https://doi.org/10.21872/2024IISE_6545

Proceedings of the IISE Annual Conference & Expo 2024
A. Brown Greer, C. Contardo, J.-M. Frayret, eds.

Explainable Nonparametric Importance Sampling in Stochastic Simulation

Chenfei Li, Jaeshin Park, Eunshin Byon
Industrial and Operations Engineering, University of Michigan
Ann Arbor, MI, 48109

Abstract

We propose a new explainable multivariate importance sampling method in stochastic simulation. Importance sampling has the potential to significantly reduce the estimation variance when its instrumental density is properly designed. Prior studies have demonstrated the advantages of using a nonparametric method when the relationship between input variables and the output variable is unknown. However, importance sampling easily faces the curse of dimensionality, as the input dimension grows. This study devises a new method that identifies crucial input variables to effectively formulate a nonparametric instrumental density and explains the importance of each variable. The wind turbine reliability case study demonstrates the method's efficiency.

Keywords: Monte Carlo sampling, reliability, variance reduction, stochastic simulation

1. Introduction

The recent advances in computing power have greatly increased the prevalence of computer simulation modeling. These simulations are frequently utilized to assess system performance metrics, including reliability, across a wide range of application areas. This study is specifically motivated to evaluate wind turbine reliability using stochastic simulators. Wind power has emerged as one of the promising renewable energy sources. Ensuring the wind turbine's reliability is a fundamental consideration to prevent potential structural failure [5]. Accordingly, the National Renewable Energy Laboratory (NREL) under the U.S. Department of Energy (DOE) has developed aeroelastic simulation models, such as TurbSim [7] and FAST [8]. These tools aid in comprehensive analysis, contributing to the enhancement of wind turbine reliability.

Studies in stochastic simulations predominantly use two approaches to estimate system performance: the Monte Carlo sampling technique and the emulator-based approach [10]. The emulator-based technique constructs a surrogate model designed to approximate the response surface, and system performance estimations are derived from the resulting surrogate model. Although this approach proves useful in modeling or characterizing the average response on the response surface, it may not be suitable for calculating the probability of rare events [1]. Hence, in this study, we employ the Monte Carlo sampling approach to estimate the failure probability.

Among several Monte Carlo sampling techniques, importance sampling (IS) is a powerful tool that is well-known for its ability to reduce estimation variance. Variance reduction in stochastic simulation is translated to enhanced computational efficiency, as the variance gets reduced with larger data size. The fundamental idea of IS method is to employ an instrumental density for simulating data samples *smartly*, thereby facilitating precise estimations. Nevertheless, IS is prone to the curse of dimensionality as the input dimension grows. Recently, Li *et al.*(2021) [10] introduced a nonparametric IS that effectively navigates the issues of multivariate input, particularly when multiple input variables interact with each other. Their approach outlines the significance of each factor, treating them distinctively when formulating an instrumental density. However, the problem is that their approach includes all input variables in building the instrumental density. It is often observed that the performance of engineering systems is primarily influenced by a selected number of critical variables. This is underscored by the concept of the *parsimony principle*. When all non-critical variables are considered collectively, the efficiency of the IS method could be deteriorated.

In this paper, we introduce a new approach aimed at identifying crucial input variables and designing an instrumental density that is formulated with these selected variables. The resulting instrumental density provides a comprehensive overview of the impact of selected variables on the final results, thereby providing an *explainability* attribute to the methodology. The variables chosen and their parameters demonstrate their significance in estimating quantities of

https://doi.org/10.21872/2024IISE_6545

interest, such as a reliability measure. We apply our method to a case study in order to evaluate wind turbine reliability. Our findings demonstrate the superiority of this method, as it successfully identifies key input variables and provides improved variance reduction capability, compared to existing strategy that considers all variables [10].

2. Problem Description

We consider a black box computer model that generates the output $Y \in \mathbb{R}^1$ at input $\mathbf{X} \in \mathcal{D} \subseteq \mathbb{R}^d$ following a distribution $f(\mathbf{X})$. Computer models can be broadly categorized into deterministic or stochastic models, depending on the nature of the output Y being fixed or with noise at a given input $\mathbf{X}$. This study is concerned with stochastic computer model, due to its growing popularity [3]. In particular, we are interested in estimating the failure probability $P(Y > l)$, where l denotes a resistance level. This failure probability is also referred to as the probability of exceedance (POE).

