



Technical Letter Report
TLR-RES/DE/REB-2021-15

Application of Probabilistic Analysis Techniques in Mechanical Integrity Assessments

Date:

September 2021

Prepared in response to Task 3 in User Need Request NRR-2016-004, by:

Nevin Martin
Lauren B. Hund
Zach Stuart
John R. Lewis
Josh Mullins
Adah Zhang
Michael J. Starr
Dusty Brooks
Aubrey C. Eckert
Remi Dingreville

Sandia National Laboratories

Patrick Raynaud
Senior Materials Engineer
Component Integrity Branch

Division of Engineering
Office of Nuclear Regulatory Research
U.S. Nuclear Regulatory Commission
Washington, DC 20555-0001

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the U.S. Government. Neither the U.S. Government nor any agency thereof, nor any employee, makes any warranty, expressed or implied, or assumes any legal liability or responsibility for any third party's use, or the results of such use, of any information, apparatus, product, or process disclosed in this publication, or represents that its use by such third party complies with applicable law.

This report does not contain or imply legally binding requirements. Nor does this report establish or modify any regulatory guidance or positions of the U.S. Nuclear Regulatory Commission and is not binding on the Commission.

Application of Probabilistic Analysis Techniques in Mechanical Integrity Assessments

Prepared by:

Nevin Martin
Lauren B. Hund
Zach Stuart
John R. Lewis
Josh Mullins
Adah Zhang
Michael J. Starr
Dusty Brooks
Aubrey C. Eckert
Remi Dingreville

Patrick Raynaud, NRC Project Manager

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico
87185 and Livermore,
California 94550

Issued by Sandia National Laboratories, operated for the United States Department of Energy by National Technology & Engineering Solutions of Sandia, LLC.

NOTICE: This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government, nor any agency thereof, nor any of their employees, nor any of their contractors, subcontractors, or their employees, make any warranty, express or implied, or assume any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represent that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government, any agency thereof, or any of their contractors or subcontractors. The views and opinions expressed herein do not necessarily state or reflect those of the United States Government, any agency thereof, or any of their contractors.

Printed in the United States of America. This report has been reproduced directly from the best available copy.

Available to DOE and DOE contractors from

U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865) 576-8401
Facsimile: (865) 576-5728
E-Mail: reports@osti.gov
Online ordering: <http://www.osti.gov/scitech>

Available to the public from

U.S. Department of Commerce
National Technical Information Service
5301 Shawnee Rd
Alexandria, VA 22312

Telephone: (800) 553-6847
Facsimile: (703) 605-6900
E-Mail: orders@ntis.gov
Online order: <https://classic.ntis.gov/help/order-methods/>



ABSTRACT

This report illustrates the application of various probabilistic analysis techniques to a simple, idealized mechanical integrity assessment problem. Its development and structure are linked to the Nuclear Regulatory Commission's (NRC) Probabilistic Fracture Mechanics (PFM) NUREG/CR-7278 by applying the methods that are introduced in Section 4 of NUREG/CR-7278 to the idealized mechanical integrity assessment problem. NUREG/CR-7278 is focused on describing the technical basis for a Regulatory Guide outlining expectation for PFM analyses conducted in support of risk-informed regulatory applications. This report is purely focused on the practical applications of analytical techniques used in probabilistic analysis.

ACKNOWLEDGEMENTS

The authors would like to thank Casey Jelsema and Matthew Martinez for their review of this document.

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

TABLE OF CONTENTS

Abstract	vi
Acknowledgements.....	vii
TABLE OF CONTENTS	viii
LIST OF FIGURES.....	xi
LIST OF Tables	xiii
ABBREVIATIONS AND ACRONYMS	xiv
1. Objective	1
2. Example problem definition.....	3
3. Separation of aleatory and epistemic uncertainty	6
3.1. Estimate a QoI without separating uncertainty.....	6
3.1.1. Process:.....	6
3.1.2. Results and potential difficulties:	7
3.1.3. Summary:.....	7
3.2. Estimate a QoI when separating aleatory and epistemic uncertainties.....	8
3.2.1. Process:.....	8
3.2.2. Results and potential difficulties:	11
3.2.3. Summary:.....	12
4. Statistical distribution fitting.....	13
4.1. Data-driven distribution estimation	13
4.1.1. Process:.....	13
4.1.2. Results and potential difficulties:	17
4.1.3. Summary:.....	18
4.2. Distribution truncation	18
4.2.1. Process:.....	18
4.2.2. Results and potential difficulties:	19
4.2.3. Summary:.....	20
5. Modeling dependence between inputs.....	21
5.1. Induce dependence between inputs using rank correlation	21
5.1.1. Process:.....	21
5.1.2. Results and potential difficulties:	22
5.1.3. Summary:.....	25
5.2. Importance sampling with correlated inputs	25
5.2.1. Process:.....	26
5.2.2. Results and potential difficulties:	27
5.2.3. Summary:.....	28
6. Sensitivity Analysis.....	29
6.1. Choose a relevant response on which to conduct sensitivity analysis	29
6.1.1. Process:.....	29
6.1.2. Results and potential difficulties:	29
6.1.3. Summary	30
6.2. Use exploratory data analysis to visualize input/output relationships.....	30
6.2.1. Process:.....	30

6.2.2.	Results and potential difficulties:	31
6.2.3.	Summary	32
6.3.	Estimate local sensitivity metrics	32
6.3.1.	Process:	32
6.3.2.	Results and potential difficulties:	33
6.3.3.	Summary	33
6.4.	Estimate global sensitivity metrics	34
6.4.1.	Process:	34
6.4.2.	Results and potential difficulties:	34
6.4.3.	Summary	35
7.	Surrogate Models	36
7.1.	Build surrogate models for forward propagation of uncertainty	36
7.1.1.	Process:	36
7.1.2.	Results and potential difficulties:	38
7.1.3.	Summary:	41
7.2.	Surrogate models for sensitivity analysis	42
7.2.1.	Process	42
7.2.2.	Results and potential difficulties:	42
7.2.3.	Summary	43
8.	Forward Propagation of Input Uncertainty	44
8.1.	Estimate the QoI distribution using SRS	44
8.1.1.	Process:	44
8.1.2.	Results and potential difficulties:	45
8.1.3.	Summary:	46
8.2.	Estimate the QoI using IS	47
8.2.1.	Process:	47
8.2.2.	Results and potential difficulties:	50
8.2.3.	Summary:	53
8.3.	Estimate the QoI using LHS	53
8.3.1.	Process:	53
8.3.2.	Results and potential difficulties:	54
8.3.3.	Summary:	54
9.	Convergence Analysis	56
9.1.	Assess convergence using closed-form statistical metrics under simple random sampling	56
9.1.1.	Process:	56
9.1.2.	Results and potential difficulties:	57
9.1.3.	Summary:	57
9.2.	Assess convergence using the statistical bootstrap to estimate convergence metrics	58
9.2.1.	Process:	58
9.2.2.	Results and potential difficulties:	59
9.2.3.	Summary:	60
9.3.	Assess convergence of the QoI as a function of the sample size	60
9.3.1.	Process:	60
9.3.2.	Results and potential difficulties:	61
9.3.3.	Summary:	62
9.4.	Assess convergence of an estimate of the QoI by performing replicate simulations	62
9.4.1.	Process:	62

9.4.2. Results and potential difficulties:	64
9.4.3. Summary:	66
10. Visualizing uncertainty in the QoI	67
10.1. Visualizing uncertainty without separation of aleatory and epistemic uncertainties	67
10.1.1. Process:	67
10.1.2. Results and potential difficulties:	67
10.1.3. Summary:	68
10.2. Visualizing uncertainty in the QoI with separation of aleatory and epistemic uncertainties	68
10.2.1. Process:	68
10.2.2. Results and potential difficulties:	69
10.2.3. Summary	69
Bibliography	70

LIST OF FIGURES

Figure 2.1. Pressurized cylindrical tank with a helical weld.	3
Figure 2.2. Distribution shape of each shear stress independent variable and the yield stress distribution of the weld.	5
Figure 3.1. Approximate distributions of the shear stress on the weld and the weld yield stress from 4×10^4 importance samples.	7
Figure 3.2. The estimated probability of failure across 1000 epistemic samples. The dashed line represents the mean estimate.	11
Figure 3.3. 100 epistemic empirical CDFs of shear stress when pressure uncertainty is aleatory (left) and epistemic (right).	12
Figure 4.1. QQ plots for evaluating distributional fit. (Top left) normal distribution; (top right) lognormal distribution; and (bottom) Weibull distribution.	16
Figure 4.2. Weibull distributions on weld yield stress defined by different one-sided confidence levels on the Weibull parameters.	17
Figure 4.3. Different distributions with a domain of finite measure, relative to the original normal distribution. The right figure is zoomed into the bottom tails of the distributions.	19
Figure 5.1. (Left) Correlated dependent inputs generated using the rank correlation method; (right) input distributions without correlation.	23
Figure 5.2. Distribution of shear stress based on uncorrelated samples (red) and correlated samples (green); the distribution of yield stress is shown in blue for reference.	24
Figure 5.3. Distribution of shear stress based on uncorrelated samples (red) and correlated samples (green) with more uncertainty in the radius; the distribution of yield stress is shown in blue for reference.	25
Figure 5.5. Distribution of correlated inputs radius and pressure when inputs are sampled from the importance distribution (left) and the target distribution (right).	27
Figure 6.1. Scatterplots for shear stress versus pressure, radius, weld thickness, and angle of the helical weld.	31
Figure 7.1. Comparing three training set designs using 1,000 training points. (Left) Root mean- squared error for shear stress. A horizontal line is drawn at the ideal $RMSE = 0$. (Right) Estimated failure probability. A horizontal line is drawn for the ideal failure probability estimate without using a surrogate.	39
Figure 7.2. Root mean-squared error (left) and failure probability estimate (right) as a function of training set size.	40
Figure 7.3. Predicted shear stress from surrogate model compared to true shear stress value from the model for: linear regression surrogate (left) and GP surrogate (right).	41
Figure 7.4. (Left) Mean-squared error of shear stress as a function of training set size. (Right) Failure probability estimate as a function of training set size.	41
Figure 7.5. Total effects sensitivity index estimates without a surrogate and using 3 different surrogates built using: (Left) 10 training points and (right) 100 training points.	43
Figure 8.1. Comparison of shear stress and yield stress distributions.	45
Figure 8.2. Boxplots for 100 replicate simulations using an SRS of sizes 1×10^7 (red) and 1×10^6 (blue).	46
Figure 8.3. Scatterplot of shear stress versus pressure.	47
Figure 8.4. The importance distributions for both the pressure (top) and weld yield stress (bottom) distributions.	49
Figure 8.5. Sampling density contours for three implementations of IS.	51

Figure 8.6. Empirical CDFs for 100 replicate simulations using 1×10^4 samples for IS on pressure, weld, and both, along with 1×10^7 samples for SRS.....	51
Figure 8.7. Importance distribution in relation to the original distribution for pressure.....	52
Figure 9.1. Histograms of the failure probability based on 1000 bootstrap samples using SRS (left) and using IS (right).	59
Figure 9.2. Probability of weld failure as a function of the sample size using SRS (left) and IS (right) with upper confidence bounds. The scales of the plots differ. A horizontal line is drawn at the true failure probability.	61
Figure 9.3. Comparison of SRS and IS failure probability estimates as the sample size increases. A horizontal line is drawn at the true failure probability.	62
Figure 9.4. Histograms of 100 failure probability estimates based on SRS (left) and IS (right) replicate simulations.....	65
Figure 9.5. Histograms of 10 failure probability estimates based on SRS (left) and IS (right) replicate simulations.....	65
Figure 10.1. Histograms of shear stress (green) and weld yield stress (purple). The right plot provides a zoomed in view of the region where shear stress and weld yield stress overlap.	68
Figure 10.2. Empirical CDF of the probability of weld failure (blue solid) with vertical line at 95th percentile (black dashed).....	69

LIST OF TABLES

Table 3.1. Classification of Inputs as Aleatory or Epistemic	10
Table 4.1. Experimental Measurements of Weld Yield Stress	13
Table 4.2. Parameter Estimates and Sampling Uncertainties	15
Table 4.3. Impact of Truncation of r on Failure Probability.....	20
Table 6.1. Set of Initial Values and Perturbations for Local Sensitivity Analysis.....	32
Table 6.2. Partial Derivatives of Shear Stress for Each Input.....	33
Table 6.3. First Order and Total Effects Indices for Each Input.....	35
Table 8.1. Failure Probability Estimates for SRS Across Replicate Simulations	46
Table 8.2. Failure Probability Estimates for Different Sampling Schemes and Importance Sampling Implementations	52
Table 8.3. Failure Probability Estimates for Different Choices of the Importance Sampling Distribution Compared to SRS.....	53
Table 8.4. Failure Probability Estimates for SRS with IS and LHS with IS and 1×10^4 Samples Across Replicate Simulations	54
Table 8.5. Failure Probability Estimates for SRS with IS and LHS with IS and 1×10^5 Samples Across Replicate Simulations	54
Table 9.1. Closed-form Convergence Metric Results	57
Table 9.2. Convergence Metrics for Both Sampling Methods Using Bootstrap Samples	59
Table 9.3. Convergence Metrics for Both Sampling Methods	65

ABBREVIATIONS AND ACRONYMS

CDF	Cumulative Distribution Function
CI	Confidence Interval
CV	Coefficient of Variation
ECDF	Empirical Cumulative Distribution Function
GP	Gaussian Process
IQR	Interquartile Range
IS	Importance Sampling
LHS	Latin Hypercube Sampling
MARS	Multivariate Adaptive Regression Splines
NRC	Nuclear Regulatory Commission
NUREG	Nuclear Regulatory Commission Report Designation
PDF	Probability Density Function
PFM	Probabilistic Fracture Mechanics
RMSE	Root Mean-Squared Error
QoI	Quantity of Interest
QQ	Quantile-Quantile
SA	Sensitivity Analysis
SRS	Simple Random Sampling
UQ	Uncertainty Quantification

1. OBJECTIVE

What is this report?

This report is intended to be a stand-alone document illustrating the application of various probabilistic analysis techniques to a simple, idealized mechanical integrity assessment problem. Though the document itself is stand-alone, its development and structure are linked to the Nuclear Regulatory Commission's (NRC) Probabilistic Fracture Mechanics (PFM) NUREG/CR-7278. Specifically, this document applies the methods that are introduced in Section 4 of NUREG/CR-7278 to the idealized mechanical integrity assessment problem. NUREG/CR-7278 is focused on describing the technical basis for a Regulatory Guide outlining expectation for PFM analyses conducted in support of risk-informed regulatory applications. This report is purely focused on the practical applications of analytical techniques used in probabilistic analysis (explained in NUREG/CR-7278, Section 4). Of course, the techniques presented can be directly mapped to a PFM analysis of the type described in NUREG/CR-7278.

Who is the intended reader?

It is expected that the primary user of this report is a solid mechanics/fracture mechanics practitioner who is seeking to gain a deeper understanding of the methods and means of leveraging probabilistic tools and techniques against a mechanical integrity assessment problem. Many of the concepts should not be foreign to the practitioner; however, the mathematical basis, application details, and usage rationale may not be well understood. In that sense, this report is intended to serve as a user's guide for fundamental, important probabilistic concepts, as applied to a representative, though accessible, mechanics problem.

Structure of the report

The report is intended to mirror the structure and sections provided in Section 4 of NUREG/CR-7278. Because the structure of NUREG/CR-7278 is specifically focused on the procedure for developing a PFM analysis in the context of a nuclear regulatory application, the order of certain probabilistic concepts in this report may not occur in the sequential order for a standard PFM analysis. Indeed, it should be noted that a PFM analysis is not a linear and sequential process. Rather a PFM analysis involves the application and iteration of several probabilistic concepts and techniques to meet the objectives associated with the mechanical integrity assessment problem at hand.

Section 2 of this report introduces the mechanical integrity assessment problem, defines the quantity of interest (QoI) in the analysis, defines all the relevant variables and their relationship to the QoI, and provides the statistics associated with those variables. Sections 3-10 illustrate eight methods (mapping to NUREG/CR-7278 given in parenthesis) used in a probabilistic analysis:

- Section 3. Separation of aleatory and epistemic uncertainty (NUREG/CR-7278, Section 4.1.1),
- Section 4. Statistical distribution fitting (NUREG/CR-7278, Section 4.2.1),
- Section 5. Modeling dependence between inputs (NUREG/CR-7278, Section 4.2.2),
- Section 6. Sensitivity analysis (NUREG/CR-7278, Sections 4.3.8, 4.3.9),
- Section 7. Surrogate models (NUREG/CR-7278, Section 4.3.10),
- Section 8. Forward propagation of input uncertainty (NUREG/CR-7278, Sections 4.3.1 – 4.3.3),

Section 9. Convergence analysis (NUREG/CR-7278, Sections 4.3.5 – 4.3.7), and
Section 10. Visualizing uncertainty in the QoI (NUREG/CR-7278, Section 4.3.11).

In each of these sections, fundamental concepts of the method are explicitly illustrated against the example mechanical integrity assessment problem described in the first section. There are multiple concepts presented for each method. Each concept is developed via step-by-step process descriptions, presentation of the results of application of the concept, discussion of common difficulties and bounds of applicability of the concept, and a final usage summary.

2. EXAMPLE PROBLEM DEFINITION

Depending on the scope and nature of each unique PFM analysis, various components need to be considered. These include:

- (i) a description of the model(s) used,
- (ii) a description of the QoI(s) to be used to evaluate the PFM analysis,
- (iii) a description of the input space to be used with this model, including which inputs are considered deterministic or probabilistic, as well as the input values for deterministic inputs, and
- (iv) a description of the input uncertainties including the classification of the uncertainties as epistemic or aleatory (if applicable), choice of sampling scheme, choice of probability distribution, etc.

Other aspects related to verification and validation of the PFM framework, definition of acceptance criteria, and risk-informed decision making against regulatory requirements are outside the scope of this document and are addressed in NUREG/CR-7278.

Model: An idealized solid mechanics representation of stresses acting on a pressurized steel tank with a helical weld.

Quantity of interest: Uncertain shear stress acting on the helical weld will be compared to an uncertain yield stress tolerance on the weld. For the purposes of this analysis, failure of the weld is defined to occur if the shear stress on the weld, τ , is greater than the yield stress threshold of the weld, Y . Figure 2.1 gives an example of the pressurized cylindrical tank with a helical weld [1].

The QoI in this example is the probability of weld failure over a population of welds. To illustrate the concepts in this report, a “benchmark” probability of weld failure was estimated by sampling from the true input distributions (typically unknown in real applications and defined below) until the probability converged. This benchmark failure probability - 8.8×10^{-7} - will be used as a reference throughout the document.

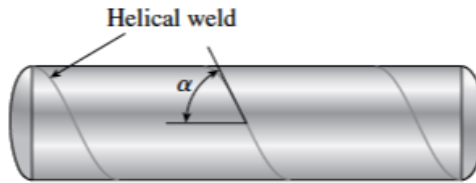


Figure 2.1. Pressurized cylindrical tank with a helical weld.

Input space: Shear stress on the weld, τ , can be calculated through the following equation:

$$\tau = \frac{\rho r}{4t} \sin(180 - 2\alpha)$$

where:

- ρ : Internal pressure (MPa)
- r : Cylinder radius (m)
- t : Wall thickness (m)
- α : Helical weld angle (degrees)

All four inputs that define shear stress are treated as uncertain, as well as the yield stress threshold of the weld.

Uncertain inputs: While typically unknown in real applications, we define uncertainties on r and t as known through the cylinder manufacturing process, α through the welding process, and ρ based on the properties of the tank and the gas contained in the tank. The radius and the angle of the weld are modeled as a truncated normal (TN) distribution where the variances are small relative to their means. The truncation occurs at ± 3 standard deviations to eliminate physically implausible tanks. Thickness is modeled with a right-skewed beta distribution because, relative to the design (average) thickness, this example tank has a tighter lower tolerance on thickness than an upper tolerance. The pressure is modeled with a normal (N) distribution. Finally, the yield stress is modeled using a Weibull distribution based on experimental data and expert judgment (discussed further in Section 4). The following is a summary of the true (the input data generating mechanism produces values that follow random samples from the distribution.) distributions of these inputs.

- $\rho \sim N(\mu = 50, \sigma = 6)$
- $r \sim TN(\mu = 0.6, \sigma = 0.012, LB = 0.564, UB = 0.636)$
- $t \sim Beta(\alpha = 2, \beta = 5, LB = 0.018, UB = 0.02)$
- $\alpha \sim TN(\mu = 55, \sigma = 1.1, LB = 51.7, UB = 58.3)$
- $Y \sim Weibull(\lambda = 600, k = 50)$

Details on estimating the parameters for these distributions is provided in Section 4. Figure 2.2 shows the probability distributions for each uncertain input. All of the analysis and visualization provided in this report were performed using the free software environment, R.

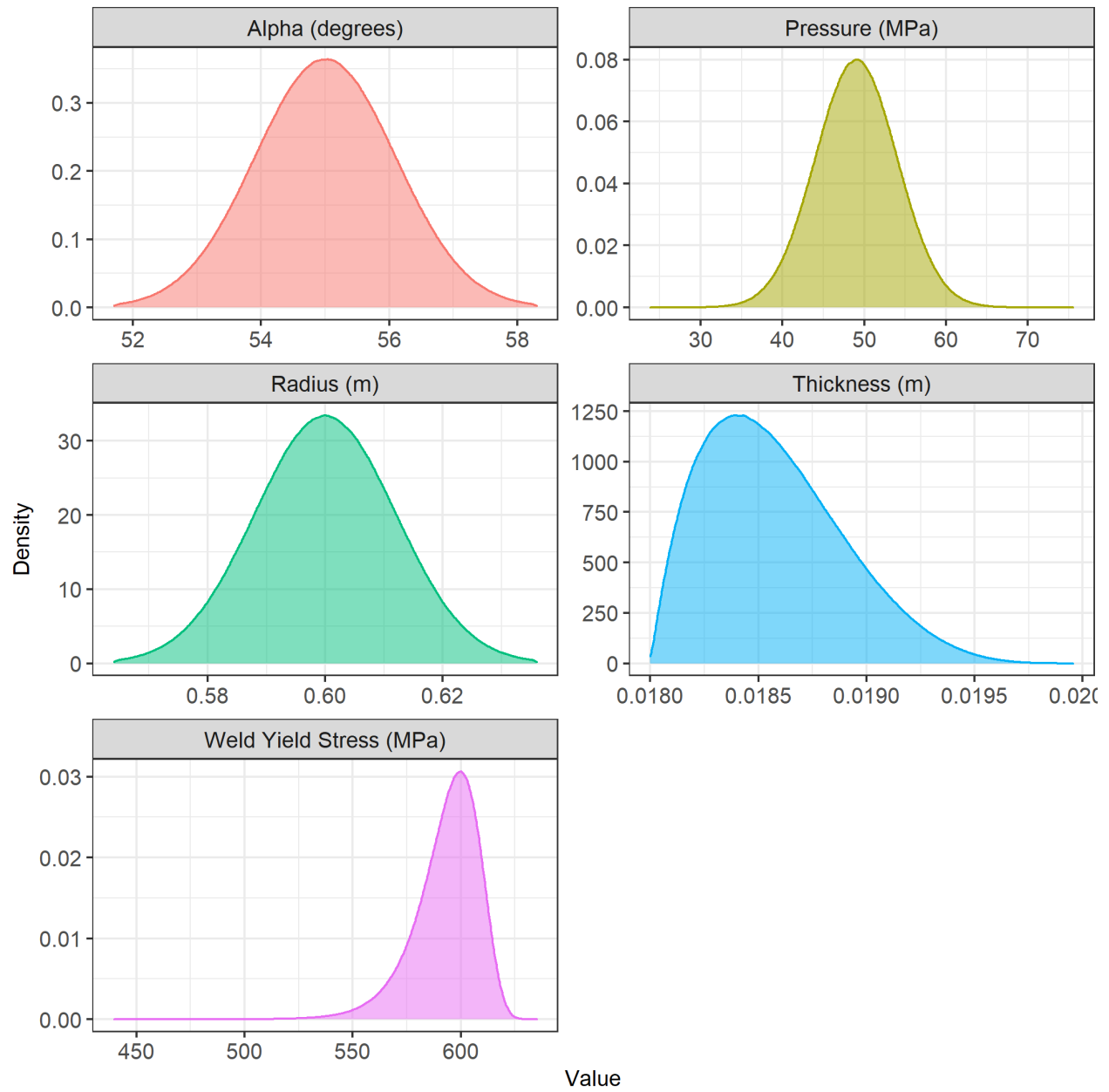


Figure 2.2. Distribution shape of each shear stress independent variable and the yield stress distribution of the weld.

