frac{m}{\epsilon^2} Var_q\left[\hat{P}_n(y) - \tilde{P}_n(y)\right]. \tag{25}$$

Now we show that the right-hand side of Eq. 25 converges to zero as $m \to \infty$. We obtain

$$\begin{aligned}&\frac{m}{\epsilon^2} Var_q\left[\hat{P}_n(y) - \tilde{P}_n(y)\right] \\ &= \frac{m}{\epsilon^2}\left(\mathbb{E}_q\left[Var\left[\hat{P}_n(y) - \tilde{P}_n(y) \mid \mathbf{X}_1, \ldots, \mathbf{X}_m\right]\right] + Var_q\left[\mathbb{E}\left[\hat{P}_n(y) - \tilde{P}_n(y) \mid \mathbf{X}_1, \ldots, \mathbf{X}_m\right]\right]\right)\end{aligned} \tag{26}$$

by variance decomposition. The second term of Eq. 26 vanishes because

$$\begin{aligned}&\mathbb{E}\left[\hat{P}_n(y) - \tilde{P}_n(y) \mid \mathbf{X}_1, \ldots, \mathbf{X}_m\right] \\ &= \mathbb{E}\left[\frac{1}{m}\sum_{i=1}^{m}\left(\frac{1}{N_i}\sum_{j=1}^{N_i}\mathbb{I}\left(Y_j^{(i)} > y\right) - \frac{1}{\tilde{N}_i}\sum_{k=1}^{\tilde{N}_i}\mathbb{I}\left(Y_k^{(i)} > y\right)\right) L_i \mid \mathbf{X}_1, \ldots, \mathbf{X}_m\right] \\ &= \frac{1}{m}\sum_{i=1}^{m}\left(s_y(\mathbf{X}_i) - s_y(\mathbf{X}_i)\right) L_i \\ &= 0.\end{aligned}$$

In the first term of Eq. 26, we obtain

$$\begin{aligned}&Var\left[\hat{P}_n(y) - \tilde{P}_n(y) \mid \mathbf{X}_1, \ldots, \mathbf{X}_m\right] \\ &= Var\left[\frac{1}{m}\sum_{i=1}^{m}\left(\frac{1}{N_i}\sum_{j=1}^{N_i}\mathbb{I}\left(Y_j^{(i)} > y\right) - \frac{1}{\tilde{N}_i}\sum_{k=1}^{\tilde{N}_i}\mathbb{I}\left(Y_k^{(i)} > y\right)\right) L_i \mid \mathbf{X}_1, \ldots, \mathbf{X}_m\right] \\ &= \frac{1}{m^2}\sum_{i=1}^{m} Var\left[\left(\frac{1}{N_i}\sum_{j=1}^{N_i}\mathbb{I}\left(Y_j^{(i)} > y\right) - \frac{1}{\tilde{N}_i}\sum_{k=1}^{\tilde{N}_i}\mathbb{I}\left(Y_k^{(i)} > y\right)\right) \mid \mathbf{X}_1, \ldots, \mathbf{X}_m\right] L_i^2 \\ &= \frac{1}{m^2}\sum_{i=1}^{m} \mathbb{E}\left[\left(\frac{1}{N_i}\sum_{j=1}^{N_i}\mathbb{I}\left(Y_j^{(i)} > y\right) - \frac{1}{\tilde{N}_i}\sum_{k=1}^{\tilde{N}_i}\mathbb{I}\left(Y_k^{(i)} > y\right)\right)^2 \mid \mathbf{X}_1, \ldots, \mathbf{X}_m\right] L_i^2.\end{aligned}$$

