C▼FER
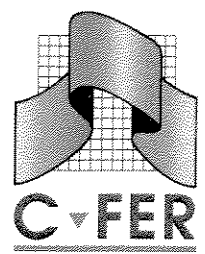
200
Karl Clark Road
Edmonton, Alberta
Canada T6N 1H2

Tel: (403) 450-3300
Fax: (403) 450-3700

Probabilistic Decision Analysis Using Influence Diagrams

PIRAMID Technical Reference Manual No. 2.1

AG

Confidential to
C-FER's Pipeline Program
Participants

Prepared by
M. A. Nessim, Ph. D., P. Eng.
and
H. P. Hong, Dr. Eng., P. Eng.

Centre For Engineering Research Inc.
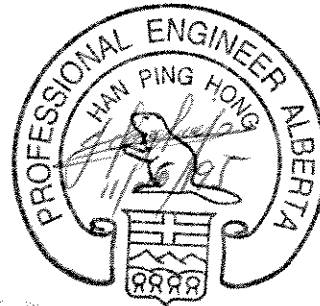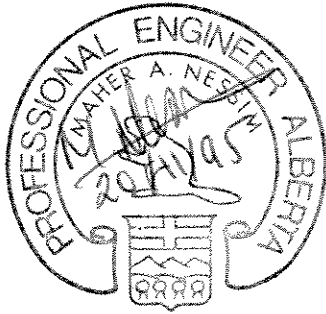
October 1995
Project 94022

Probabilistic Decision Analysis Using Influence Diagrams
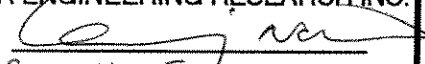
PIRAMID Technical Reference Manual No. 2.1

Confidential to
C-FER's Pipeline Program
Participants

Prepared by
M. A. Nessim, Ph. D., P. Eng.

and
H. P. Hong, Dr. Eng., P. Eng.

PERMIT TO PRACTICE
CENTRE FOR ENGINEERING RESEARCH INC.
Signature:
Date: 20· 11 ·95
PERMIT NUMBER: P4487
The Association of Professional Engineers,
Geologists and Geophysicists of Alberta

October 1995
Project 94022

NOTICE

Restriction on Disclosure

TABLE OF CONTENTS

LIST OF FIGURES

1.0 INTRODUCTION

1.1 Background

This document is part of the deliverables of C-FER's joint industry research program on risk-based optimization of pipeline integrity maintenance. The goal of the research program is to develop models and software tools that can assist pipeline operators in making optimal decisions regarding integrity maintenance activities for a given pipeline or pipeline segment. The software resulting from this joint industry program is called PIRAMID (Pipeline Risk Analysis for Maintenance and Inspection Decisions). This document forms part of the technical reference manuals for the program.

The risk associated with a pipeline failure depends on a number of interrelated uncertain parameters. To begin with, there is uncertainty about whether or not the pipeline will fail. If a failure occurs, there is uncertainty regarding the failure location and weather conditions at the time of failure. These factors determine the amount of product released, the types of hazard that may develop (such as jet fires, vapour cloud fires or explosions), and ultimately the costs, number of fatalities and environmental impact associated with the failure.

The benefits associated with a given integrity maintenance action are measured in terms of its impact on the risk associated with operating the pipeline. In order to find the optimal set of integrity maintenance actions in the presence of the above-mentioned uncertainties, a probabilistic optimization methodology is required. The basic elements of this methodology are:

1. a method to derive the probability distributions of the major consequence parameters (*i.e.*, costs, losses in life and spill volumes) from the probability distributions of the failure type, environmental conditions, and line operating parameters; and

2. a model to use the information from (1) to calculate the total expected value associated with a given choice, so that the choice with the highest expected value can be adopted.

The probabilistic optimization methodology selected for this purpose is decision influence diagrams (Shachter 1986 and Call and Miller 1990). An introduction to this methodology and an explanation of the reasons for its selection are given in PIRAMID Technical Reference Manual No. 1.2 (Stephens *et al.* 1995a).

Introduction

The influence diagram methodology, as it exists in the literature, was developed in the field of management science; it has some limitations that affect its applicability to engineering problems (see Section 2.0 for a description of these limitations). As part of the present program, new developments were undertaken to extend and generalize the methodology. This work resulted in a new influence diagram methodology that is more suited to risk-based optimization of pipeline systems. The new influence diagram methodology is used in the present program as a tool for all probabilistic inference and risk-based optimization problems. As such, an understanding of the methodology is essential for effective use of the software products developed.

1.2   Objective and Scope

This report describes the influence diagram methodology and its use as a basis for probabilistic inference and risk-based optimization. The original influence diagram methodology as it exists in the literature is outlined, and the extensions developed by C-FER are described. The steps involved in solving an influence diagram are explained using a simplified representation of a pipeline risk-based optimization problem. This example problem demonstrates the inputs needed to characterize a risk-based decision problem and the outputs obtained from the influence diagram.

