TCH-13-190: Supplements

In this supplementary document, Section1 includes the derivations for the optimal solutions of SIS1 and SIS2 with a multivariate input vector. Sections 2 and 3 present the numerical examples with a univariate input variable and a multivariate input vector, respectively, which are used to investigate the impacts of various factors on the performance of the proposed methods. Section 4 discusses the implementation details with the wind turbine simulators.

1 Derivations for Optimal Stochastic Important Sampling

This section details the derivations of the optimal allocation size, $N_i, i = 1, 2, \cdots, M$, and the optimal IS density, q_{SIS1}, for SIS1 and the optimal IS density, q_{SIS2}, for SIS2, presented in Section 3 of the paper. In the sequel, we consider a multivariate input vector, $\mathbf{X} \in \mathbb{R}^p$. Note that a univariate input vector is a special case with $p = 1$.

1.1 Optimal Important Sampling density and allocations in SIS1

First, we consider the SIS1 estimator,

$$\begin{aligned}\hat{P}_{SIS1} &= \frac{1}{M}\sum_{i=1}^{M}\hat{P}\left(Y > l \mid \mathbf{X}_i\right)\frac{f(\mathbf{X}_i)}{q(\mathbf{X}_i)} \\ &= \frac{1}{M}\sum_{i=1}^{M}\left(\frac{1}{N_i}\sum_{j=1}^{N_i}\mathbb{I}\left(Y_j^{(i)} > l\right)\right)\frac{f(\mathbf{X}_i)}{q(\mathbf{X}_i)}. \end{aligned} \tag{S.1}$$

We decompose the variance of this estimator into two components, the expectation of conditional variance and the variance of conditional expectation, as

$$\begin{aligned}Var\left[\hat{P}_{SIS1}\right] &= Var\left[\frac{1}{M}\sum_{i=1}^{M}\hat{P}\left(Y > l \mid \mathbf{X}_i\right)\frac{f(\mathbf{X}_i)}{q(\mathbf{X}_i)}\right] \\ &= \frac{1}{M^2}E_q\left[Var\left[\sum_{i=1}^{M}\hat{P}\left(Y > l \mid \mathbf{X}_i\right)\frac{f(\mathbf{X}_i)}{q(\mathbf{X}_i)}\,\middle|\,\mathbf{X}_1, \cdots, \mathbf{X}_M\right]\right] \\ &\quad + \frac{1}{M^2}Var_q\left[E\left[\sum_{i=1}^{M}\hat{P}\left(Y > l \mid \mathbf{X}_i\right)\frac{f(\mathbf{X}_i)}{q(\mathbf{X}_i)}\,\middle|\,\mathbf{X}_1, \cdots, \mathbf{X}_M\right]\right]. \end{aligned} \tag{S.2}$$

For simplicity, let $s(\mathbf{X})$ denote the conditional POE, $P(Y > l \mid \mathbf{X})$. Using the fact that $\mathbf{X}_i \overset{i.i.d}{\sim} q$ for $i = 1, 2, \cdots, M$, we simplify $Var\left[\hat{P}_{SIS1}\right]$ in (S.2) to

$$\begin{aligned}
Var\left[\hat{P}_{SIS1}\right] &= \frac{1}{M^2} E_q\left[Var\left[\sum_{i=1}^{M}\left(\frac{1}{N_i}\sum_{j=1}^{N_i}\mathbb{I}\left(Y_j^{(i)} > l\right)\right)\frac{f(\mathbf{X}_i)}{q(\mathbf{X}_i)} \,\middle|\, \mathbf{X}_1, \cdots, \mathbf{X}_M\right]\right] \\
&\quad + \frac{1}{M^2} Var_q\left[\sum_{i=1}^{M} s(\mathbf{X}_i)\frac{f(\mathbf{X}_i)}{q(\mathbf{X}_i)}\right] \\
&= \frac{1}{M^2} E_q\left[\sum_{i=1}^{M}\left(\frac{1}{N_i^2}\sum_{j=1}^{N_i} s(\mathbf{X}_i)\left(1 - s(\mathbf{X}_i)\right)\right)\frac{f(\mathbf{X}_i)^2}{q(\mathbf{X}_i)^2}\right] + \frac{1}{M} Var_q\left[s(\mathbf{X})\frac{f(\mathbf{X})}{q(\mathbf{X})}\right] \\
&= \frac{1}{M^2} E_q\left[\sum_{i=1}^{M}\frac{1}{N_i} s(\mathbf{X}_i)\left(1 - s(\mathbf{X}_i)\right)\frac{f(\mathbf{X}_i)^2}{q(\mathbf{X}_i)^2}\right] + \frac{1}{M} Var_q\left[s(\mathbf{X})\frac{f(\mathbf{X})}{q(\mathbf{X})}\right]. \qquad \text{(S.3)}
\end{aligned}$$

We express the allocation size, N_i, at $\mathbf{X}_i$ as a proportion of the total simulation budget, N_T,

$$N_i = N_T \cdot \frac{c(\mathbf{X}_i)}{\sum_{j=1}^{M} c(\mathbf{X}_j)}, \qquad i = 1, 2, \cdots, M, \qquad \text{(S.4)}$$

where $c(\mathbf{X})$ is a non-negative function. Lemma 1 presents the optimal assignment of simulation replications, N_i, to each $\mathbf{X}_i$ for given q.

Lemma 1 *Given q, the variance in (S.3) is minimized if and only if*

$$N_i = \frac{\sqrt{s(\mathbf{X}_i)\left(1 - s(\mathbf{X}_i)\right)} f(\mathbf{X}_i)/q(\mathbf{X}_i)}{\sum_{j=1}^{M}\sqrt{s(\mathbf{X}_j)\left(1 - s(\mathbf{X}_j)\right)} f(\mathbf{X}_j)/q(\mathbf{X}_j)} \cdot N_T \quad \textit{for } i = 1, 2, \cdots, M. \qquad \text{(S.5)}$$

Proof. *We want to find N_i, $i = 1, 2, \cdots, M$, that minimizes the variance in* (S.3) *for any given function, $q(\mathbf{X})$. Note that the second term in* (S.3) *is constant, provided that the function $q(\mathbf{X})$ is given, and the other functions, $f(\mathbf{X})$ and $s(\mathbf{X})$, are fixed. Thus, we find N_i*

that minimizes the first term in (S.3),

$$
\begin{aligned}
&\frac{1}{M^2} E_q \left[\sum_{i=1}^{M} \frac{1}{N_i} s(\mathbf{X}_i)\,(1 - s(\mathbf{X}_i)) \frac{f(\mathbf{X}_i)^2}{q(\mathbf{X}_i)^2} \right] \\
&= \frac{1}{M^2} \sum_{i=1}^{M} E_q \left[\frac{1}{N_i} s(\mathbf{X}_i)\,(1 - s(\mathbf{X}_i)) \frac{f(\mathbf{X}_i)^2}{q(\mathbf{X}_i)^2} \right] \\
&= \frac{1}{M} E_q \left[\frac{1}{N_1} s(\mathbf{X}_1)\,(1 - s(\mathbf{X}_1)) \frac{f(\mathbf{X}_1)^2}{q(\mathbf{X}_1)^2} \right] && \text{(S.6)} \\
&= \frac{1}{M} \frac{1}{N_T} E_q \left[\frac{\sum_{j=1}^{M} c(\mathbf{X}_j)}{c(\mathbf{X}_1)} s(\mathbf{X}_1)\,(1 - s(\mathbf{X}_1)) \frac{f(\mathbf{X}_1)^2}{q(\mathbf{X}_1)^2} \right] && \text{(S.7)} \\
&= \frac{1}{M} \frac{1}{N_T} \left(\sum_{j=1}^{M} E_q \left[\frac{c(\mathbf{X}_j)}{c(\mathbf{X}_1)} s(\mathbf{X}_1)\,(1 - s(\mathbf{X}_1)) \frac{f(\mathbf{X}_1)^2}{q(\mathbf{X}_1)^2} \right] \right) \\
&= \frac{1}{M} \frac{1}{N_T} \left(E_q \left[s(\mathbf{X}_1)\,(1 - s(\mathbf{X}_1)) \frac{f(\mathbf{X}_1)^2}{q(\mathbf{X}_1)^2} \right] + \sum_{j=2}^{M} E_q \left[\frac{c(\mathbf{X}_j)}{c(\mathbf{X}_1)} s(\mathbf{X}_1)\,(1 - s(\mathbf{X}_1)) \frac{f(\mathbf{X}_1)^2}{q(\mathbf{X}_1)^2} \right] \right) \\
&= \frac{1}{M} \frac{1}{N_T} \left(E_q \left[s(\mathbf{X})\,(1 - s(\mathbf{X})) \frac{f(\mathbf{X})^2}{q(\mathbf{X})^2} \right] + (M-1) \cdot E_q \left[c(\mathbf{X}) \right] \cdot E_q \left[\frac{1}{c(\mathbf{X})} s(\mathbf{X})\,(1 - s(\mathbf{X})) \frac{f(\mathbf{X})^2}{q(\mathbf{X})^2} \right] \right) && \text{(S.8)} \\
&\geq \frac{1}{M} \frac{1}{N_T} \left(E_q \left[s(\mathbf{X})\,(1 - s(\mathbf{X})) \frac{f(\mathbf{X})^2}{q(\mathbf{X})^2} \right] + (M-1) \cdot \left(E_q \left[\sqrt{s(\mathbf{X})\,(1 - s(\mathbf{X}))} \frac{f(\mathbf{X})}{q(\mathbf{X})} \right] \right)^2 \right) && \text{(S.9)}
\end{aligned}
$$

The equalities in (S.6) and (S.8) are due to the fact that $\mathbf{X}_i$, $i = 1, 2, \cdots, M$, *is independent and identically distributed. We use the definition in (S.4) for (S.7). The inequality in* (S.9) *follows by applying the Cauchy-Schwarz inequality to the second term in (S.8). The equality in* (S.9) *holds if and only if* $c(\mathbf{X}) = k\sqrt{s(\mathbf{X})\,(1 - s(\mathbf{X}))} f(\mathbf{X}) / q(\mathbf{X})$, *where* k *is a positive constant. Therefore, by the definition in* (S.4), *the optimal allocation size in (S.5) follows.* □