Here, the conditional expectation in the last equation can be simplified as follows:

$$
\begin{aligned}
&\mathbb{E}\left[\left(\frac{1}{N_i}\sum_{j=1}^{N_i}\mathbb{I}\left(Y_j^{(i)}>y\right)-\frac{1}{\tilde{N}_i}\sum_{k=1}^{\tilde{N}_i}\mathbb{I}\left(Y_k^{(i)}>y\right)\right)^2 \mid \mathbf{X}_1,\ldots,\mathbf{X}_m\right]\\
&=\mathbb{E}\left[\frac{1}{N_i^2}\sum_{j=1}^{N_i}\mathbb{I}\left(Y_j^{(i)}>y\right)+\frac{2}{N_i^2}\sum_{k=1}^{N_i}\sum_{l>k}^{N_i}\mathbb{I}\left(Y_k^{(i)}>y\right)\mathbb{I}\left(Y_l^{(i)}>y\right)\right.\\
&\quad+\frac{1}{\tilde{N}_i^2}\sum_{j=1}^{\tilde{N}_i}\mathbb{I}\left(Y_j^{(i)}>y\right)+\frac{2}{\tilde{N}_i^2}\sum_{k=1}^{\tilde{N}_i}\sum_{l>k}^{\tilde{N}_i}\mathbb{I}\left(Y_k^{(i)}>y\right)\mathbb{I}\left(Y_l^{(i)}>y\right)\\
&\quad\left.-\frac{2}{N_i\tilde{N}_i}\sum_{k=1}^{N_i}\sum_{l=1}^{\tilde{N}_i}\mathbb{I}\left(Y_k^{(i)}>y\right)\mathbb{I}\left(Y_l^{(i)}>y\right)\mid \mathbf{X}_1,\ldots,\mathbf{X}_m\right]\\
&=\frac{1}{N_i}\left(s_y(\mathbf{X}_i)+(N_i-1)\,s_y^2(\mathbf{X}_i)\right)+\frac{1}{\tilde{N}_i}\left(s_y(\mathbf{X}_i)+\left(\tilde{N}_i-1\right)s_y^2(\mathbf{X}_i)\right)\\
&\quad-\frac{2}{N_i\tilde{N}_i}\left(\min(N_i,\tilde{N}_i)s_y(\mathbf{X}_i)+\left(N_i\tilde{N}_i-\min(N_i,\tilde{N}_i)\right)s_y^2(\mathbf{X}_i)\right)\\
&=s_y(\mathbf{X}_i)\left(1-s_y(\mathbf{X}_i)\right)\frac{N_i+\tilde{N}_i-2\min(N_i,\tilde{N}_i)}{N_i\tilde{N}_i}\\
&=s_y(\mathbf{X}_i)\left(1-s_y(\mathbf{X}_i)\right)\frac{\left|N_i-\tilde{N}_i\right|}{N_i\tilde{N}_i}\\
&=s_y(\mathbf{X}_i)\left(1-s_y(\mathbf{X}_i)\right)\left|\frac{1}{N_i}-\frac{1}{\tilde{N}_i}\right|
\end{aligned}
\tag{27}
$$

Therefore, the equation in Eq. 26 is simplified as

$$
\begin{aligned}
&\frac{m}{\epsilon^2}Var_q\left[\hat{P}_n(y)-\tilde{P}_n(y)\right]\\
&=\frac{m}{\epsilon^2}\mathbb{E}_q\left[\frac{1}{m^2}\sum_{i=1}^{m}s_y(\mathbf{X}_i)\left(1-s_y(\mathbf{X}_i)\right)\left|\frac{1}{N_i}-\frac{1}{\tilde{N}_i}\right|L_i^2\right]\\
&=\frac{1}{\epsilon^2 m}\sum_{i=1}^{m}\mathbb{E}_q\left[s_y(\mathbf{X}_i)\left(1-s_y(\mathbf{X}_i)\right)\left|\frac{1}{N_i}-\frac{1}{\tilde{N}_i}\right|L_i^2\right]\\
&=\frac{1}{\epsilon^2}\mathbb{E}_q\left[s_y(\mathbf{X}_1)\left(1-s_y(\mathbf{X}_1)\right)\left|\frac{1}{N_1}-\frac{1}{\tilde{N}_1}\right|L_1^2\right],
\end{aligned}
\tag{28}
$$

where the last equation in Eq. 28 holds because $X_1,\ldots,X_m$ are identically distributed.