Noting that $P(Y > l) = \int_{\chi_f} P(Y > l|\mathbf{X} = \mathbf{x}) f(\mathbf{x}) d\mathbf{x}$ with χ_f denoting the domain of $\mathbf{X}$, the crude Monte Carlo (CMC) method samples the input $\mathbf{X}_i$ from f and runs the computer model at $\mathbf{X}_i$ to get Y_i. Repeating the procedure, CMC estimates the POE by

$$\hat{P}_{CMC} = \frac{1}{N_T} \sum_{i=1}^{N_T} \mathbb{I}(Y_i > l). \tag{1}$$

Here, N_T denotes the total number of samples.

On the other hand, IS samples inputs from an instrumental density q instead of f and estimates the POE by

$$\hat{P}_{IS} = \frac{1}{N_T} \sum_{i=1}^{N_T} \mathbb{I}(Y_i > l) \frac{f(\mathbf{X}_i)}{q(\mathbf{X}_i)}. \tag{2}$$

With the stochastic computer model, $\hat{P}_{IS}$ is an unbiased estimator when the support of q, denoted as $supp\{q(\mathbf{x})\}$, includes $supp\{P(Y > l \mid \mathbf{X} = \mathbf{x}) f(\mathbf{x})\}$ [12]. In [3], it has been shown that the optimal IS density that minimizes the variance of the failure probability estimator $\hat{P}_{IS}$ is given by

$$q^*_{SIS}(\mathbf{x}) = \frac{1}{C_q} f(\mathbf{x}) \sqrt{s(\mathbf{x})}, \tag{3}$$

where $s(\mathbf{x}) = P(Y > l|\mathbf{X} = \mathbf{x})$ is the conditional failure probability at $\mathbf{x}$ and $C_q = \int_{\chi_f} f(\mathbf{x}) \sqrt{s(\mathbf{x})} d\mathbf{x}$ is the normalizing constant. The subscript *SIS* stands for the importance sampling for the stochastic computer model, or shortly, stochastic importance sampling [9]. With q^*_{SIS}, $\hat{P}_{IS}$ in (2) is an unbiased estimator [3].

While the density illustrated in (3) represents a theoretically optimal instrumental density, it cannot be straightforwardly implemented. This challenge arises because $s(\mathbf{x})$, a crucial component of the model, is unknown. This uncertainty is primarily due to the black-box nature of the computer model. Consequently, our study will primarily focus on identifying an effective estimation method for $s(\mathbf{x})$. Several approaches have been proposed for approximating $s(\mathbf{x})$ (or q^*_{SIS} more broadly), including parametric, cross-entropy, and nonparametric approaches [2, 4, 10]. Li *et al.*(2021) [10] discussed that parametric and cross-entropy approaches encounter difficulties when the input dimension is three or higher and proposed a new nonparametric technique as a remedy. However, the nonparametric method in [10] still demonstrates its own limitations in managing multivariate input as the input dimension increases. In the following section, we introduce a new method to discern key input variables vital to obtaining the instrumental density.

3. Methodology

To estimate $s(\mathbf{x})$ in (3), we utilize kernel regression [10], as follows.

$$\hat{s}(\mathbf{x}) = \frac{\sum_{i=1}^{n} K_d(\frac{x_1 - X_{i1}}{h_1}, \ldots, \frac{x_d - X_{id}}{h_d}) Z_i}{\sum_{i=1}^{n} K_d(\frac{x_1 - X_{i1}}{h_1}, \ldots, \frac{x_d - X_{id}}{h_d})}, \tag{4}$$

with $Z_i = \mathbb{I}(Y_i > l|\mathbf{X} = \mathbf{X}_i)$ and $\mathbf{x} = [x_1, x_2, \cdots, x_d]^T$, where K_d represents a d-dimensional kernel function. There are several potential kernel functions that could be used for K_d, each with its own advantages and disadvantages. The additive kernel, which aggregates all univariate kernels, offers a dimension-scalable methodology. However, its limitation lies in failing to recognize interaction effects among input variables. In contrast, the full multiplicative kernel, which involves the multiplication of all univariate kernels, accounts for all interaction effects [6, 13]. However, this approach necessitates a sufficiently large data set for kernel construction and may risk data sparsity issues when the input dimension is three or higher.

