Reliability Engineering and System Safety 107 (2012) 3–15

Contents lists available at ScienceDirect

Reliability Engineering and System Safety

journal homepage: www.elsevier.com/locate/ress

Development of a localized probabilistic sensitivity method to determine random variable regional importance ☆

Harry Millwater *, Gulshan Singh, Miguel Cortina

Department of Mechanical Engineering, University of Texas at San Antonio, San Antonio, TX, USA

ARTICLE INFO

Article history:
Received 14 October 2010
Received in revised form
17 March 2011
Accepted 29 April 2011
Available online 7 May 2011

Keywords:
Probabilistic sensitivities
Score Function sensitivities
Localized sensitivity

ABSTRACT

There are many methods to identify the important variable out of a set of random variables, i.e., "inter-variable" importance; however, to date there are no comparable methods to identify the "region" of importance within a random variable, i.e., "intra-variable" importance. Knowledge of the critical region of an input random variable (tail, near-tail, and central region) can provide valuable information towards characterizing, understanding, and improving a model through additional modeling or testing. As a result, an intra-variable probabilistic sensitivity method was developed and demonstrated for independent random variables that computes the partial derivative of a probabilistic response with respect to a localized perturbation in the CDF values of each random variable. These sensitivities are then normalized in absolute value with respect to the largest sensitivity within a distribution to indicate the region of importance. The methodology is implemented using the Score Function kernel-based method such that existing samples can be used to compute sensitivities for negligible cost. Numerical examples demonstrate the accuracy of the method through comparisons with finite difference and numerical integration quadrature estimates.

1. Introduction

Probabilistic sensitivity analysis is often a critical component of a risk assessment. Its purpose is traditionally to identify the important variables in an analysis in order to focus resources, computational, experimental, or both, on the parameters that most affect the system response. The word "important" has different meanings in different contexts. Within a probabilistic analysis, it usually refers to variables that are modeled as random whose variation has the largest affect on the response.

Frey and Patil [1] provide an overview article discussing ten sensitivity methods, both probabilistic and deterministic, such as automatic differentiation, regression, scatter plots, ANOVA, and others. Similarly, Hamby [2] discusses fourteen different sensitivity methods including partial derivatives, variation in inputs by 1 standard deviation, regression, Smirnov test, Cramer-von Mises test, and others.

Scatter plots and correlation coefficients are a straightforward and low-cost method to define importance [1,3]. Variables that show a clear "relationship" between the variable and the response are important. Variables for which the scatter plots largely reflect the marginal distribution are not important. Often, the correlation coefficient is used to obtain a numerical value of the relationship.

Linear regression is a well-known method to assess the importance of a random variable [3,4]. The standardized regression coefficients indicate the amount of variance of the response explained by each variable and the amount of the response variance defined by the entire linear model and groups of variables. Stepwise regression is particularly useful to determine the parsimonious model that best accounts for the response variance given a fixed number of variables, e.g., the best linear regression model given *k*-out-of-*N* variables, where *N* is the total number of random variables.

Variance-based sensitivity methods are powerful in that these methods identify the amount of the total variance that can be attributed to each input random variable and the amount the variance would be reduced if a particular random variable were to be fixed at a specific value [4–6]. Main effects, higher order, and interaction effects can be explored. Other similar sensitivity methods based on the Kullback–Liebler divergence are similar in spirit to variance-based sensitivity metrics but allow consideration of differences higher order than second moments [7].

A number of sensitivity methods are available for the First Order Reliability Method (FORM). Sensitivity factors (derivatives

☆ The results indicate that accurate localized sensitivities can be obtained for the dominant random variables as long as sufficient samples are available.

* Corresponding author. Department of Mechanical Engineering, University of Texas at San Antonio, One UTSA Circle, San Antonio, TX 78249, USA.
Tel.: +1 210 458 4481 (V); fax: +1 210 458 6504 (F).
E-mail address: harry.millwater@utsa.edu (H. Millwater).

0951-8320/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ress.2011.04.003

of the safety index with respect to the random variables) [8], derivatives of the probability of failure with respect to the random variable parameters, e.g., $\partial P/\partial \mu$, $\partial P/\partial \sigma$ [8], and omission factors [9] are computed as by-products of an analysis.

Various authors develop and discuss the "Score Function (SF)" method for the computation of partial derivatives of a probabilistic performance function (probability-of-failure or response moment) with respect to parameters of the underlying input probability distributions [10–18]. This method provides local partial derivatives of the probability-of-failure or response moments with respect to the parameters of the input PDFs, e.g., $\partial P/\partial \mu$, $\partial P/\partial \sigma$. Implementation of the methodology is convenient using sampling methods. A significant advantage is that negligible additional computing time is required to determine the sensitivities since the same samples used to compute the probabilistic response can be reused to compute the sensitivities; however, this assumes that sufficient samples are already available in order to obtain convergence of the sensitivity estimates. Wu and Mohanty [15] use the SF method to compute the partial derivative of the response mean with respect to the inputs and combine this information with hypothesis testing to identify the important variables. Sues and Cesare [16] use the SF method to compute the partial derivative of the response standard deviation with respect to the parameters of the input PDFs. Millwater et al. [18] have extended the method to input random variables with arbitrary dimensioned correlated multivariate normal distributions, including providing sensitivities with respect to the correlation coefficients.