We show that the expectation in Eq. 28 converges to zero as $m\to\infty$. By the continuous mapping theorem and Lemma 1, we obtain

$$
s_y(\mathbf{X}_1)\left(1-s_y(\mathbf{X}_1)\right)\left|\frac{1}{N_1}-\frac{1}{\tilde{N}_1}\right|L_1^2\xrightarrow{P}0
$$

as $m\to\infty$. Because

$$
s_y(\mathbf{X}_1)\left(1-s_y(\mathbf{X}_1)\right)\left|\frac{1}{N_1}-\frac{1}{\tilde{N}_1}\right|L_1^2\le 2s_y(\mathbf{X})L^2 \tag{29}
$$

and $\mathbb{E}_q\left[s_y(\mathbf{X})L^2\right] < \infty$ by Assumption 2, the dominated convergence theorem yields that the expectation in Eq. 28 converges to zero as $m \to \infty$. Because the right-hand side of Eq. 25 converges to zero, we complete the proof of Eq. 24, which implies (22).

To prove the second main step's result in Eq. 23,

$$\sqrt{\frac{m}{\sigma_y^2}}\left(\tilde{P}_n(y) - p_y\right) \xrightarrow{d} N(0, 1),$$

we use the Lindeberg—Lévy central limit theorem. For the theorem to hold, we verify its conditions as follows. First, $\tilde{P}_n(y)$ in Eq. 20 is the sample mean of

$$\tilde{Z}_i \equiv \left(\frac{1}{\tilde{N}_i}\sum_{j=1}^{\tilde{N}_i}\mathbb{I}\left(Y_j^{(i)} > y\right)\right)L_i, \quad i = 1, \ldots, m, \tag{30}$$

which are i.i.d. with

$$\begin{aligned}
\mathbb{E}_q\left[\tilde{Z}_i\right] &= \mathbb{E}_q\left[\left(\frac{1}{\tilde{N}_i}\sum_{j=1}^{\tilde{N}_i}\mathbb{I}\left(Y_j^{(i)} > y\right)\right)L_i\right] \\
&= \mathbb{E}_q\left[\frac{1}{\tilde{N}_i}\sum_{j=1}^{\tilde{N}_i}\mathbb{E}\left[\mathbb{I}\left(Y_j^{(i)} > y\right) \mid \mathbf{X}_i\right]L_i\right] \\
&= \mathbb{E}_q\left[\mathbb{P}\left(Y > y \mid \mathbf{X}_i\right)L_i\right] \\
&= p_y,
\end{aligned} \tag{31}$$

where the last equality holds by Assumption 1.

Next, we obtain $Var_q\left[\tilde{Z}_i\right] = \sigma_y^2 < \infty$ because

$$\begin{aligned}
Var_q\left[\tilde{Z}_i\right] &= \mathbb{E}_q\left[\tilde{Z}_i^2\right] - \left(\mathbb{E}_q\left[\tilde{Z}_i\right]\right)^2 \\
&= \mathbb{E}_q\left[\frac{1}{\tilde{N}_i^2}\left(\sum_{j=1}^{\tilde{N}_i}\mathbb{I}\left(Y_j^{(i)} > y\right)^2 + 2\sum_{k=1}^{\tilde{N}_i}\sum_{l>k}^{\tilde{N}_i}\mathbb{I}\left(Y_k^{(i)} > y\right)\mathbb{I}\left(Y_l^{(i)} > y\right)\right)L_i^2\right] - p_y^2 \\
&= \mathbb{E}_q\left[\mathbb{E}\left[\frac{1}{\tilde{N}_i^2}\left(\sum_{j=1}^{\tilde{N}_i}\mathbb{I}\left(Y_j^{(i)} > y\right) + 2\sum_{k=1}^{\tilde{N}_i}\sum_{l>k}^{\tilde{N}_i}\mathbb{I}\left(Y_k^{(i)} > y\right)\mathbb{I}\left(Y_l^{(i)} > y\right)\right)L_i^2 \mid \mathbf{X}_i\right]\right] - p_y^2 \\
&= \mathbb{E}_q\left[\frac{1}{\tilde{N}}s_y(\mathbf{X})L^2 + \frac{\tilde{N}-1}{\tilde{N}}s_y(\mathbf{X})^2L^2\right] - p_y^2 \\
&= \mathbb{E}_q\left[\frac{1}{\tilde{N}}s_y(\mathbf{X})\left(1 - s_y(\mathbf{X})\right)L^2\right] + \mathbb{E}_q\left[s_y(\mathbf{X})^2L^2\right] - p_y^2 \\
&= \sigma_y^2
\end{aligned} \tag{32}$$