Specifically, the purpose of this document is to explain the process of building, solving, and interpreting the results of an influence diagram to a user of software products that use influence diagrams as a method of risk-based optimization. Such users are assumed to be familiar with the basic concepts of risk analysis and probability distributions, but knowledge of the details of probabilistic analysis methods is not assumed. The document is therefore written in an informal manner, focusing on the concepts and avoiding mathematical details. In addition, it only covers aspects of the methodology that are used in PIRAMID. The actual algorithms are described in Appendices A and B, and a more formal description of the original influence diagram methodology can be found in Shachter (1986).

2.0   DECISION ANALYSIS USING INFLUENCE DIAGRAMS

2.1   Introduction

This section provides a general description of an influence diagram and outlines how it is used as a basis for optimization under uncertainty. The example problem shown in Figure 2.1 is used as a basis for the discussion. This problem deals with optimizing corrosion integrity maintenance actions for a segment of gas pipeline. It is noted at the outset that this influence diagram is only a small subset of the diagram that would be required to provide a realistic solution to a practical problem. However, it contains the main elements of the influence diagram methodology and is therefore suitable as an illustrative example. More complete influence diagrams have been developed and used as a basis for decision making for actual pipeline problems as described in PIRAMID Technical Reference Manual No. 3.1 (Stephens *et al.* 1995b).

2.2   Influence Diagram Representation and Terminology

The basic elements of the diagram are the *nodes* (square, circles and rounded square) and *arrows*. The nodes contain parameters that are relevant to the decision problem, whereas the arrows represent dependence relationships between these parameters. The influence diagram is explained below using the example in Figure 2.1 as a basis. The notation and terminology used are summarized in Figure 2.2, and these will be explained more fully as they arise in the discussion.

(1) Integrity Action

The first node in the diagram is the *decision node* (denoted by a square). This means that the value of the node parameter is to be selected by the decision maker. The node parameter (decision) in this case is called Integrity Action, and it can take one of three values representing the available choices. These values are No Action, Run Pig and CP Survey as indicated in the box adjacent to node (1) in Figure 2.1. It is assumed that problems identified by the pig or from the CP Survey will be corrected.

Decision Analysis Using Influence Diagrams

(2) Pipe Performance

The second node in the diagram pertains to the parameter referred to as Pipe Performance. This parameter is assumed to take one of two values, namely Safe or Failure. Because it cannot be determined with certainty whether or not the pipe is going to fail, this parameter is uncertain (or random). The node is therefore called a *chance node* and is denoted by a circle. Because the parameter can take one of two discrete values, it is represented by a discrete probability distribution (see box adjacent to node 2 in Figure 2.1).

The arrow starting at node (1) and ending at node (2) indicates that Pipe Performance is dependent on the Integrity Action selected. The node at which an arrow originates is referred to as a *direct predecessor* to the node at which the arrow ends. Conversely, the node at which the arrow ends is referred to as a *direct successor* of the node at which the arrow originates. Therefore, Pipe Performance is a direct successor of Integrity Action, whereas Integrity Action is a direct predecessor of Pipe Performance.

The solid arrow (as opposed to the dashed arrows in Figure 2.1), connecting Integrity Action and Pipe Performance denotes *conditional dependence*. This means that the direct successor node is probabilistically conditional on the direct predecessor node, so that knowledge of the value of the direct predecessor does not determine the value of the direct successor, but determines its *probability distribution*. In this case, knowledge of the Integrity Action adopted does not determine with certainty whether or not the pipe will fail, but it can be used to assign the probabilities of Failure and Safe pipe. So, if a more reliable integrity action is adopted, the benefits will be reflected in a lower failure probability. A chance node that receives only conditional arrows is referred to as a *conditional node*.

(3) Cost

Cost is treated as a random variable (as indicated by the circular chance node) to account for uncertainties regarding actual costs of integrity maintenance actions and the costs of failure. One difference between this node and the Pipe Performance chance node, is that the cost is a *continuous parameter* that can take any value in a given range, not a discrete parameter that can take specific values. Cost is therefore represented by a continuous probability distribution as indicated in the box adjacent to the node in Figure 2.1. The original influence diagram methodology does not

4

allow for continuous parameters and this is one of the extensions that were developed under the present study.

The total cost is the sum of integrity action costs and failure costs (if failure occurs), and therefore the Cost node is dependent on Integrity Action and Pipe Performance as indicated by the arrows. In standard influence diagram terminology this can be stated as: the Cost node has Integrity Action and Pipe Performance as direct predecessors. Solid line arrows are used, indicating conditional dependence, which again means that the probability distribution of the total cost can be assigned for given combinations of a specific integrity action and a specific pipe performance.