The Score Function method has some advantages and disadvantages over variance-based importance measures. One advantage is that the quantity of interest can be defined and focused upon, e.g., the probability-of-failure or the mean response. For example, the sensitivities of the mean of the response may be completely different than the sensitivities of the probability-of-failure. Although the results are local and depend on the values used when computing the sensitivities, sometimes this is just what is needed, for example, as in reliability-based design. A disadvantage is that a large number of samples may be needed in order to obtain convergence in the sensitivity estimates.

In all the methods listed above, the purpose is to identify the important variable or groups of variables among all variables. There is no focus on which part of a variable is important such as the left or the right tails, center region, near center, etc. This information can be useful in a number of contexts. For example, in consideration of experimental design, it would be useful to know where to tailor experiments if possible, how much data is needed to characterize the important region, and what computational strategies may suffice during analysis. Therefore, a methodology was developed and is presented here to compute the partial derivative of the probabilistic response (probability-of-failure, response mean, or standard deviation) with respect to a set of discretized CDF values that span the distribution. The methodology is developed for independent random variables and folded into the Score Function approach such that existing samples can be reused to estimate the sensitivities.

2. Methodology

The basic concept is to discretize each random variable CDF into regions using discretization points $X_{i,j}$ (the jth location of random variable i), introduce a local disturbance into the CDF for each region centered at $X_{i,j}$, then determine the relative change in a probabilistic response: probability-of-failure, P_f, or response moments (mean, μ_g, and standard deviation, σ_g), for each regional disturbance.

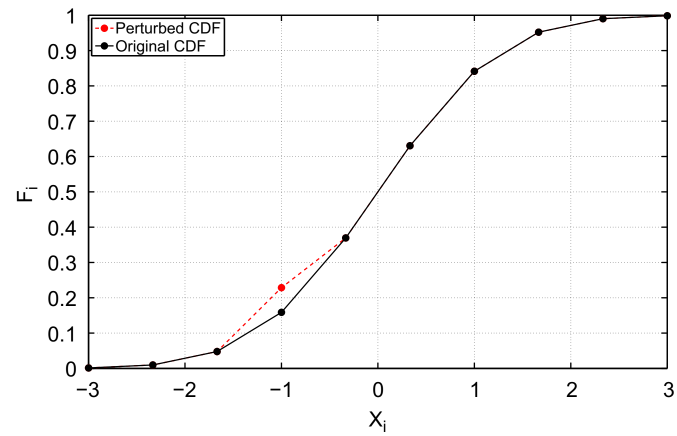

Fig. 1. Schematic of the effect of the localized sensitivity method on the CDF.

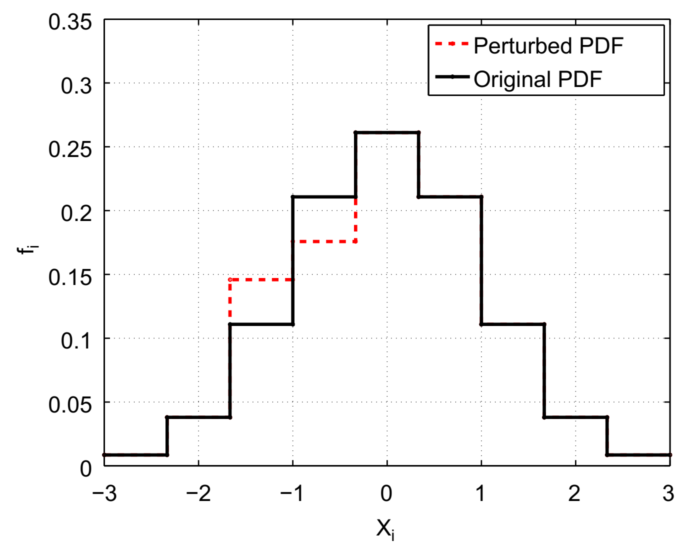

Fig. 2. Schematic of the effect of the localized sensitivity method on the PDF.

The concept is shown in Fig. 1, whereby the CDF, F_i, for random variable X_i is discretized at discrete points $x_{i,j}$, then a local disturbance is input into the CDF at a discretization point, see red dashed line in Fig. 1, yielding $\hat{F}_i$. The perturbation at x_j for random variable i only extends over the range $x_{j-1} < x < x_{j+1}$. The points at which to discretize the CDF are arbitrary and user defined. The effect of the perturbation on the PDF is given in Fig. 2.

An estimate to the partial derivative of a probabilistic response, L, with respect to a CDF of random variable X_i at a specific location x_j, $F_i(x_j)$, can then be obtained using the finite difference method, namely

$$\frac{\partial L}{\partial F_i(x_j)} \approx \frac{\Delta L}{\Delta F_i(x_j)} = \frac{L(\hat{F}_i(x_j)) - L(F_i(x_j))}{\Delta F_i(x_j)} \qquad (1)$$

where L denotes either the probability-of-failure, P_f, the mean of the response, μ_g, or the standard deviation of the response, σ_g. The response is defined by an arbitrary function $g(\mathbf{x})$ of the random variables $\mathbf{X}$ with failure defined when $g(\mathbf{x}) \le 0$.

Calculation of the sensitivity using the finite difference method, while possible, is arduous in that multiple analyses are required ($K+1$ for K discretized regions for each random variable) and, if sampling is used, a large number of samples are required