Seeking a balanced evaluation between the two extremes of fully additive and fully multiplicative kernels, Li et al. [10] put forth a weighted additive multivariate kernel for stochastic importance sampling method, referred to as WAMK-

https://doi.org/10.21872/2024IISE_6545

SIS. This method employs all of the $\binom{d}{2}$ bivariate kernels for estimating $s(\mathbf{x})$ in (3) with the d-dimensional input. The use of bivariate kernels enables the characterization of interaction effects present between any pair of variables. The fundamental premise of WAMK-SIS is the assumption of high-order interactions being negligible, with a belief that a two-way interaction is fully capable of depicting output behavior. Specifically, WAMK-SIS employs the following estimator for $s(\mathbf{x})$.

$$\hat{s}(\mathbf{x})_{full} = \sum_{1\le p<q\le d} w_{pq} B(x_p, x_q), \tag{5}$$

with $\sum_{1\le p<q\le d} w_{pq} = 1$, where the bivariate kernel $B(x_p, x_q)$ is defined by

$$B(x_p, x_q) = \frac{\sum_{i=1}^{n} K_2(\frac{x_p - X_{pi}}{h_p}, \frac{x_q - X_{qi}}{h_q}) Z_i}{\sum_{i=1}^{n} K_2(\frac{x_p - X_{pi}}{h_p}, \frac{x_q - X_{qi}}{h_q})}. \tag{6}$$

Here X_{pi} and X_{qi} denote the ith random samples of the two input variables x_p and x_q, respectively, and K_2 implies a 2-dimensional multiplicative kernel and h_p and h_q denote the bandwidths of two variables x_p and x_q, respectively. In (5), w_{pq} denotes the weight of $B(x_p, x_q)$, quantifying the importance of $B(x_p, x_q)$, defined by the following cross-entropy error function

$$w_{pg} = \frac{1/e(p,q)}{\sum_{1\le p'<q'\le d} 1/e(p',q')}, \tag{7}$$

with

$$e(p,q) = -\sum_{i=1}^{n} \{Z_i \ln(B(X_{pi}, X_{qi})) + (1 - Z_i) \ln(1 - B(X_{pi}, X_{qi}))\}. \tag{8}$$

Note that $e(p,q)$ diminishes as $B(x_p, x_q)$ approaches Z_i more closely. In this sense, a smaller $e(p,q)$ indicates good estimation of the bivariate kernel $B(x_p, x_q)$, and hence, should be correspondingly assigned a greater weight, as shown in (7).

While the WAMK-SIS demonstrates superiority over CMC and other parametric methodologies [10], it still has limitations. Firstly, as the dimension d grows, the number of bivariate kernels increases significantly, specifically in the order of $\binom{d}{2}$. This increase results in unwarranted complexity in the form of both the estimated conditional failure probability and the instrumental distribution q. Secondly, the parsimony principle often holds in engineering systems, suggesting that a system's functioning is typically governed by a few crucial factors. According to this principle, a simpler model could be sought. Thirdly, the comprehensive implementation of bivariate kernels could lead to redundant kernels. For instance, consider two bivariate kernels, $B(x_1, x_2)$ and $B(x_1, x_3)$. Suppose x_2 and x_3 are of minimal significance and their interactions with x_1 are negligible; under these circumstances, the two kernels might closely resemble each other, that is, $B(x_1, x_2) \approx B(x_1, x_3)$. This indicates that $\hat{s}(\mathbf{x})_{full}$ in (5) could contain redundant terms.

To improve the WAMK-SIS, our approach is to identify important bivariate kernels. The idea is similar to feature selection (or subset selection) in machine learning. Note that, with $\binom{d}{2}$ bivariate kernels, there are $2^{\binom{d}{2}} - 1$ possible estimators for approximating $s(\mathbf{x})$ in (3). To choose the best one among them, our primary analytical tool is a modified version of cross-validation (CV) technique which is customized for the IS context. Suppose we have a pilot dataset or dataset obtained in previous iterations in the iterative procedure. We partition this dataset into two exclusive sets, training and validation sets. We systematically navigate through every subset of $\binom{d}{2}$ bivariate kernels. For each subset, we construct a POE estimator anchored on the training set, and then gauge its performance using the validation set. This CV process is performed multiple times, each utilizing randomly partitioned training and validation sets. Finally, we identify the estimator with the minimum variance of the POE estimations.