(4) Failure Location

Failure Location is a chance node with a continuous parameter. This parameter represents the kilometer post at which the failure occurs along the segment. The probability distribution given for this parameter in Figure 2.1 is uniform indicating the assumption that a failure is equally likely to occur anywhere along the pipeline segment. Failure Location has no predecessors and is therefore referred to as an *orphan node*.

(5) Number of Fatalities

The Number of Fatalities is an uncertain parameter that is assigned a chance node. It is directly dependent on Pipe Performance and Failure Location as indicated by the arrows in Figure 2.1. The dashed arrow lines indicate *functional dependence*. This means that the Number of Fatalities is defined as a deterministic function of the Pipe Performance (P) and Failure Location (L), so that if P and L are known the Number of Fatalities can be calculated without any uncertainty. This does not mean that the Number of Fatalities in a deterministic parameter. It is in fact an uncertain parameter because both of the parameters on which it depends (namely P and L) are uncertain - hence the probability distribution in the adjacent box in Figure 2.1. It is recognized that the Number of Fatalities depends on many other including wind direction, atmospheric stability, temperature, product type, pipe diameter and operating pressure. Only two parameters (namely, P and L) are used in this illustrative example for simplicity, but the concept is applicable to an arbitrary number of parameters. The concept of functional dependence is not part of the original influence diagram methodology and was developed by C-FER specifically for this project.

Decision Analysis Using Influence Diagrams

It is also noted that the Number of Fatalities is a discrete parameter and its treatment as a continuous parameter is an approximation.

Distinction can now be made between *predecessors* and direct predecessors. By examining Figure 2.1 it can be seen that the Number of Fatalities is dependent upon the decision node. This is because the Number of Fatalities depends on Pipe Performance, which in turn depends on the decision node. Therefore, Integrity Action is a predecessor of the Number of Fatalities (although not a direct predecessor). More formally, node A is said to be a predecessor of node B if a path can be found (through other nodes) from B to A. Recall that if A immediately precedes B (*i.e.*, the path does not contain any other nodes) then A is a direct predecessor of B. An analogous distinction can also be made between successor and direct successor nodes.

(6) Value

The Value node is a chance node that contains the parameter being optimized in the decision problem. Use of the "rounded square" distinguishes the Value node from other chance nodes. Value in this example is defined as functionally dependent on Cost and Number of Fatalities.

2.3   Optimization Using Influence Diagrams

The basic premise of decision theory is that a given decision maker can define a Value function such that different choices can be ranked according to the *expectation* of the Value function (where expectation is defined as the mean or expected value). As was shown in Section 2.2, the outcome of a given choice depends on a sequence of uncertain parameters and therefore the Value associated with a given decision is not known with certainty at the time the decision is being made. This is why the decision is made on the basis of optimizing the *expectation* of the Value function. The expectation is calculated as the sum of all possible Value outcomes, each weighted by its probability of occurrence.

For the influence diagram in Figure 2.1, the Value (V) is defined as a deterministic function of the Cost (C) and the Number of Fatalities (N). A common simple approach is to define the value function as

$$V = -(C + aN)$$

[2.1]

6

Decision Analysis Using Influence Diagrams

where a is a constant that converts losses in life to financial costs. Because of the negative sign in this equation, maximizing the expectation of V corresponds to minimizing the total expected cost including the cost associated with losses in life. This approach implies that the decision maker can express all consequences in terms of monetary costs, and that the best choice is the one that minimizes the expected cost. It can be shown that this approach is not always appropriate, and other types of functions may better represent the true preferences of decision makers (see e.g., Keeney and Raiffa 1976). The selection of an appropriate Value function is discussed more fully in PIRAMID Technical Reference Manual No. 3.1 (Stephens *et al.* 1995b). It is sufficient for the purpose of this report to assume that the Value (V) is defined as a deterministic function of the Cost (C) and the number of Fatalities (N).

Based on the foregoing discussion, the solution of the influence diagram consists of finding the expectation of the Value function for each possible decision. Once this is done, the optimal choice is identified as the one yielding the highest expected Value.

2.4 Comments on Influence Diagram Implementation

2.4.1 Including Deterministic Parameters

It is important to note that the influence diagram in Figure 2.1 includes only two types of parameters. These are decision parameters and uncertain parameters (of which the Value parameter is a special case). In addition to these two types, most problems will involve deterministic parameters such as the population density and its variation along the route, the type and physical properties of the product in the pipeline, and the pipeline diameter.