3. SEPARATION OF ALEATORY AND EPISTEMIC UNCERTAINTY

This section applies methods for the separation of aleatory (irreducible randomness) and epistemic (lack of knowledge) uncertainty (NUREG/CR-7278, Section 4.1.1) to the pressurized tank example. Two concepts are illustrated:

1. Estimate a QoI without separating uncertainty.
2. Estimate a QoI when separating aleatory and epistemic uncertainties.

When to apply this method. The choice of whether to separate aleatory and epistemic uncertainties should be tied to the analysis objectives. This decision should be made early in the process as it can affect how input distributions are defined (Section 4), as well as methods for uncertainty propagation (Section 8) and assessing convergence (Section 9). The choice will also dictate how uncertainty in the result is interpreted and visualized (Section 10).

3.1. Estimate a QoI without separating uncertainty

This example illustrates the process of estimating the probability of weld failure without separating by type of uncertainty.

3.1.1. Process:

1. Determine sample size and sampling scheme

The sample size must be sufficiently large relative to the size of the probability of weld failure. To illustrate the importance for determining a sufficient target sample size up front, Section 9 (Convergence Analysis) under importance sampling (IS) of pressure and weld yield stress shows that a sample size of 4×10^4 is sufficiently large to get a stable estimate of the probability of failure. If there's an insufficient sample size, a QoI will not sufficiently converge, and a simulation and analysis may need to be redone. As will be discussed in Section 8, IS was performed on pressure and weld yield stress as they were found to contribute to most of the uncertainty in the probability of failure.

2. Simulate stress values by sampling from input probability distributions

Pressure, radius, wall thickness, angle of the helical weld, and weld yield stress values are sampled based on the distributions defined in the problem statement (Section 2) and IS is used on pressure and weld yield stress since these variables are expected to have a larger impact on the QoI than the other inputs. No distinction is made in the classification of aleatory and epistemic uncertainties for these inputs. 4×10^4 samples of the inputs are taken and the resulting stress samples are used to approximate the shear stress and weld yield stress distributions, as shown in Figure 3.1.

3. Compare stress values to calculate the probability of weld failure

The probability of failure of the weld can be estimated using the following steps:

- i. For the i^{th} sample, let $X_i = 1$ if the shear stress is larger than the weld yield stress and 0 otherwise.
- ii. The estimated probability of failure is

$$\hat{P}(\text{failure}) = \frac{1}{n} \sum_{i=1}^n X_i$$

where n is the sample size.

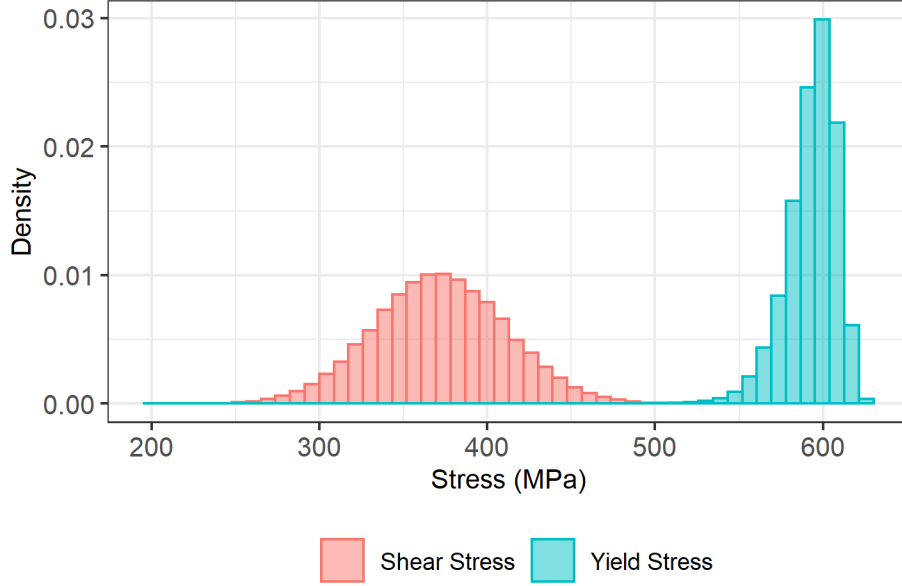


Figure 3.1. Approximate distributions of the shear stress on the weld and the weld yield stress from 4×10^4 importance samples.

3.1.2. Results and potential difficulties:

Using 4×10^4 samples with IS, the probability of weld failure is estimated to be 8.3×10^{-7} . Note that this is relatively close to the benchmark (true) probability of failure of 8.8×10^{-7} , but there is a difference due to sampling uncertainty. Since aleatory and epistemic uncertainties were not separated, this estimate does not distinguish between the likelihood of failure (i.e., aleatory uncertainty) and the confidence of failure (i.e., epistemic uncertainty). Consider a probability of failure threshold of 1×10^{-6} . Looking at the estimated probability of failure alone gives the impression that the true probability is below that threshold. However, since uncertainties were not separated, we are unable to assess the confidence of that probability with respect to the epistemic uncertainty of the model inputs.

When uncertainties are not separated, pure sampling uncertainty (i.e., the uncertainty due to a finite sample size) can be estimated using techniques described in Section 9. Sampling uncertainty is one type of epistemic uncertainty, though it is distinguished from the epistemic uncertainty of the model inputs.

3.1.3. Summary:

- Without the separation of uncertainties, a point estimate of the failure probability can be calculated, but the confidence of that probability cannot be estimated without violating assumptions.

3.2. Estimate a Qol when separating aleatory and epistemic uncertainties

This example illustrates the process of separating aleatory and epistemic uncertainties to calculate a point estimate of the probability of weld failure along with a measure of the epistemic uncertainty. This process is described more completely in the Nuclear Regulatory Commission's (NRC) Probabilistic Fracture Mechanics (PFM) NUREG/CR-7278.

3.2.1. Process:

1. Categorize uncertainties as aleatory or epistemic

All four uncertainties associated with the calculation of the shear stress and the uncertainty in the weld yield stress are separated into aleatory and epistemic based on the analyst's domain knowledge and the goals of the analysis.

Since the goal is to estimate the failure probability across a population of welds, the most natural classification in this example would be to categorize all variables as aleatory. The radius, wall thickness, angle of the helical weld, and weld yield stress are all physical properties associated with manufacturing tolerances. These uncertainties are intrinsic to the tank building and welding processes and will vary randomly across the population of tanks. Additionally, pressure also varies across multiple pressurized tanks with different known gas properties, providing evidence that pressure uncertainty is aleatory.

Although classifying all uncertainties as aleatory is reasonable in this example, weld yield stress will be classified as epistemic to illustrate how to do the analysis with both types of uncertainty.

Table 3.1 gives the classification for each uncertain input.

In practice, variables need not be entirely epistemic or aleatory. For example, weld yield stress could be modeled as both aleatory and epistemic, with an aleatory component that represents the uncertainty across a population of welds, and an epistemic component that represents the uncertainty in the parameters of the weld yield stress distribution. The epistemic component can additionally include uncertainty in the distribution for weld yield stress. There can be uncertainty in this distribution as it is often chosen empirically from limited measured data with statistical distribution fitting techniques (see Section 4).

If the analyst was only interested in calculating the probability of failure for a single weld, rather than the probability of failure in a population of welds, then classifying these uncertainties as epistemic would be appropriate. This is because a single weld has one set of true, but unknown, values for these parameters and the goal is to assess the probability for a single weld. Collecting more information (e.g., measured data) on these quantities for the single weld would reduce the uncertainty.

Table 3.1. Classification of Inputs as Aleatory or Epistemic

Input	Classification
Pressure, ρ	Aleatory
Radius, r	Aleatory
Wall Thickness, t	Aleatory
Angle of Helical Weld, α	Aleatory
Weld Yield Stress, Y	Epistemic

2. Determine sample size (and sampling scheme):

Computational costs will be more prevalent when uncertainties are separated. In this case, a “double-loop” (also referred to as “nested loop”) sampling procedure is performed:

- i. A set of epistemic variables is sampled randomly from their probability distributions. If epistemic uncertainty is not defined probabilistically, a discrete worst-case value can be used instead.
- ii. Run a fixed epistemic set with n_a samples of the aleatory variables through a model in order to estimate the failure probability across the aleatory samples. Here, n_a is the aleatory sample size.
- iii. Steps 1 and 2 are repeated as many times as the epistemic sample size (n_e). The separation of the results by the epistemic uncertainty is maintained. It is recommended that the random seed be changed with each epistemic sample.

This will require a total of $n_a \times n_e$ realizations. The aleatory sample size (n_a) must be large with respect to the probability of weld failure and the epistemic sample size (n_e) must be chosen so that there is a strong degree of confidence in the estimation. In practice, the final choice of sample size is generally determined using iterative convergence analyses. More information on convergence analyses can be found in Section 9.

With these considerations, the following sample sizes were used with IS on pressure and weld yield stress:

- Aleatory sample size, $n_a = 4 \times 10^4$
- Epistemic sample size, $n_e = 1 \times 10^3$

3. Use a double-loop process to simulate the distribution of the failure probability:

The double-loop procedure (as described in Step 1) is implemented as follows:

- i. 1×10^3 epistemic samples of weld yield stress are generated by IS using the importance distribution defined in Section 8.2.
- ii. For each sample, 4×10^4 sample sets of the aleatory variables (cylinder radius, thickness, weld angle and pressure) are used to estimate the distribution of the shear stress calculated in the weld. Using IS, pressure is sampled using the distribution defined in Section 8.2.

- iii. The shear stress distribution is compared to the weld yield stress to estimate the probability of weld failure. This results in 1×10^3 estimates of the probability of failure, where each epistemic realization represents the epistemic uncertainty in the failure probability. The distribution of these estimates is shown in Figure 3.2.

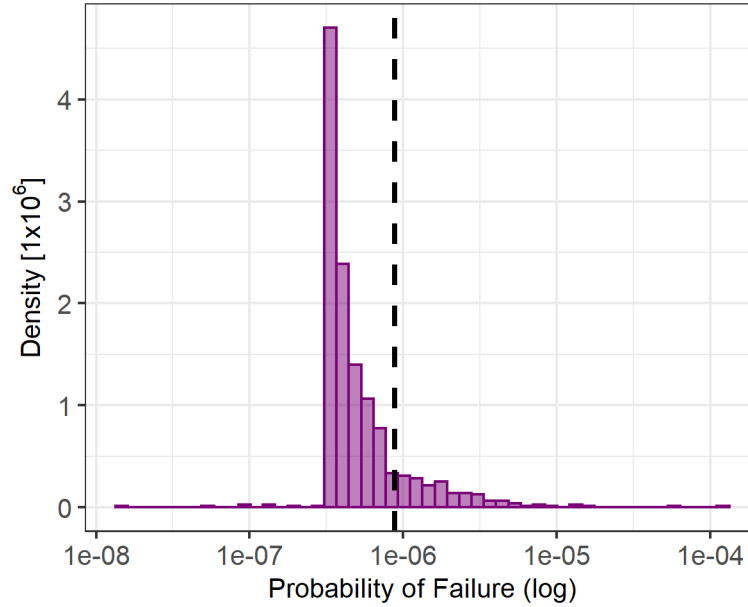


Figure 3.2. The estimated probability of failure across 1000 epistemic samples. The dashed line represents the mean estimate.

3.2.2. Results and potential difficulties:

The mean estimate of the probability of failure is 8.2×10^{-7} , approximately the same as in the case when aleatory and epistemic uncertainties were not separated (8.3×10^{-7}), and still relatively close to the true failure probability (8.8×10^{-7}). However, interpretation of the confidence in that estimate has changed. Even though the mean estimate is below the threshold of 1×10^{-6} also considered in Section 3.1, 13.5% of the epistemic samples had estimated failure probabilities that were greater than 1×10^{-6} . The 95th percentile of the probability estimates is 1.99×10^{-6} , which can be considered an upper 95% epistemic confidence bound on the failure probability.

When uncertainties are separated into aleatory and epistemic, confidence in inferences about the failure probability can change [2]. These insights cannot be seen if uncertainties are not separated before the simulation is performed.

While the 95% bound does not include epistemic sampling uncertainty since the epistemic sample set was held fixed, it does include two other types of uncertainty: the uncertainty due to the epistemic variables, and sampling uncertainty from the aleatory variables. It is important that the aleatory sample size be large enough to minimize the aleatory sampling uncertainty; otherwise, the two types of uncertainties will be confounded.

In the above example, all uncertainties associated with shear stress are aleatory. Therefore, the distribution of shear stress does not change across epistemic samples, as shown in the left plot of Figure 3.3. However, if pressure had been classified as epistemic instead of aleatory, significantly different distributions for resulting shear stress would be generated for each epistemic sample, as seen

in the right plot of Figure 3.3. These results emphasize two points. First that, the characterization of aleatory and epistemic uncertainties can potentially affect the results of an analysis, and second, when epistemic uncertainty is large, estimates of aleatory distributions conditioned on epistemic variables can vary significantly.

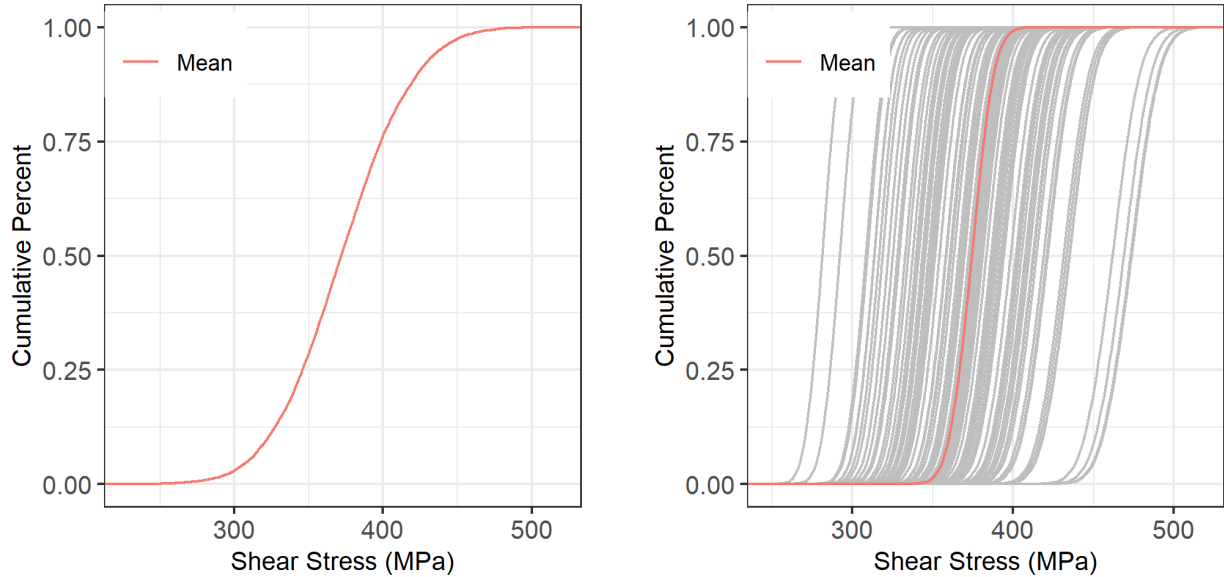


Figure 3.3. 100 epistemic empirical CDFs of shear stress when pressure uncertainty is aleatory (left) and epistemic (right).

3.2.3. Summary:

- Characterization of uncertainty types depends on both the goal of the analysis and the analyst's domain knowledge.
- The double-looping procedure requires a larger sample size than when uncertainties are not separated.
- The classification of aleatory and epistemic will not change the point estimate of the failure probability (as long as the sample size is large enough), but it will affect the interpretation of the uncertainty associated with the failure probability.
- When epistemic uncertainty is large, estimates of the failure probability for different epistemic values can vary significantly.

4. STATISTICAL DISTRIBUTION FITTING

This section applies methods for statistical distribution fitting (NUREG/CR-7278, Section 4.2.1) to the pressurized tank example. Two concepts are illustrated:

1. Estimating an input distribution from measured data.
2. Restricting the range of an input through distribution truncation.

When to apply this method. Statistical distribution fitting should occur in a PFM analysis after a list of uncertain inputs has been established and those inputs have been classified as aleatory or epistemic (Section 3). Once distributions are fit to the data, they can be used in the forward propagation of uncertainty (Section 8).

4.1. Data-driven distribution estimation

This example illustrates the process of estimating the probability distribution of weld yield stress from available measured data.

4.1.1. Process:

There are 5 steps in input distribution specification:

1. Determine relevant data.

Suppose there are 75 experimental measurements of weld yield stress in MPa (Table 4.1):

Table 4.1. Experimental Measurements of Weld Yield Stress

Measurement	Weld Yield Stress [MPa]	Measurement	Weld Yield Stress [MPa]	Measurement	Weld Yield Stress [MPa]
1	608.64	26	593.23	51	606.33
2	586.67	27	600.12	52	599.93
3	587.70	28	602.95	53	612.45
4	618.17	29	605.02	54	614.56
5	607.84	30	582.91	55	581.05
6	597.89	31	612.83	56	604.24
7	590.48	32	598.86	57	606.36
8	604.67	33	562.70	58	598.88
9	596.59	34	614.95	59	604.53
10	583.76	35	575.00	60	588.70
11	613.08	36	618.82	61	593.54
12	593.97	37	602.99	62	601.27
13	589.40	38	594.36	63	614.19
14	598.95	39	597.73	64	597.00

Measurement	Weld Yield Stress [MPa]	Measurement	Weld Yield Stress [MPa]	Measurement	Weld Yield Stress [MPa]
15	600.38	40	597.91	65	601.50
16	600.55	41	600.23	66	616.28
17	614.83	42	592.72	67	581.08
18	580.26	43	596.97	68	593.01
19	550.59	44	537.39	69	595.47
20	578.81	45	582.17	70	581.57
21	604.76	46	602.87	71	601.59
22	600.94	47	597.86	72	608.42
23	562.31	48	601.44	73	590.71
24	595.49	49	601.31	74	608.50
25	591.08	50	574.12	75	603.58

For the data to be relevant, these measurements should be collected from a representative sample of relevant welds. Furthermore, testing should not introduce any measurement bias in the data. Specifically, the weld yield stress measurements should not be systematically biased high or low; if there is bias, measurements should be first calibrated to remove the bias. If there is random measurement noise in the data (not systematic), then this measurement error can be accounted for in the analysis.

If the data are not sufficiently relevant, expert judgment can be used to determine how the input distribution should be adjusted to account for any biases in the data.

2. Select candidate probability distributions.

For illustrative purposes, the normal, lognormal, and Weibull distributions are considered as candidate probability distributions. These distributions are characterized by a shape and scale parameter. More flexible distributions can be considered.

3. Fit the distributions to the data.

Estimation: The parameters for the distributions are estimated using maximum likelihood estimation, which is an optimization procedure for finding the most likely values of the model parameters given the available data. For the normal and lognormal distributions, these parameters are a special case of simply the sample mean and standard deviation of the original and log-transformed data, respectively.

Uncertainty quantification: Because only 75 data points representing a sample from the population are available, the estimated distribution parameters will contain uncertainty. The statistical bootstrap (NUREG/CR-7278 Section 4.3.7) can be used to characterize sampling uncertainty in the distribution parameters (other methods for quantifying sampling uncertainty could have been applied, such as the asymptotic standard error based on maximum likelihood estimation). One-sided 90% confidence bounds are calculated for the parameters, calculating sampling uncertainty in the “conservative direction”. Specifically, for the normal and lognormal distributions, the probability of failure increases

with the standard deviation and decreases with the mean. For the Weibull distribution, the probability of failure decreases with both the shape and scale parameters.

Parameter estimates using maximum likelihood estimation and uncertainties using statistical bootstrapping are displayed in Table 4.2.

Table 4.2. Parameter Estimates and Sampling Uncertainties

Distribution	Parameter	Point-estimate	Parameter uncertainty One-sided 90% confidence bound
Normal	Mean	596.4	594.0 (lower bound)
	Standard deviation	12.5	16.3 (upper bound)
Lognormal	Log-mean	6.4	6.4 (lower bound)
	Log-standard deviation	0.02	0.03 (upper bound)
Weibull	Shape	56.7	49.6 (lower bound)
	Scale	602.3	600.7 (lower bound)

4. Evaluate the fit of the distributions to the data.

QQ plots (Figure 4.1) are used to graphically evaluate the statistical distribution fit. A straight line between the empirical and model-based quantiles provides evidence that the model fits the data. A two-sided 95% confidence interval [3] around the line is also plotted as the shaded region to graphically visualize how much sampling uncertainty in the quantile comparisons is expected when $n = 75$. There is evidence of a deviation from the straight line for both the normal and lognormal fitted models. While the QQ plots for the normal and lognormal fitted models look similar, the two confidence intervals on the normal and lognormal QQ plots have several points that fall outside the interval. The Weibull model appears to follow the straight line more closely, though due to the limited sample size, it is difficult to assess the distributional fit in the tails of the distribution.