Plugging N_i's in (S.5) into the estimator variance in (S.3) leads to

$$
\begin{aligned}
Var\left[\hat{P}_{SIS1}\right] &= \frac{1}{M}\frac{1}{N_T}\left(E_q\left[s(\mathbf{X})\left(1-s(\mathbf{X})\right)\frac{f(\mathbf{X})^2}{q(\mathbf{X})^2}\right] + (M-1)\left(E_f\left[\sqrt{s(\mathbf{X})\left(1-s(\mathbf{X})\right)}\right]\right)^2\right) \\
&\quad + \frac{1}{M}Var_q\left[s(\mathbf{X})\frac{f(\mathbf{X})}{q(\mathbf{X})}\right] \qquad \text{(S.10)} \\
&= \frac{1}{M}\frac{1}{N_T}\left(E_f\left[s(\mathbf{X})\left(1-s(\mathbf{X})\right)\frac{f(\mathbf{X})}{q(\mathbf{X})}\right] + (M-1)\left(E_f\left[\sqrt{s(\mathbf{X})\left(1-s(\mathbf{X})\right)}\right]\right)^2\right) \\
&\quad + \frac{1}{M}\left(E_q\left[s(\mathbf{X})^2\frac{f(\mathbf{X})^2}{q(\mathbf{X})^2}\right] - \left(E_q\left[s(\mathbf{X})\frac{f(\mathbf{X})}{q(\mathbf{X})}\right]\right)^2\right) \\
&= \frac{1}{M}\frac{1}{N_T}\left(E_f\left[s(\mathbf{X})\left(1-s(\mathbf{X})\right)\frac{f(\mathbf{X})}{q(\mathbf{X})}\right] + (M-1)\left(E_f\left[\sqrt{s(\mathbf{X})\left(1-s(\mathbf{X})\right)}\right]\right)^2\right) \\
&\quad + \frac{1}{M}\left(E_f\left[s(\mathbf{X})^2\frac{f(\mathbf{X})}{q(\mathbf{X})}\right] - P\left(Y>l\right)^2\right), \qquad \text{(S.11)}
\end{aligned}
$$

where we obtain the equation in (S.10) using the expression in (S.9). Please note that $E_q\left[\sqrt{s(\mathbf{X})\left(1-s(\mathbf{X})\right)}\frac{f(\mathbf{X})}{q(\mathbf{X})}\right] = E_f\left[\sqrt{s(\mathbf{X})\left(1-s(\mathbf{X})\right)}\right]$.

Recall that $s(\mathbf{X})$ denotes $P\left(Y>l \mid \mathbf{X}\right)$. Thus, only the following terms in (S.11) contain q,

$$
\begin{aligned}
&\frac{1}{M}\frac{1}{N_T}E_f\left[s(\mathbf{X})\left(1-s(\mathbf{X})\right)\frac{f(\mathbf{X})}{q(\mathbf{X})}\right] + \frac{1}{M}E_f\left[s(\mathbf{X})^2\frac{f(\mathbf{X})}{q(\mathbf{X})}\right] \\
&= \frac{1}{M}\int_{\mathcal{X}_f}\left(\frac{1}{N_T}s(\mathbf{x})\cdot\left(1-s(\mathbf{x})\right) + s(\mathbf{x})^2\right)\frac{f^2(\mathbf{x})}{q(\mathbf{x})}\,\mathrm{d}\mathbf{x}, \qquad \text{(S.12)}
\end{aligned}
$$

where $\mathcal{X}_f = \{\mathbf{x}\in\mathbb{R}^p : f\left(\mathbf{x}\right) > 0\}$ is the support of f. Finding q that minimizes (S.12) is a functional minimization problem. To specify the boundary conditions, we define the joint cumulative distribution function (CDF) of $\mathbf{X}\in\mathbb{R}^p$ with the IS density, q, as

$$
Q\left(x_1, x_2, \cdots, x_p\right) \equiv \int_{-\infty}^{x_1}\int_{-\infty}^{x_2}\cdots\int_{-\infty}^{x_p} q\left(\tilde{x}_1, \tilde{x}_2, \cdots, \tilde{x}_p\right)\,\mathrm{d}\tilde{x}_1\mathrm{d}\tilde{x}_2\cdots,\mathrm{d}\tilde{x}_p.
$$

Then, we impose the boundary conditions,

$$
\begin{aligned}
Q\left(-\infty, -\infty, \cdots, -\infty\right) &= 0, \\
Q\left(\infty, \infty, \cdots, \infty\right) &= 1.
\end{aligned}
$$

Therefore, we minimize the functional in (S.12) over the set of functions,

$$
\{q : Q\left(-\infty, -\infty, \cdots, -\infty\right) = 0;\ Q\left(\infty, \infty, \cdots, \infty\right) = 1;\ q\left(\mathbf{x}\right) \geq 0, \forall\mathbf{x}\in\mathbb{R}^p\}.
$$

In the following, we use principles of the calculus of variations. The integrand in (S.12) is the Lagrangian function, $\mathcal{L}(x_1, x_2, \cdots, x_p, q)$. The optimal q should satisfy the Euler-Lagrange

equation [3],

$$
\begin{aligned}
0 &= (-1)^p \frac{\partial^p}{\partial x_1 \partial x_2 \cdots \partial x_p} \left(\frac{\partial \mathcal{L}}{\partial q}(x_1, x_2, \cdots, x_p, q) \right) \\
&= (-1)^p \frac{\partial^p}{\partial x_1 \partial x_2 \cdots \partial x_p} \left(-\frac{\mathcal{L}(x_1, x_2, \cdots, x_p, q)}{q(x_1, x_2, \cdots, x_p)} \right).
\end{aligned}
$$

This Euler-Lagrange equation is satisfied if the function q satisfies

$$
C_{q1}^2 = \left(\frac{1}{N_T} s(\mathbf{x}) \left(1 - s(\mathbf{x})\right) + s(\mathbf{x})^2 \right) \frac{f^2(\mathbf{x})}{q^2(\mathbf{x})},
$$

where C_{q1} is a positive constant. Rearranging the above equation gives

$$
q(\mathbf{x}) = \frac{1}{C_{q1}} f(\mathbf{x}) \sqrt{\frac{1}{N_T} s(\mathbf{x}) \left(1 - s(\mathbf{x})\right) + s(\mathbf{x})^2}. \tag{S.13}
$$

This function q also satisfies the boundary conditions on Q by setting C_{q1} to satisfy the normalizing constraint of the joint IS density, q, as follows.

$$
\begin{aligned}
C_{q1} &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(\mathbf{x}) \sqrt{\frac{1}{N_T} s(\mathbf{x}) \cdot \left(1 - s(\mathbf{x})\right) + s(\mathbf{x})^2} \, \mathrm{d}x_1 \mathrm{d}x_2 \cdots \mathrm{d}x_p \\
&\equiv \int_{\mathcal{X}_f} f(\mathbf{x}) \sqrt{\frac{1}{N_T} s(\mathbf{x}) \cdot \left(1 - s(\mathbf{x})\right) + s(\mathbf{x})^2} \, \mathrm{d}\mathbf{x}.
\end{aligned} \tag{S.14}
$$

To guarantee that the resulting q is a minimizer of the functional in (S.12), we verify that the following second variation [3] is positive definite,

$$
J[Q; R] = \int_{\mathcal{X}_q} R^2 \frac{\partial^2 \mathcal{L}}{\partial Q^2} + 2Rr \frac{\partial^2 \mathcal{L}}{\partial Q \partial q} + r^2 \frac{\partial^2 \mathcal{L}}{\partial q^2} \, \mathrm{d}\mathbf{x}, \tag{S.15}
$$

where $\mathcal{X}_q = \{\mathbf{x} \in \mathbb{R}^p : q(\mathbf{x}) > 0\}$ is the support of q. The function, $R(x_1, x_2, \cdots, x_p)$, in (S.15) represents a variation that should satisfy the boundary conditions,

$$
\begin{aligned}
R(-\infty, -\infty, \cdots, -\infty) &= 0, \\
R(\infty, \infty, \cdots, \infty) &= 0,
\end{aligned}
$$

so that the varied function, $\tilde{Q}(x_1, x_2, \cdots, x_p) \equiv Q(x_1, x_2, \cdots, x_p) + R(x_1, x_2, \cdots, x_p)$, satisfies the prescribed boundary conditions,

$$
\begin{aligned}
\tilde{Q}(-\infty, -\infty, \cdots, -\infty) &= 0, \\
\tilde{Q}(\infty, \infty, \cdots, \infty) &= 1.
\end{aligned}
$$

The function, $r\left(x_1, x_2, \cdots, x_p\right)$, in (S.15) is

$$r\left(x_1, x_2, \cdots, x_p\right) \equiv \frac{\partial^p R}{\partial x_1 \partial x_2 \cdots \partial x_p}\left(x_1, x_2, \cdots, x_p\right).$$

We note that

$$\begin{aligned}
\frac{\partial^2 \mathcal{L}}{\partial Q^2}(x_1, x_2, \cdots, x_p, q) &= 0, \\
\frac{\partial^2 \mathcal{L}}{\partial Q \partial q}(x_1, x_2, \cdots, x_p, q) &= 0, \\
\frac{\partial^2 \mathcal{L}}{\partial q^2}(x_1, x_2, \cdots, x_p, q) &= 2\left(\frac{1}{N_T} s(\mathbf{x})\left(1-s(\mathbf{x})\right)+s(\mathbf{x})^2\right) \frac{f^2(\mathbf{x})}{q^3(\mathbf{x})} \\
&\quad > 0 \text{ for all } \mathbf{x} \in \mathcal{X}_q = \{\tilde{\mathbf{x}} \in \mathbb{R}^p : q(\tilde{\mathbf{x}}) > 0\}.
\end{aligned}$$

Therefore, the second variation in (S.15) is reduced to

$$J[Q; R] = \int_{\mathcal{X}_q} r^2 \frac{\partial^2 \mathcal{L}}{\partial q^2} \, \mathrm{d}\mathbf{x},$$

where $\frac{\partial^2 \mathcal{L}}{\partial q^2}$ is positive. Therefore, $J[Q; R]$ vanishes if and only if $r(\mathbf{x}) = 0$ for all $\mathbf{x} \in \mathcal{X}_q$. The latter condition implies that $R(\mathbf{x})$ is a constant function of 0, since $R(\mathbf{x})$ is 0 at $(x_1, x_2, \cdots, x_p) = (-\infty, -\infty, \cdots, -\infty)$ and $(x_1, x_2, \cdots, x_p) = (\infty, \infty, \cdots, \infty)$. Therefore, for all allowable nonzero variations, $R(\mathbf{x})$, the second variation is positive definite (i.e., $J[Q; R] > 0$). This verifies that the IS density, q, in (S.13) with the normalizing constant in (S.14) is the minimizing function of the variance in (S.3). We also plug this q into (S.5) to obtain the optimal allocation size, which leads to Theorem 1.