Each of these deterministic parameters is included as part of the specific node for which it is used. For example, the population density and its variation along the route is required to evaluate the Number of Fatalities from the Pipe Performance and the Failure Location. It would therefore be included as part of the information required to characterize the Number of Fatalities node. Deterministic parameters therefore do not appear explicitly in a standard influence diagram representation. Influence diagrams deal only with decision and uncertain (random) parameters, and this should be taken into consideration in interpreting an influence diagram.

Decision Analysis Using Influence Diagrams

2.4.2 Building Influence Diagrams

With reference to the discussion in Section 2.4.1, it is recognized that there is a degree of uncertainty associated with almost all parameters, and that the treatment of any specific parameter as probabilistic or deterministic is a choice made on a case-by-case basis, depending largely on the relative magnitude of the associated uncertainty. If, for example, the population density is shown to have significant uncertainty, then it should be included as a node in the influence diagram and the appropriate arrows added to characterize its relationships to other diagram nodes. In addition, there may be different models available to calculate a given node, which use different input parameters. This is why a unique influence diagram does not necessarily exist for any particular problem

The above argument suggests that it is desirable for the software to be flexible with respect to building and modifying influence diagrams for particular problems. To achieve this, PIRAMID includes separate modules that create an influence diagram, lead the user through definition of the input parameters for each node in the appropriate sequence, solve for node parameters by calling specific node algorithms, and solve the diagram to obtain the optimal choice. A new diagram can be easily built by creating initialization code that defines the diagram nodes and their interdependencies (arrows).

It is possible to develop a user interface that allows users to define the initialization data mentioned in the previous paragraph, and thus build their own influence diagrams. However, the level of effort required for this is beyond the scope of the present program. In addition, if users define their own influence diagrams, they would also have to be responsible for defining the specific node algorithms. For example, the diagram in Figure 2.1 assumes that the Number of Fatalities is defined as a function of Pipe Performance and Failure Location. If this diagram were modified to include an additional parameter such as wind speed, then a new model that uses wind speed in calculating the Number of Fatalities would be expected. Therefore, much of the technical modeling required to solve the problem would be left to the user.

Based on the above, the approach adopted in PIRAMID is to develop specific influence diagrams. Once these diagrams are approved by the program Steering Committee, they are used in the analysis. As mentioned earlier in this section the program has been designed in such a manner as to make it easy for C-FER to modify these diagrams if required.
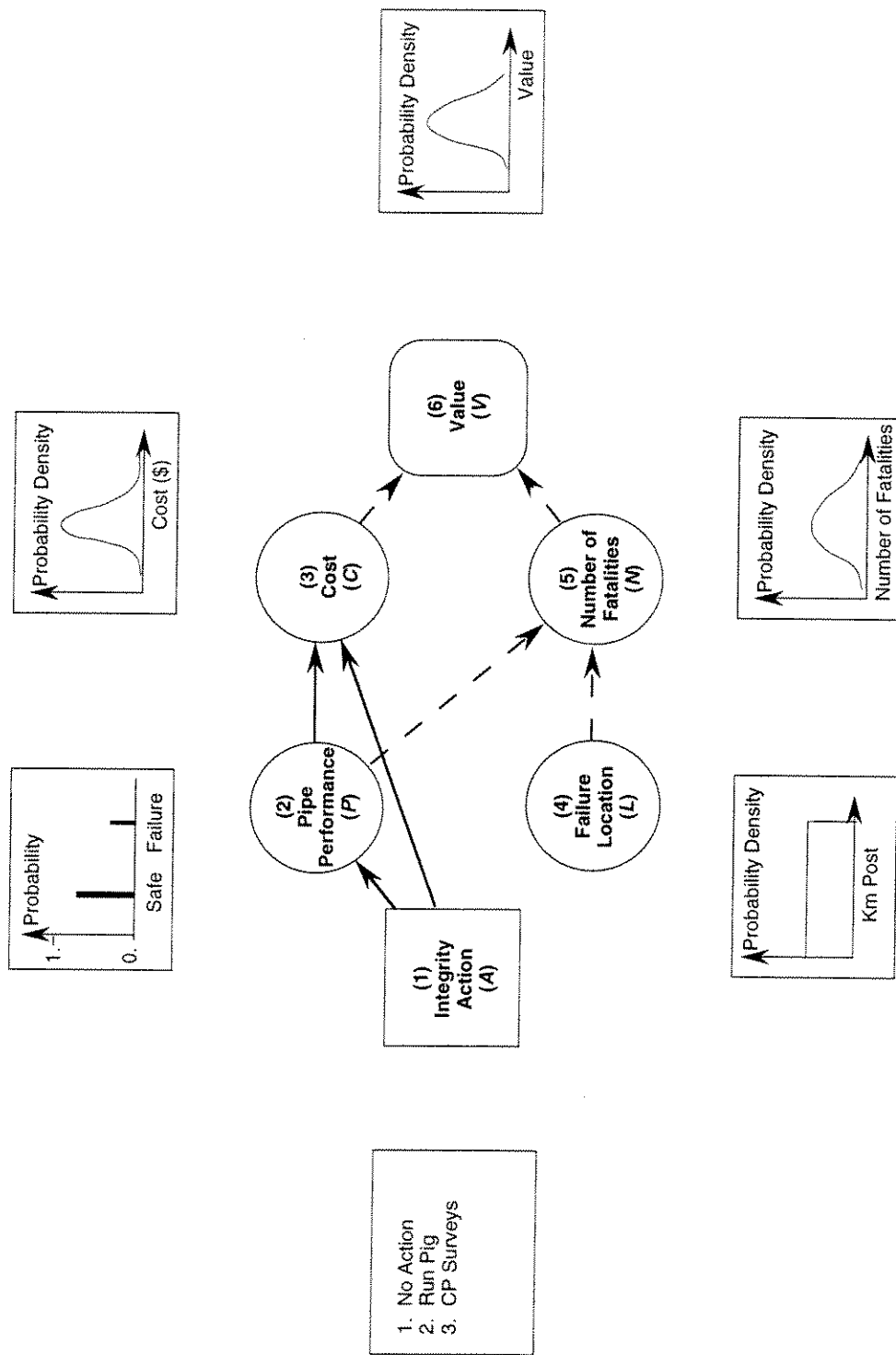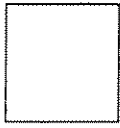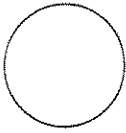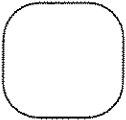
Figures