Note that applying the CV procedure for a total of $2^{\binom{d}{2}} - 1$ possible candidates could be computationally burdensome. Thus, before applying the CV procedure, we first pre-screen the candidates so we can apply the CV to those selected candidates only. Specifically, let Ω_k denote the kth nonempty subset that contains the indices of bivariate kernels for $k = 1, \cdots, 2^{\binom{d}{2}} - 1$. For instance, for the k^{th} conditional POE estimator $\hat{s}_k(\mathbf{x})$ with $B(x_1, x_2)$ and $B(x_1, x_3)$, we have $\Omega_k = \{(1,2),(1,3)\}$. We define the estimator $\hat{s}_k(\mathbf{x})$ by

$$\hat{s}_k(\mathbf{x}) = \sum_{(p,q)\in\Omega_k} w_{pq} B(x_p, x_q), \tag{9}$$

where $B(x_p, x_q)$ can be obtained using (6), with w_{pg} defined in (7)-(8). To obtain $\hat{s}_k(\mathbf{x})$, we use the total data available to us, either from pilot sampling or previous iterations.

After obtaining all possible estimators, we choose the top candidates using the cross-entropy error function. The cross-entropy error function in (8) is used to quantify the importance of each bivariate kernel. Now, we replace the bivariate

https://doi.org/10.21872/2024IISE_6545

kernel function value with each candidate estimator $\hat{s}_k(\mathbf{x})$ to get its cross-entropy function as follows.

$$CE_k = -\sum_{i=1}^{n} log(\hat{s}_k(\mathbf{X}_i))Z_i + (1 - log(\hat{s}_k(\mathbf{X}_i)))(1 - Z_i), \tag{10}$$

where n is the number of data used to build $\hat{s}_k(\mathbf{x})$. Note that an estimator with small CE_k implies that $\hat{s}_k(\mathbf{X}_i)$ is close to 1 when there is a failure (or exceedance event) at $\mathbf{X} = \mathbf{X}_i$ (i.e., $Z_i = \mathbb{I}(Y_i > l) = 1$), whereas it is close to 0 when there is no failure. Thus, we choose estimators with *small* CE_k values. In our case study, we choose ten such estimators.

Let $\hat{s}_{(k)}(\mathbf{x})$, $k = 1, \cdots, 10$, denote the kth pre-screened candidate. Then we employ the CV procedure to choose the best one among them. As a remark, one may suggest using (10), instead of employing the CV procedure, to choose the best estimator. However, doing so could lead to an overfitting issue. Moreover, while the cross-entropy error function quantifies the prediction accuracy of each estimator for the conditional POE $s(\mathbf{x})$, our ultimate goal is to minimize the variance of the POE estimator. For the CV implementation, we divide the entire data set into the training and validation sets. Next, we construct $\hat{s}_{(k)}(\mathbf{x})$ using the training set for each pre-screened candidate and evaluate its performance using the validation set. Let $q_{(k)}$ denote the instrumental density with the trained $\hat{s}_{(k)}(\mathbf{x})$, i.e.,

$$q_{(k)}(\mathbf{x}) = \frac{1}{C_{q_{(k)}}} f(\mathbf{x}) \sqrt{\hat{s}_{(k)}(\mathbf{x})}. \tag{11}$$

With $q_{(k)}$, the POE estimated in the validation set becomes

$$\hat{P}_{(k)} = \frac{1}{n_{vld}} \sum_{i=1}^{n_{vld}} \mathbb{I}(Y_i > l) \frac{f(\mathbf{X}_i)}{q_{(k)}(\mathbf{X}_i)}, \tag{12}$$

where n_{vld} is the number of data sampled from $q_{(k)}$ to test its performance in the validation set. Note that, while we use the notation $q_{(k)}(\mathbf{X}_i)$ for the trained instrumental density, it is formulated with the selected bivaraiate kernels only.