Theorem 1 *(a) The variance of the estimator in* (S.1) *is minimized if the following IS density and the allocation size are used.*

$$q_{SIS1}(\mathbf{x}) = \frac{1}{C_{q1}} f(\mathbf{x}) \sqrt{\frac{1}{N_T} s(\mathbf{x})\left(1-s(\mathbf{x})\right)+s(\mathbf{x})^2}, \tag{S.16}$$

$$N_i = N_T \frac{\sqrt{\frac{N_T(1-s(\mathbf{x}_i))}{1+(N_T-1)s(\mathbf{x}_i)}}}{\sum_{j=1}^{M} \sqrt{\frac{N_T(1-s(\mathbf{x}_j))}{1+(N_T-1)s(\mathbf{x}_j)}}}, \quad i = 1, 2, \cdots, M, \tag{S.17}$$

where C_{q1} *is* $\int_{\mathcal{X}_f} f(\mathbf{x}) \sqrt{\frac{1}{N_T} s(\mathbf{x}) \cdot (1-s(\mathbf{x})) + s(\mathbf{x})^2} \, \mathrm{d}\mathbf{x}$ *and* $s(\mathbf{x})$ *is* $P(Y > l | \mathbf{X} = \mathbf{x})$.
(b) With q_{SIS1} *and* $N_i, i = 1, 2, \cdots, M$, *the estimator in* (S.1) *is unbiased.*
Proof. *(a) We already derived the optimal* q_{SIS1} *in* (S.16) *from the above discussion.*

Plugging the optimal q_{SIS1} into the formula of N_i in (S.5) *gives*

$$\begin{aligned}
N_i &\propto \sqrt{s(\mathbf{x}_i)\left(1-s(\mathbf{x}_i)\right)}\frac{f(\mathbf{x}_i)}{q_{SIS1}(\mathbf{x}_i)} \\
&= \sqrt{s(\mathbf{x}_i)\left(1-s(\mathbf{x}_i)\right)}f(\mathbf{x}_i)\left(\frac{1}{C_{q1}}f(\mathbf{x}_i)\sqrt{\frac{1}{N_T}s(\mathbf{x}_i)\left(1-s(\mathbf{x}_i)\right)+s(\mathbf{x}_i)^2}\right)^{-1} \\
&\propto \sqrt{\frac{s(\mathbf{x}_i)\left(1-s(\mathbf{x}_i)\right)}{\frac{1}{N_T}s(\mathbf{x}_i)\left(1-s(\mathbf{x}_i)\right)+s(\mathbf{x}_i)^2}} \\
&= \sqrt{\frac{N_T(1-s(\mathbf{x}_i))}{1-s(\mathbf{x}_i)+N_T s(\mathbf{x}_i)}} \\
&= \sqrt{\frac{N_T(1-s(\mathbf{x}_i))}{1+\left(N_T-1\right)s(\mathbf{x}_i)}}.
\end{aligned}$$

By imposing the normalizing constraint of $N_T = \sum_{i=1}^{M} N_i$, the expression of the optimal allocation size in (S.17) *follows.*

(b) The estimator in (S.1) *is unbiased if $q_{SIS1}(\mathbf{x}_i) = 0$ implies*

$$\begin{aligned}
\hat{P}\left(Y > l \mid \mathbf{X}=\mathbf{x}_i\right) f(\mathbf{x}_i) &= \left(\frac{1}{N_i}\sum_{j=1}^{N_i}\mathbb{I}\left(Y_j^{(i)} > l\right)\right) f(\mathbf{x}_i) \\
&= 0
\end{aligned}$$

for any $\mathbf{x}_i$. Note that $q_{SIS1}(\mathbf{x}) = 0$ holds only if $f(\mathbf{x}) = 0$ or $s(\mathbf{x}) = 0$. If $s(\mathbf{x}) = 0$, then $\hat{P}\left(Y > l \mid \mathbf{X}=\mathbf{x}\right) = 0$. Therefore, if $q_{SIS1}(\mathbf{x}) = 0$, then $\hat{P}\left(Y > l \mid \mathbf{X}=\mathbf{x}\right) f(\mathbf{x}) = 0$, which concludes the proof. □

1.2 Optimal Important Sampling density in SIS2

Now we consider the SIS2 estimator with a multivariate input vector, $\mathbf{X} \in \mathbb{R}^p$,

$$\hat{P}_{SIS2} = \frac{1}{N_T}\sum_{i=1}^{N_T}\mathbb{I}\left(Y_i > l\right)\frac{f(\mathbf{X}_i)}{q(\mathbf{X}_i)}, \tag{S.18}$$

where Y_i is an output at $\mathbf{X}_i$, $i = 1, 2, \cdots, N_T$. Theorem 2 presents the optimal IS density, q, for the estimator in (S.18). Similar to the derivation of q_{SIS1} in (S.16), we first decompose the estimator variance and apply the principles of the calculus of variation.

Theorem 2 *(a) The variance of the estimator in (S.18) is minimized with the density,*

$$q_{SIS2}(\mathbf{x}) = \frac{1}{C_{q2}}\sqrt{s(\mathbf{x})}f(\mathbf{x}), \tag{S.19}$$

where C_{q2} is $\int_{\mathcal{X}_f} f(\mathbf{x})\sqrt{s(\mathbf{x})}\mathrm{d}\mathbf{x}$ and $s(\mathbf{x})$ is $P(Y > l|\mathbf{X}=\mathbf{x})$. (b) With q_{SIS2}, the estimator

in (S.18) is unbiased.

Proof. *(a)*

$$
\begin{aligned}
Var\left[\hat{P}_{SIS2}\right] &= Var\left[\frac{1}{N_T}\sum_{i=1}^{N_T}\mathbb{I}\left(Y_i > l\right)\frac{f(\mathbf{X}_i)}{q(\mathbf{X}_i)}\right] \\
&= \frac{1}{N_T^2}E_q\left[Var\left[\sum_{i=1}^{N_T}\mathbb{I}\left(Y_i > l\right)\frac{f(\mathbf{X}_i)}{q(\mathbf{X}_i)}\,\middle|\,\mathbf{X}_1,\cdots,\mathbf{X}_{N_T}\right]\right] \\
&\quad + \frac{1}{N_T^2}Var_q\left[E\left[\sum_{i=1}^{N_T}\mathbb{I}\left(Y_i > l\right)\frac{f(\mathbf{X}_i)}{q(\mathbf{X}_i)}\,\middle|\,\mathbf{X}_1,\cdots,\mathbf{X}_{N_T}\right]\right] \\
&= \frac{1}{N_T^2}E_q\left[\sum_{i=1}^{N_T}s(\mathbf{X}_i)\cdot\left(1-s(\mathbf{X}_i)\right)\frac{f(\mathbf{X}_i)^2}{q(\mathbf{X}_i)^2}\right] \\
&\quad + \frac{1}{N_T^2}Var_q\left[\sum_{i=1}^{N_T}s(\mathbf{X}_i)\frac{f(\mathbf{X}_i)}{q(\mathbf{X}_i)}\right] \\
&= \frac{1}{N_T}E_q\left[s(\mathbf{X})\cdot\left(1-s(\mathbf{X})\right)\frac{f(\mathbf{X})^2}{q(\mathbf{X})^2}\right] \\
&\quad + \frac{1}{N_T}\left(E_q\left[s(\mathbf{X})^2\frac{f(\mathbf{X})^2}{q(\mathbf{X})^2}\right] - \left(E_q\left[s(\mathbf{X})\frac{f(\mathbf{X})}{q(\mathbf{X})}\right]\right)^2\right) \\
&= \frac{1}{N_T}E_q\left[s(\mathbf{X})\frac{f(\mathbf{X})^2}{q(\mathbf{X})^2}\right] - \frac{1}{N_T}\left(E_q\left[s(\mathbf{X})\frac{f(\mathbf{X})}{q(\mathbf{X})}\right]\right)^2 \\
&= \frac{1}{N_T}E_f\left[s(\mathbf{X})\frac{f(\mathbf{X})}{q(\mathbf{X})}\right] - \frac{1}{N_T}P\left(Y > l\right)^2. \qquad \text{(S.20)}
\end{aligned}
$$

To find the optimal IS density which minimizes the functional in (S.20), we apply the similar procedure discussed for SIS1. Since only the first term of (S.20) *involves* q*, we consider the following Lagrangian function,*

$$\mathcal{L}(\mathbf{x}, q) = s(\mathbf{x})\frac{f^2(\mathbf{x})}{q(\mathbf{x})}.$$

Note that the Lagrangian function for SIS2 replaces

$$\left(\frac{1}{N_T}s(\mathbf{x})\cdot\left(1-s(\mathbf{x})\right) + s(\mathbf{x})^2\right)$$

in the Lagrangian function for SIS1 (i.e., the integrand in (S.12)) with $s(\mathbf{x})$*. Therefore, the Euler-Lagrange equation and the second variation for SIS2 are analogous to those for SIS1. They lead to the minimizing function in* (S.19) *for SIS2, which is also analogous to the minimizing function in* (S.16) *for SIS1.*

(b) The estimator in (S.18) *is unbiased if* $q_{SIS2}(\mathbf{x}) = 0$ *implies* $\mathbb{I}\left(Y > l\right)f(\mathbf{x}) = 0$ *for any* $\mathbf{x}$*. Note that* Y *is an output corresponding to* $\mathbf{x}$*.* $q_{SIS2}(\mathbf{x}) = 0$ *holds only if* $f(\mathbf{x}) = 0$ *or*

$s(\mathbf{x}) = 0$. *Also, if* $s(\mathbf{x}) = 0$, *then* $\mathbb{I}(Y > l) = 0$. *Therefore, it follows that* $\mathbb{I}(Y > l) f(\mathbf{x}) = 0$ *if* $q_{SIS2}(\mathbf{x}) = 0$. $\square$

2 Univariate Example

To design a univariate stochastic example, we take a deterministic simulation example in Cannamela et al. [2] and modify it to have stochastic elements. Specifically, we have the following data generating structure.

$$X \sim N(0,1)\,,$$
$$Y|X \sim N\big(\mu(X)\,,\sigma^2(X)\big)\,,$$

where the mean, $\mu(X)$, and the standard deviation, $\sigma(X)$, are

$$\mu(X) = 0.95\delta X^2\left(1 + 0.5\cos(10\kappa X) + 0.5\cos(20\kappa X)\right),$$
$$\sigma(X) = 1 + 0.7\,|X| + 0.4\cos(X) + 0.3\cos(14X),$$

respectively. The metamodel of the conditional distribution, $Y|X$, is $N(\hat{\mu}(X)\,,\hat{\sigma}^2(X))$, where

$$\hat{\mu}(X) = 0.95\beta\delta X^2\left(1 + 0.5\rho\cos(10\kappa X) + 0.5\rho\cos(20\kappa X)\right),$$
$$\hat{\sigma}(X) = 1 + 0.7\,|X| + 0.4\rho\cos(X) + 0.3\rho\cos(14X).$$

We vary the following parameters to test different aspects of our proposed methods compared to alternative methods.