Figure 2.1 Simplified influence diagram for a pipeline corrosion integrity maintenance problem

Node Notation

☐	Decision node:	Indicates a choice to be made
◯	Chance node:	Indicates uncertain parameter or event (discrete or continuous)
◯	Value node:	Indicates the criterion used to evaluate consequences

Arrow Notation

──▶	Solid Line arrow:	Indicates probabilistic dependence
─ ─▶	Dashed line arrow:	Indicates functional dependence
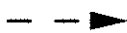
Other Terminology

Predecessor to node A :   Node from which a path leading to A begins

Successor to node A: Node to which a path leading to A begins

Functional predecessor: Predecessor node from which a functional arrow emanates

Conditional predecessor: Predecessor node from which a conditional arrow emanates

Direct predecessor to A: Predecessor node that immediately precedes A
                (i.e. the path from it to A does not contain any other nodes)

Direct successor to A: Successor node that immediately succeeds A
                (i.e. the path from A to it does not contain any other nodes)

Direct conditional predecessor to A: A predecessor node from which the path to node A contains
(A must be a functional node)   only one conditional arrow (may contain functional arrows)

Functional node: A chance node that receives only functional arrows

Conditional node: A chance node that receives only conditional arrows

Orphan node: A node that does not have any predecessors

Figure 2.2   Influence diagram notation and terminology

3.0   SOLVING AN INFLUENCE DIAGRAM

3.1   Introduction

The influence diagram discussed in Section 2.0 is a qualitative representation of the decision problem. It specifies all decision and uncertain parameters influencing the optimization, and describes the interdependence between these parameters. In order to use this diagram to solve the optimization problem, each diagram node must be quantified. Quantifying a node means defining the conditional probability distributions of the node parameter for all possible values of its conditional predecessor nodes. Once all nodes are defined, the diagram can be solved to calculate the expected value for each choice, giving the basis for selecting the optimal solution.

The diagram solution algorithm is essentially a method that identifies the conditional predecessors of any node, derives its conditional probability distributions and calculates the expectations of node parameters conditional on the choices. The following sections describe the influence diagram solution algorithm used in the software delivered under this program. This description is based on the simplified example problem given in Section 2.0 (Figure 2.1). The actual algorithm used is described in Appendix A, with specific mention of the parts that were developed under this project.

The algorithm given here has some extensions over the standard influence diagram methodology as mentioned in Section 2.2. On the other hand, some aspects of the standard algorithm were omitted because they are not used in the present program. A complete description of the standard influence diagram solution algorithm can be found in Shachter (1986).

3.2   Sequence of Node Definition

Due to the interdependence between diagram nodes, there are constraints regarding the sequence in which the different diagram nodes can be defined. For example, the probability distribution of Pipe Performance depends on the Integrity Action choice made (see Section 2.1), and therefore it is not possible to define the Pipe Performance node until the Integrity Action node is defined. In general, a node becomes accessible only after all of its direct predecessors are fully defined. The term *defined* is used to indicate a node for which all the required probability distributions have been defined.