and Assumption 2 ensures that the expectation terms in σ_y^2 are finite:

$$\mathbb{E}_q\left[\frac{1}{\tilde{N}}s_y(\mathbf{X})\left(1 - s_y(\mathbf{X})\right)L^2\right] \le \mathbb{E}_q\left[s_y(\mathbf{X})\left(1 - s_y(\mathbf{X})\right)L^2\right] \le \mathbb{E}_q\left[s_y(\mathbf{X})L^2\right] < \infty, \tag{33}$$

$$\mathbb{E}_q\left[s_y(\mathbf{X})^2L^2\right] \le \mathbb{E}_q\left[s_y(\mathbf{X})L^2\right] < \infty. \tag{34}$$

Thus, $Var_q\left[\tilde{Z}_i\right] = \sigma_y^2 < \infty$ follows, completing the proof of Eq. 23 by the Lindeberg—Lévy central limit theorem.

Finally, by applying Slutsky's theorem to Eq. 21 based on Eqs. 22 and 23, we complete the proof of Eq. 15.

References

Ankenman B, Nelson BL, Staum J (2010) Stochastic kriging for simulation metamodeling. Oper Res 58(2):371–382

Asmussen S, Binswanger K, Højgaard B (2000) Rare events simulation for heavy-tailed distributions. Bernoulli 6(2):303–322

Balesdent M, Morio J, Brevault L (2016) Rare event probability estimation in the presence of epistemic uncertainty on input probability distribution parameters. Methodol Comput Appl Probab 18(1):197–216

Blanchet J, Lam H (2012) State-dependent importance sampling for rare-event simulation: An overview and recent advances. Surv Oper Res Manag Sci 17(1):38–59

Blanchet J, Lam H (2014) Rare-event simulation for many-server queues. Math Oper Res 39(4):1142–1178

Blanchet J, Glynn P, Lam H (2009) Rare event simulation for a slotted time M/G/s model. Queueing Syst 63(1-4):33–57

Botev ZI, Kroese DP (2008) An efficient algorithm for rare-event probability estimation, combinatorial optimization, and counting. Methodol Comput Appl Probab 10(4):471–505

Byon E, Ntaimo L, Ding Y (2010) Optimal maintenance strategies for wind power systems under stochastic weather conditions. IEEE Trans Reliab 59(2):393–404

Byon E, Choe Y, Yampikulsakul N (2016) Adaptive learning in time-variant processes with application to wind power systems. IEEE Trans Autom Sci Eng 13(2):997–1007

Cannamela C, Garnier J, Iooss B (2008) Controlled stratification for quantile estimation. Ann Appl Stat 2(4):1554–1580

Chang CS, Heidelberger P, Juneja S, Shahabuddin P (1994) Effective bandwidth and fast simulation of ATM intree networks. Perform Eval 20(1):45–65

Choe Y, Byon E, Chen N (2015) Importance sampling for reliability evaluation with stochastic simulation models. Technometrics 57(3):351–361

Choe Y, Pan Q, Byon E (2016) Computationally efficient uncertainty minimization in wind turbine extreme load assessments. ASME J Solar Energy Eng 138(4):041,012–041,012–8

Coles SG (2001) An introduction to statistical modeling of extreme values. Springer, London

Gatto R, Baumgartner B (2016) Saddlepoint approximations to the probability of ruin in finite time for the compound Poisson risk process perturbed by diffusion. Methodol Comput Appl Probab 18(1):217–235

Geweke J (2005) Contemporary Bayesian econometrics and statistics. Wiley, Hoboken

Glasserman P, Li J (2005) Importance sampling for portfolio credit risk. Manag Sci 51(11):1643–1656

Glynn PW, Iglehart DL (1989) Importance sampling for stochastic simulations. Manag Sci 35(11):1367–1392

Heidelberger P (1995) Fast simulation of rare events in queueing and reliability models. ACM Trans Model Comput Simul (TOMACS) 5(1):43–85

Hoeffding W, Robbins H (1948) The central limit theorem for dependent random variables. Duke Math J 15(3):773–780