- P_T, the magnitude of target failure probability: By varying $P_T = P(Y > l)$, where l is a threshold for the system failure, we want to see how the proposed methods perform at different levels of P_T. Based on 1 million CMC simulation replications, we decide l that corresponds to the target failure probability, P_T. We consider the three levels of P_T, namely, 0.10, 0.05, and 0.01.

- δ, the difference between the original input density, f, and the optimal IS density, q_{SIS1} (or q_{SIS2}): We want to investigate how the computational gains of SIS1 and SIS2 change when the optimal IS density is more different from the original input density, f. Note that the original input density, f, is a standard normal density with a mode at 0. When $\delta = 1$, q_{SIS1} and q_{SIS2} will focus their sampling efforts on the input regions far from 0, since the response variable, Y, tends to be large in such regions due to the term, $0.95X^2$, in $\mu(X)$. Conversely, when $\delta = -1$, q_{SIS1} and q_{SIS2} will focus their sampling efforts on the regions close to 0.

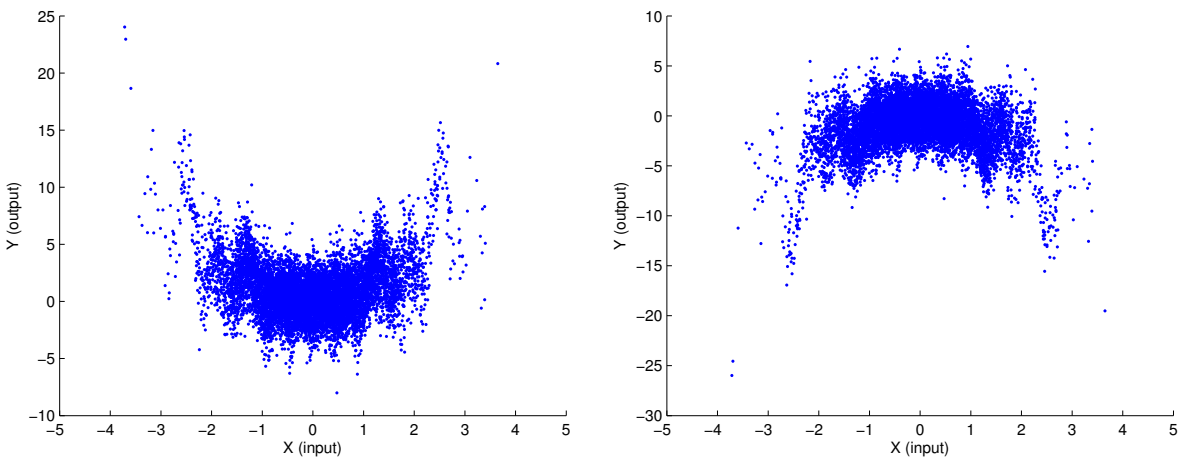
- ρ, the metamodeling accuracy for the oscillating pattern: We vary ρ in $\hat{\mu}(X)$ and $\hat{\sigma}(X)$ to control the quality of the metamodel in capturing the oscillating pattern of the true model with $\mu(X)$ and $\sigma(X)$. We consider ρ of 0, 0.5, and 1. When $\rho = 1$, the metamodel mimics the oscillating pattern perfectly in both the mean and standard deviation, whereas $\rho = 0$ means that the metamodel fails to capture the oscillating pattern.

- β, the metamodeling accuracy for the global pattern: We consider a variation of β in $\hat{\mu}(X)$ with five levels, $\beta = 0.90$, 0.95, 1, 1.05, and 1.10. Note that when $\beta = 1$ (and $\rho = 1$), the metamodel perfectly mimics the true model.

- M/N_T, the ratio of the input sample size to the total number of simulation replications: We consider various choices of M/N_T including 10%, 30%, 50%, 70%, and 90% to see how sensitive the performance of SIS1 is to the choice of M/N_T.

- κ, the locality (or roughness, nonlinearity) of the location function, $\mu(X)$: We consider the three levels of $\kappa = 0$, 0.5, and 1. When κ is far from zero, the cosine terms in $\mu(X)$ add locality, roughness, or nonlinearity to the shape of $\mu(X)$. On the other hand, when $\kappa = 0$, the location function, $\mu(X)$, simply becomes a quadratic function of X.

We use the following setup as a baseline and vary each parameter to see its effect on the performances of the proposed methods: $P_T = 0.01$, $\delta = 1$, $M/N_T = 30\%$, $\rho = 1$, $\beta = 1$, and $\kappa = 0.5$. Figure 1 illustrates the scatter plots at the baseline setup with $\delta = 1$ and -1.



(a) $\delta = 1$ (b) $\delta = -1$

Figure 1: Scatter plots of the baseline univariate example with different δ

We set N_T, the total simulation replications, as 1,000. To obtain the sample average and standard error of each method's POE estimation, we repeat the experiment 500 times.

2.1 Effects of P_T and δ

Table 1 summarizes the effects of P_T and δ. Except these two, we keep the other parameters at their baseline values. We use the perfect metamodel (i.e., $\rho = 1$, $\beta = 1$) so that we can examine the main effect of P_T and δ without any interaction effects with the metamodel quality.

We compute the relative ratio, $N_T/N_T^{(CMC)}$, as follows. Let $N_T^{(CMC)}$ denote the number of CMC simulation replications to achieve the same standard error of each method in the table. With $N_T^{(CMC)}$ replications, the standard error of the CMC failure probability estimator is $\sqrt{P_T\left(1-P_T\right)/N_T^{(CMC)}}$. Table 1 shows that the relative ratios of SIS1 and SIS2 are

comparable to each other and clearly better than BIS, and that they generally decrease as P_T gets smaller. That is, the efficiencies of the SIS methods against CMC improve as P_T gets close to zero. For example, when $\delta = 1$ and P_T are 0.10, 0.05, and 0.01, SIS1 requires 51%, 32%, and 2.5% of the CMC simulation efforts to achieve the same estimation accuracy, respectively (in other words, CMC needs about twice, three times, and forty times more simulation efforts than SIS1, respectively.) These remarkable computational savings are also observed in our case study with the wind turbine simulators (see Table 5 of the paper). Specifically, SIS1 and SIS2 respectively lead to 4.9% and 5.9% of the relative ratios for edgewise bending moments with l = 9,300 kNm. Note that the corresponding sample averages, namely 0.00992 and 0.01005, are close to the failure probability of $P_T = 0.01$.

Table 1 also shows that the computational gains of SIS1 and SIS2 are much more significant when $\delta = 1$ (i.e., when f and q_{SIS1} (or q_{SIS2}) are quite different) than when $\delta = -1$. This finding is intuitive and also consistent with the observation in the wind turbine simulation that the computational gains of SIS1 and SIS2 for the edgewise bending moments are much more remarkable than those for the flapwise bending moments. Interestingly, when $\delta = -1$, BIS has no advantage over CMC, whereas the proposed methods still lead to lower standard errors than CMC.

Table 1: POE estimation results with different δ and P_T

Method		$\delta = 1$			$\delta = -1$		
		P_T			P_T		
		0.10	0.05	0.01	0.10	0.05	0.01
SIS1	Sample Average	0.1004	0.0502	0.0100	0.1001	0.0500	0.0100
	Standard Error	0.0068	0.0039	0.0005	0.0090	0.0062	0.0026
	Relative Ratio	51%	32%	2.5%	90%	81%	68%
SIS2	Sample Average	0.0999	0.0501	0.0100	0.1001	0.0500	0.0099
	Standard Error	0.0069	0.0042	0.0006	0.0086	0.0064	0.0028
	Relative Ratio	53%	37%	3.6%	82%	86%	79%
BIS	Sample Average	0.1002	0.0505	0.0101	0.1009	0.0503	0.0101
	Standard Error	0.0089	0.0068	0.0014	0.0095	0.0067	0.0031
	Relative Ratio	88%	97%	20%	100%	95%	97%
CMC	Sample Average	0.1005	0.0506	0.0100	0.1005	0.0498	0.0100
	Standard Error	0.0092	0.0070	0.0030	0.0096	0.0071	0.0031

2.2 Effects of metamodel accuracy

Now, we consider how computational efficiency varies when the metamodel accuracy changes. First, we study the effect of ρ, the metamodeling accuracy for the oscillating pattern. We keep all other parameters at their baseline values. The results in Table 2 suggest that the standard errors for SIS1, SIS2, and BIS increase as ρ decreases (i.e., the metamodel quality deteriorates). However, the standard errors for both SIS1 and SIS2 increase more slowly than for BIS as ρ decreases. Also, SIS1 and SIS2 produce lower standard errors than BIS by 50-85% and CMC by 40-85%. Interestingly, the increase of the SIS2's standard error is minimal,

indicating that SIS2 is the least sensitive to the metamodel quality. The performance of BIS differs significantly depending on the metamodel quality, and BIS generates an even higher standard error than CMC when $\rho = 0$.

Table 2: POE estimation results with different ρ

Method		ρ 1.00	0.50	0
SIS1	Sample Average	0.0100	0.0100	0.0101
	Standard Error	0.0005	0.0008	0.0017
SIS2	Sample Average	0.0100	0.0101	0.0100
	Standard Error	0.0006	0.0007	0.0010
BIS	Sample Average	0.0101	0.0100	0.0102
	Standard Error	0.0014	0.0018	0.0063
CMC	Sample Average	0.0099	0.0099	0.0099
	Standard Error	0.0030	0.0030	0.0030

Second, we consider the effect of β, the metamodeling accuracy for the global pattern. We keep all other parameters at their baseline values. The results in Table 3 do not show any clear patterns to explain the impact of the metamodel accuracy of the global pattern on the performances of SIS1 and SIS2. However, in all cases, SIS1 and SIS2 outperform BIS and CMC, reducing the standard errors by 45-70% and 80-85%, respectively.