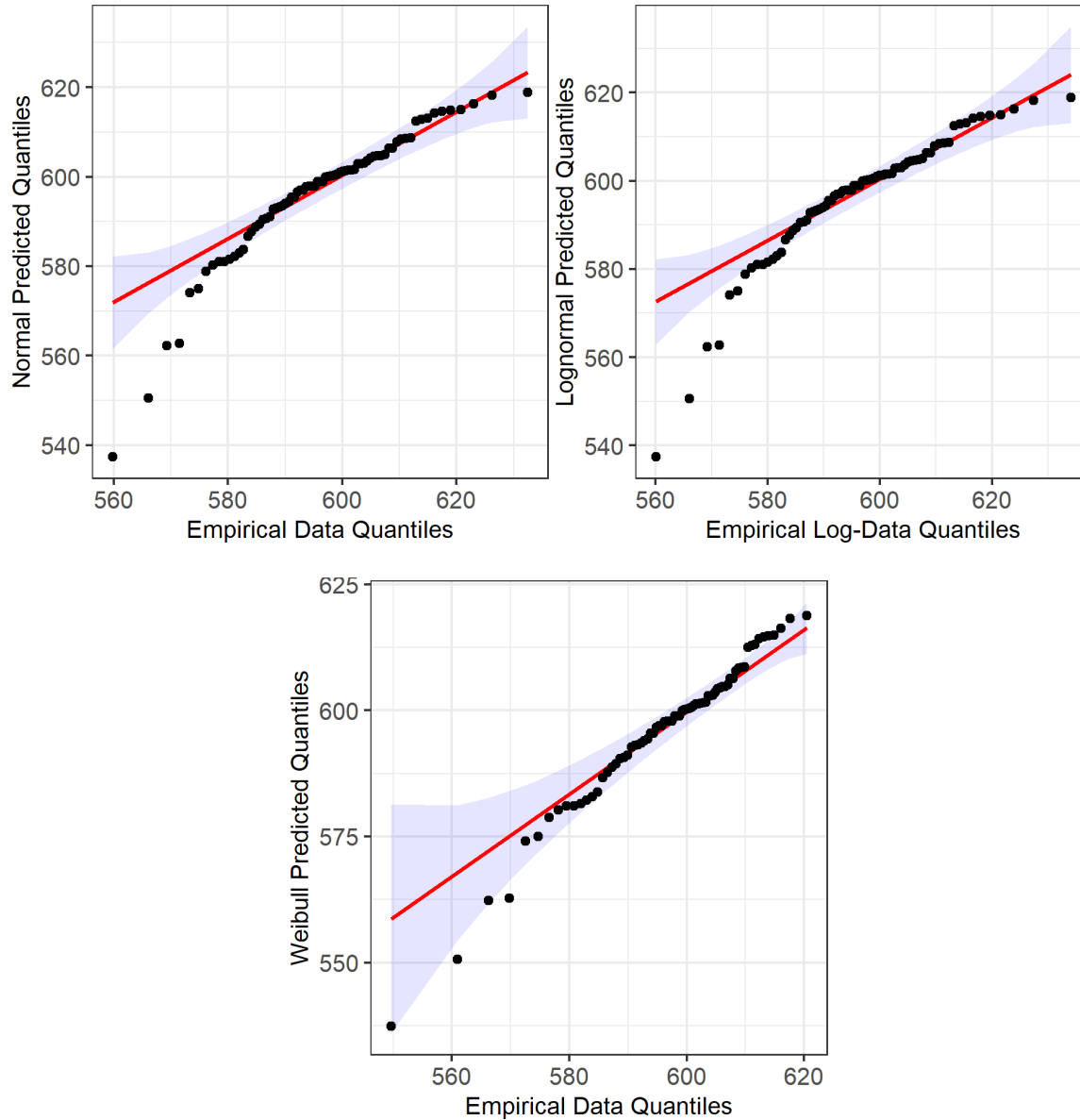


Figure 4.1. QQ plots for evaluating distributional fit. (Top left) normal distribution; (top right) lognormal distribution; and (bottom) Weibull distribution.

Additionally, the Anderson-Darling statistical goodness of fit test [4] suggests that there is enough evidence in the data to conclude a lack of fit at the 10% significance level for the log-normal distribution. This is consistent with the results in the QQ plots. Unfortunately, the nature of goodness of fit tests makes them unable to positively identify a distribution, but they can help to eliminate poor-fitting distributions from consideration. In this example, the Anderson-Darling test eliminated the lognormal distribution.

5. Select a final input distribution model.

The Weibull distribution produces a reasonable fit to the data, but the sample size is limited. Therefore, the final input distribution should be supported using subject-matter judgment. For example:

It is possible for the distribution of weld yield stress to be left-skewed. Yield strength in steels (all metals) depends on its microstructural characteristics and is therefore considered a structure sensitive property. For a given class of steel (grade, composition, heat treatment, grainsize), there is an upper bound on its ability to work harden, resulting in yield data with left-skewness. Other types of skew could also exist in a population. During mill inspections, material that does not fulfill a specified strength generally is detected in routine control tests and is not included in the sample. The Weibull distribution is the only candidate distribution that allows for left-skewness; therefore, the Weibull distribution is selected as most appropriate.

There is no additional expert judgment about the location and scale parameter values. Therefore, one option is to apply a confidence bound on the parameters (e.g. 90% one-sided confidence bounds (Figure 4.2)) that represents parameter sampling uncertainty. Using a higher confidence level would result in more conservatism, and the selection of the confidence level is somewhat arbitrary based on the desired level of conservatism and amount of available domain knowledge. Therefore, the final model selection should use expert judgment to ensure the selected distribution adequately represents uncertainty.

Note that there are many other ways to incorporate parameter sampling uncertainty. For instance, sampling uncertainty can be considered as a source of epistemic uncertainty and looped over in the analysis.

To adequately represent the uncertainty, the final input distribution for weld yield stress is $Y \sim \text{Weibull}(\lambda = 600, k = 50)$. This corresponds to the 90% one-sided confidence bound on the Weibull parameters.

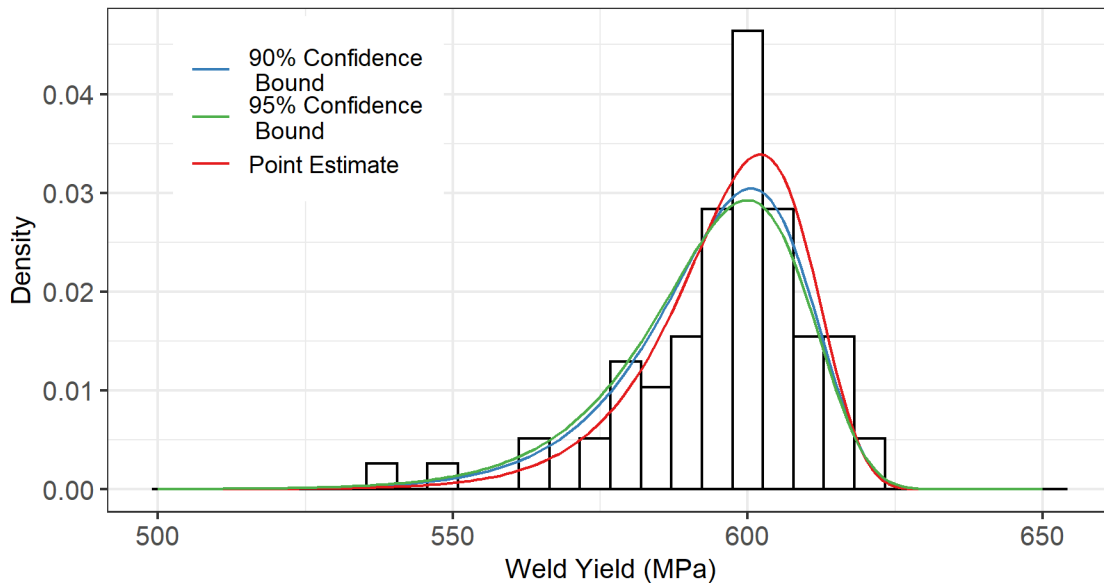


Figure 4.2. Weibull distributions on weld yield stress defined by different one-sided confidence levels on the Weibull parameters.

4.1.2. Results and potential difficulties:

Input distribution estimation can be inaccurate with limited data because:

- The exact form of the input distribution cannot be identified with limited data.
- Even if the distributional form is correct, parameter uncertainty can be underestimated.
- It is difficult to characterize the tails of the distribution, and when estimating failure probabilities, the tails can drive the results.

Therefore, it is important to consider the above sources of uncertainty when estimating input distributions; furthermore, the analyst should be cognizant that excessive conservatism is not a prudent solution and can produce inaccurate results. Accurately capturing the tails of probability distributions with limited data requires expert judgment and cannot be a solely data-driven activity.

The same process described above is performed for pressure and thickness and their final distributions are:

$$\rho \sim N(\mu = 50, \sigma = 6)$$

$$t \sim \text{Beta}(\alpha = 2, \beta = 5, LB = 0.018, UB = 0.02)$$

4.1.3. Summary:

- Evaluating distributional fits using limited data cannot be solely data-driven and requires expert judgment, particularly when the tails of the distribution are important to the analysis results.
- Failure probability calculations are sensitive to the tails of the input distributions. Sensitivity studies should be applied to important inputs when there is uncertainty about the probability distribution.
- The ‘best-estimate’ of an input distribution should account for sampling uncertainty when data are limited. Sampling uncertainty in probability distribution parameters can increase the input parameter uncertainty.

4.2. Distribution truncation

This example illustrates the process of restricting the range of an input through truncation and potential difficulties.

4.2.1. Process:

There are several steps in specifying truncated distributions.

1. Determine the range of the distribution.

Suppose that any cylinders with radius less than 0.564 m or greater than 0.636 m were scrapped during manufacturing, such that r is in $[0.564, 0.636]$ m. Modeling $r \sim N(0.6, 0.012)$ is not appropriate as the normal distribution can take on values $(-\infty, \infty)$, but r is restricted to the range $[0.564, 0.636]$ m.

2. Select a distributional form given the distributional range.

The new distribution should be restricted to the range [0.564, 0.636] m. Two candidate distributions are a shifted and scaled beta distribution or a truncated normal distribution (Figure 4.3). The beta distribution is a continuous distribution with a domain of finite measure that can take on a variety of shapes, and therefore is a good candidate for modeling inputs with a bounded range. The truncated normal distribution is a normal distribution with restricted upper and/or lower bounds. When truncation occurs, the distribution is normalized so that the area of the non-truncated region integrates to one [5].

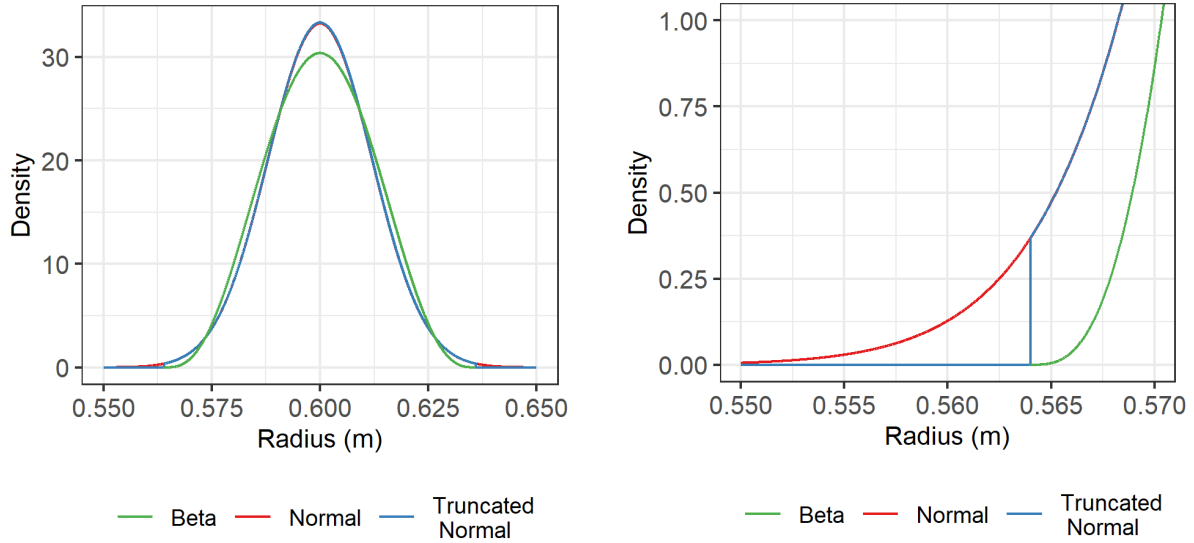


Figure 4.3. Different distributions with a domain of finite measure, relative to the original normal distribution. The right figure is zoomed into the bottom tails of the distributions.

Because the data were generated by scrapping parts with radii that were outside the valid range, the truncated normal distribution is likely more appropriate for this application. The uncertainty in the radius can be estimated as:

$$r \sim TN(\mu = 0.6, \sigma = 0.012, LB = 0.564, UB = 0.636)$$

where the mean is 0.6, standard deviation is 0.012, and upper and lower truncation limits are 0.564 and 0.636, respectively.

4.2.2. Results and potential difficulties:

The first row of Table 4.3 shows the analysis results for radius with and without truncation using the limits that were previously set. In this case, the truncation does not remove a significant amount of area under the normal distribution, so the estimated failure probability is similar with and without truncation.

Now, consider a case where the lower and upper bounds of the radius were 0.590 and 0.610. Performing the same analysis gives the results in the second row of Table 4.3. With these new limits, more area under the curve is truncated, resulting in a failure probability that is smaller. If the truncated distribution was the correct distribution, then the failure probability would be over-estimated by failing to truncate.

Table 4.3. Impact of Truncation of r on Failure Probability

Bounds (m)	With Truncation	Without Truncation
[0.564, 0.636]	8.6×10^{-7}	8.8×10^{-7}
[0.590, 0.610]	6.4×10^{-7}	8.8×10^{-7}

A similar process is performed for the angle of the helical weld and its distribution is chosen as:

$$\alpha \sim TN(\mu = 55, \sigma = 1.1, LB = 51.7, UB = 58.3)$$

4.2.3. Summary:

- Identifying limits on important inputs can help prevent non-physical realizations and can result in more appropriate failure probability estimates.

5. MODELING DEPENDENCE BETWEEN INPUTS

This section applies methods for modeling dependence between inputs (NUREG/CR-7278, Section 4.2.2) to the pressurized tank example. Two concepts are illustrated:

1. Induce dependence between inputs using rank correlation.
2. Importance sampling with correlated inputs.

When to apply this method. In PFM analyses, some uncertain model inputs may be correlated. For instance, the inner and outer diameter of a pipe may be uncertain, and these diameters will be positively correlated because, by definition, the inner diameter must be smaller than the outer diameter. When propagating input uncertainty through the model (Section 8), these correlations must be properly accounted for to obtain valid uncertainties on the model outputs.

5.1. Induce dependence between inputs using rank correlation

Using the welded pressure vessel, this example illustrates the process of incorporating dependence between inputs by inducing correlation on the ranks of the inputs [6]. This method can be used with both simple random sampling (SRS) and Latin hypercube sampling (LHS).

Most input variables in a PFM analysis will be uncorrelated, such that changes in one input do not affect the values of other inputs. However, some inputs will exhibit either positive or negative correlations. Positive correlations occur when one variable increases and another increases; negative correlations occur when one variable increases and another decreases. Measures of correlation take on values between -1 and 1, where -1 and 1 represent perfect negative and positive correlation between two variables, respectively. Perfect correlation occurs when there is a deterministic linear relationship between two variables. Uncorrelated variables have a correlation of 0.

The pressurized tank example is modified slightly in this section to induce correlation between two variables for pedagogical reasons. Specifically, suppose that tanks with a smaller radius, r , have higher internal pressure, p . That is, the radius and pressure inputs are negatively correlated.

5.1.1. Process:

There are four primary steps for inducing dependence via the rank correlation method:

1. **Sample the full set of realizations for all inputs independently from their respective uncertainty distributions, initially ignoring correlation.** These independent inputs are later transformed to induce correlation.
2. **Specify correlation between the ranks of the inputs.** In this method, correlation is specified on the ranks of the inputs, rather than on the inputs themselves. The rank of a continuous input is simply an integer-value representing the order of all of its values. Assume in this example that the rank correlation between radius and pressure is -0.9 (this correlation value induces a strong negative correlation between these inputs). The correlation matrix C is the matrix defining the correlations between input ranks. As an example, for the ranks of radius and pressure, this matrix is $C = \begin{pmatrix} 1 & -0.9 \\ -0.9 & 1 \end{pmatrix}$. The diagonal of the matrix is 1, representing the fact that the correlation

of an input with itself is 1; the off-diagonal elements of the matrix correspond to the rank correlation between radius and pressure.

3. Transform the independent samples to induce correlation.

The premise of the rank correlation method is that if correlation is induced between the ranks of the inputs, then the inputs are reordered to preserve this rank correlation. For more details and methodological extensions, see [6].

Specific steps are:

- i. Calculate the Cholesky decomposition P of the rank correlation matrix C , such that $PP^T = C$.
- ii. Construct a matrix R with columns comprised of scores, defined here as a random permutation of $\Phi^{-1}[i/(n+1)]$ for $i = 1, \dots, n$, where n is the total number of samples and Φ is the cumulative distribution function for the normal distribution. For more information about score selection, see [6].
- iii. Calculate $R^* = RP^T$. The correlation matrix for R^* will be close to C .
- iv. Reorder the columns of the independently sampled inputs to have the same ordering as the columns of R^* .

4. Verify correlation between dependent inputs.

To verify that the correlation method has worked as intended, the correlation between the dependent inputs, as well as the correlation between the ranks of the dependent inputs, can be calculated after applying the rank transformation. In this example, the correlation can be estimated using the standard formula for correlation:

$$\hat{r} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 (y_i - \bar{y})^2}}$$

where \bar{x}, \bar{y} are the sample means of variables x and y . In the example, the correlation between the ranks of pressure and radius should be approximately -0.9 and the correlation between the raw values of pressure and radius should be negative.

5. Propagate uncertainty through the model using the dependent inputs.

The distribution of weld shear stress and the corresponding failure probability can be estimated based on the sampled dependent inputs.

5.1.2. Results and potential difficulties:

The rank correlation method is used to induce correlation between r and ρ . The distributions of the individual inputs (i.e., marginal distributions) are the same in the correlated and uncorrelated samples, but the joint distribution clearly differs and reflects the negative correlation using the rank correlation method (Figure 5.1). The estimated empirical correlation between the radius and pressure after applying the rank correlation method is -0.90 and the estimated correlation between the ranks is -0.89, suggesting the method worked as intended to induce a negative correlation of -0.9 between the ranks of the inputs.

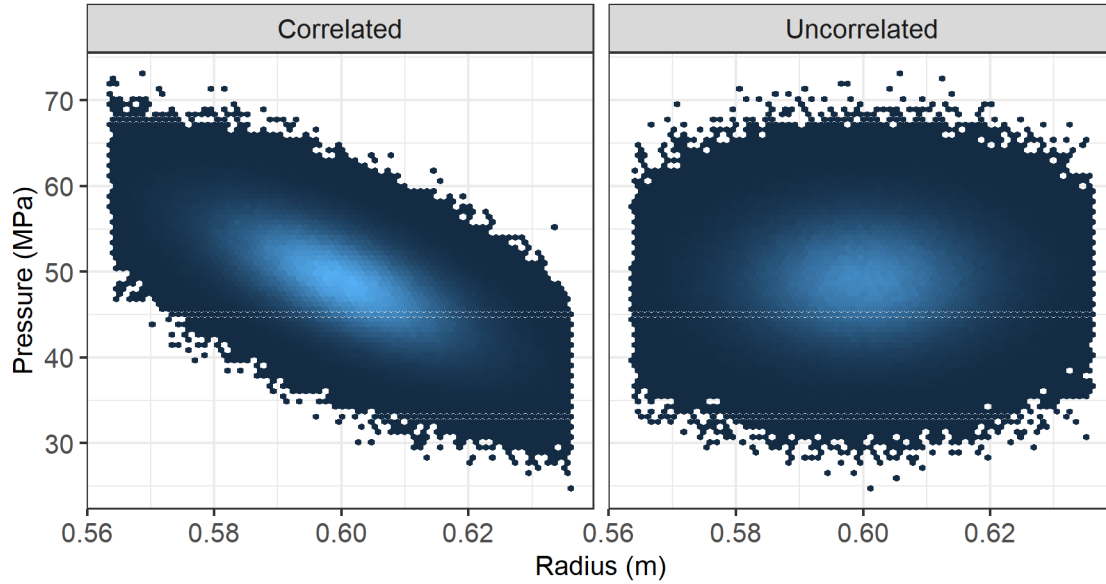


Figure 5.1. (Left) Correlated dependent inputs generated using the rank correlation method; (right) input distributions without correlation.

When propagated through the model, the correlated inputs result in a narrower shear stress distribution as compared to uncorrelated inputs (Figure 5.2). Given the relationship between shear stress, pressure, and radius, this change in distribution is expected. Increasing both radius and pressure increases shear stress. When radius and pressure are negatively correlated, extreme values of shear stress are less likely because high values of pressure are accompanied by low values of radius. Therefore, the probability of failure decreases with this correlation; the estimated probability of failure is $\sim 3 \times 10^{-8}$ after inducing correlation, compared to $\sim 6 \times 10^{-7}$ with independent samples. While the probability of failure clearly decreases, the difference between the distributions of shear stress using correlated and uncorrelated samples is difficult to discern because uncertainty in the radius has little impact on the shear stress distribution (due to the narrow range of uncertainty in the radius distribution).

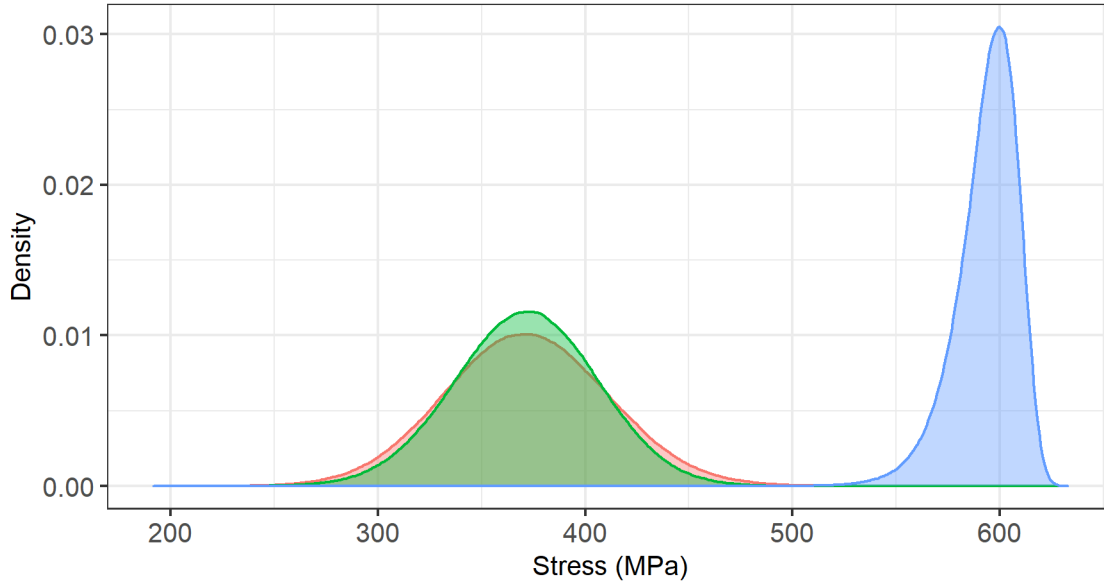


Figure 5.2. Distribution of shear stress based on uncorrelated samples (red) and correlated samples (green); the distribution of yield stress is shown in blue for reference.

If the radius had more uncertainty, accounting for dependence between inputs would have a larger impact on the estimated failure probability. To illustrate this point, the distribution of the radius is changed from:

$$r \sim TN(\mu = 0.6, \sigma = 0.012, LB = 0.564, UB = 0.636)$$

to:

$$r \sim TN(\mu = 0.6, \sigma = 0.05, LB = 0.5, UB = 0.7)$$

With more uncertainty in the radius, the dependence between the inputs results in a much narrower stress distribution than the stress distribution without accounting for input dependence (Figure 5.3).