Solving an Influence Diagram

Figure 3.1 shows the sequence of node definition for the example problem. Initially, only the nodes that do not have any predecessors (the orphan nodes) can be defined. These are nodes 1 and 4 (see Figure 3.1a). Once nodes 1 and 4 are defined, node 2 becomes available (Figure 3.1b), and when 2 is defined, 3 and 5 become available (Figure 3.1c). Finally, when nodes 3 and 5 are defined the Value node becomes available (Figure 3.1d), and the whole diagram is completely defined (Figure 3.1e). It is noted that the sequence in which the nodes are defined is not unique. Possible sequences include 1-2-3-4-5-6, 4-1-2-5-3-6 or 1-4-2-3-5-6.

3.3   Defining Influence Diagram Nodes

3.3.1   Node Conditional Distributions

To understand the process of solving an influence diagram, it must be recognized that each chance node (including the value node) in the influence diagram embodies a series of conditional probability distributions, corresponding to specific values of its conditional predecessor nodes. In this context, a *conditional probability distribution* is defined as the probability distribution of the node parameter *given* a certain combination of values of its predecessor nodes. This can be explained by examining the Pipe Performance node in Figure 3.1. This node is conditional on the decision (Integrity Action node), which means that for each choice, there is a different conditional probability distribution of the Pipe Performance. In this case, there are three conditional probability distributions corresponding to the three possible choices, namely No Action, Pig Run and CP Survey.

Based on the above, the definition of each node involves the following two steps:

1. Identify the node's direct conditional predecessors and find all combinations of possible outcomes of these predecessors. The number of these combinations is equal to the number of conditional distributions required to define the node.

2. Define the conditional probability distributions of the node parameters for all combinations identified in (1).

These steps are described in more detail for different nodes in Sections 3.3.2 and 3.3.3.

Solving an Influence Diagram

3.3.2 Defining Orphan and Conditional Nodes

Orphan and conditional nodes are usually defined by direct input to the influence diagram. This is the case for orphan nodes (Nodes 1 and 4 in Figure 3.2) because they do not have any predecessors from which they can be calculated. Conditional nodes (nodes 2 and 3 in Figure 3.2) correspond to basic parameters for which the probability distributions are usually defined as input.

The definition of orphan and conditional nodes in the diagram is described as follows:

1. Integrity Action (the decision node). The input required for complete definition of the decision node is the number of choices and the designation (or title) of each choice.

2. Pipe Performance. Pipe Performance is conditional only on the decision node. Its input therefore consists of three conditional probability distributions, one for each Integrity Action (see Figure 3.2). Each conditional probability distribution of Pipe Performance is defined by the probability of Failure and the probability of Safe performance for the duration considered in the analysis (*e.g.*, 1 year). These two probabilities must add up to 1.0.

    Instead of defining the conditional probability distributions of the node directly, it may be convenient to calculate them from other inputs. For example, pipe performance is often characterized by the failure rate per km.year. It may therefore be more convenient to define the failure rate for each Integrity Action option and use it to calculate the required probability distribution. In this case the probability of Failure can be calculated as the failure rate multiplied by the segment length in kilometers and the duration of the analysis in years. The Probability of Safe performance can then be obtained by subtracting the probability of Failure from 1.0. This does not change the basic concept, since the input in this case is just a different form of the required probability distribution.

3. Cost. The Cost node is conditional on both Integrity Action and Pipe Performance. Therefore, a conditional distribution is needed for each combination of these two parameters (see Figure 3.2). There are three possible Integrity Actions (No Action, Run Pig and CP Survey) and two possible Pipe Performance levels (Safe and Failure), leading to 2 x 3 = 6 input conditional distributions for the Cost.

4. Failure Location. Failure Location is an orphan chance node that does not have any predecessors. Its required input consists of one probability distribution representing the location of a failure if one occurs (see Figure 3.2).

Solving an Influence Diagram

3.3.3 Defining Functional Nodes

The two functional nodes in the influence diagram are node 5 representing the Number of Fatalities and node 6 representing the Value. As discussed earlier, these nodes are defined as functions of their direct predecessors. For example, the Number of Fatalities (N) is a function of the Pipe Performance (P) and Failure Location (L) in the form

$$N = g(P, L)$$

[3.1]

where g denotes an arbitrary function. Equation [3.1] along with the probability distributions of P and L, can be used to calculate the probability distribution of N. This calculation can be based on any standard probability integration method such as Monte Carlo simulation or First and Second Order Reliability Methods (see Appendix B).

The number of conditional distributions required for node 5 (Number of Fatalities) can be determined by examining its direct predecessors. It can be seen from Figure 3.2 that the Failure Location (node 4) has only one probability distribution, whereas the Pipe Performance (node 2) is defined by three conditional probability distributions. The Number of Fatalities is therefore characterized by three probability distributions, each of which resulting from using one of the conditional distributions of node 2 with the distribution of node 4. Since each distribution in node 2 is conditional on a given option in the decision node, the resulting probability distribution of node 5 will be conditional on the same option (see Figure 3.2). The Number of Fatalities is therefore also conditional on the Integrity Action (decision) node.