Table 3: POE estimation results with different β

Method		β 0.90	0.95	1.00	1.05	1.10
SIS1	Sample Average	0.0101	0.0101	0.0100	0.0101	0.0101
	Standard Error	0.0005	0.0005	0.0005	0.0005	0.0005
SIS2	Sample Average	0.0101	0.0100	0.0100	0.0100	0.0101
	Standard Error	0.0006	0.0006	0.0006	0.0006	0.0006
BIS	Sample Average	0.0101	0.0100	0.0101	0.0101	0.0101
	Standard Error	0.0013	0.0016	0.0014	0.0013	0.0011
CMC	Sample Average	0.0100	0.0100	0.0099	0.0100	0.0099
	Standard Error	0.0031	0.0031	0.0030	0.0030	0.0030

Third, we investigate the effect of the metamodel quality on the computational gains of the proposed methods as the failure probability gets smaller, when the metamodel is poor. Specifically, we consider the cases of $(\rho = 0.5, \beta = 1)$, $(\rho = 0, \beta = 0.6)$, and $(\rho = 0, \beta = 1.2)$. We keep all other parameters at their baseline values. Table 4 shows that the computational efficiencies of SIS1 and SIS2 are substantially better than BIS in all cases. Similar to the pattern in Table 2, SIS2 tends to perform better than SIS1 when the metamodel is inaccurate, and when P_T changes from 0.10 to 0.01, the efficiencies of SIS1 and SIS2 improve remarkably. However, we note that there are some cases (e.g., SIS1 with $\rho = 0, \beta = 1.2$ and SIS2 with $\rho = 0, \beta = 0.6$) where the efficiency slightly diminishes when P_T changes from 0.10 to 0.05.

This result indicates that if the metamodel is inaccurate, the efficiencies of SIS1 and SIS2 do not necessarily improve when smaller P_T is estimated. Even so, SIS1 and SIS2 perform much better than BIS.

Table 4: POE estimation results with different ρ and β

Method		$\rho = 0.5, \beta = 1$			$\rho = 0, \beta = 0.6$			$\rho = 0, \beta = 1.2$		
		P_T			P_T			P_T		
		0.10	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01
SIS1	Ave.	0.0998	0.0503	0.0100	0.0998	0.0503	0.0100	0.1001	0.0503	0.0102
	S.E.	0.0080	0.0046	0.0008	0.0104	0.0066	0.0016	0.0120	0.0090	0.0024
	Ratio	71%	44%	6.4%	120%	91%	26%	160%	170%	58%
SIS2	Ave.	0.0999	0.0503	0.0101	0.0999	0.0506	0.0100	0.0993	0.0503	0.0101
	S.E.	0.0068	0.0045	0.0007	0.0082	0.0064	0.0009	0.0078	0.0054	0.0010
	Ratio	51%	42%	4.9%	75%	86%	8.1%	67%	61%	10%
BIS	Ave.	0.1007	0.0502	0.0100	0.1014	0.0493	0.0105	0.1028	0.0511	0.0105
	S.E.	0.0134	0.0078	0.0018	0.0355	0.0086	0.0082	0.0665	0.0184	0.0104
	Ratio	199%	128%	32%	1398%	155%	673%	4905%	710%	1082%
CMC	Ave.	0.1004	0.0506	0.0099	0.1005	0.0504	0.0100	0.1001	0.0504	0.0099
	S.E.	0.0091	0.0071	0.0030	0.0093	0.0071	0.0030	0.0093	0.0070	0.0030

Notes: 'Ave.' denotes the sample average, 'S.E.' denotes the standard error, and 'Ratio' denotes the relative ratio of N_T/N_T^{CMC}.

2.3 Effects of the ratio, M/N_T

Here, we want to see how sensitive SIS1 is to the choice of M/N_T. We keep all other parameters at their baseline values. The results in Table 5 suggest that the standard error of the SIS1 estimator is generally insensitive to the choice of M/N_T. This result is consistent with the result of the wind turbine simulations. Note that the standard error in Table 5 is presented up to 5 digits (not 4 digits).

Table 5: Effect of different M/N_T ratios in the univariate example

M/N_T	Sample Average	Standard Error
10%	0.0100	0.00055
30%	0.0100	0.00050
50%	0.0101	0.00059
70%	0.0101	0.00063
90%	0.0100	0.00076

2.4 Effects of locality, κ

We consider the effect of κ, the locality (or roughness, nonlinearity) of the location function, $\mu(X)$. We keep all other parameters at their baseline values. The results in Table 6 show that the standard errors for SIS1 and SIS2 slightly increase as κ increases. However, regardless of κ, SIS1 and SIS2 outperform BIS and CMC, lowering the standard errors by 30-65% and 75-90%, respectively.

Table 6: POE estimation results with different κ

Method		κ		
		0	0.50	1.00
SIS1	Sample Average	0.0100	0.0100	0.0100
	Standard Error	0.0004	0.0005	0.0007
SIS2	Sample Average	0.0100	0.0100	0.0101
	Standard Error	0.0005	0.0006	0.0007
BIS	Sample Average	0.0100	0.0101	0.0100
	Standard Error	0.0008	0.0014	0.0010
CMC	Sample Average	0.0100	0.0099	0.0099
	Standard Error	0.0031	0.0030	0.0031

2.5 Effects of variation of ϵ

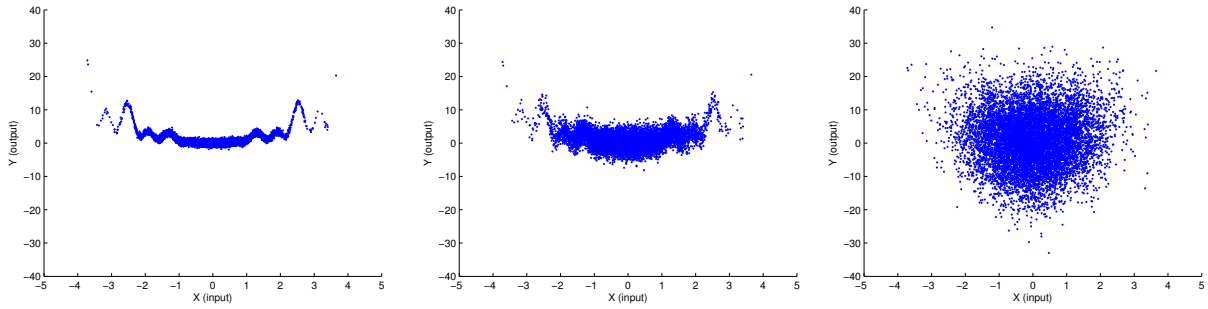
Theoretically, SIS1 and SIS2 are reduced to deterministic importance sampling (DIS) when the simulator is deterministic. Recall that the standard error for DIS with q_{DIS} is zero. Thus, in a stochastic computer model, if the uncontrollable randomness represented by ϵ has a smaller level of variation, then the standard errors for SIS1 and SIS2 will get closer to zero. We conduct a numerical study to illustrate the impact of the variance of ϵ. We consider the same data generating structure as before except that the variance of ϵ does not depend on the input, X, but is constant:

$$\sigma^2(X) = \tau^2.$$

Equivalently, we consider $Y = \mu(X) + \epsilon$, where ϵ follows a normal distribution with mean zero and standard deviation, τ. We use the optimal IS densities for SIS1 and SIS2 with the perfect knowledge of $s(X)$. We consider τ of 0.5, 1, 2, 4, and 8. In Figure 2, we can see the scatter plots of Y versus X for τ of 0.5, 2, and 8, by which the variation of Y given X is controlled. We set all other parameters at their baseline values: $P_T = 0.01$, $\delta = 1$, $M/N_T = 30\%$, and $\kappa = 0.5$.

Table 7 shows that as τ gets close to zero (please see from right to left), so do the standard errors of SIS1 and SIS2. The results indicate that the proposed methods practically reduce to DIS, since the optimal DIS density makes the standard error zero for the deterministic simulation (i.e., the case with $\tau = 0$).

Also, Figure 3 illustrates that the optimal SIS1 and SIS2 densities are almost the same as the BIS density when the variation of ϵ is very small (in the figure, we use $\tau = 0.5$).



(a) $\tau = 0.5$ (b) $\tau = 2$ (c) $\tau = 8$

Figure 2: Scatter plots of the baseline case with different τ

Table 7: POE estimation results with different $\tau = \mathrm{Var}(\epsilon)$

Method		τ				
		0.50	1.00	2.00	4.00	8.00
SIS1	Sample Average	0.0102	0.0101	0.0101	0.0102	0.0100
	Standard Error	0.0001	0.0001	0.0005	0.0021	0.0028
SIS2	Sample Average	0.0102	0.0101	0.0101	0.0104	0.0100
	Standard Error	0.0001	0.0002	0.0006	0.0023	0.0028
BIS	Sample Average	0.0102	0.0101	0.0100	0.0103	0.0101
	Standard Error	0.0002	0.0003	0.0010	0.0033	0.0033
CMC	Sample Average	0.0100	0.0100	0.0099	0.0101	0.0101
	Standard Error	0.0030	0.0031	0.0030	0.0032	0.0031

Notes: SIS1's standard errors for $\tau = 0.50$ and $\tau = 1.00$ are 0.00007 and 0.00013, respectively, in one more digit.

Since the BIS density theoretically reduces to the DIS density for deterministic simulation and closely mimics the DIS density when τ is negligibly small, we can see that the proposed methods practically reduce to DIS when the variation of the uncontrollable randomness is very small.

2.6 Precision of numerical integration

When we use the numerical integration to compute the normalizing constant of an IS density, we make sure that the numerical precision is accurate enough so that the POE estimation accuracy is unaffected. We present POE estimation results up to 5 digits after the decimal point. Given that we bound the numerical error by -7 orders of magnitude or smaller, the numerical integration does not contribute to the error of POE estimation. To check the precision, we also conduct numerical studies with the same data generating structure used in Section 2.5. In Table 8, the sample averages and standard errors are based on 500 POE estimates. The POEs in the last column are estimated by CMC with 100 million replications. We note that the estimated POE values from SIS1 and SIS2 are the same as the values from CMC.