Hong LJ, Juneja S, Liu G (2017) Kernel smoothing for nested estimation with application to portfolio risk measurement. Oper Res 65(3):657–673

Ibragimov IA (1975) A note on the central limit theorems for dependent random variables. Theory Probab Appl 20(1):135–141

International Electrotechnical Commission (2005) IEC/TC88, 61400-1 ed. 3, Wind Turbines - Part 1: Design Requirements

Jonkman BJ (2009) TurbSim user's guide: version 1.50. Technical Report NREL/TP-500-46198, National Renewable Energy Laboratory. Golden, Colorado

Jonkman JM, Buhl ML Jr (2005) FAST User's Guide. Technical Report NREL/EL-500-38230, National Renewable Energy Laboratory. Golden, Colorado

Juneja S, Shahabuddin P (2006) Rare-event simulation techniques: an introduction and recent advances. Handbooks Oper Res Manag Sci 13:291–350

Kahn H, Marshall AW (1953) Methods of reducing sample size in Monte Carlo computations. J Oper Res Soc Amer 1(5):263–278

Kawai R (2008) Adaptive Monte Carlo variance reduction for Lévy processes with two-time-scale stochastic approximation. Methodol Comput Appl Probab 10(2):199–223

Koopman SJ, Shephard N, Creal D (2009) Testing the assumptions behind importance sampling. J Econ 149(1):2–11

Kroese DP, Taimre T, Botev ZI (2011) Handbook of Monte Carlo methods. Wiley, New York
Lee G, Byon E, Ntaimo L, Ding Y (2013) Bayesian spline method for assessing extreme loads on wind turbines. Ann Appl Stat 7(4):2034–2061
Manuel L, Nguyen HH, Barone MF (2013) On the use of a large database of simulated wind turbine loads to aid in assessing design standard provisions In: Proceedings of the 51st AIAA Aerospace Sciences Meeting including the New Horizons Forum and Aerospace Exposition. Grapevine, Texas
Moriarty P (2008) Database for validation of design load extrapolation techniques. Wind Energy 11(6):559–576
Plumlee M, Tuo R (2014) Building accurate emulators for stochastic simulations via quantile kriging. Technometrics 56(4):466–473
Rigby RA, Stasinopoulos DM (2005) Generalized additive models for location, scale and shape. J Royal Stat Soc: Ser C (Appl Stat) 54(3):507–554
Rosenblatt M (1956) A central limit theorem and a strong mixing condition. Proc Natl Acad Sci 42(1):43–47
Roussas GG (1994) Asymptotic normality of random fields of positively or negatively associated processes. J Multivar Anal 50(1):152–173
Rubinstein R (1999) The cross-entropy method for combinatorial and continuous optimization. Methodol Comput Appl Probab 1(2):127–190
Sadowsky JS (1991) Large deviations theory and efficient simulation of excessive backlogs in a GI/GI/m queue. IEEE Trans Autom Control 36(12):1383–1394
Schwarz G (1978) Estimating the dimension of a model. Ann Stat 6(2):461–464
Soleimanzadeh M, Wisniewski R, Kanev S (2012) An optimization framework for load and power distribution in wind farms. J Wind Eng Ind Aerodyn 107-108:256–262
Staid A, Guikema SD, Nateghi R, Quiring SM, Gao MZ (2014) Simulation of tropical cyclone impacts to the US power system under climate change scenarios. Clim Chang 127(3-4):535–546
Steckley SG, Henderson SG, Ruppert D, Yang R, Apley DW, Staum J (2016) Estimating the density of a conditional expectation. Electron J Stat 10(1):736–760
Sun Y, Apley DW, Staum J (2011) Efficient nested simulation for estimating the variance of a conditional expectation. Oper Res 59(4):998–1007
Withers CS (1981) Central limit theorems for dependent variables. I. Z Wahrscheinlichkeitstheorie verwandte Geb 57(4):509–534
Zhang N, Apley DW (2014) Fractional Brownian fields for response surface metamodeling. J Qual Technol 46(4):285
Zhang N, Apley DW (2016) Brownian integrated covariance functions for Gaussian process modeling: Sigmoidal versus localized basis functions. J Amer Stat Assoc 111(515):1182–1195