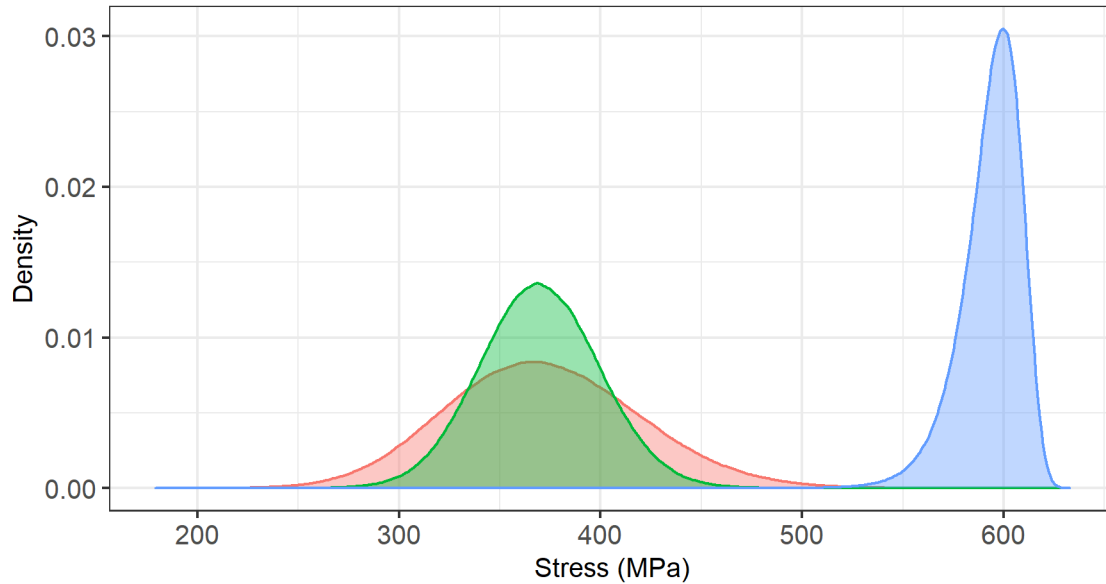


Figure 5.3. Distribution of shear stress based on uncorrelated samples (red) and correlated samples (green) with more uncertainty in the radius; the distribution of yield stress is shown in blue for reference.

5.1.3. Summary:

- Accounting for dependence between inputs using rank correlation is straightforward for many sampling schemes and can substantively impact final analysis results.
- Failing to include correlation between inputs that should be correlated can be misleading

5.2. Importance sampling with correlated inputs

In the pressurized tank problem, IS on pressure can improve failure probability estimation (see Section 8.2 for more information). Recall that IS is a variance reduction technique where the input sampling distribution is changed to focus on the failure region and samples are weighted accordingly.

If pressure is correlated with the radius of the cylinder, as in the previous example, then the correlation between the pressure and radius should be considered when using IS. The rank correlation method from the previous concept cannot be applied under IS because the inputs are sampled from the importance distribution, rather than the marginal distributions of the inputs. Inducing correlation between inputs sampled from their importance distribution would not necessarily produce the correct correlation structure on the target distribution.

In this example, pressure and radius are sampled using IS while accounting for their correlation. To use IS with dependent inputs, the joint distribution of the inputs must first be identified, unlike in the rank correlation method.

5.2.1. Process:

There are three primary steps for inducing dependence when inputs are sampled using IS:

1. **Identify the joint distribution of the correlated inputs that will be importance sampled.**

In this case, pressure and radius are sampled from a joint truncated multivariate normal distribution with correlation -0.9. The joint truncated multivariate normal distribution is a generalization of the univariate normal distribution for multiple dimensions. It is formed using the normal marginal distribution for pressure and the truncated normal marginal distribution for radius. This distribution allows for the specification of the correlation between the inputs. Figure 5.4 shows a notional example of a highly negatively correlated bivariate normal distribution between two inputs.

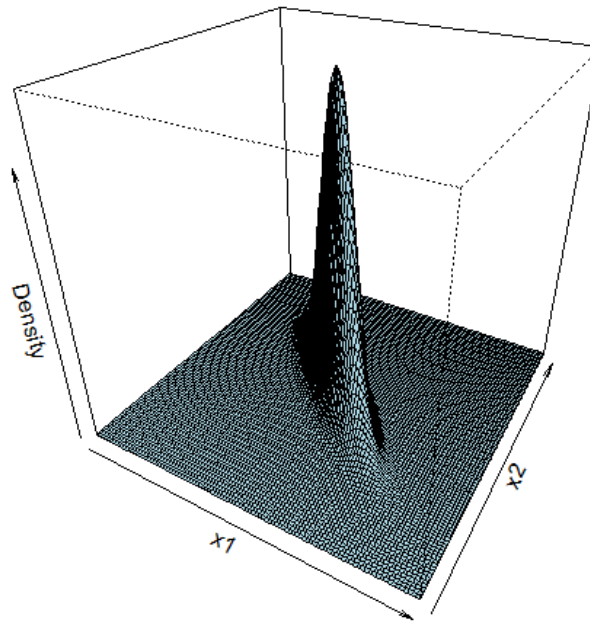


Figure 5.4 Notional example of a highly negatively correlated bivariate normal joint distribution between two variables, x_1 and x_2 .

2. **Identify an importance distribution for the correlated inputs.**

The selection of the importance distribution should consider both the importance region and the correlation between inputs. Failing to include correlation between inputs in the importance distribution could decrease efficiency of failure probability estimation.

The selected importance distribution in this example is again a joint truncated multivariate normal distribution with correlation -0.9. The mean values of the pressure and radius distributions are incremented by +3 and -3 standard deviations, respectively; and the standard deviations of both pressure and radius are increased by a factor of 1.5. These changes shift the input distribution

toward the important failure region while preserving the relationship between pressure and radius (Figure 5.5).

Note that the importance weights are now constructed based on the joint distribution of radius and pressure before and after IS. Specifically, define $p(r_i, \rho_i)$ and $q(r_i, \rho_i)$ as the joint probability density functions for the target distribution and importance distribution, respectively, for radius and pressure evaluated at r_i, ρ_i . Then, the importance weights are:

$$w_i = \frac{p(r_i, \rho_i)}{q(r_i, \rho_i)}$$

3. Sample inputs and propagate uncertainty through the model.

For more information about estimating a failure probability based on an importance sample, see Section 8.2.

5.2.2. Results and potential difficulties:

The original target and importance sampled distributions for pressure and radius are plotted in Figure 5.5. Note that the importance distribution covers the failure region (high pressures) more densely than the original distribution. Additionally, the importance distribution preserves the correlation between radius and pressure. Using 10^5 importance samples (IS on pressure and radius, as well as weld yield stress), the estimated failure probability is 8×10^{-8} , an order of magnitude lower than the estimated failure probability without dependent inputs (8×10^{-7}).

This example considers IS with only two correlated inputs; specifying an importance distribution that increases efficiency in failure probability estimation becomes more difficult as the number of correlated inputs increases.

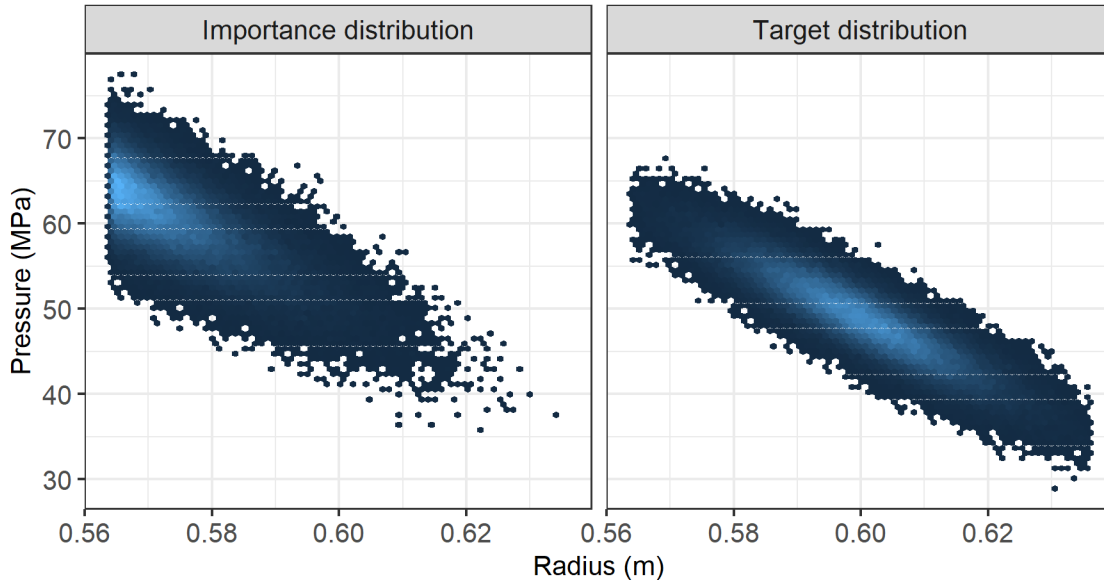


Figure 5.5. Distribution of correlated inputs radius and pressure when inputs are sampled from the importance distribution (left) and the target distribution (right).

5.2.3. Summary:

- When dependent inputs are importance sampled, the joint distribution of the inputs must be considered when determining how to importance sample.

6. SENSITIVITY ANALYSIS

This section applies methods for analyzing sensitivity (NUREG/CR-7278, Section 4.2.3 and 4.2.4) to the pressurized tank example. Four concepts are illustrated:

1. Choose a relevant response on which to conduct sensitivity analysis.
2. Use exploratory data analysis to visualize input/output relationships.
3. Estimate local sensitivity metrics.
4. Estimate global sensitivity metrics.

When to apply this method. Sensitivity analysis (SA) helps identify which uncertain inputs explain a high proportion of the uncertainty in the output. SA is typically performed iteratively with the forward propagation of uncertainty (Section 8). SA can take place after an initial set of inputs have been propagated through the model and the results of the SA may be used to determine an updated sampling scheme before assessing for convergence (Section 9).

6.1. Choose a relevant response on which to conduct sensitivity analysis

Prior to starting a sensitivity analysis, it is important to decide whether the failure probability or a model output will be used in the analysis.

6.1.1. Process:

Determine which output(s) to use for sensitivity analysis

In the pressurized tank example, the model returns an indicator value that is equal to 1 if shear stress is larger than weld yield stress and 0 otherwise. Using this indicator for sensitivity analysis poses two problems:

- The probability of failure is rare (i.e., most indicator values will be equal to 0), so many samples will be needed to quantify the input/output relationship.
- Binary/categorical responses in regression inherently contain less information than continuous responses. Often a binary response is a function of continuous variables. These continuous variables can provide better information on directional effects with fewer samples.

Because of these issues, it is beneficial to use a continuous variable for sensitivity analysis when possible.

6.1.2. Results and potential difficulties:

In this example, shear stress will be used because it is used in the calculation of the indicator value, and it is known that higher values of shear stress result in more failures. Note that weld yield stress is not considered in the sensitivity analysis as it is clear in this example problem that weld yield stress has a large impact on the probability of failure.

There could be cases where using a continuous response in place of an indicator can pose problems. For example, if there are conjoint influences on the indicator, there may not be a single continuous

response that is sufficiently correlated with the indicator. In that case, the indicator would need to be used for the sensitivity analysis and a larger sample size would likely be required.

6.1.3. Summary

- A model output may be better suited for sensitivity analysis than a failure probability.

6.2. Use exploratory data analysis to visualize input/output relationships

A scatterplot is a visualization tool that can be used to understand the relationship between model inputs and outputs. They can be used as a first step to assess the nature and magnitude of the input/output relationships.

6.2.1. Process:

1. Determine which inputs and outputs should be plotted

The choice of which inputs and outputs should be plotted depends on the number of each:

- If there are a small number of input and output variables, scatterplots can be produced for each combination of input and output. In this problem, there are only 4 uncertain inputs and 1 output, so scatterplots are produced for each combination.
- With many inputs/outputs, visualizations can be selectively chosen based on subject matter knowledge or by first calculating sensitivity metrics to get an estimate of which input/output relationships are most important.

2. Sample from the uncertain inputs and calculate the output

- 1×10^2 simple random sample (SRS) values of pressure, radius, wall thickness, and angle of the helical weld are sampled from the distributions defined in the Problem Statement (Section 1) and the model is applied to generate a distribution of shear stress. Note that the goal here is to understand the input/output relationship. With a continuous outcome (shear stress), fewer samples are generally needed to estimate this relationship as compared to estimating a rare probability.

3. Generate scatterplots and assess the input/output relationships

Scatterplots can be used for many purposes, including:

- **Determining whether relationships are roughly linear, monotonic (i.e., entirely increasing or decreasing), or more complex.** This can help hypothesize statistical models that may be used to generate the global sensitivity metrics. For example, if the scatterplots show that the input/output relationships are roughly linear, a simple linear regression may be appropriate for calculating the metrics. However, if more complex relationships are present, more flexible models may be required.
- **Identifying two-way interactions between inputs.** A two-way interaction is present when the effect one dependent variable has on the response depends on the value of another dependent variable. An option for generating a two-way interaction with continuous inputs is to create a heat map with the inputs along the x and y axis, and the output representing the heat of the map. When

there are many uncertain inputs, subject matter expertise can be leveraged to select candidate inputs to plot.

- **Confirming results that are generated during a global sensitivity analysis.** If possible, results from a global sensitivity analysis should be confirmed to be as expected visually via scatterplots. For example, if a visual trend is present in a scatterplot, it will likely indicate that the output is sensitive to that input.

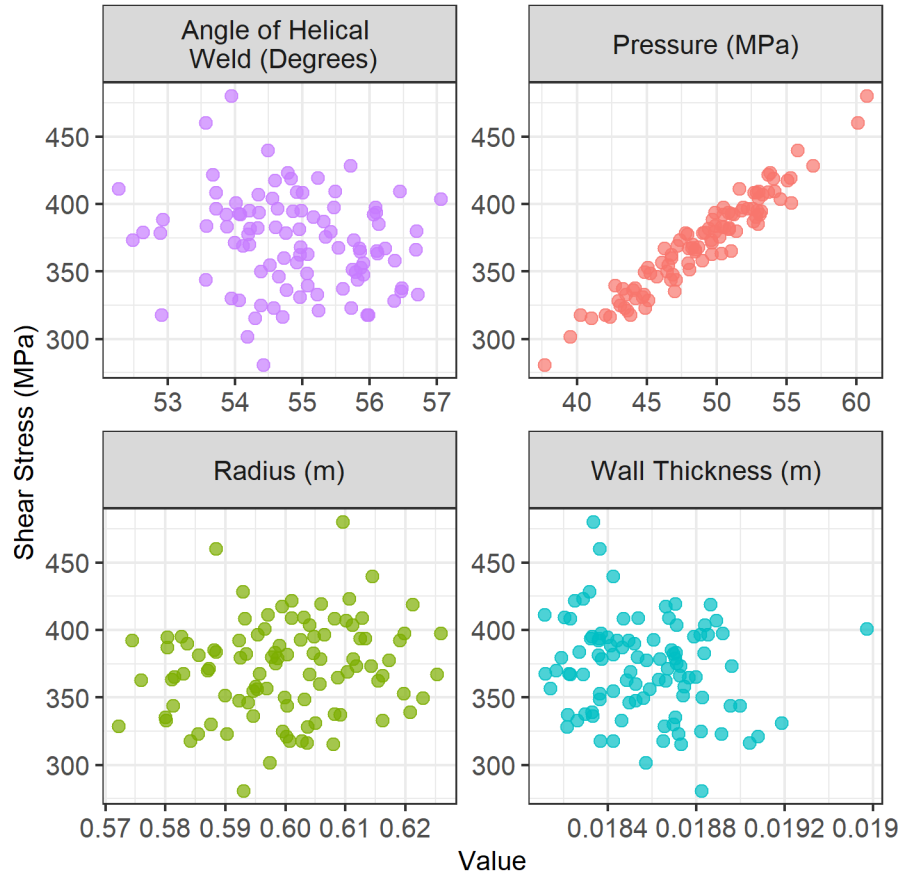


Figure 6.1. Scatterplots for shear stress versus pressure, radius, weld thickness, and angle of the helical weld.

6.2.2. Results and potential difficulties:

Figure 6.1 shows the scatterplots for shear stress versus pressure, radius, wall thickness, and angle of the helical weld. Here, it appears that pressure has a strong, approximately linear relationship with shear stress, while radius, wall thickness, and angle of helical weld do not have clear relationships with the output.

The main difficulties in sensitivity analysis are failing to visualize important relationships and misinterpreting the visualizations. More complex relationships, such as interactions amongst the input variables, will usually not be obvious in univariate scatterplots such as those in Figure 6.1. Some relationships are inherently more difficult to visualize, and often statistical modelling is needed to understand if a relationship exist. Estimating global sensitivity metrics (Section 6.4), often relies on statistical modeling.

6.2.3. Summary

- Scatterplots are a simple way to understand the nature and magnitude of the relationship between uncertain inputs and the output.

6.3. Estimate local sensitivity metrics

Local sensitivity analysis is a tool to assess how outputs of interest are affected by changes to each input at or near a specific reference point in the input space [7]. This contrasts with global sensitivity analysis, which looks at how outputs of interest are affected across the entire input space. Local sensitivity analysis can be used as a first step in a sensitivity analysis, or as a computationally inexpensive (but limited) alternative to global sensitivity analysis.

6.3.1. Process:

1. Execute the model using a fixed set of inputs

For this example, local sensitivity on shear stress is assessed for an initial set of input variables. The choice of these variables is not trivial and may depend on the regions of the input space of most concern for the application. In practice, multiple sets should generally be explored, particularly if the input/output relationship is expected to be non-linear. For this example, the mean of each input is used as the initial value, as shown in the second column of Table 6.1.

2. Perturb a single input and execute the model again, keeping all other model inputs fixed. Repeat this for each input.

The amount of the perturbation should be large enough so that a change in the output can be observed, but small enough to stay within the region of concern in the input space. The choice of perturbation depends on the objective of the analysis. If the goal is to rank the *influence* of inputs, a percentage (e.g., $\pm 1\%$) of the initial value for each input can be used as the amount of perturbation. However, if the goal is to assess the *importance* of an input (as is often the case), the perturbation should incorporate a measure of the uncertainty of each input. A variable's influence only concerns its effect on the output as it is varied. Importance also incorporates the assumed uncertainty of an input variable. For example, changing an input could have a large impact on the output. Such a variable would have high influence. However, if the true value of the variable is precisely known, its impact within its range of uncertainty can be small. Such a variable would have low importance. In this example, perturbations of both $\pm 1\%$ and $\pm 0.1\sigma$ were used for each input, shown in the third and fourth columns of Table 6.1, respectively.

Table 6.1. Set of Initial Values and Perturbations for Local Sensitivity Analysis

Input	Initial Value	$\pm 1\%$ Perturbation	$\pm 0.1\sigma$ Perturbation
Pressure, p	50	0.50	0.6
Radius, r	0.6	0.006	0.0012
Wall Thickness, t	0.0186	0.00186	3.2×10^{-5}
Angle of Helical Weld, α	55	0.55	0.11

3. Measure the change in the output by estimating the partial derivative for each input

The partial derivative of the output with respect to each input can be approximated using the finite difference formula provided in NUREG/CR-7278 Section 4.2.4.

4. Compare the partial derivative across inputs to assess variable influence/importance

Inputs that have a partial derivative with a large magnitude are influential/important (depending on the type of perturbation used) at the local region where the model was executed in Step 1. The results of the local sensitivity analysis depend on the initial values that were used, and the local sensitivity of an output to an input can change significantly over the domain of interest.

6.3.2. Results and potential difficulties:

The partial derivatives for each input are given in Table 6.2. Looking at the case where $\pm 1\%$ perturbation was used, it appears that pressure, radius, and wall thickness have similar influence on shear stress. This result is unsurprising considering how shear stress is calculated. However, when uncertainty information is included in the perturbation (i.e., when $\pm 0.1\sigma$ perturbation is used), variable importance can be assessed, and pressure stands out as the most important variable. This is because the uncertainty with pressure is large relative to the uncertainty associated with radius and wall thickness.

Table 6.2. Partial Derivatives of Shear Stress for Each Input

Input	Partial Derivative ($\pm 1\%$ Perturbation)	Partial Derivative ($\pm 0.1\sigma$ Perturbation)
Pressure, ρ	7.49	9.12
Radius, r	7.49	1.47
Wall Thickness, t	-7.34	-1.26
Angle of Helical Weld, α	-5.15	-1.03

Since local sensitivity analysis only requires two model realizations per input (at a minimum), it is a relatively inexpensive way to assess variable importance. That said, as mentioned above, the local sensitivity can change across the input space. Even if local sensitivity analysis is conducted at multiple initial values, a pitfall is the potential to miss regions of high importance/influence. This can especially occur if the model is complex with a high dimensional input space that is difficult to understand. Additionally, these techniques assume the relationship is locally-linear and results can be misleading if this assumption is not a good approximation of the local behavior. The use of the mean value as an initial value may in some cases focus the analysis on an inappropriate region; the mean value can be an improbable point in the tail of a distribution.

6.3.3. Summary

- Local sensitivity analysis is a relatively inexpensive and effective first step toward learning the important inputs of a model.
- The local sensitivity to an input can change significantly over the domain of interest.

6.4. Estimate global sensitivity metrics

Global sensitivity analysis attempts to quantify the amount of output uncertainty attributed to uncertainty in input variables across the entire input space. It can be a useful tool for understanding which inputs drive the problem. This understanding can prompt the refinement of those input distributions, as well as determine candidate inputs for IS. As global sensitivity analysis accounts for the uncertainty in an input, it provides a measure of variable importance (as described in Section 6.3).

6.4.1. Process:

1. Choose method for estimating metrics

Global sensitivity metrics are often estimated using regression techniques (e.g., linear regression) or variance-based methods. For this example, variance-based indices are used as they can handle complex relationships and they are easy to interpret. The calculation of variance-based indices generally requires a Monte-Carlo integration approach (NUREG/CR-7278 Section 4.2.3) which may require the use of surrogate models to reduce the computational cost. However, since this is a simple problem, a surrogate model is not required.

2. Sample the uncertain inputs and calculate the output

1×10^3 values of pressure, radius, wall thickness, angle of the helical weld and weld yield stress are sampled from the distributions defined in the Problem Statement (Section 1). These values are used as a base sample that is then used to estimate the sensitivity metrics following the Monte-Carlo approach described in [8]. The Monte Carlo approach recommends that this base sample, N , is sufficiently large (i.e., 500 samples or more). The approach then uses this base sample to generate a total of $N(k + 2)$ realizations of the model, with k being the number of parameters.

3. Calculate global sensitivity metrics for each input/output

The first order and total effects indices [8] [9] are the most commonly used variance-based sensitivity metrics. They are calculated (as described in NUREG/CR-7278 Section 4.2.3) for each uncertain input and shear stress. Recall that first order indices (S_j) provide a measure of the proportion of the total output uncertainty that is explained by the uncertainty in the j^{th} input alone. Total effects indices (T_j) give the fraction of output uncertainty that is explained by the uncertainty in the j^{th} input and its interactions with other inputs.

4. Compare the global sensitivity metrics across inputs to assess variable importance

Higher values of first order and total effects indices correspond to variables that contribute the most to the uncertainty in the output. Variables that are high contributors may be candidates for IS or input uncertainty refinement.

6.4.2. Results and potential difficulties:

Table 6.3 provides the first order and total effects variance-based indices for each input. Pressure alone contributes to 92% of the uncertainty in shear stress, while the remaining three inputs contribute only marginal amounts. These results are consistent with what was seen in Sections 6.2 and 6.3.