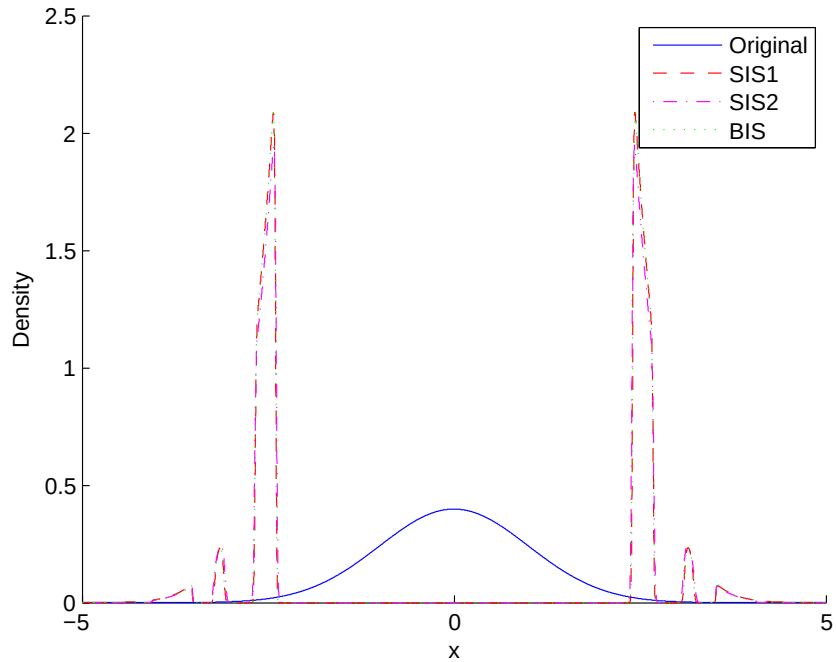


Figure 3: Density plots for SIS1, SIS2, and BIS optimal densities when $\tau = 0.50$ along with the original input density

3 Multivariate Example

We also design a multivariate stochastic example. We take an example in Huang et al. [4], which adds a normal stochastic noise to a deterministic example originally in Ackley [1]. We slightly revise the example in Huang et al. [4] by adding more complexity to the stochastic elements, and use the following data generating structure where the input vector, $\mathbf{X} = (X_1, X_2, X_3)$, follows a multivariate normal distribution:

$$\mathbf{X} \sim MVN(0, I_3)\,,$$
$$Y|\mathbf{X} \sim N\big(\mu(\mathbf{X})\,, \sigma^2(\mathbf{X})\big)\,,$$

where the mean, $\mu(\mathbf{X})$, and the standard deviation, $\sigma(\mathbf{X})$, are

$$\mu(\mathbf{X}) = 20\delta\left(1 - \exp\left(-0.2\sqrt{\frac{1}{3}\|\mathbf{X}\|^2}\right)\right) + \delta\left(\exp(1) - \exp\left(\frac{1}{3}\sum_{i=1}^{3}\cos(2\pi\kappa X_i)\right)\right),$$
$$\sigma(\mathbf{X}) = 1 + 0.7\sqrt{\frac{1}{3}\|\mathbf{X}\|^2} + 0.4\left(\frac{1}{3}\sum_{i=1}^{3}\cos(3\pi X_i)\right),$$

respectively. The output, Y, with the above $\mu(\mathbf{X})$ and $\sigma(\mathbf{X})$ presents a very complicated pattern over the input domain. The metamodel for the conditional distribution, $Y|\mathbf{X}$, is

Table 8: POE estimation results for SIS1 and SIS2, compared to the POE estimated by CMC with 100 million replications, for different $\tau = \mathrm{Var}(\epsilon)$

τ	Sample Average (Standard Error)		CMC
	SIS1	SIS2	
0.50	0.0102 (0.0001)	0.0102 (0.0001)	0.0102
1.00	0.0101 (0.0001)	0.0101 (0.0002)	0.0101
2.00	0.0101 (0.0005)	0.0101 (0.0006)	0.0101

$N(\hat{\mu}(\mathbf{X}), \hat{\sigma}^2(\mathbf{X}))$, where

$$\hat{\mu}(\mathbf{X}) = 20\beta\delta\left(1 - \exp\left(-0.2\sqrt{\frac{1}{3}\|\mathbf{X}\|^2}\right)\right) + \rho\delta\left(\exp(1) - \exp\left(\frac{1}{3}\sum_{i=1}^{3}\cos(2\pi\kappa X_i)\right)\right),$$

$$\hat{\sigma}(\mathbf{X}) = 1 + 0.7\sqrt{\frac{1}{3}\|\mathbf{X}\|^2} + 0.4\rho\left(\frac{1}{3}\sum_{i=1}^{3}\cos(3\pi X_i)\right).$$

The parameters in the above equations take similar roles in the univariate example. We use the same baseline setup we used in the univariate example, namely, $P_T = 0.01$, $\delta = 1$, $M/N_T = 30\%$, $\rho = 1$, $\beta = 1$, and $\kappa = 0.5$. We explain each parameter as follows.

- P_T, the magnitude of target failure probability: Based on 10 million CMC simulation replications, we decide l that corresponds to the target failure probability, $P_T = P(Y > l)$. We consider the three levels of P_T, 0.10, 0.05, and 0.01.

- δ, the difference between the original input density, f, and the optimal IS density, q_{SIS1}(or q_{SIS2}): We consider δ of 1 or -1. The densities, f and q_{SIS1} (or q_{SIS2}), are more different from each other when $\delta = 1$ than when $\delta = -1$. Note that the original input density, f, has the highest likelihood at the origin. When $\delta = 1$, q_{SIS1} and q_{SIS2} will focus their sampling efforts on the regions far from the origin, since the response variable, Y, tends to be large in such regions due to the term, $20\left(1 - \exp\left(-0.2\sqrt{\frac{1}{3}\|\mathbf{X}\|^2}\right)\right)$, in $\mu(\mathbf{X})$. Conversely, when $\delta = -1$, q_{SIS1} and q_{SIS2} will focus their sampling efforts on the regions close to the origin.

- ρ, the metamodeling accuracy for the oscillating pattern: We consider ρ of 0, 0.5, and 1. When $\rho = 1$, the metamodel mimics the oscillating pattern perfectly, whereas $\rho = 0$ implies that the metamodel captures no oscillating term.

- β, the metamodeling accuracy for the global pattern: We consider $\beta = 0.95$, 1, and 1.05. Note that when $\beta = 1$ (and $\rho = 1$), the metamodel perfectly mimics the true model.

- M/N_T, the ratio of the input sample size to the total number of simulation replications: We consider M/N_T of 10%, 30%, 50%, 70%, and 90%.

- κ, the locality (or roughness, nonlinearity) of the location function, $\mu(\mathbf{X})$: We consider the four levels of κ, 0, 0.5, 1, and 2. When κ is far from zero, the cosine terms in $\mu(\mathbf{X})$ add locality, roughness, or nonlinearity to the shape of $\mu(\mathbf{X})$. On the other hand, when $\kappa = 0$, the location function, $\mu(\mathbf{X})$, simply becomes a monotonically increasing function of $\|\mathbf{X}\|$.

3.1 Effects of P_T and δ

Table 9 summarizes the effects of P_T and δ. We keep all other parameters at their baseline values. Similar to the univariate example, the experiment results suggest that the computational gains of SIS1 and SIS2 against CMC increase as P_T gets smaller. We also see that the computational gains of SIS1 and SIS2 are more significant when $\delta = 1$ (i.e., f and q_{SIS1} (or q_{SIS2}) are quite different) than when $\delta = -1$. In all cases, SIS1 and SIS2 perform better than BIS and CMC.

We note that when $\delta = 1$, the relative ratios of SIS1 and SIS2 decrease more slowly than the univariate input example results in Table 1. Specifically, for $P_T = 0.01$, SIS1 and SIS2 yield 2.5% and 3.6% of the relative ratios in Table 1; but, both methods give 29% of the relative ratio in Table 9. We attribute such differences in the two example results to the differences in the data generating structures. The data generating structure of the univariate example in Section 2 and the multivariate example in Section 3 are different not only in the input dimension but also in the mean function, $\mu(\mathbf{x})$, and the standard deviation function, $\sigma(\mathbf{X})$. We detail this point in Section 3.5.

Table 9: POE estimation results with different δ and P_T for the multivariate example

Method		$\delta = 1$			$\delta = -1$		
		P_T			P_T		
		0.10	0.05	0.01	0.10	0.05	0.01
SIS1	Sample Average	0.1002	0.0501	0.0100	0.1000	0.0500	0.0100
	Standard Error	0.0070	0.0046	0.0017	0.0072	0.0051	0.0020
	Relative Ratio	54%	45%	29%	58%	55%	40%
SIS2	Sample Average	0.1002	0.0499	0.0100	0.1001	0.0499	0.0100
	Standard Error	0.0070	0.0048	0.0017	0.0078	0.0050	0.0020
	Relative Ratio	54%	49%	29%	68%	53%	40%
BIS	Sample Average	0.1000	0.0500	0.0100	0.1001	0.0500	0.0102
	Standard Error	0.0082	0.0062	0.0026	0.0096	0.0069	0.0036
	Relative Ratio	75%	81%	68%	102%	100%	131%
CMC	Sample Average	0.0997	0.0500	0.0101	0.0998	0.0499	0.0101
	Standard Error	0.0094	0.0069	0.0031	0.0093	0.0069	0.0031

3.2 Effects of metamodel accuracy

We consider the effect of ρ, the metamodeling accuracy for the oscillating pattern. We keep all other parameters at their baseline values. Similar to the univariate example, Table 10 shows that the standard errors of the SIS1 and SIS2 estimators increase as ρ decreases. Also, the standard error for SIS2 increases more slowly than that for SIS1, which shows that SIS2 is less sensitive than SIS1 to the metamodel quality. It appears that the performance of BIS is the most sensitive to the metamodel quality. In all cases, SIS1 and SIS2 lead to smaller standard errors than BIS and CMC by 20–60% and 20–50%, respectively.

Table 10: POE estimation results with different ρ in the multivariate example

Method		ρ		
		1.00	0.50	0
SIS1	Sample Average	0.0100	0.0101	0.0100
	Standard Error	0.0017	0.0019	0.0024
SIS2	Sample Average	0.0100	0.0100	0.0099
	Standard Error	0.0016	0.0018	0.0020
BIS	Sample Average	0.0100	0.0100	0.0098
	Standard Error	0.0022	0.0040	0.0047
CMC	Sample Average	0.0101	0.0102	0.0101
	Standard Error	0.0031	0.0031	0.0031

Notes. At $\rho = 1$, standard errors for SIS1 and SIS2 are 0.00167 and 0.00163, respectively, in one more digit.