Let $\Omega_{(k)}$ denote the kth set of bivariate kernels associated with $\hat{s}_{(k)}(\mathbf{x})$. Ideally, we should choose the best subset $\Omega_{(k)}$ that leads to the smallest variance of $\hat{P}_{(k)}$ in (12). The challenge is that $\hat{P}_{(k)}$ is not readily available. Note that, in order to evaluate the performance of $q_{(k)}$, $\mathbf{X}_i$ should be sampled from $q_{(k)}$ and Y_i should be obtained by running the computer model at the sampled $\mathbf{X}_i$. However, running the computer model n_{vld} times for all candidates $q_{(k)}$ is computationally demanding. On the other hand, without running the computer model, Y_i in (12) is not obtained.

We address the challenge by approximating $\mathbb{I}(Y_i > l)$ in (12) with the estimated conditional POE. Here, if we use $\hat{s}_{(k)}(\mathbf{x})$ in (9), which is learned with the training set, to approximate $\mathbb{I}(Y_i > l)$, it will incur the overfitting issue. Moreover, it violates the underlying principal of CV: train a model in the training set and test the resulting model in a separate validation set. Our remedy is to re-learn the conditional POE using the validation set data. Then the new conditional POE will play a role of proxy that replaces $\mathbb{I}(Y_i > l)$ in (12).

Let $\hat{s}^{vld}_{(k)}(\mathbf{x})$ denote the estimated conditional POE with the bivariate kernels, each with the input pair in $\Omega_{(k)}$, using the validation set. We then approximate $\hat{P}_{(k)}$ in (12) as follows.

$$\tilde{P}_{(k)} = \frac{1}{n_{vld}} \sum_{i=1}^{n_{vld}} \hat{s}^{vld}_{(k)}(\mathbf{x}) \frac{f(\mathbf{X}_i)}{q_{(k)}(\mathbf{X}_i)}, \tag{13}$$

Note that $q_{(k)}(\mathbf{X}_i)$ is the instrumental density constructed with the training set. We evaluate it using the validation set via $\hat{s}^{vld}_{(k)}(\mathbf{x})$. It has been known that the normalizing constant $C_{q_{(k)}}$ in (11) could be unstable for multivariate instrumental density when the input dimension is three or higher [10]. Thus, we instead use the self-normalizing constant that replaces n_{vld} with $\sum_{i=1}^{n_{vld}} f(\mathbf{X}_i)/q_{(k)}(\mathbf{X}_i)$ in our implementation.

We repeat this CV procedure multiple times with randomly partitioned training and validation sets and compute the sample variance of the POE estimator $\tilde{P}_{(k)}$ in (13). Finally, we select the best subset $\Omega_{(k)}$ that produces the smallest sample variance. Our approach identifies key input variables and produces more explainable instrumental density. Accordingly, we refer our method as explainable IS, shortly X-SIS.

4. Wind Turbine Case Study

We use the NREL's aeroelastic wind turbine simulators, including TurbSim (version 1.50) [7] and FAST (version 7.01.00a-bjj) [8], to estimate the load exceedance probability $P(Y > l)$. Specifically, we consider the 10-minute maximum tip deflection as the response variable Y, which is one of the essential load responses related to the wind turbine stability [11]. We set the resistance level at $l = 2.06$. We consider the following five input variables as in [10], including the wind speed V, turbulence intensity TI, wind shear S, wind vertical angle VA, and surface roughness length, SR. Therefore, we have $\mathbf{X} = (V, TI, S, VA, SR)$. For their original input distribution f, we use the same setting

https://doi.org/10.21872/2024IISE_6545

as outlined in [10]. We conduct 25 experiments. Each experiment draws 600 pilot samples from uniform distribution and includes 10 iterations, that each iteration generates 200 samples from the NREL simulators.