Table 6.3 also shows that the uncertainty in shear stress is not heavily driven by the interactions between the input variables as the total effects indices are similar in magnitude to the first order indices for each input.

Table 6.3. First Order and Total Effects Indices for Each Input

Input	Shear Stress, S_j	Weld Yield Stress, T_j
Pressure, ρ	0.92	0.93
Radius, r	0.03	0.03
Wall Thickness, t	0.02	0.03
Angle of Helical Weld, α	0.02	0.02

Since pressure is an important variable, it may be a good candidate for IS. Looking at the scatterplot in Figure 6.1, it is apparent that higher values of pressure result in higher values of shear stress and consequently, a higher probability of weld failure. Therefore, IS on the upper tail of the pressure distribution may improve the estimation of the probability of weld failure. The results of IS on pressure are shown in Section 8.2.

Several difficulties in estimating global sensitivity metrics are worth noting:

- Most importantly, the metrics often depend on an assumed statistical model. If this model poorly fits the data, the metrics can be misleading. Model checking to assess the quality of the statistical fit is an important aspect of obtaining meaningful metrics.
- Interpretation of the sensitivity metrics is often nuanced and depends on the types of interactions the assumed statistical model can estimate.
- Since these metrics are calculated from sample data, they contain statistical uncertainty. Confidence intervals [10] can be computed using bootstrapping or other techniques to help assess if the variable effects are real. However, when estimated metrics are close between two or more variables, accurate ordering of variable importance is not straightforward, and likely not meaningful. It is often sufficient to group variables with similar metric values as having similar importance.
- Results should be interpreted within the context of the physical model and should make sense to the analyst. If the results do not make sense, it is possible the physical model is not being implemented correctly or there was a mistake in conducting the sensitivity analysis. This pitfall is related to all sensitivity analysis (SA) techniques, both local and global.

6.4.3. Summary

- Global sensitivity analysis is more computationally expensive, but it allows for the quantification of the effects of uncertain inputs across the entire input space.
- Global sensitivity analysis can be helpful for determining candidate inputs for IS.

7. SURROGATE MODELS

This section applies methods for fitting surrogate models (NUREG/CR-7278, Section 4.2.5) to the pressurized tank example. Two concepts are illustrated:

1. Surrogate models for forward propagation of uncertainty.
2. Surrogate models for sensitivity analysis.

When to apply this method. PFM models are often computationally expensive to run. In this case, using an inexpensive computer model to approximate the full model is often useful when conducting sensitivity analyses (Section 6) or when propagating input uncertainty to the model outputs (Section 8). Unlike the full computer model, predictions from the inexpensive model are computationally cheap, allowing the number of model evaluations to increase by many orders of magnitude. Empirical statistical models, called surrogate models, are often used as an inexpensive model approximation.

7.1. Build surrogate models for forward propagation of uncertainty

In this section, a surrogate model is constructed for shear stress and is used to estimate the failure probability. The surrogate model is constructed for shear stress, rather than for the failure binary outcome, (i.e. an indicator for whether shear stress exceeds weld yield stress). Statistical surrogate models can be built directly for binary outcomes to predict a probability across the input space. However, substantially more training points are needed for binary outputs compared to continuous outputs, because continuous data contains more statistical information than binary data; further, the number of required training points increases as the failure probability decreases. Hence, building surrogates for binary data will often not save computing time, since training set sizes must be large. Additionally, building flexible surrogates for binary data is often more challenging than building flexible surrogates for continuous outcomes. For instance, standard software programs often do not have options to build flexible Gaussian process (GP) surrogates for binary data. Failing to achieve surrogate model convergence is another common issue. Therefore, in this section, surrogates are considered for continuous outcomes only.

7.1.1. Process:

There are four primary steps for using surrogate models for forward propagation of uncertainty:

1. Select a training set of inputs for building the surrogate model.

The training set is the set of model inputs and outputs used to construct a surrogate model. The training set inputs should cover the model input space thoroughly to ensure that the full model is accurately approximated by the surrogate in important regions (i.e., regions of the input space where the probability of failure is non-negligible). Therefore, using the original input distributions is often not a prudent choice for sampling training points for surrogate modeling, as demonstrated in the example below.

2. Build the surrogate model using the training set.

There are many ways to build surrogate models. GPs for example are a common choice due to their flexibility [11]. Linear and non-linear regression models are also commonly used as surrogates, as they provide a simple analytical form for the surrogate. An insufficiently flexible

surrogate model will bias estimation of the failure probability, as demonstrated in the example below.

When the number of model inputs is high, including all model inputs in the surrogate model is often unnecessary. Inputs with negligible influence on the model output can be omitted from the model; these inputs can be identified using a global sensitivity analysis (Section 6).

3. Evaluate the accuracy of the surrogate model.

Evaluating a surrogate using a test set. Surrogates are often evaluated by using a test set of sampled inputs that are propagated through the full model. A test set is a set of model inputs and outputs used to evaluate the surrogate's accuracy. Typically, the test set inputs are sampled using the sampling scheme selected for forward propagation of uncertainty. Using the test set, the computer model outputs are compared to predictions from the surrogate model to determine how accurately the surrogate approximates the full model. The comparison is often made both graphically (by plotting surrogate predictions versus observed model outputs) and by calculating metrics.

Two metrics are considered for evaluating surrogate accuracy:

- Estimated failure probability from the surrogate, \tilde{p} , compared to the estimated failure probability without the surrogate, \hat{p} , based on n test set observations.
- Root mean-squared error (RMSE) in shear stress, defined as:

$$\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

where y_i is the shear stress from the original computer model realization for test set observation i and \hat{y}_i is the corresponding predicted value from the surrogate model.

In this example, the test set is comprised of 1×10^5 inputs sampled using IS. Note that, for practical applications where the model is expensive, the test set size will be much smaller. With a small test set, comparing the estimated failure probability from the surrogate to the estimate without a surrogate may be more difficult, since estimates of this probability may be inaccurate in small samples. A large test set is used here to illustrate how surrogate accuracy can vary across different surrogate models, training point designs, and numbers of training points.

Evaluating a surrogate using statistical convergence metrics. If a test set is not available for evaluating surrogate accuracy (e.g., if the full model is very expensive and all realizations must be used in the training set), then sampling uncertainty associated with a surrogate model can be estimated using statistical theory. For example, if a linear regression model is used, confidence intervals on the regression model coefficients can be calculated and the width of these intervals can be assessed to determine whether there is sufficiently low sampling uncertainty. If the sampling uncertainty is too high for the application, additional samples will be needed in the training set. However, this approach relies on the assumption that the surrogate model is an accurate approximation to the full model.

In this case, sampling uncertainty can be summarized using the same types of convergence metrics as described in Section 9, including standard errors, coefficients of variation, and confidence interval widths.

4. Estimate the failure probability based on the surrogate model.

The surrogate model is used in place of the full computational model to predict model outputs and estimate the failure probability. Because the surrogate model is computationally inexpensive, the number of samples for forward propagation can be increased. In the examples below, 1×10^5 samples are generated for the surrogate model inputs using IS. These inputs are propagated over the surrogate to estimate shear stress, which is ultimately used to estimate the failure probability.

7.1.2. Results and potential difficulties:

In this example, three different surrogate models are evaluated using a test set:

1. A linear regression model, where shear stress is assumed to be a linear function of the inputs. This linear model is clearly inadequate and is included solely to illustrate how an inadequate surrogate model can lead to poor failure probability estimates [8].
2. A GP model that interpolates shear stress over the input space [12].
3. A multivariate adaptive regression splines (MARS) model, where shear stress is modeled as a flexible function of the inputs [13].

Each surrogate model includes all uncertain model inputs that impact shear stress: internal pressure (r), cylinder radius (r), wall thickness (t), and helical weld angle (a).

Training set selection. LHS was used to select training points for the surrogate. When selecting training points, directly using the input distributions would result in very few training points placed in the important region of the input space, where failure is most likely to occur. As an alternative, uniform sampling over the inputs was used to oversample the important region. The distribution of the internal pressure p was unbounded, so bounds were constructed for p based on knowledge of the problem; specifically, p was bounded at 2 standard deviations below its mean and 7 standard deviations above its mean, since the important region was known to lie above the mean of p . Choosing 7 standard deviations above the mean resulted in reasonable confidence that the surrogate would predict accurately over the important region.

Poor selection of training points can result in bad surrogate predictions. The implications of training point selection are illustrated using three different training set designs: (1) the design described above; (2) input distributions corresponding to the importance distribution (Section 8.2); and (3) input distributions corresponding to the original input distributions. For each set of training points, a GP surrogate was constructed and tested on an importance sample of shear stress. The training point designs that cover the important region of the input space (i.e., the uniform design and IS design) have superior performance to the design based on the original input distributions in terms of both shear stress RMSE and failure probability estimation (Figure 7.1).

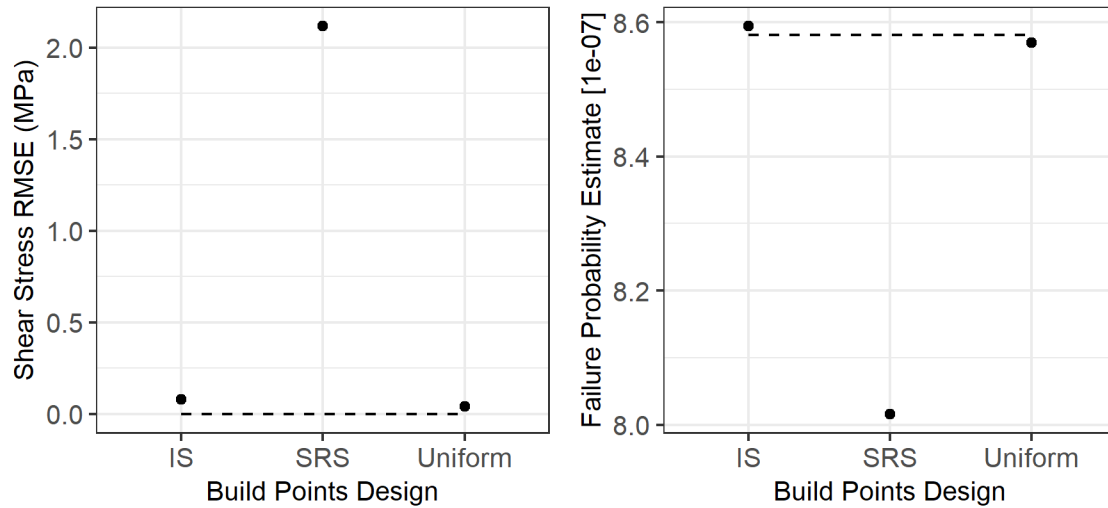


Figure 7.1. Comparing three training set designs using 1,000 training points. (Left) Root mean-squared error for shear stress. A horizontal line is drawn at the ideal RMSE = 0. (Right) Estimated failure probability. A horizontal line is drawn for the ideal failure probability estimate without using a surrogate.

Size of training set. Determining the number of training points used to train the surrogate model requires a trade-off between the computational burden of executing the computer model and the desired accuracy of the surrogate. Using an insufficient number of training points will result in poor failure probability estimates. Figure 7.2 illustrates how the predictive accuracy of the GP surrogate increases with the number of training points. Specifically, the failure probability estimate using the surrogate gets closer to the failure probability estimate without a surrogate as the number of training points increases. However, the GP is quite accurate even with 100 training points, likely due to the simplicity of the true ‘computer model’ in this case. The number of training points will need to grow with the number of model inputs and the complexity of the input-output relationships.

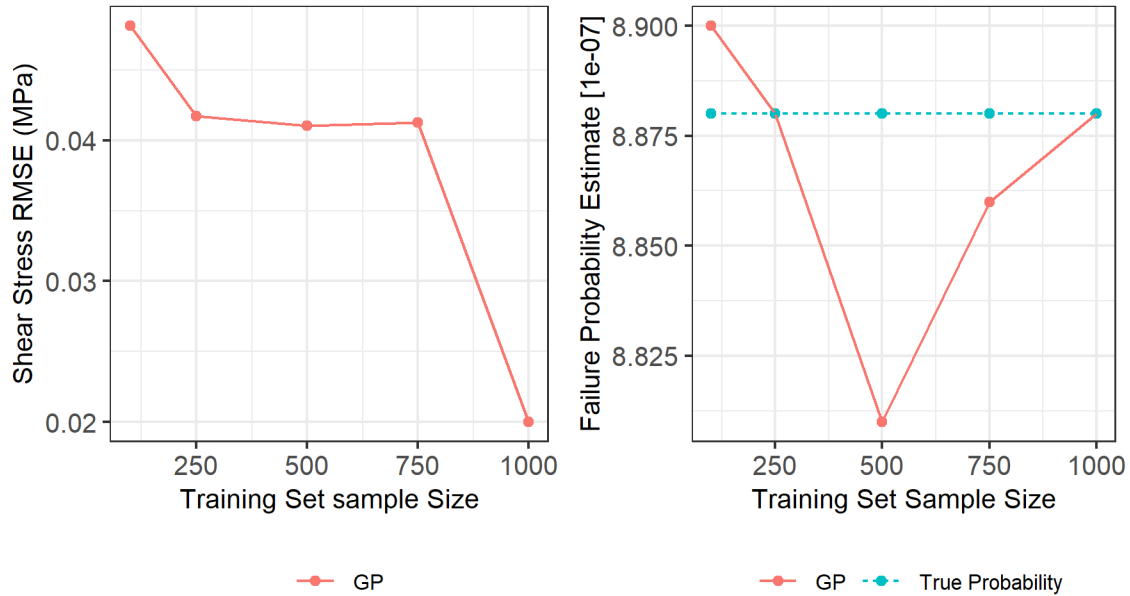


Figure 7.2. Root mean-squared error (left) and failure probability estimate (right) as a function of training set size.

Inadequate model flexibility. Surrogate models should allow for flexible relationships between inputs and outputs, including non-linearities and interactions. GPs are commonly used surrogates because they are flexible interpolators. Linear and non-linear regression in contrast do not offer as much flexibility in their functional form. Using an insufficient model approximation reduces accuracy in failure probability estimates. In Figure 7.3, predictions from a GP surrogate and linear regression surrogate built using 1,000 training points are compared to the true model output, illustrating how the increased flexibility in the GP model results in improved prediction, particularly in the upper tail of shear stress. Figure 7.4 compares the predictive accuracy of three different surrogates: a GP model, a MARS model, and a linear regression model. The root mean-squared error of shear stress is much higher for the linear model than the MARS and GP models; the GP model provides the most accurate estimate of the failure probability. Clearly, the linear regression model is not flexible enough to accurately capture the input-output relationship. The GP outperforms MARS, though the predictive accuracy of the MARS model may still be sufficient depending on the desired accuracy. Linear regression requires explicitly specifying an input-output relationship; in this case, the model assumed the output is a linear function of the inputs. If this is substantively wrong, the model will not perform well. On the other hand, flexible models like GP and MARS require less prior knowledge on the input-output relationship and flexibly adapt to the data.

Excessive model flexibility. Surrogate models can also be too flexible. When a model is too flexible, it may predict trends that are spurious and lead to misleading results. The tendency for some models to overfit can be controlled using model fitting options and separate training and validation data sets. Additionally, trends predicted by a surrogate model that do not make physical sense in the context of the problem are suspect and should be further investigated.

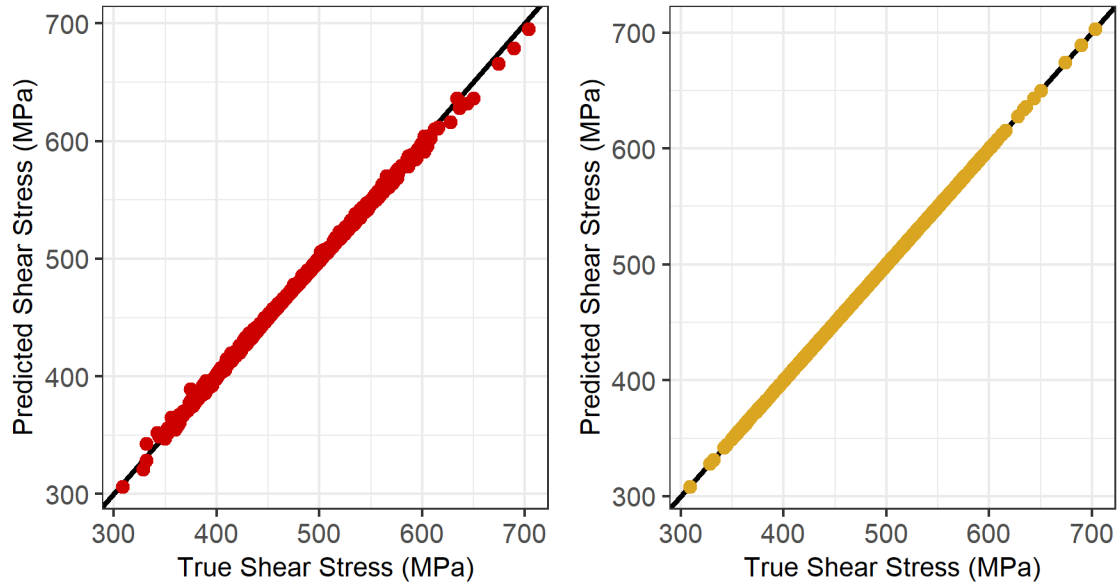


Figure 7.3. Predicted shear stress from surrogate model compared to true shear stress value from the model for: linear regression surrogate (left) and GP surrogate (right).

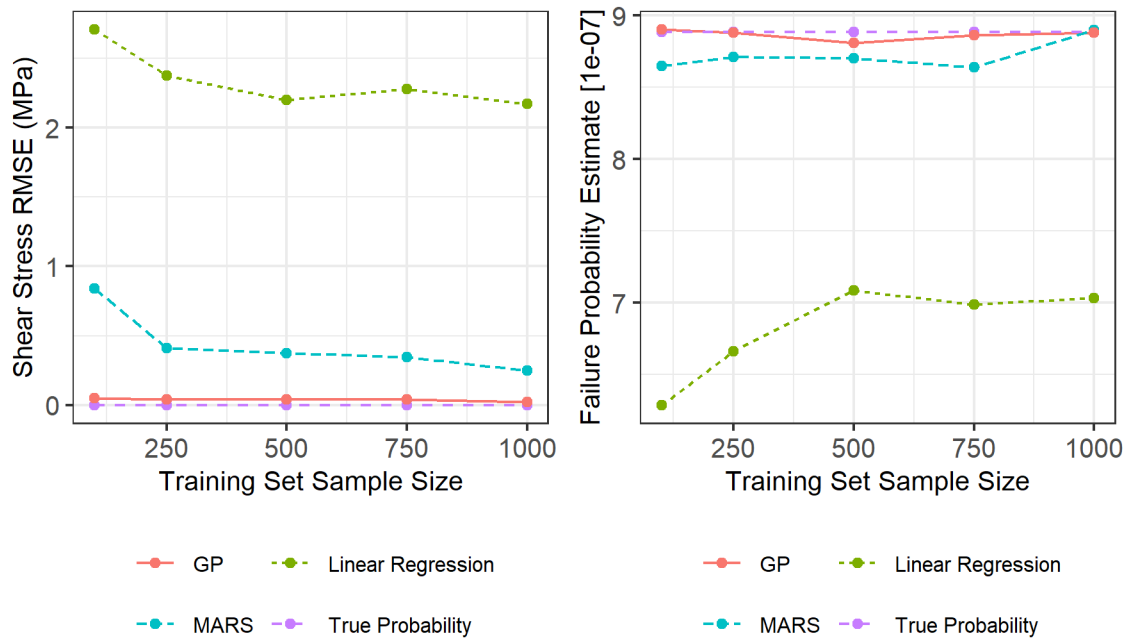


Figure 7.4. (Left) Mean-squared error of shear stress as a function of training set size. (Right) Failure probability estimate as a function of training set size.

7.1.3. Summary:

- Surrogate models can serve as computationally efficient approximations to a full computational model.

- Training points for a surrogate should be selected such that the surrogate model predicts accurately over all important regions of the input parameter space. The number of training points will need to grow with the number of model inputs and the complexity of the input-output relationships.
- Surrogate models will be inadequate approximations of the full model if the number of training points is insufficient and/or the surrogate model is not sufficiently flexible.

7.2. Surrogate models for sensitivity analysis

While surrogate models often need to be very accurate for failure probability estimation, surrogate models for sensitivity analysis often can be rough approximations to the full model. The surrogate must be sufficiently accurate to identify sensitive model parameters but does not necessarily need to accurately predict the model output. In this case, sensitivity analysis results reflect trends but cannot be interpreted as quantifying the proportion of variance in the QoI that is accounted for by each input.

7.2.1. Process

There are three primary steps for using surrogate models for sensitivity analysis. As in Concept 1, we build a surrogate model for shear stress.

1. Select a training set of inputs for building the surrogate model.

The training set, in general, can be smaller for sensitivity analysis, compared to forward propagation of uncertainty. Due to the simplicity of the model in this example, training sets of size 10 and 100 are considered. Training points are selected using uniform distributions over the inputs as described in Concept 1.

2. Build the surrogate model using the training set.

Three surrogate models are again considered: GP, MARS, and linear model. See NUREG/CR-7278 Section 4.2.5 for details.

3. Estimate sensitivity metrics based on the surrogate model.

The surrogate model is used in place of the full computational model to predict sensitivity indices because the surrogate model is computationally inexpensive. Sensitivity indices are estimated based on the original input parameter distributions.

7.2.2. Results and potential difficulties:

When building the surrogates for the example problem using 100 training points, the total effects sensitivity index estimates are essentially indistinguishable from the estimates without a surrogate (Figure 7.5); even using as few as 10 training points to build the surrogate appears to adequately approximate the total effects sensitivity indices in this example (Figure 7.5). Note that the surrogate performs well with few training points for several reasons:

1. The model is very simple, including only 4 input parameters.
2. Of the 4 input parameters, only one of the parameters (ρ) drives the variability in the output.

3. The relationship between the sensitive parameter and model output is approximately linear and there are no strong interactions/synergies between inputs. In the presence of such nonlinearities and interactions, the surrogate model would need to be sufficiently flexible to capture this behavior.

As in Concept 1, the number of training points will need to grow with the number of model inputs, number of sensitive parameters, and complexity of the input-output relationships.

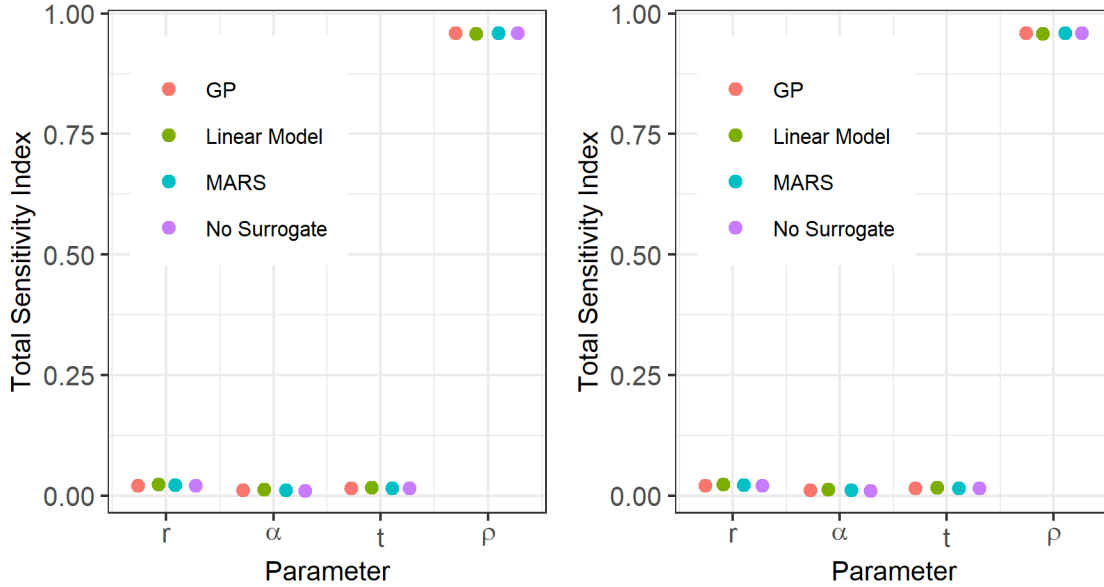


Figure 7.5. Total effects sensitivity index estimates without a surrogate and using 3 different surrogates built using: (Left) 10 training points and (right) 100 training points.