We consider the effect of β, the metamodeling accuracy for the global pattern. We keep all other parameters at their baseline values. Table 11 shows that the standard errors of the SIS1 and SIS2 estimators do not vary significantly, so the performances of SIS1 and SIS2 are insensitive to the metamodeling accuracy for the global pattern in this example. In all cases, SIS1 and SIS2 outperform BIS and CMC, providing lower standard errors than BIS and CMC by 25–40% and 45-50%, respectively. .

Table 11: POE estimation results with different β in the multivariate example

Method		β		
		0.95	1.00	1.05
SIS1	Sample Average	0.0099	0.0099	0.0100
	Standard Error	0.0016	0.0017	0.0017
SIS2	Sample Average	0.0100	0.0100	0.0100
	Standard Error	0.0017	0.0017	0.0017
BIS	Sample Average	0.0101	0.0100	0.0100
	Standard Error	0.0025	0.0023	0.0023
CMC	Sample Average	0.0101	0.0101	0.0101
	Standard Error	0.0031	0.0031	0.0031

3.3 Effects of the ratio, M/N_T

We want to see how sensitive SIS1 is to the choice of M/N_T. We keep all other parameters at their baseline values. The results in Table 12 suggest that the standard error of the SIS1 estimator is generally insensitive to the choice of M/N_T as we observed in the univariate example and the wind turbine simulations. Note that the standard error is presented up to 5 digits (not 4 digits).

Table 12: Effect of different M/N_T ratios in the multivariate example

M/N_T	Sample Average	Standard Error
10%	0.0100	0.00168
30%	0.0100	0.00167
50%	0.0100	0.00168
70%	0.0100	0.00173
90%	0.0100	0.00185

3.4 Effects of locality, κ

We consider the effect of κ, the locality (or roughness, nonlinearity) of the location function, $\mu(\mathbf{X})$. We keep all other parameters at their baseline values. The results in Table 13 suggest that κ has little effect on the standard errors of the SIS1 and SIS2 estimators in this specific example. For all κ values, SIS1 and SIS2 lead to smaller standard errors than BIS and CMC by 20–50% and 45–50%, respectively.

Table 13: POE estimation results with different κ in the multivariate example

Method		κ			
		0	0.5	1	2
SIS1	Sample Average	0.0099	0.0100	0.0101	0.0099
	Standard Error	0.0017	0.0017	0.0016	0.0016
SIS2	Sample Average	0.0101	0.0100	0.0100	0.0100
	Standard Error	0.0017	0.0017	0.0016	0.0016
BIS	Sample Average	0.0101	0.0100	0.0101	0.0100
	Standard Error	0.0031	0.0022	0.0033	0.0031
CMC	Sample Average	0.0102	0.0101	0.0102	0.0103
	Standard Error	0.0031	0.0031	0.0031	0.0032

3.5 Analysis with the univariate input

In Section 3.1, as P_T gets smaller, we observe that the relative ratio of SIS1 and SIS2 with $\delta = 1$ decreases more slowly than those in the univariate example (see Tables 1 and 9 with $\delta = 1$). These different patterns in the two numerical examples are mainly due to the different data generating structure not only in the input dimension but also in the mean

function, $\mu(\mathbf{x})$, and the standard deviation function, $\sigma(\mathbf{X})$. For the univariate example in Section 2, we take a deterministic simulation example in Cannamela *et al.* [2] and modify it by adding stochastic elements to it, whereas for the multivariate example in Section 3, we add a normal stochastic noise to a deterministic multivariate example originally in Ackley [1]. In the sequel, we call these univariate and multivariate examples as Cannamela1D and Ackley3D, respectively, based on their respective sources [1, 2].

To clarify the different patterns in Cannamela1D and Ackley3D, we devise a new univariate example which is one-dimensional version of Ackley3D, and we call this new example as Ackley1D. Specifically, we consider the following data generating structure:

$$\begin{aligned} X &\sim N(0,1)\,, \\ Y|X &\sim N\big(\mu(X)\,,\sigma^2(X)\big)\,, \end{aligned}$$

where the mean, $\mu(X)$, and the standard deviation, $\sigma(X)$, are

$$\begin{aligned} \mu(X) &= 20\delta\left(1-\exp\left(-0.2|X|\right)\right)+\delta\left(\exp\left(1\right)-\exp\left(\cos(2\pi\kappa X)\right)\right), \\ \sigma(X) &= 1+0.7|X|+0.4\left(\cos(3\pi X)\right), \end{aligned}$$

respectively.

The metamodel for the conditional distribution, $Y|X$, is $N(\hat{\mu}(X)\,,\hat{\sigma}^2(X))$, where

$$\begin{aligned} \hat{\mu}(X) &= 20\beta\delta\left(1-\exp\left(-0.2|X|\right)\right)+\rho\delta\left(\exp\left(1\right)-\exp\left(\cos(2\pi\kappa X)\right)\right), \\ \hat{\sigma}(X) &= 1+0.7|X|+0.4\rho\left(\cos(3\pi X)\right). \end{aligned}$$

For the experiments of Ackley1D, we use the same baseline setup used in Cannamela1D and Ackley3D, namely, $P_T = 0.01$, $\delta = 1$, $M/N_T = 30\%$, $\rho = 1$, $\beta = 1$, and $\kappa = 0.5$. Note that $\rho = 1$, $\beta = 1$ imply that the metamodel is perfect so that the optimal IS densities and allocations can be used.

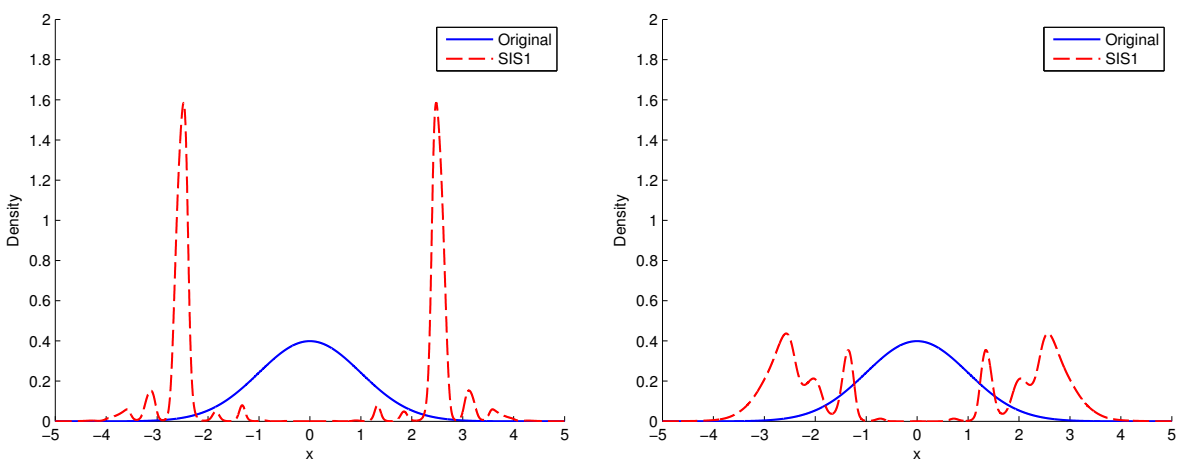
Table 14 below compares the results of Ackley1D and Ackley3D. We note that the relative ratios of SIS1 and SIS2 for Ackley1D, namely, 15% and 17%, are smaller than those of Ackley3D, namely, 29% and 29%. Yet, the performances for Ackley1D are not as remarkable as those for Cannamela1D in Table 1, namely, 2.5% and 3.6%. Such performance differences in Cannamela1D and Ackley1D can be explained mainly by the difference in their underlying data generating structures: See Figure 4 below, where we plot the optimal SIS1 density along with the original input density for both examples. Apparently, the optimal SIS1 density for Cannamela1D is deviating much more from the original input density than that for Ackley1D is. We observe the similar pattern for SIS2. This explains the better performances of SIS1 and SIS2 for Cannamela1D.

Obviously, the computational gain of SIS1 and SIS2 over CMC largely depends on the general trend represented by the location parameter functions, $\mu(X)$. In addition, the scale parameter functions, $\sigma(X)$, also make a difference in the performance of SIS1 and SIS2 for Cannamela1D and Ackley1D. We plot 20,000 input-output pairs, (X, Y)'s, generated from the baseline setups for Cannamela1D and Ackley1D in Figures 5(a) and (b), respectively. We draw the solid horizontal line in each plot to indicate the resistance level, l, corresponding to $P_T = 0.01$. We observe that the location parameter functions, $\mu(X)$, in both examples

tend to have large values at the regions where $f(X)$ is small. However, the scale parameter functions, $\sigma(X)$, lead to a major difference around the region, $(-2,-1)\cup(1,2)$, where $\mu(X)$ itself is not yet close to l but many responses of Ackley1D in Figure 5(b) exceed l unlike Cannamela1D in Figure 5(a). Accordingly, we observe the relevant peaks at $(-2,-1)\cup(1,2)$ in Figure 4(b), which disperse sampling efforts in a larger input area and make q_{SIS1} (and q_{SIS2}) more overlapped with f for Ackley1D.