Algorithm 1 Subset selection procedure in X-SIS

Input: Pilot samples or data from previous iterations
Obtain $\hat{s}_k(\mathbf{x})$ for each k, $k = 1, \cdots, 2^{\binom{d}{2}} - 1$.
Calculate the cross-entropy error function value in (10) for each $\hat{s}_k(\mathbf{x})$
Choose the top ten estimators $\hat{s}_{(k)}(\mathbf{x})$ for $k = 1, \cdots, 10$ that have the ten smallest cross-entropy errors.
for $iter = 1, 2, \cdots, 20$ **do**
    Randomly divide the entire dataset into training and validation sets
    for $k = 1, \cdots, 10$ **do**
        Using the training set, obtain $\hat{s}_{(k)}(\mathbf{x})$ using (9) to construct the instrumental density $q_{(k)}$ in (11).
        Sample $\mathbf{X}_i$ for $i = 1, 2, \cdots, n_{vld}$ from $q_{(k)}$.
        With data in the validation set, obtain $\hat{s}^{vld}_{(k)}(x)$ using (9) and calculate $\tilde{P}_{(k)}$ using (13).
    end for
end for
Find the standard error of $\tilde{P}_{(k)}$ for $k = 1, 2, \cdots, 10$ and choose the best $q^*_{(k)}$ with the smallest standard error of $\tilde{P}_{(k)}$.
Choose the best $q^*_{(k)}$ that leads to the smallest standard error of $\tilde{P}_{(k)}$.
Output: $\Omega_{(k)}$ associated with $q^*_{(k)}$.

We compare our approach X-SIS with the WAMK-SIS. In addition to assessing the standard error (SE) of POE estimates from 25 experiments, we further evaluate the estimator performance using relative ratio (RR), which is defined as $RR = N_T / N_{CMC}$. Here, N_{CMC} denotes the number of CMC replications necessary to attain a similar SE, given by $N_{CMC} = P_T(1 - P_T)/SE^2$ with $P_T = P(Y > l)$ [3]. Here P_T is the true failure probability, which is unknown. Thus, we use the estimated POE in computing RR.

Table 1 summarizes the results. The means of 25 POE estimates in both methods are similar. However, the results suggest that the X-SIS estimator achieves smaller SE than the WAMK-SIS. It uses only 36% computational resources required by CMC, implying that the CMC needs about three times larger repetitions to the same SE level of 0.0029 from the X-SIS. While WAMK-SIS reduces the variance over CMC, it underperforms the X-SIS.

Table 1: **Comparison results from 25 experiments**

Estimator	Mean	SE	RR
X-SIS	0.0495	0.0029	35.75%
WAMK-SIS	0.0495	0.0035	52.07%

Table 2 reports the total number and percentage of bivariate kernels chosen in X-SIS over 10 iterations in 25 experiments. In all cases, the X-SIS chooses two or three bivariate kernels, which is very few compared to the WAMK-SIS estimator that uses 10 bivariate kernels. Table 3 further summarizes the selected bivariate kernels. In most cases, the bivariate kernel $B(V,TI)$ is employed in constructing the instrumental density. This implies that these two variables are the most important among five input variables and their interaction is significant. Out of the candidates, the three kernels $B(V,TI)$, $B(V,S)$, and $B(TI,VA)$ are employed to form the instrumental density most frequently.

Table 2: **Number of bivariate kernels employed in X-SIS estimator**

Number of bivariate kernels	2	3
Frequency	73	177
Percent	29.2%	70.8%

Notably, the estimators that employ the three bivariate kernels primarily encompass $B(V,TI)$ and $B(V,S)$ in most cases, suggesting that the three variables, V, TI, and S, are considered vital input variables impacting wind turbine reliability. These estimators, however, often include an additional bivariate kernel, as shown in Table 3. We believe that this inclusion is an attempt to prevent overfitting through the CV. For instance, consider the estimators in the first two rows in Table 3. While $B(V,TI)$ and $B(V,S)$ can characterize the exceedance pattern well, their sole inclusion might result in the instrumental density overly concentrating on a specific region, which could potentially induce estimation bias and escalated variance. The CV process aids in circumventing such unfavor-

Table 3: **Frequency of selected bivariate kernels**

Estimator	Frequency	Percent
$B(V,TI), B(V,S), B(TI,VA)$	58	23.2%
$B(V,TI), B(V,S), B(TI,S)$	37	14.8%
$B(V,TI), B(S,VA)$	33	13.2%
$B(V,TI), B(V,S), B(TI,SR)$	29	11.6%
$B(V,TI), B(VA,SR)$	19	7.6%
$B(V,TI), B(V,S), B(S,VA)$	19	7.6%

https://doi.org/10.21872/2024IISE_6545

able scenarios. A similar phenomenon can be observed with estimators that incorporate $B(V,TI)$, the most significant bivariate kernel (refer to the third and fifth row estimators). In addition to $B(V,TI)$, these estimators add an extra kernel such as $B(S,VA)$ and $B(VA,SR)$. It is noteworthy that these supplementary kernels can be considered as less critical, even less so than the third kernel in the estimators consisting of three bivariate kernels, e.g., $B(TI,VA)$ in the first row estimator. By coupling these kernels with the most important kernel $B(V,TI)$, the resulting instrumental density can effectively circumvent overfitting issues.