More formally, it can be stated that Integrity Action is a *direct conditional predecessor* of the Number of Fatalities. By definition, a direct conditional predecessor of a functional node (A) is a predecessor node from which the path to A contains only one conditional arrow (it can contain any number of functional arrows). Based on this definition, the results reached by inspection in the previous paragraph can be reproduced using the following systematic procedure:

1. Identify the direct conditional predecessors of the functional node. In this case the Number of Fatalities has only one direct conditional predecessor, namely the Integrity Action node.

2. The functional node *inherits* the direct conditional predecessor from each of its direct (functional) predecessors. In the example problem, the Number of Fatalities inherits conditional dependence on Integrity Action from Pipe Performance.

12

Solving an Influence Diagram

3. For each combination of outcomes of the direct conditional predecessors, the probability distribution of the functional node being considered is evaluated by using the corresponding distributions of the direct (functional) predecessor nodes in the deterministic functional relationship. This means that the Number of Fatalities for the No Action outcome of Integrity Action node is obtained by using the probability distribution of Pipe Performance for the No Action case with the probability distribution the Failure Location in Equation [3.1].

Once the Number of Fatalities node is defined as discussed above, the influence diagram can be simplified as shown in Figure 3.3b. The simplification is possible because once the function has been used to calculate the probability distribution of the Number of Fatalities, the functional arrows are no longer needed in the diagram (unless the node needs to be re-evaluated due to changes in the diagram input). Elimination of the functional arrow between nodes 4 and 5 causes node 4 to be completely isolated from the diagram. This reflects the fact that the information embodied in node 4 was passed to node 5. Node 4 can therefore be eliminated from the influence diagram. The last modification to the diagram in Figure 3.3b is to add the inherited conditional arrow from node 1 to 5.

The Value node is another functional node for which the function takes the form

$$V = g(C, N)$$

[3.2]

where C is the Cost and N is the Number of Fatalities. Calculation of the Value node follows the same procedure described earlier for the Number of Fatalities. The direct conditional predecessors are Pipe Performance (inherited from the Cost node) and Integrity Action (inherited from both the Cost and Number of Fatalities). Based on this the Value node is defined by 6 distributions, one for each combination of Integrity Action and Pipe Performance (see Figure 3.2). As before, these distributions can be calculated by using the corresponding conditional distributions of the Cost and Number of Fatalities and Equation [3.2] in an appropriate probability integration algorithm.

Definition of the Value node results in simplifying the diagram by eliminating the functional arrows, adding the inherited conditional arrows and eliminating disconnected nodes. Figure 3.3c gives the simplified diagram, which shows that the Number of Fatalities and Cost nodes are eliminated. The Number of Fatalities is eliminated because it is disconnected from the diagram once the functional arrow from node 5 to 6 is eliminated and the conditional arrow from 1 to 5 is

Solving an Influence Diagram

replaced by a conditional arrow from 1 and 6. A similar logic leads to elimination of node 3 and creation of the conditional arrow from 2 to 6.

3.3.4 Calculating the Expectation of the Value Node

The influence diagram resulting from solving functional nodes contains only conditional nodes (see Figure 3.3c). This can be solved using the original influence diagram algorithm (Shachter 1986), which was developed for conditional dependence only. The solution involves calculating the expectation of the Value, and as mentioned in Section 2.4, the optimal choice is identified based on maximizing this quantity.

Calculation of the Value expectation is achieved by eliminating diagram nodes until only the decision and Value nodes remain (see Figure 3.4). The final diagram (Figure 3.4b) gives the probability distribution of the Value node conditional only on the decision, and this implies that the mean of each conditional distribution is the required Value expectation for the corresponding decision.

Node elimination is essentially a process of weighting conditional distributions to calculate distributions that are conditional on fewer predecessor nodes. This can be explained in relation to Figure 3.4. In Figure 3.4a the Value (V) has six distributions conditional on all combinations of Integrity Action (A) and Pipe Performance (P). This means that the probability distribution of V is conditional on both A and P. In standard probability notation this can be written as $f_{V|A,L}(v)$. The only node that needs to be eliminated in this case is the Pipe Performance node. Elimination of this node means finding a probability distribution of V that is not conditional on P (i.e., $f_{V|A}(v)$). Using the total probability rule, this can be calculated as follows:

$$f_{V|A}(v) = f_{V|Failure,A}(v)\, p(Failure|A) + f_{V|Safe,A}(v)\, p(Safe|A)$$

[3.3]

where $p(Failure\,|A)$ and $p(Safe|A)$ represent the probability of Failure and Safe given a certain choice A. Equation [3.3] simply states that the probability distribution of the Value for a given choice can be obtained as a weighted sum of the probability distributions corresponding to Safe and Failure for the same choice. Because the resulting distribution in unconditional on P, the arrow from P to V can be eliminated and the node P can also be eliminated because it no longer affects the Value node. This results in the diagram in Figure 3.4b, in which the Value has three probability distributions conditional only on the decision node. The means of the distributions in Figure 3.4b