7.2.3. Summary

- Surrogate models built for sensitivity analysis do not need to be as accurate when predicting the model outcome as surrogate models built for forward propagation of uncertainty.
- When building a surrogate model for sensitivity analysis, the number of training points will need to grow with the number of model inputs, the number of sensitive parameters, and the complexity of the input-output relationships.

8. FORWARD PROPAGATION OF INPUT UNCERTAINTY

This section applies methods for sampling from uncertain inputs (NUREG/CR-7278, Section 4.3) to the pressurized tank example. Three concepts are illustrated:

1. Estimating the QoI using SRS.
2. Estimating the QoI using IS.
3. Estimating the QoI using LHS.

When to apply this method. The forward propagation of uncertainty is performed iteratively with several steps in the PFM analysis process. It is used when conducting sensitivity analyses (Section 6), fitting a surrogate model (Section 7), and producing a converged estimate of the QoI (Section 9).

8.1. Estimate the QoI distribution using SRS

8.1.1. Process:

There are four steps for taking a SRS:

1. Specify probability distributions for the uncertain inputs

This was performed in Section 4.

2. Choose the sample size

As mentioned in Section 2, the true probability of failure, p , in this example is known to be 8.8×10^{-7} , which is on the order of 1×10^{-6} . Since the SRS sample size should be at least 10 to 20 times larger than $1/p$, ideally [14], the sample size would be at least 1×10^7 . However, computational costs can often prohibit such a large sample size. To illustrate the effect of a smaller sample size on SRS estimates (and the potential benefit of using alternative sampling options), a sample size of $n = 1 \times 10^6$ will be compared to a sample size of $n = 1 \times 10^7$.

3. Implement SRS by randomly sampling the inputs from their probability distributions

The implementation of SRS is straightforward. A single value is sampled from each of the following distributions¹:

$$\rho \sim N(\mu = 49, \sigma = 5)$$

$$r \sim TN(\mu = 0.6, \sigma = 0.012, LB = 0.564, UB = 0.636)$$

$$t \sim Beta(\alpha = 2, \beta = 5, LB = 0.018, UB = 0.02)$$

$$\alpha \sim TN(\mu = 55, \sigma = 1.1, LB = 51.7, UB = 58.3)$$

$$Y \sim Weibull(\lambda = 600, k = 50)$$

The first four inputs are used to calculate a single shear stress value while the last input is the corresponding weld yield stress value. This process is repeated n times and the sampled points are

¹ Parameter definitions can be found in the problem definition, Section 2.

then used to approximate the shear stress and weld yield stress distributions, as shown in Figure 8.1.

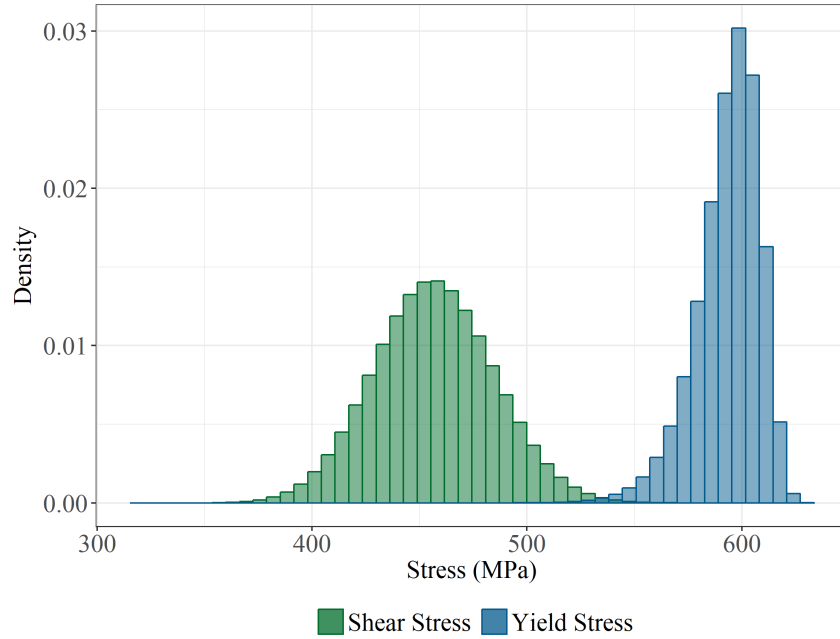


Figure 8.1. Comparison of shear stress and yield stress distributions.

4. Estimate the probability of failure

The probability of failure of the weld can be estimated using the following steps:

- i. Let $X_i = 1$ if the i^{th} shear stress is larger than the i^{th} weld yield stress and 0 otherwise.
- ii. The estimated probability of failure is $\hat{P}(failure) = \frac{1}{n} \sum_{i=1}^n X_i$ where n is the sample size.

8.1.2. Results and potential difficulties:

Replicate simulations are performed to characterize the sampling uncertainty in the failure probability under this sampling scheme and sample size. Specifically, 100 different failure probabilities are estimated using SRSs of size 1×10^6 and 1×10^7 , and the variability in the failure probability is characterized over these replicate estimates (the method of using replicate simulations to characterize sampling uncertainty is described in more detail in Section 9).

Figure 8.2 provides boxplots of the failure probability estimates under SRS with sample sizes of 1×10^7 (red) and 1×10^6 (blue). The upper and lower horizontal edges of the boxplot represent the 75th and 25th percentiles of the SRS samples, also called the third quartile (Q3) and first quartile (Q1), respectively. The difference between Q3 and Q1 is called the interquartile range (IQR). The ends of the vertical lines represent the effective minimum and maximum estimates, which is the smallest sample greater than $Q1 - 1.5 \times IQR$, and the largest sample less than $Q3 + 1.5 \times IQR$. The points that fall outside of these lines are considered to be outliers based on the definition of a boxplot. With the smaller sample size, the failure probability boxplot has a larger IQR, as shown in Figure 8.2 as distance between the upper and lower edges of the boxplot.

When the sample size is increased, the IQR decreases, giving a more precise estimate. Table 8.1 includes the mean probability of weld failure across all replicate simulations, along with the median IQR, and effective minimum and maximum. While the mean probability estimate is similar for both sample sizes, the median and IQR using 1×10^6 samples are larger than the values estimated using 1×10^7 samples. Additionally, the estimate using the larger sample size is closer to the true probability of failure, 8.8×10^{-7} .

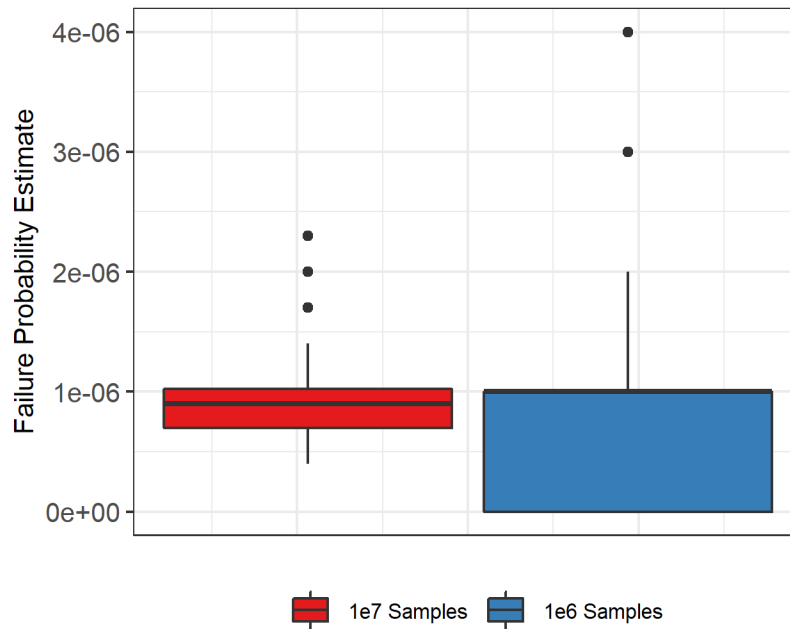


Figure 8.2. Boxplots for 100 replicate simulations using an SRS of sizes 1×10^7 (red) and 1×10^6 (blue).

Table 8.1. Failure Probability Estimates for SRS Across Replicate Simulations

Sample Size	Average Failure Probability	Median	IQR	Effective Minimum	Effective Maximum
1×10^6	7.9×10^{-7}	1×10^{-6}	1×10^{-6}	0	2×10^{-6}
1×10^7	9.0×10^{-7}	9×10^{-7}	3.3×10^{-7}	4×10^{-7}	1.4×10^{-6}

8.1.3. Summary:

- SRS is relatively easy to implement.
- If the sample size is not sufficiently large with respect to the failure probability, SRS estimates of a failure probability can be imprecise. If more precision is needed for a given sample size, alternative sampling methods may be needed.

8.2. Estimate the Qol using IS

This example illustrates estimating the probability of weld failure through IS, a sampling technique that can increase precision in the estimated probability.

8.2.1. Process:

There are 4 steps for performing an IS:

1. Determine the most sensitive inputs

Because the probability of failure occurs at the intersection of the distribution tails of shear stress and weld yield stress, efficiently sampling in those tails will produce more failures (and consequently, a more accurate probability of failure estimate). Importance sampling from the weld yield stress tail is straightforward, since weld yield stress is a single distribution, and it is known that smaller values of weld yield stress result in more failures. On the other hand, the shear stress distribution is a composite of four different uncertain inputs.

To sample from the tail of the shear stress, two steps are required:

- i. Determine which inputs (α , ρ , r , and t) contribute the most to the uncertainty in shear stress using a sensitivity analysis (as described in Section 6). In this example, pressure explains a large fraction of the variance of shear stress.
- ii. Establish which region of the pressure distribution results in high shear stress (i.e., the upper tail of the shear stress distribution). A scatterplot of pressure vs. shear stress, as shown in Figure 8.3, indicates that higher pressures result in high values of shear stress. Therefore, higher values of pressure should be targeted for IS.

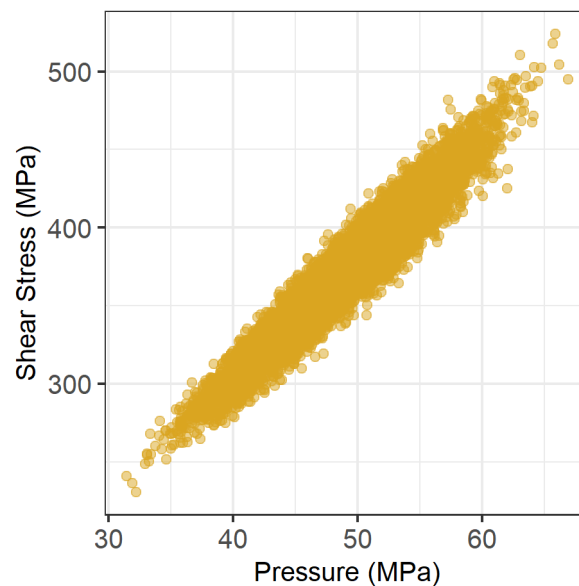


Figure 8.3. Scatterplot of shear stress versus pressure.

2. Choose an importance distribution for the sensitive inputs

Importance distributions must now be defined for weld yield stress and pressure. Importance distributions should be chosen such that the probability density function (PDF) for the importance distribution is greater than the original PDF of the input at all points in the important region of the uncertain inputs (e.g., the upper tail of the pressure distribution and lower tail of the weld yield stress distribution). If this property does not hold, then the areas in which the importance distribution is smaller than the original distribution will be sampled less efficiently than they would under SRS. Additionally, the support (bounds of the sampling region) of the importance distributions must contain the support of the original distribution. Without this property, theoretical properties that allow IS to work will not hold.

For this example, normal distributions were chosen as the importance distributions for both pressure and weld yield stress. As can be seen in Figure 8.1, the overlap region between shear stress and weld yield stress is small. The importance distributions were centered around the 99th percentile of the pressure distribution and the 1st percentile of the weld yield stress distribution, and the standard deviation was chosen to ensure that the important region (i.e. the region of overlap between shear stress and weld yield stress, was sufficiently covered):

$$\rho_{IS} \sim N(\mu = 60.61, \sigma = 7.5) \quad Y_{IS} \sim N(\mu = 547.2, \sigma = 22.5)$$

Figure 8.4 shows the importance distributions in relation to the original distributions of pressure (left) and weld yield stress (right). There may be more efficient importance distributions for this application, and more sophisticated approaches (e.g., adaptive IS, [15]) could be applied to find a near-optimal importance distribution. However, as illustrated below, this selection of importance distribution is sufficient to produce a reasonably precise estimate of the failure probability.

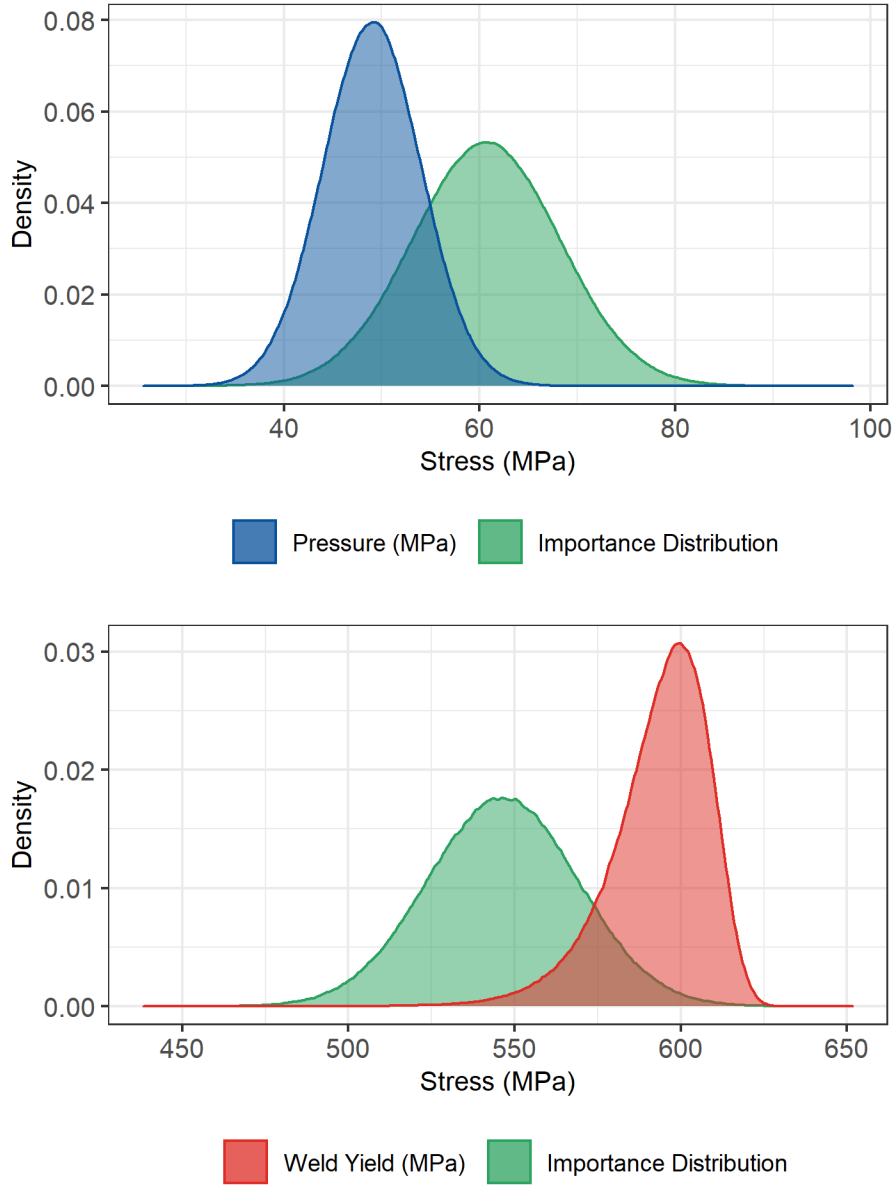


Figure 8.4. The importance distributions for both the pressure (top) and weld yield stress (bottom) distributions.

3. Sample inputs from the importance distribution and execute the model at these inputs

The process of simulating shear stress and weld yield stress values is the same as in SRS. However, for pressure and weld yield stress, the importance distribution is sampled. All remaining inputs are sampled from their original input distributions using SRS. For this example, 1×10^4 samples were used.

4. Estimate the failure probability

The estimated probability of failure is now defined as $\hat{p} = \frac{1}{n} \sum_{i=1}^n w_i X_i$, where w_i is the i^{th} IS weight, $X_i = 1$ if the i^{th} shear stress is larger than the i^{th} weld yield stress and 0 otherwise, and n is the sample size. When a single input is importance sampled, the weights are defined as:

$$w_i = \frac{\pi(u_i)}{h(u_i)}$$

where u_i is the i^{th} sampled value of the input, h is the probability density of the importance distribution from which u_i is sampled, and π is the original probability density.

If m inputs are importance sampled independently, then the importance weights will simply be

$$w_i = \prod_{j=1}^m \frac{\pi_j(u_{ji})}{h_j(u_{ji})}$$

where u_{ji} is the i^{th} sampled value of the j^{th} input. In this example, the importance weights are:

$$w_i = \frac{\pi_\rho(\rho_i)\pi_y(y_i)}{h_\rho(\rho_i)h_y(y_i)}$$

where π_ρ is the original input distribution for pressure, π_y is the original input distribution for weld yield stress, h_ρ is the importance distribution for pressure, and h_y is the importance distribution for weld yield stress.

8.2.2. Results and potential difficulties:

Three implementations of IS were considered: IS on the pressure distribution, IS on the weld yield stress distribution, and IS on both distributions. Figure 8.5 shows the approximation of the bivariate density of weld yield stress and shear stress for each of these implementations. Here, IS on both pressure and weld yield stress results in samples that are in the failure region, whereas IS on either input alone results in few samples in the failure region.

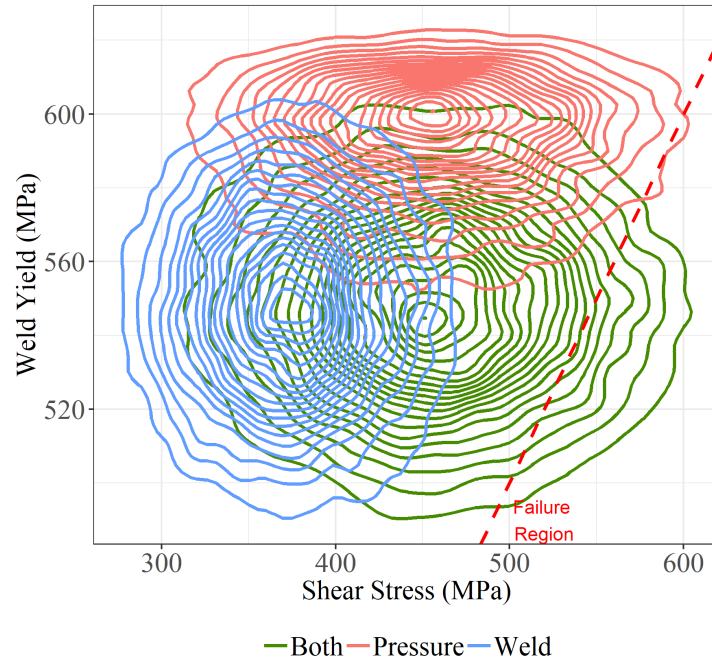


Figure 8.5. Sampling density contours for three implementations of IS.

The empirical CDFs of replicated failure probability estimates for all three IS implementations using 1×10^4 samples are compared to SRS using 1×10^7 samples (Figure 8.6). The most precise estimate occurred when both pressure and weld yield stress were importance sampled, and this estimate is like the one that is produced using SRS. Similarly, the results in Table 8.2 show that when IS is applied to both pressure and weld yield stress, the IQR of the estimates is smaller than that of SRS. For this example, using IS provides a more precise failure probability estimate using many fewer samples than SRS. Note that while in this case it is beneficial to importance sample on both pressure and weld yield stress, IS on too many inputs can be inefficient.

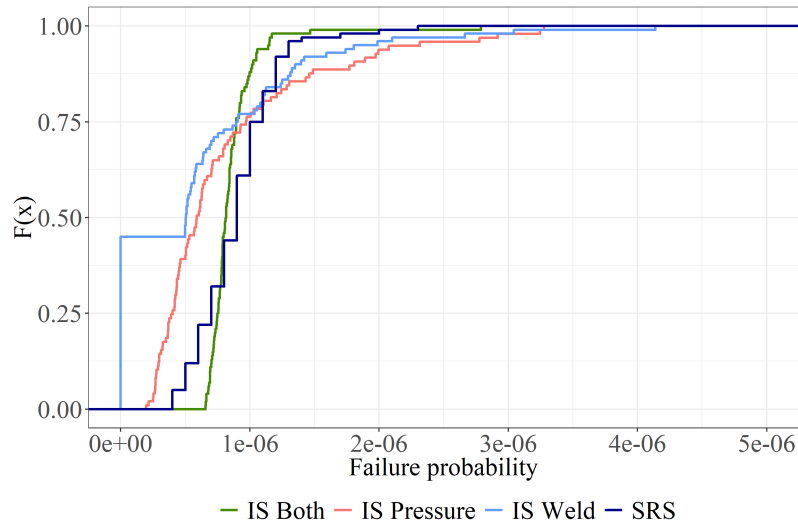


Figure 8.6. Empirical CDFs for 100 replicate simulations using 1×10^4 samples for IS on pressure, weld, and both, along with 1×10^7 samples for SRS.

Table 8.2. Failure Probability Estimates for Different Sampling Schemes and Importance Sampling Implementations

Sampling Method	Sample Size	Average Failure Probability	Median	IQR
IS Weld	1×10^4	5.7×10^{-7}	5.0×10^{-7}	9.0×10^{-7}
IS Pressure	1×10^4	1.0×10^{-6}	6.1×10^{-7}	6.0×10^{-7}
IS Both	1×10^4	8.6×10^{-7}	8.1×10^{-7}	1.4×10^{-7}
SRS	1×10^7	9.0×10^{-7}	9×10^{-7}	3.3×10^{-7}

Assessing the importance distribution. It is important to carefully choose the importance distribution, as using an inappropriate distribution will make IS less efficient [15]. To illustrate this, 1×10^4 samples will be generated using IS on both pressure and weld, but now the IS on pressure will have the following distribution:

$$\rho_{IS} \sim N(\mu = 55.4, \sigma = 3)$$

This distribution is now centered around the 90th percentile of the pressure distribution and the standard deviation has been reduced. Figure 8.7 illustrates this importance distribution in relation to the original distribution. The right plot of this figure is zoomed into the upper tail region; here, the importance distribution density is not greater than the original distribution density in this important region. A summary of replicated failure probability estimates is given in Table 8.3, which show that by choosing a poor importance distribution, IS can be less efficient than SRS.