Additionally, the proposed method effectively steers clear of repetitive terms in estimating $s(\mathbf{x})$. It is worth noting that the selected estimator does not incorporate multiple kernels that comprise the same significant variable intertwined with other less important variables. For instance, consider the first and fourth row estimators that include $B(TI,VA)$ and $B(TI,SR)$, respectively. $B(TI,VA)$ and $B(TI,SR)$ mirror a univariate kernel with only TI, because VA and SR are insignificant and thus, their bandwidths are much wider than that of TI. Accordingly, no estimators include both $B(TI,VA)$ and $B(TI,SR)$ simultaneously. We observe a similar trend across other estimators as well.

In summary, the X-SIS estimator ensures the chosen kernels encompass key input variables. This estimator frequently employs $B(V,TI)$ and $B(V,S)$, underscoring the significance of the three variables along with their interactions. In addition to choosing vital kernels, it effectively circumvents overfitting issues, thus averting the potential problem of the instrumental density focusing excessively on overly narrow regions.

5. Summary

This study devises a new nonparametric importance sampling method designed for assessing reliability with multivariate inputs. Our primary focus is on explaining the significance of each input, facilitating the construction of an instrumental density with enhanced effectiveness. To address the challenge of overfitting, we propose a modified CV procedure within the framework of IS. Notably, our new CV procedure avoids additional computational burden for generating new data samples while effectively mitigating the risk of overfitting. It achieves this by guiding the selection of suitable bivariate kernels, ultimately reducing the variance in the resulting estimation. In the future, we plan to extend the approach to estimate extreme loads in wind turbine application [4].

References

[1] Claire Cannamela, Josselin Garnier, and Bertrand Iooss. Controlled stratification for quantile estimation. *Annals of Applied Statistics*, 2(4):1554–1580, 2008.

[2] Quoc Dung Cao and Youngjun Choe. Cross-entropy based importance sampling for stochastic simulation models. *Reliability Engineering & System Safety*, 191:106526, 2019.

[3] Youngjun Choe, Eunshin Byon, and Nan Chen. Importance sampling for reliability evaluation with stochastic simulation models. *Technometrics*, 57(3):351–361, 2015.

[4] Youngjun Choe, Qiyun. Pan, and Eunshin Byon. Computationally efficient uncertainty minimization in wind turbine extreme load assessments. *Journal of Solar Energy Engineering*, 138(4):041012, 2016.

[5] Yu Ding. *Data science for wind energy*. CRC Press, 2019.

[6] Jianqing Fan. *Local polynomial modelling and its applications: monographs on statistics and applied probability 66*. Routledge, 2018.

[7] Bonnie J Jonkman. Turbsim user's guide v2. 00.00. Technical report, National Renewable Energy Lab.(NREL), Golden, Colorado, 2009.

[8] Jason Mark Jonkman and Marshall L Buhl Jr. Fast user's guide. Technical report, National Renewable Energy Lab.(NREL), Golden, Colorado, 2005.

[9] Young Myoung Ko and Eunshin Byon. Optimal budget allocation for stochastic simulation with importance sampling: Exploration vs. replication. *IISE Transactions*, 54(9):881–893, 2022.

[10] Shuoran Li, Young Myoung Ko, and Eunshin Byon. Nonparametric importance sampling for wind turbine reliability analysis with stochastic computer models. *The Annals of Applied Statistics*, 15(4):1850–1871, 2021.

[11] Leon Mishnaevsky Jr. Root causes and mechanisms of failure of wind turbine blades: Overview. *Materials*, 15 (9):2959, 2022.

[12] Qiyun Pan, Eunshin Byon, Young Myoung Ko, and Henry Lam. Adaptive importance sampling for extreme quantile estimation with stochastic black box computer models. *Naval Research Logistics (NRL)*, 67(7):524–547, 2020.

[13] Larry Wasserman. *All of nonparametric statistics*. Springer Science & Business Media, 2006.