Table 14: POE estimation results with different input dimension and target failure probability, P_T, for the numerical examples based on Ackley [1]

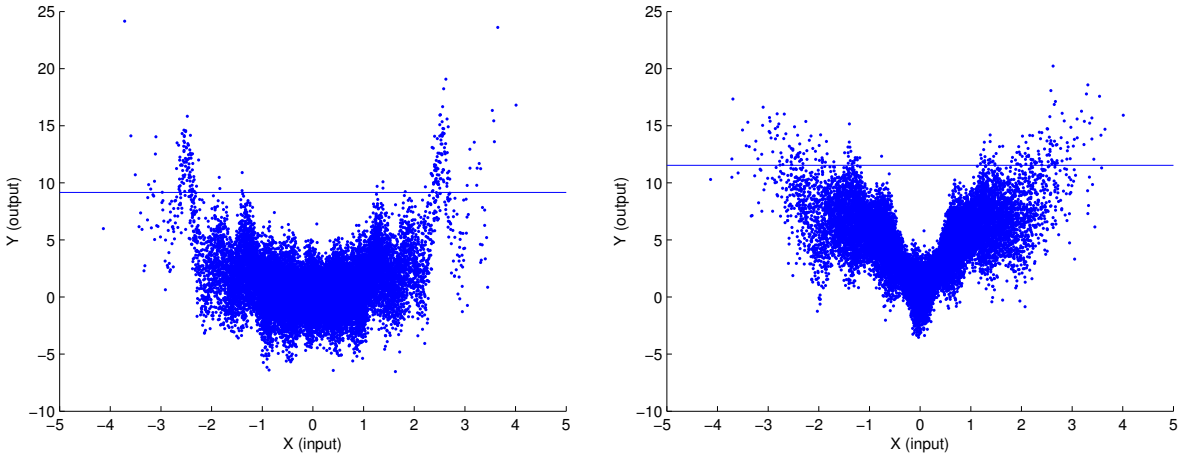
Method		Ackley1D			Ackley3D		
		P_T			P_T		
		0.10	0.05	0.01	0.10	0.05	0.01
SIS1	Sample Average	0.1001	0.0501	0.0100	0.1002	0.0501	0.0100
	Standard Error	0.0059	0.0038	0.0012	0.0070	0.0046	0.0017
	Relative Ratio	39%	30%	15%	54%	45%	29%
SIS2	Sample Average	0.0998	0.0501	0.0100	0.1002	0.0499	0.0100
	Standard Error	0.0060	0.0040	0.0013	0.0070	0.0048	0.0017
	Relative Ratio	40%	34%	17%	54%	49%	29%
BIS	Sample Average	0.1000	0.0499	0.0100	0.1000	0.0500	0.0100
	Standard Error	0.0072	0.0052	0.0027	0.0082	0.0062	0.0026
	Relative Ratio	58%	57%	74%	75%	81%	68%
CMC	Sample Average	0.1001	0.0501	0.0100	0.0997	0.0500	0.0101
	Standard Error	0.0098	0.0071	0.0031	0.0094	0.0069	0.0031



(a) Cannamela1D (b) Ackley1D

Figure 4: Comparison of the optimal SIS1 density and the original input density for the two examples

In summary, the performance of the proposed methods will depend on the characteristics of the simulation model. Note that the variances of the proposed estimators depend only on



(a) Cannamela1D (b) Ackley1D

Figure 5: Scatter plots of the data generated from the baseline data generating structures: the solid horizontal line is the quantile, l, corresponding to $P_T = 0.01$

the functions, $s(\mathbf{x})$ and $f(\mathbf{x})$, according to Theorems 1 and 2 (note that the SIS1 and SIS2 densities are also expressed in $s(\mathbf{x})$ and $f(\mathbf{x})$) and both functions are determined by the true data generating structure. Lastly, we remark that the higher relative ratios of SIS1 and SIS2 for Ackley3D compared to those for Ackley1D should not be generalized as that the input dimension negatively affects the performance of SIS1 and SIS2. In the case of Ackley3D, due to the highly oscillating response over the *three* dimensional input space, the sampling efforts are more distributed in the larger input space and the resulting q_{SIS1} (and q_{SIS2}) is more overlapped with f, compared to the case of Ackley1D. However, even with high dimensional input vectors, significant computational reduction can be achieved when the joint density of the input vector, f, and the optimal SIS1 (and SIS2) density, q_{SIS1} (and q_{SIS2}), are different.

3.6 Summary

Overall, we observe similar patterns both in the univariate example and the multivariate example. These patterns are also consistent with the wind turbine simulation results. For a wide range of parameter settings, the performances of SIS1 and SIS2 are superior to BIS and CMC.

4 Implementation Details with Wind Turbine Simulators

In this section, we present the implementation details with wind turbine simulators.

4.1 NREL simulators and the original input distribution

The NREL simulators used in this study include TurbSim [6] and FAST [7]. Given a wind condition (e.g., 10-minute average wind speed), TurbSim produces a three-dimensional stochastic wind profile. FAST, taking the generated wind profile as an input, simulates load responses (or loads) at turbine subsystems such as blades and shafts. Noting that there are many types of load responses, we limit our study to consider edgewise and flapwise bending moments at a blade root as output variables, where edgewise (flapwise) bending moments imply structural loads parallel (perpendicular) to the rotor span at a blade root. These two load types are of great concern in ensuring a wind turbine's structural reliability [10].

As in [10], we use the same turbine specification for an onshore version of an NREL 5-MW baseline wind turbine [8]. The target turbine operates within a specified wind speed range between the cut-in speed, $x_{in} = 3$ meter per second (m/s), and the cut-out speed, $x_{out} = 25\ m/s$. Following wind industry practice and the international standard, IEC 61400-1 [5], we use a 10-minute average wind speed as an input, X, to the simulators. We use a Rayleigh density for X with a truncated support of $[x_{in}, x_{out}]$ as in [10]:

$$f(x) = \frac{f_R(x)}{F_R(x_{out}) - F_R(x_{in})},$$

where $F_R(x) = 1 - e^{-x^2/2\tau^2}$ denotes the cumulative distribution function of Rayleigh distribution with a scale parameter, $\tau = \sqrt{2/\pi} \cdot 10$ (unit: m/s). Also, f_R denotes the Rayleigh density function with the same scale parameter.

4.2 Acceptance rates of the acceptance-rejection algorithm

We use the acceptance-rejection algorithm in the implementation. The algorithm's acceptance rate is equal to the normalizing constant of each IS density because we use the original input density, f, as an instrumental (or auxiliary) density for the algorithm [9]. Note that the normalizing constants are C_{q1} for SIS1, C_{q2} for SIS2, and $P(Y > l)$ for BIS.

The acceptance rates differ, depending on POE, $P(Y > l)$. In our implementation, when POE is around 0.05 (i.e., edgewise moments with $l = 8{,}600\ kNm$ or flapwise moments with $l = 13{,}800\ kNm$), the acceptance rates are 5–21%. When POE is around 0.01 (i.e., edgewise moments with $l = 9{,}300\ kNm$ or flapwise moments with $l = 14{,}300\ kNm$), the acceptance rates are 1–14%. In practice, the computational cost of the acceptance-rejection algorithm would be insignificant. For example, sampling thousands of inputs from the IS densities is a matter of seconds, whereas thousands of the NREL simulation replications can take days.

4.3 Goodness-of-fit test for the model

In constructing the metamodel, we assume no prior information on important area, so sampling X from the uniform distribution would be generally suitable. We use the GEV distribution for approximating the conditional POE given X regardless of the choice of distribution for X, and the GEV distribution is employed over the entire input space with varying location and scale parameters, $\mu(X)$ and $\sigma(X)$. In our implementation, we use the metamodel based on the GEV distribution to approximate the theoretically optimal IS density, q_{SIS1} (or

q_{SIS2}). That is, the GEV distribution is used as a means to find the good IS density. Then, we run the real simulators (not the metamodel) to gather Y for each X sampled from q_{SIS1} (or q_{SIS2}).

Obviously, the metamodel quality affects the performance of the proposed approach. Therefore, in our study, we used the GEV goodness-of-fit to check if the GEV provides a good approximation of the conditional distribution over the entire input space as shown in the paper. In this section, we additionally check if the GEV is suitable in the area where X is likely sampled. Noting that high edgewise (flapwise) bending loads are most likely observed when wind speeds are between 17 and 25 (11 and 19), we take 50 observations each at $17, 19, \cdots, 25$ $(11, 13, \cdots, 19)$ m/s and conduct Kolmogorov-Smirnov (KS) tests to assess the goodness-of-fit of the GEV distribution at each wind speed. The results in Table 15 below support the use of GEV distribution for edgewise and flapwise bending moments, as the p-values are greater than a reasonable significance level, say, 5%.

Table 15: KS tests for GEV at imporant wind speeds

Edgewise bending moments		Flapwise bending moments	
x (m/s)	p-value	x (m/s)	p-value
17	0.34	11	0.31
19	0.60	13	0.52
21	0.89	15	0.35
23	0.19	17	0.57
25	0.64	19	0.36

4.4 CMC simulations

We want to ensure that the estimations of $N_T^{(CMC)}$ are accurate, which are used to compute the relative ratios in Tables 5 and 6 of the paper. Thus, we run CMC simulations with $N_T^{(CMC)}$ corresponding to SIS1 and SIS2 for the flapwise moment with $l = 13,800$ kNm and compute the standard errors based on 50 repetitions. The corresponding $N_T^{(CMC)}$ for SIS1 and SIS2 are 6,219 and 4,762, respectively. In addition, we run simulations with $N_T^{(CMC)} = 5,000$ and $N_T^{(CMC)} = 6,000$. With $N_T^{(CMC)}$ of 6,000 and 6,219, we obtain the CMC's standard error of 0.0028, which is the same with the SIS1's standard error with $N_T = 2,000$. With $N_T^{(CMC)} = 4,762$ and $N_T^{(CMC)} = 5,000$, we obtain the CMC's standard errors of 0.0036 and 0.0033, respectively, which are close to the SIS2's standard error of 0.0032. We omit the CMC implementation for other cases due to the intensive computational requirement.

References

[1] Ackley, D. H. (1987). *A connectionist machine for genetic hillclimbing.* Boston: Kluwer Academic Publishers.

[2] Cannamela, C., Garnier, J., and Iooss, B. (2008). Controlled stratification for quantile estimation. *The Annals of Applied Statistics*, 2(4):1554–1580.

[3] Courant, R. and Hilbert, D. (1989). *Methods of Mathematical Physics.* New York: Wiley.

[4] Huang, D., Allen, T. T., Notz, W. I., and Zeng, N. (2006). Global optimization of stochastic black-box systems via sequential kriging meta-models. *Journal of global optimization*, 34(3):441–466.

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

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

[7] Jonkman, J. M. and Buhl Jr., M. L. (2005). FAST User's Guide. Technical Report NREL/EL-500-38230, National Renewable Energy Laboratory, Golden, Colorado.

[8] Jonkman, J. M., Butterfield, S., Musial, W., and Scott, G. (2009). Definition of a 5-MW reference wind turbine for offshore system development. Technical Report NREL/TP-500-38060, National Renewable Energy Laboratory, Golden, Colorado.

[9] Kroese, D. P., Taimre, T., and Botev, Z. I. (2011). *Handbook of Monte Carlo Methods.* New York: John Wiley and Sons.

[10] Moriarty, P. (2008). Database for validation of design load extrapolation techniques. *Wind Energy*, 11(6):559–576.