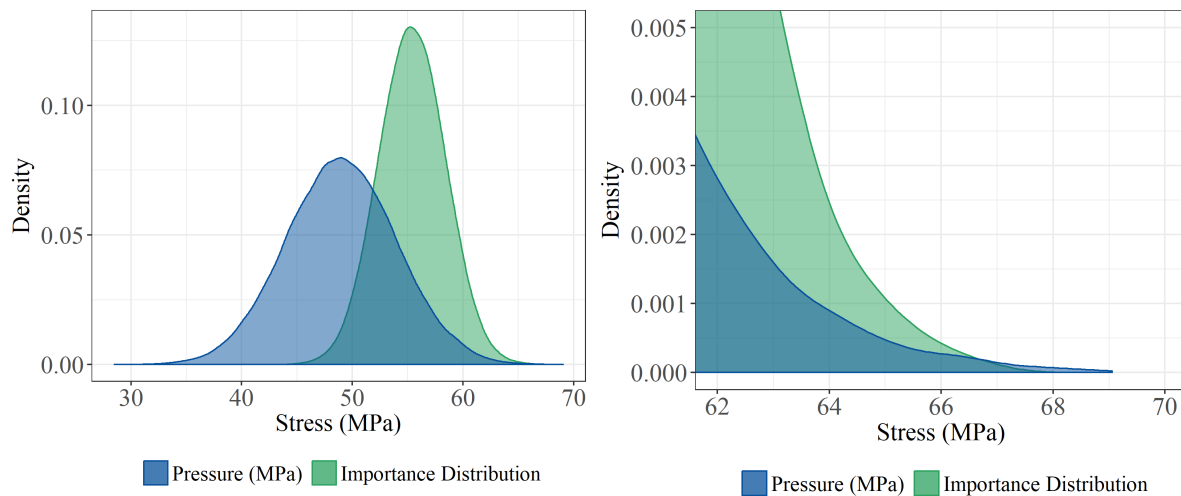


Figure 8.7. Importance distribution in relation to the original distribution for pressure.

Table 8.3. Failure Probability Estimates for Different Choices of the Importance Sampling Distribution Compared to SRS

Sampling Method	Sample Size	Average Failure Probability	Median	IQR
IS Both (Good IS Distribution)	1×10^4	8.4×10^{-7}	8.2×10^{-7}	1.7×10^{-7}
IS Both (Poor IS Distribution)	1×10^4	7.7×10^{-7}	4.9×10^{-7}	4.2×10^{-7}
SRS	1×10^7	9.0×10^{-7}	9.0×10^{-7}	3.3×10^{-7}

8.2.3. Summary:

- When the importance distribution is chosen properly, IS can significantly decrease the variance in failure probability estimates.
- Once an importance distribution is selected, IS is not much more difficult to implement than SRS.
- Poor implementation of IS can increase the variance of an estimate, resulting in less precision.

8.3. Estimate the Qol using LHS

This example illustrates estimating the probability of weld failure when using Latin Hypercube Sampling (LHS), a sampling technique designed to reduce variation in the estimated probability [16] [17].

8.3.1. Process:

Most standard statistical and engineering software packages contain implementations of LHS, where the user only must supply the uncertain input distributions. Manual implementation of LHS requires four steps:

1. **Stratify the input space by dividing the range of each input, x_j , into n disjoint intervals of equal probability.** 1×10^4 samples using LHS with IS will be compared to 1×10^4 samples using SRS with IS; therefore, the distribution of each uncertain input is partitioned into 1×10^4 disjoint intervals of equal probability. Since IS is used on pressure and weld, the importance distributions for those variables are partitioned.
2. **For each input, sample a single value from each interval, resulting in n sampled values for each input.** For a given input and interval, the sample is taken from the conditional distribution of the input on the interval. For each uncertain model input, a value is sampled from each of the 1×10^4 intervals (because they were defined such that they have equal probability).
3. **Randomly combine samples without replacement.** The 1×10^4 samples from each input are randomly combined into 1×10^4 tuples, defined as a vector of model inputs. For example, a possible tuple may look like this:

$$(\alpha = 53.84, \rho = 42.56, r = 0.59, t = 0.019, y = 606.30)$$

4. **Calculate the failure probability:** For a given tuple, α , ρ , r , and t are used to calculate a single shear stress value which is then compared to the weld yield stress value. Once the 1×10^4 shear stress and weld yield stress values have been calculated, the probability of weld failure is estimated the same way as it is calculated when using SRS with IS.

8.3.2. Results and potential difficulties:

Table 8.4 provides the average failure probability estimate, median and IQR for LHS and IS with 1×10^4 samples compared to SRS and IS with 1×10^4 samples. Here, the average failure probability estimate is relatively close to the estimate found under SRS, though the IQR is slightly smaller with LHS, indicating higher precision.

One thing to note is that the improvement of using LHS over SRS decreases as the sample size increases. To show this, the same analysis was performed using 1×10^5 samples for both SRS and LHS, and the results are given in

Table 8.5. With 1×10^4 samples, the percent change in the IQR from SRS to LHS is 11.8%. When the sample size is increased to 1×10^5 , the percent change is 7.8%, showing that the variance reduction obtained when using LHS decreases as sample size increases.

Table 8.4. Failure Probability Estimates for SRS with IS and LHS with IS and 1×10^4 Samples Across Replicate Simulations

Sampling Scheme	Sample Size	Average Failure Probability	Median	IQR
SRS	1×10^4	8.4×10^{-7}	8.2×10^{-7}	1.7×10^{-7}
LHS	1×10^4	8.6×10^{-7}	8.4×10^{-7}	1.5×10^{-7}

Table 8.5. Failure Probability Estimates for SRS with IS and LHS with IS and 1×10^5 Samples Across Replicate Simulations

Sampling Scheme	Sample Size	Average Failure Probability	Median	IQR
SRS	1×10^5	8.5×10^{-7}	8.5×10^{-7}	6.4×10^{-8}
LHS	1×10^5	8.7×10^{-7}	8.6×10^{-7}	5.9×10^{-8}

The results of this section showed that LHS reduces the variance in the estimate of a failure probability compared to SRS. Though the amount of variance reduction will depend on the sample size, LHS will typically be more efficient than SRS.

8.3.3. Summary:

- LHS is typically easy to implement.

- LHS will generally be more efficient than SRS, though the improvement in efficiency will decrease with sample size.

9. CONVERGENCE ANALYSIS

This section applies methods for assessing convergence of an estimated QoI (NUREG/CR-7278, Section 4.3.5) to the pressurized tank example. Four concepts are illustrated:

1. Assess convergence using a closed-form statistical metric under simple random sampling.
2. Assess convergence using the statistical bootstrap.
3. Assess convergence of an estimated QoI as a function of the sample size.
4. Assess convergence of an estimate of the QoI by performing replicate simulations.

When to apply this method. When propagating input uncertainty through a model (Section 8), the user must decide the sampling scheme and how many random samples to propagate (i.e., the sample size). Convergence analysis is used to evaluate whether the selected sampling scheme and sample size are sufficient for estimating a converged QoI.

9.1. Assess convergence using closed-form statistical metrics under simple random sampling

This example illustrates the process of estimating the sampling variance of the estimated probability of weld failure using closed-form statistical metrics [18] when the model outputs are generated from an SRS.

9.1.1. Process:

The steps for assessing convergence of SRS using closed-form statistical metrics are:

1. **Decide the sample size, use SRS to sample values from the uncertain inputs, and estimate the failure probability.**

The selected SRS sample size is $n = 1 \times 10^7$.

2. **Use closed-form statistical metrics to assess the amount of sampling variability in the failure probability estimate.**

Common statistical metrics for measuring sampling variability include the standard error, coefficient of variation, and statistical confidence intervals:

- i. Standard Error:

$$se_{\hat{p}} = \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}$$

where:

- n : The sample size
- \hat{p} : The estimated probability of failure

- ii. Coefficient of Variation:

$$CV_{\hat{p}} = \frac{se_{\hat{p}}}{\hat{p}}$$

- iii. Confidence Interval: there are numerous ways to calculate confidence intervals for a failure probability, such as:
 - Large-sample approximation: the $(1 - \alpha) * 100\%$ confidence interval is defined as: $\hat{p} \pm z_{\alpha/2} * se_{\hat{p}}$ where $z_{\alpha/2}$ is the $1 - \alpha$ critical value of the standard normal distribution. Here, α is the significance level of the confidence interval. Note that this confidence interval relies on the assumption that n is large relative to \hat{p} , such as $(n \times \hat{p} \geq 5)$ and $n(1 - \hat{p}) \geq 5$.
 - Exact binomial confidence interval: an exact confidence interval for p can be constructed using the fact that the number of observed failures follows a binomial distribution with sample size n and probability of success p [18]. When n is small relative to p ($n \times \hat{p} < 5$ or $n(1 - \hat{p}) < 5$), exact confidence intervals are conservative and are preferable to large-sample approximations that can underestimate uncertainty.

9.1.2. Results and potential difficulties:

The convergence estimate metrics are shown in Table 9.1. The analyst can use these metrics to determine whether sampling uncertainty is sufficiently low, keeping in mind that sampling uncertainty can be underestimated when convergence has not yet been achieved. For instance, the approximate 95% confidence interval gives a plausible uncertainty interval for the true probability of weld failure. However, this interval relies on the SRS sample size being “large enough;” a common rule of thumb for SRS samples is $n\hat{p} \geq 5$ and $n(1 - \hat{p}) \geq 5$; in this case, $n\hat{p} = (1 \times 10^7)(6 \times 10^{-7}) = 6$, which meets the sample size criteria. This rule of thumb should *not* be considered a universal criterion and should be used with caution.

Table 9.1. Closed-form Convergence Metric Results

Failure Estimate	6.00×10^{-7}
Standard Error	2.45×10^{-7}
Coefficient of Variation	0.410
95% Confidence Interval – Exact	$(2.20 \times 10^{-7}, 1.31 \times 10^{-6})$
95% Confidence interval – Approx.	$(1.20 \times 10^{-7}, 1.08 \times 10^{-6})$

9.1.3. Summary:

- Closed-form metrics for SRS are simple to calculate on a replicate using SRS.

9.2. Assess convergence using the statistical bootstrap to estimate convergence metrics

This example uses the statistical bootstrap [19] to estimate convergence metrics for the failure probability estimate. Unlike the closed-form metrics in Section 9.1, the bootstrap can also be used to estimate convergence metrics under IS. The bootstrap can be applied to many, but not all, sampling schemes; for instance, the bootstrap fails for LHS.

9.2.1. Process:

There are 4 steps for assessing convergence with bootstrap statistical metrics:

1. **Decide the sample size, sample values from the uncertain inputs, and execute the model to estimate the failure probability.** For this example, both SRS and IS are implemented. The SRS sample size is $n = 1 \times 10^7$, and the IS sample size is $n = 1 \times 10^4$.
2. **Using the generated samples, re-sample the model outputs with replacement and calculate the failure probability over the re-sampled values.** In the bootstrap procedure, the outputs are re-sampled *with replacement*, to generate a new bootstrap sample of size n . The failure probability is then re-estimated in the newly generated bootstrap sample. When IS is used, the output is resampled with replacement *in proportion to its weight*. Since the bootstrap relies on resampling from the model outputs *after executing the model*, the bootstrap is computationally tractable.
3. **Repeat step 2 many times to generate many bootstrapped estimates of the failure probability.** In this example, 1,000 different bootstrap samples, m^* , of sample size n were generated by resampling from the model outputs with replacement.
4. **Use the bootstrap samples to estimate convergence metrics.** The standard error can be estimated from the bootstrap samples using the formula:

$$\widehat{se}_{\hat{p}} = \sqrt{\frac{\sum_{i=1}^n (\hat{p}_i^* - \hat{p}^*)^2}{m^* - 1}}$$

- m^* : The number of bootstrap samples
- \hat{p}_i^* : Estimated probability of failure for the i -th bootstrap sample.
- $\hat{p}^* = \frac{1}{m^*} \sum_{i=1}^{m^*} \hat{p}_i^*$

Given the standard error, the coefficient of variation can also be calculated.

Bootstrap confidence intervals can be estimated in many ways; the simplest method is to use quantiles of the empirical distribution of the bootstrapped estimates.

9.2.2. Results and potential difficulties:

Distributions of the bootstrapped failure probability estimates are shown in Figure 9.1. Under SRS, the convergence metric estimates using the bootstrap (Table 9.2) are similar to the closed form metrics (Table 9.1).

The bootstrap generates ‘resamples’ of the failure probability from the model outputs to estimate uncertainty in the failure probability. The bootstrap is more accurate with larger sample sizes, because the model output distribution is better approximated as the sample size increases. Under SRS, considering the value of $n \times p$, as in Section 9.1, is again helpful to gauge the accuracy of the bootstrap.

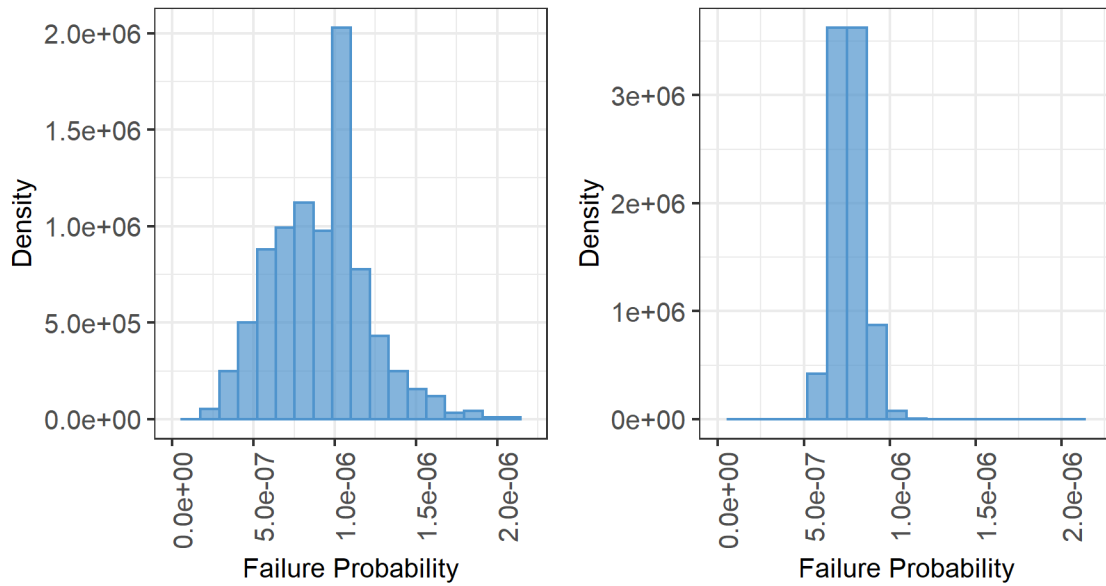


Figure 9.1. Histograms of the failure probability based on 1000 bootstrap samples using SRS (left) and using IS (right).

Table 9.2. Convergence Metrics for Both Sampling Methods Using Bootstrap Samples

	SRS	IS
Estimated failure probability	6.00×10^{-7}	8.01×10^{-7}
Standard Error	2.95×10^{-7}	8.87×10^{-8}
Coefficient of Variation	0.322	0.114
95% Confidence Interval	$(4.00 \times 10^{-7}, 1.50 \times 10^{-6})$	$(6.14 \times 10^{-7}, 9.71 \times 10^{-7})$

9.2.3. Summary:

- The statistical bootstrap can be used to estimate convergence metrics under different sampling schemes, including SRS and IS.
- Since the bootstrap relies on resampling from the model outputs *after executing the model*, the bootstrap is computationally tractable.

9.3. Assess convergence of the QoI as a function of the sample size

This example illustrates the process of iteratively increasing the sample size until the variability in the calculated failure probability is sufficiently low. The sample size is gradually increased, and convergence metrics are calculated as the sample size grows. The method is illustrated using both SRS and IS but can also be applied with LHS. Note that differentiating between types of uncertainty (aleatory vs. epistemic uncertainty) may result in additional steps in the process depending on the goals of a convergence analysis (NUREG/CR-7278, Section 4.1.1, 4.3.11, 4.3.5).

9.3.1. Process:

There are two steps for assessing convergence as a function of the sample size:

1. **Determine the sequence of sample sizes to use, sample values from the uncertain inputs, and estimate the failure probability.** This step can be further broken down into three steps:
 - i. Determine the starting sample size
 - ii. Determine the tentative maximum size
 - iii. Determine the step size between sample sizes

In this example, the starting sample size for SRS is 1×10^5 . This selection should be driven by a best-guess about the order of magnitude of the true failure probability. A step size of 1×10^5 was chosen. The maximum sample size should be determined based on when the model has converged. However, in some cases, the maximum sample size will be determined by computational feasibility. If the maximum sample size is reached and convergence has not yet been achieved, a different, more efficient sampling scheme should be used. For this example, 1×10^8 was chosen as the maximum sample size.

The starting sample size for IS is 1×10^3 ; step size is 1×10^3 ; and maximum sample size is 1×10^5 .

2. **Evaluate convergence of the failure probability estimate as the sample size increases.** For the IS example, 1×10^3 new samples are added to the previous samples at each increment; the failure probability and statistical convergence metrics are then calculated with the additional new samples. Specifically, the standard error, coefficient of variation, and statistical confidence intervals are estimated for the failure probability estimate.

Plotting the failure probability estimate as the sample size increases demonstrates evidence of convergence. If evidence of convergence is strong at a specific sample size, then sampling can stop and results can be presented for the converged failure probability estimate. If the process

ends at the selected maximum sample size and it is not clear whether the estimate has converged, then the analyst has several options:

- i. Increase the maximum sample size
- ii. Choose a different sampling method

9.3.2. Results and potential difficulties:

To visualize evidence of convergence, the failure probability estimate and upper confidence interval were plotted as a function of the sample size (Figure 9.2). For SRS, exact and asymptotic confidence intervals were calculated (Section 9.1); for IS, bootstrap confidence intervals were calculated (Section 9.2). For SRS, the conservatism of the exact confidence interval relative to the asymptotic interval is clear.

As expected, variability in the failure probability estimate decreases as the sample size increases, but the gains in efficiency decrease as the sample size grows (Figure 9.2). For SRS, the probability estimate seems to stabilize around 5×10^7 samples; for IS, the estimate stabilizes around 4×10^4 samples. IS converges much faster than SRS (Figure 9.3) in this example.

When using LHS, it is often of interest to augment the existing LHS sample with additional points, as opposed to resampling an entirely new LHS sample with a larger sample size. The LHS design depends on a known sample size, and therefore, care must be taken to preserve the properties of an LHS if the sample size increases. Incremental LHS is a common approach for adding samples to an existing LHS [20].

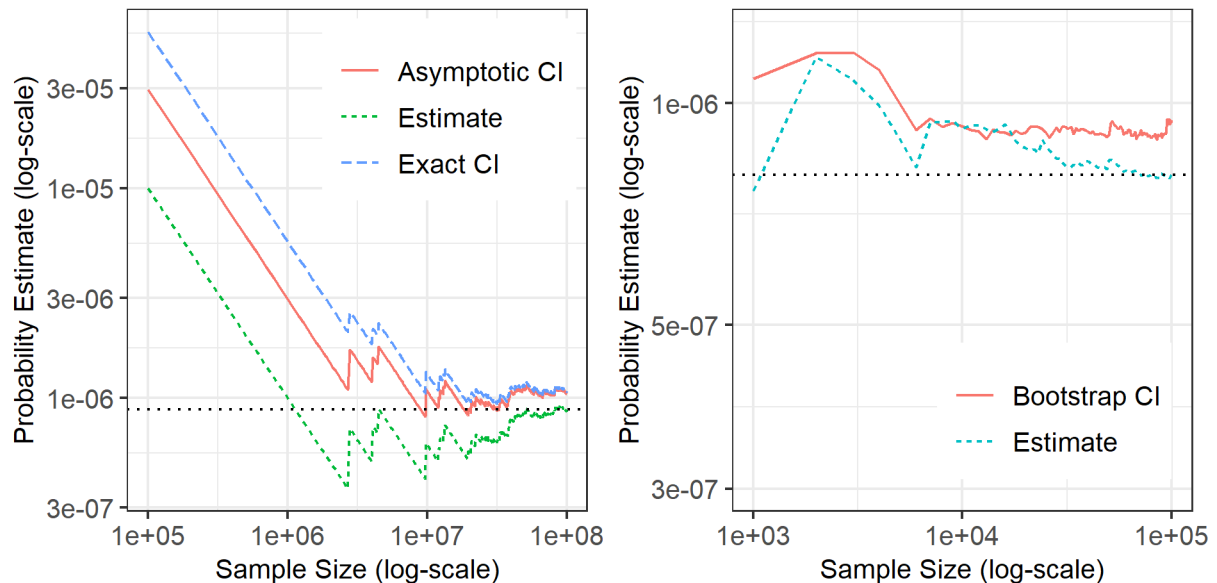


Figure 9.2. Probability of weld failure as a function of the sample size using SRS (left) and IS (right) with upper confidence bounds. The scales of the plots differ. A horizontal line is drawn at the true failure probability.

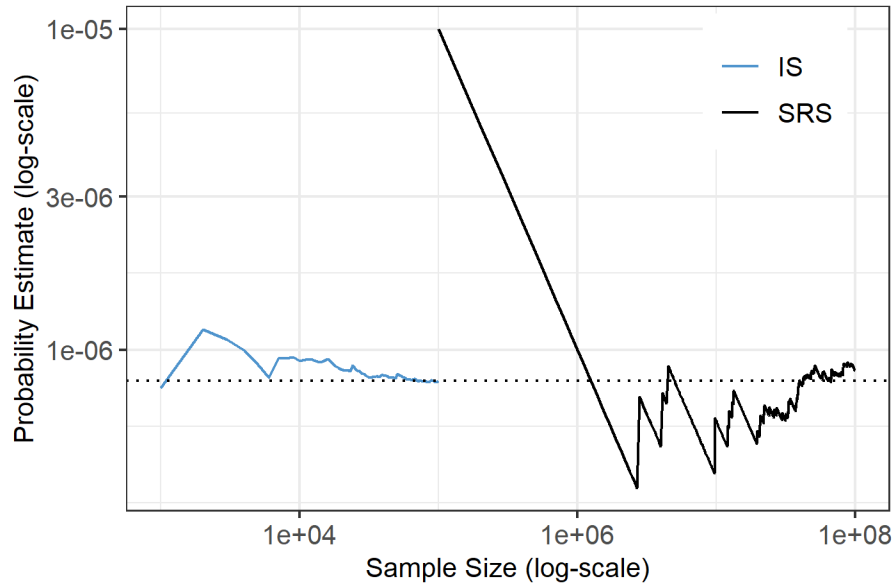


Figure 9.3. Comparison of SRS and IS failure probability estimates as the sample size increases. A horizontal line is drawn at the true failure probability.

9.3.3. Summary:

- Plotting the failure probability as a function of the sample size is a valuable visual diagnostic for convergence.
- If it is unclear whether convergence has been obtained, more samples should be obtained, or a different sampling method applied.

9.4. Assess convergence of an estimate of the QoI by performing replicate simulations

This example illustrates the process of estimating the sampling variance of the estimated probability of weld failure by performing many replicate simulations. Replicate simulations are simply repeated simulations with a fixed sample size, regardless of whether the variable uncertainty is considered to be aleatory or epistemic. When the failure probability is estimated many times in replicate simulations of the same sample size, the variation in the estimate can be assessed.

Conducting replicate simulations will often not be a practical solution. Specifically, executing replicates may not be feasible for computationally expensive models. For computationally inexpensive problems, the model can be executed enough times that demonstrating convergence should not be an issue. Further limitations of conducting replicate simulations are discussed below.

In this section, convergence metrics for replicate simulations are calculated under two different sampling schemes: SRS and IS.

9.4.1. Process:

There are three primary steps for performing the replicate simulations:

1. **Determine the sample size of a single simulation.** For this example, the selected sample size for SRS is 1×10^7 . Since IS reduces variation in the estimate significantly, a smaller sample size of 1×10^4 was selected for IS. These sample sizes were selected based on knowledge that the probability of failure is expected to be very small (close to 1×10^{-6}).
2. **Estimate the failure probability based on the above sample size in a single simulation. Then repeat this step over many simulations to produce replicate estimates of the failure probability.** Determining the number of replicate simulations involves a trade-off between accuracy and computation time; more replicate simulations will produce more accurate convergence metrics, but the computational burden is higher. If the computational burden is not a consideration, a large number of replicates could be executed. However, in many applications, executing even a single replicate may be too expensive. In this example, a large number of replicates (100) as well as a more moderate number (10) are considered to illustrate how the convergence metrics perform on different replicate numbers. Using 100 replicates, each with 1×10^7 samples, was feasible in this example because the model is computationally inexpensive.
3. **Use convergence metrics to assess the amount of sampling variability.** Different convergence metrics can be applied to the replicate simulations. Commonly used metrics include:

i. Standard Error:

$$se_{\hat{p}} = \sqrt{\frac{\sum_{i=1}^m (\hat{p}_i - \bar{p})^2}{m - 1}}$$

- m : The number of replicate simulations
- \hat{p}_i : Estimated probability of failure for the i -th replicate simulation.
- $\bar{p} = \frac{1}{m} \sum_{i=1}^m \hat{p}_i$

ii. Coefficient of Variation:

$$CV_{\hat{p}} = \frac{se_{\hat{p}}}{\bar{p}}$$

- By dividing the standard error by the estimated failure probability, the standard error is scaled by the estimate and is a unitless quantity.

iii. Normal Distribution Prediction Interval:

$$CI_{1-\alpha} = \bar{p} \pm t_{1-\alpha; m-1} \times se_{\hat{p}} \sqrt{1 + 1/m}$$

- The prediction interval is distinct from the confidence interval. It is effectively wider than confidence intervals because it attempts to predict the range in which a future individual observation will fall, while confidence interval shows the likely range of values associated with statistical parameter of the data.
- $1 - \alpha$: The confidence level for the prediction interval; 95% confidence is a common choice.

- $t_{1-\alpha; m-1}$: The $1 - \alpha$ critical value of a t distribution with $m - 1$ degrees of freedom.
- Note that the distribution of \hat{p}_i should be approximately normal to use this interval. The validity of the metric will decrease as the distribution becomes more non-normal. The analyst should assess the distribution of the estimates of p over the replicates to determine if this is a reasonable assumption.

The metrics can be compared to a pre-determined allowable value. For instance, standard error is an estimate of the standard deviation in \hat{p}_i over replicate simulations; if the analyst has an idea of what an acceptable standard error is, then the estimated standard error can be compared to this threshold to assess convergence.

9.4.2. Results and potential difficulties:

After conducting replicate simulations, the estimates of the failure probabilities can be plotted in a histogram to assess the shape of the sampling distribution of the failure probabilities (Figure 9.4). The sampling distribution based on SRS is more symmetric than the sampling distribution using IS (Figure 9.4). However, there is less variation in the IS estimates than the SRS estimates despite the sample size of each replicate being far less for IS (1×10^4) versus SRS (1×10^7).

In many scenarios, executing 100 replicates is not feasible. Therefore, consider the distribution of the failure probabilities with only 10 replicate simulations (which may still be a high number of replicates for some applications). Figure 9.5 shows histograms of the failure probability estimates over 10 replicate simulations. In this case, the histograms do not capture the shape of the true sampling distributions due to the limited number of replicates; however, they do still provide valuable information about variability in the failure probability estimates across replicates (Table 9.3). Because the shape of the sampling distribution is more accurately estimated as the number of replicates increases from 10 to 100, the estimated convergence metrics (which measure properties of this sampling distribution) will also be more accurate with a larger number of replicates.

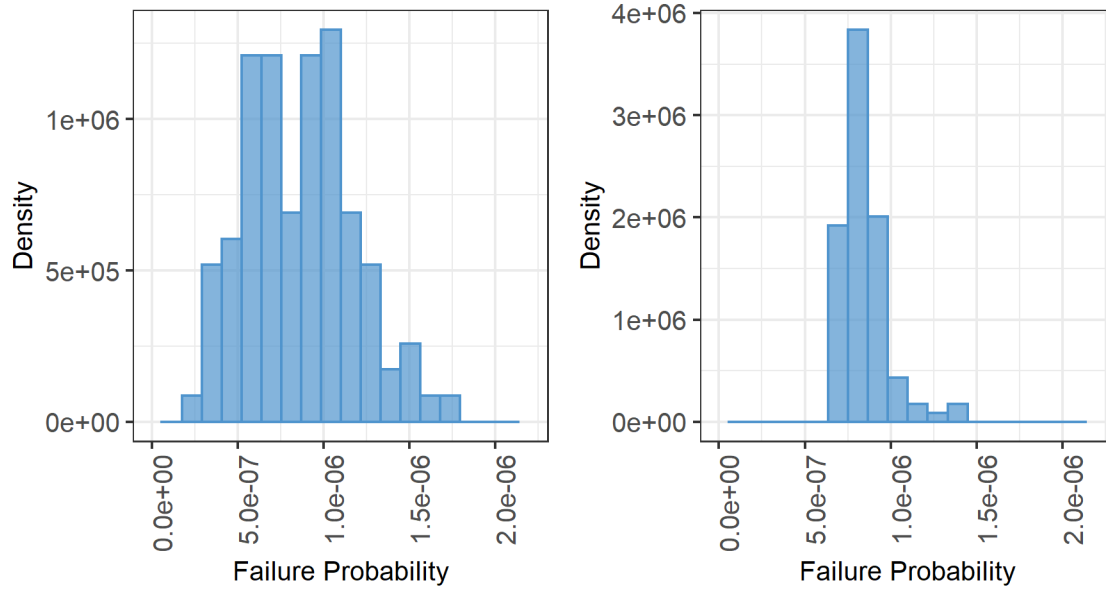


Figure 9.4. Histograms of 100 failure probability estimates based on SRS (left) and IS (right) replicate simulations.

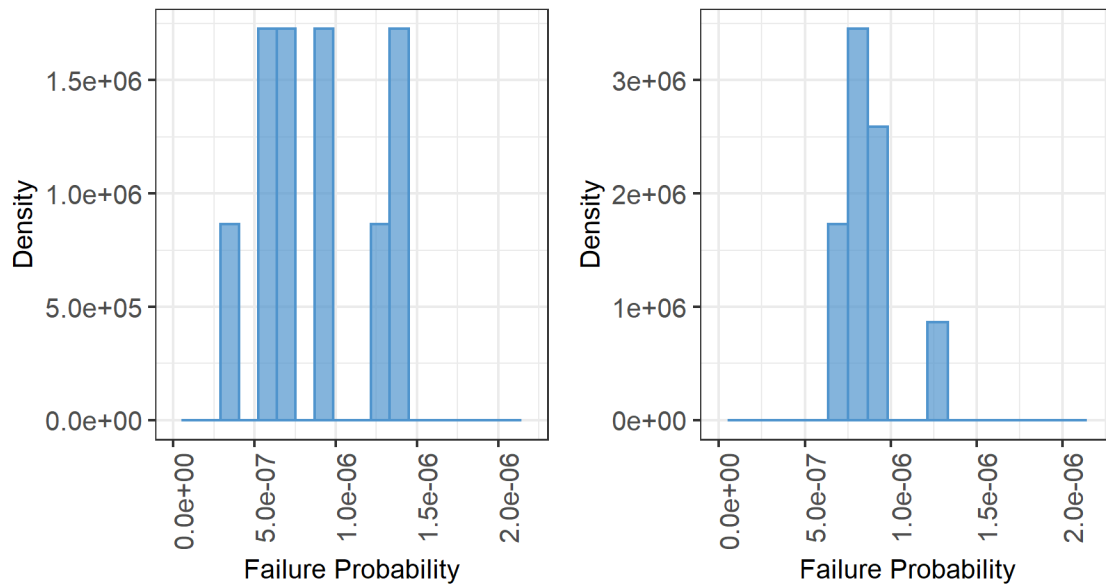


Figure 9.5. Histograms of 10 failure probability estimates based on SRS (left) and IS (right) replicate simulations.

Table 9.3. Convergence Metrics for Both Sampling Methods

	Number of replicates: 10		Number of replicates: 100	
	SRS	IS	SRS	IS
Mean estimate	8.90×10^{-7}	8.46×10^{-7}	8.68×10^{-7}	8.60×10^{-7}

Standard Error	3.60×10^{-7}	1.55×10^{-7}	3.16×10^{-7}	1.66×10^{-7}
Coefficient of Variation	0.405	0.183	0.364	0.192
Prediction interval	$[7.47 \times 10^{-8}, 1.71 \times 10^{-6}]$	$[4.95 \times 10^{-7}, 1.20 \times 10^{-6}]$	$[2.41 \times 10^{-7}, 1.49 \times 10^{-6}]$	$[5.32 \times 10^{-7}, 1.19 \times 10^{-6}]$

For SRS, the coefficients of variation are approximately 0.4; that is, the standard deviation of the estimated failure probability is 0.4 times the failure probability. This coefficient of variation is large, suggesting the sample size may need to be increased. For IS, the coefficient of variation is lower and prediction intervals are narrower. Even though the sample size is three orders of magnitude smaller, this highlights the benefits of IS.

The prediction interval provides an indication of the range of variability in the failure probability estimate for a single replicate with a given sample size. However, the prediction interval relies on a normality assumption that is violated for the IS case, decreasing the accuracy of the prediction interval. Plotting the data to look for skewness and outliers is informative when interpreting the convergence metrics, particularly the prediction interval. The non-normality in the IS distribution was more evident with 100 replicate simulations than with only replicate 10 simulations.

Limitations of replicate simulations. Conducting m different replicate simulations at a sample size of n is typically equivalent to (or less efficient than) simply conducting a single simulation of size $n \times m$. Therefore, the final estimate of the failure probability can be based on all $n \times m$ outputs, increasing the stability in the failure probability estimate far beyond the stability with sample size n . Hence, replicate simulations require a lot of additional computing to evaluate whether a sample of size n is sufficient, and, after conducting replicate simulations, the analyst has essentially increased the sample size to $n \times m$. However, replicate simulations may still be a useful tool. For instance, in some cases, a replicate simulation convergence analysis may be conducted on a single nominal case and the results applied to other scenarios where replicates are not conducted.

9.4.3. Summary:

- Performing replicate simulations is computationally expensive.
- Convergence metrics are estimated over the replicate simulations and are more accurate with more replicates.

10. VISUALIZING UNCERTAINTY IN THE QOI

This section applies methods for visualizing uncertainty in the QoI (NUREG/CR-7278, Section 4.3.8) to the pressurized tank example. Two concepts are illustrated:

1. Visualizing uncertainty without separation of aleatory and epistemic uncertainties.
2. Visualizing uncertainty with separation of aleatory and epistemic uncertainties and IS.

When to apply this method. Uncertainty can be visualized after it has been propagated through the model (Section 8).

10.1. Visualizing uncertainty without separation of aleatory and epistemic uncertainties

Recall that the QoI in this problem is the probability of weld failure. When aleatory and epistemic uncertainties are separated, there is epistemic uncertainty in the failure probability estimate that can be visualized. When separation does not occur, visualizing uncertainty in one or more continuous model outputs that contribute to the failure probability is a reasonable choice.

10.1.1. Process:

1. **Determine what should be visualized:** When there is no separation of aleatory and epistemic uncertainties, each realization will produce values of shear stress and weld yield stress, as well as a discrete indicator variable that is equal to 1 if shear stress is larger than weld yield stress and 0 otherwise. Taking the average of these indicator variables across all realizations provides a single probability of failure estimate.

Plotting a single probability of failure estimate is not informative, so another option in this case is to plot the two model outputs that determine the probability of failure, namely shear stress and weld yield stress, as weld failure occurs in the region where shear stress is higher than weld yield stress. If the probability of failure estimate changes across some dependent variable (such as time), an option for visualization would be to plot the failure estimates across the dependent variable with a confidence interval that represents the sampling uncertainty.

2. **Choose an appropriate plot based on Step 1:** SRS is used to generate 1×10^7 samples of shear stress and weld yield stress. The resulting histograms for both are plotted.

10.1.2. Results and potential difficulties:

Figure 10.1 gives histograms of shear stress (green) and weld yield stress (purple). The area where the two histograms overlap (as shown in the right plot in Figure 10.1) provides the region where weld failure will occur.

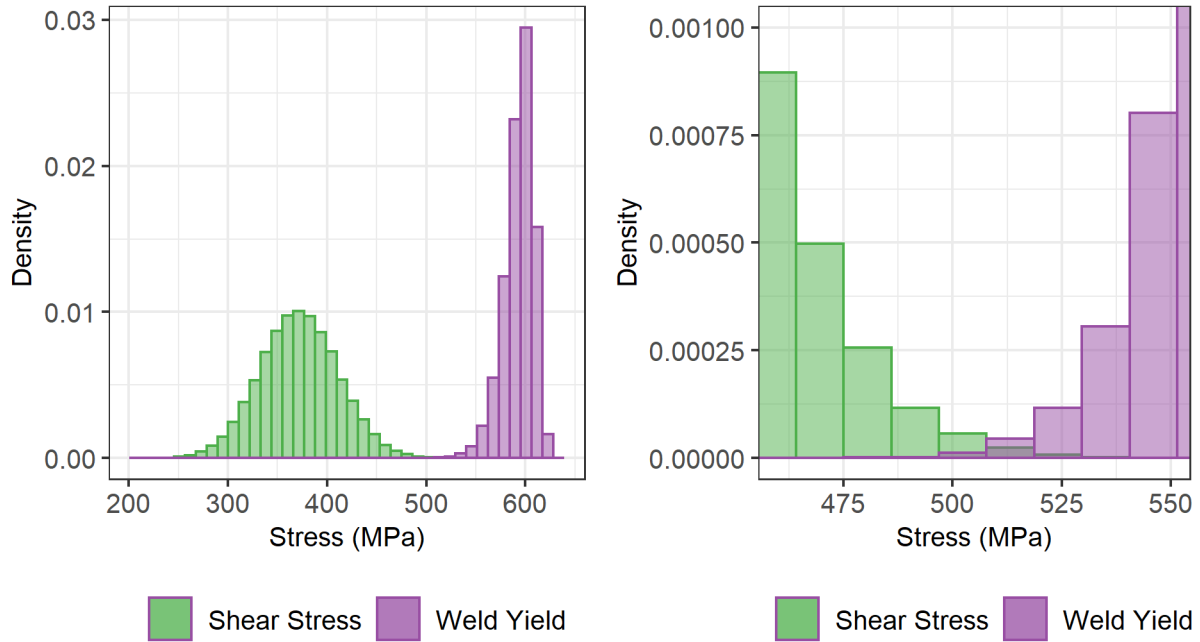


Figure 10.1. Histograms of shear stress (green) and weld yield stress (purple). The right plot provides a zoomed in view of the region where shear stress and weld yield stress overlap.

10.1.3. Summary:

- Visualizing uncertainty in important model outputs provides useful information when the QoI is a failure probability.

10.2. Visualizing uncertainty in the QoI with separation of aleatory and epistemic uncertainties

When uncertainties are separated, n_e estimates of the failure probability are calculated: one for each set of aleatory samples. The differences between these estimates can be plotted to visualize the epistemic uncertainty in the failure probability estimate.

10.2.1. Process:

1. **Determine the output to be visualized:** For this concept, pressure, radius, wall thickness and angle of the helical weld are categorized as aleatory and weld yield stress is categorized as epistemic. The aleatory and epistemic sample sizes are 4×10^4 and 1×10^3 , respectively, based on the results found in Section 9. Importance sampling is used on pressure and weld yield stress using the importance distributions that are described in Section 8.2. Since there is a separation of uncertainty, 1×10^3 probability of failure estimates will be plotted.
2. **Estimate the failure probability for each epistemic sample.** The probability of weld failure is estimated for each of the 1×10^3 epistemic samples. When IS is used, recall that the inputs are no longer equally probable, and each aleatory sample that indicates failure (indicator function = 1) must be reweighted when calculating the failure probability (Section 8.2).

- 3. Choose an appropriate plot based on Step 1.** An empirical CDF of the probability of weld failure across epistemic samples is plotted along with a measure of the epistemic uncertainty (e.g., the 95th percentile across the epistemic samples). A histogram of the failure probability could also be plotted, but a CDF was chosen because percentiles of the failure probability estimate can easily be seen from the plot.

10.2.2. Results and potential difficulties:

Figure 10.2 shows the empirical CDF of the estimated probability of weld failure (blue solid line) for each epistemic sample, along with the 95th percentile (black dashed line). As described in Section 3, when input uncertainties are separated, the output contains two types of uncertainty: 1) epistemic uncertainty dictated by the epistemic uncertainty of weld yield stress and 2) epistemic uncertainty due to a finite aleatory sample size. The 95th percentile in Figure 10.2 includes both uncertainties. Therefore, it is important that the aleatory sample size be large enough that it has a negligible effect on the overall output uncertainty. Otherwise, the uncertainty due to the epistemic uncertainty of only weld yield stress will not be accurately assessed.

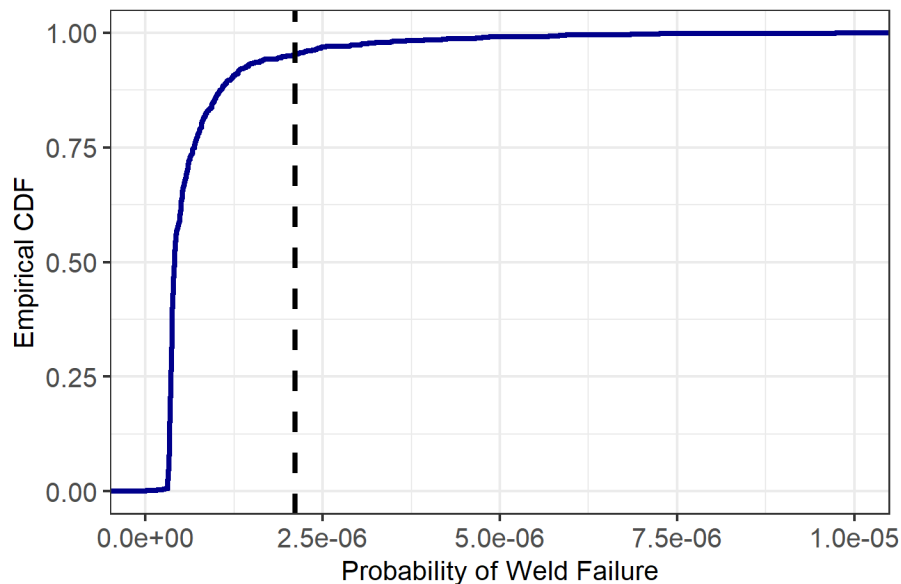


Figure 10.2. Empirical CDF of the probability of weld failure (blue solid) with vertical line at 95th percentile (black dashed).

10.2.3. Summary

- The failure probability can be plotted across epistemic samples.
- The aleatory sample size should be large enough so that uncertainty due to the aleatory sample size does not significantly contribute to the uncertainty in the failure probability estimate.

BIBLIOGRAPHY

- [1] J. M. Gere and B. J. Goodno, *Mechanics of Materials*, Eighth Edition, Cengage Learning, 2012.
- [2] G. Apostolakis, "The Distinction Between Aleatory and Epistemic Uncertainties is Important: an example from the inclusion of aging effects into PSA," in *Proceedings of PSA '99, International Topical Meeting on Probabilistic Safety Assessment*, 1999.
- [3] S. Aldor-Norman, "The Power to See: A New Graphical Test of Normality," *The American Statistician*, vol. 67, no. 4, 2013.
- [4] T. W. Anderson and D. A. Darling, "A test of goodness of fit," *Journal of the American Statistical Association*, vol. 49, no. 268, pp. 765-769, 1954.
- [5] N. L. Johnson, S. Kotz and N. Balakrishnan, *Continuous Univariate Distributions*, New York: Wiley, 1994.
- [6] R. L. Imam and W. J. Conover, "A distribution-free approach to inducing rank correlation among input variables," *Communication in Statistics - Simulation and Computation*, vol. 11, no. 3, pp. 311-334, 1982.
- [7] R. C. Smith, *Uncertainty Quantification Theory, Implementation, and Applications*, Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 2014.
- [8] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Carboni, D. Gatelli and S. Tarantola, *Global Sensitivity Analysis. The Primer*, John Wiley and Sons, Ltd., 2008.
- [9] I. M. Sobol', "On sensitivity estimation for nonlinear mathematical models," *Matematicheskoe Modelirovanie*, vol. 2, no. 1, pp. 112-118, 1990.
- [10] G. Archer, A. Saltelli and I. M. Sobol, "Sobol, sensitivity measures, anova-like techniques and the use of bootstrap," *Journal of Statistical Computation and Simulation*, vol. 58, no. 2, pp. 99-120, 1997.
- [11] J. L. Loeppky, J. Sacks and W. J. Welch, "Choosing the sample size of a computer experiment: A practical guide," *Technometrics*, vol. 51, no. 4, pp. 366-376, 2009.
- [12] C. E. Rasmussen and C. Williams, *Gaussian Processes for Machine Learning*, Cambridge, MA: MIT Press, 2006.
- [13] J. Friedman, "Multivariate adaptive regression splines (with discussion)," *Annals of Statistics*, pp. 1-141, 1991.
- [14] A. Haldar and S. Mahadevan, *Probability, Reliability, and Statistical Methods in Engineering Design*, John Wiley and Sons, Inc., 2000.
- [15] A. Karamchandani, P. Bjerager and A. C. Cornell, "Adaptive Importance Sampling," in *Proceedings, International Conference on Structural Safety and Reliability (ICOSSAR)*, San Francisco, CA, 1989.
- [16] M. D. McKay, R. J. Beckman and W. J. Conover, "Comparison of three methods for selecting values of input variables in the analysis of output from a computer code," *Technometrics*, vol. 21, no. 2, pp. 239-245, 1979.
- [17] J. C. Helton and F. J. Davis, "Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems," *Reliability Engineering & System Safety*, vol. 81, no. 1, pp. 23-69, 2003.
- [18] L. D. Brown, T. T. Cai and A. DasGupta, "Interval estimation for a binomial proportion," *Statistical Science*, vol. 16, no. 2, pp. 101-133, 2001.

- [19] A. C. Davison and D. V. Hinkley, Bootstrap Methods and Their Application (Vol. 1), Cambridge University Press, 1997.
- [20] C. Tong, "Refinement strategies for stratified sampling methods," *Reliability Engineering & System Safety*, vol. 91, pp. 1257-1265, 2006.

This page left blank