Solving an Influence Diagram

represent the required Value expectations, and by comparing them the choice leading to the highest expected Value can be identified.
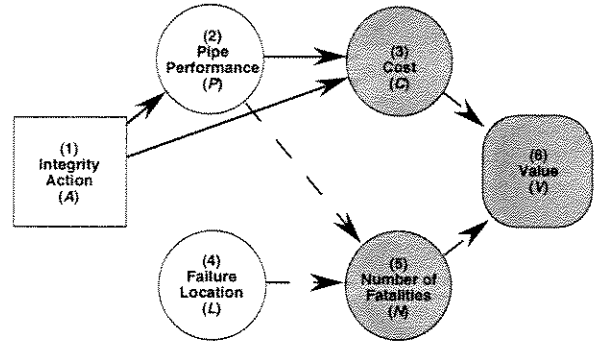
After evaluation of the functional nodes and elimination of the functional arrows (as discussed in Section 3), the influence diagram will generally contain more than one node to be eliminated. The Shachter (1986) algorithm includes a method for identifying the sequence in which these nodes are to be eliminated.

Figures

(a)

(b)

(c)

(d)

(e)

Specified

Accessible
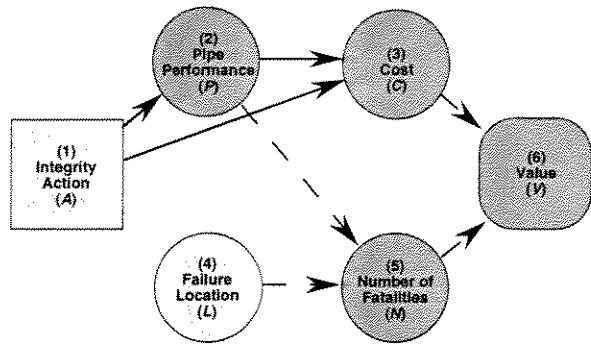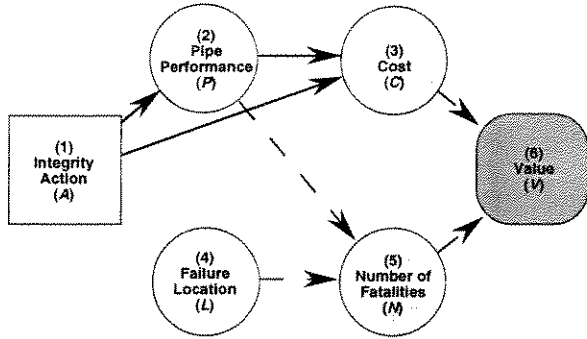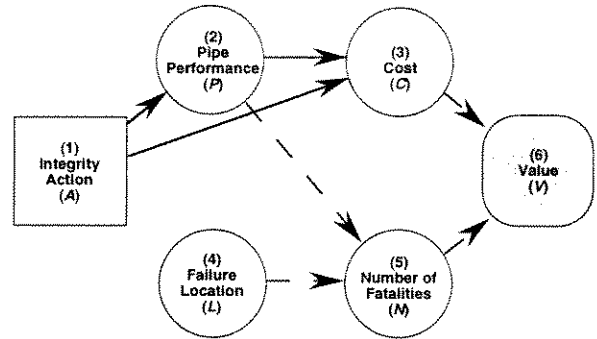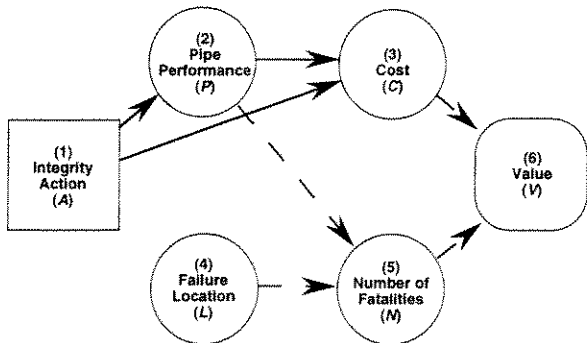
Inaccessible
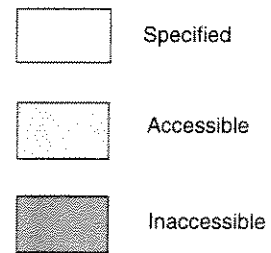
Figure 3.1 Illustration of one possible sequence of specifying the influence diagram
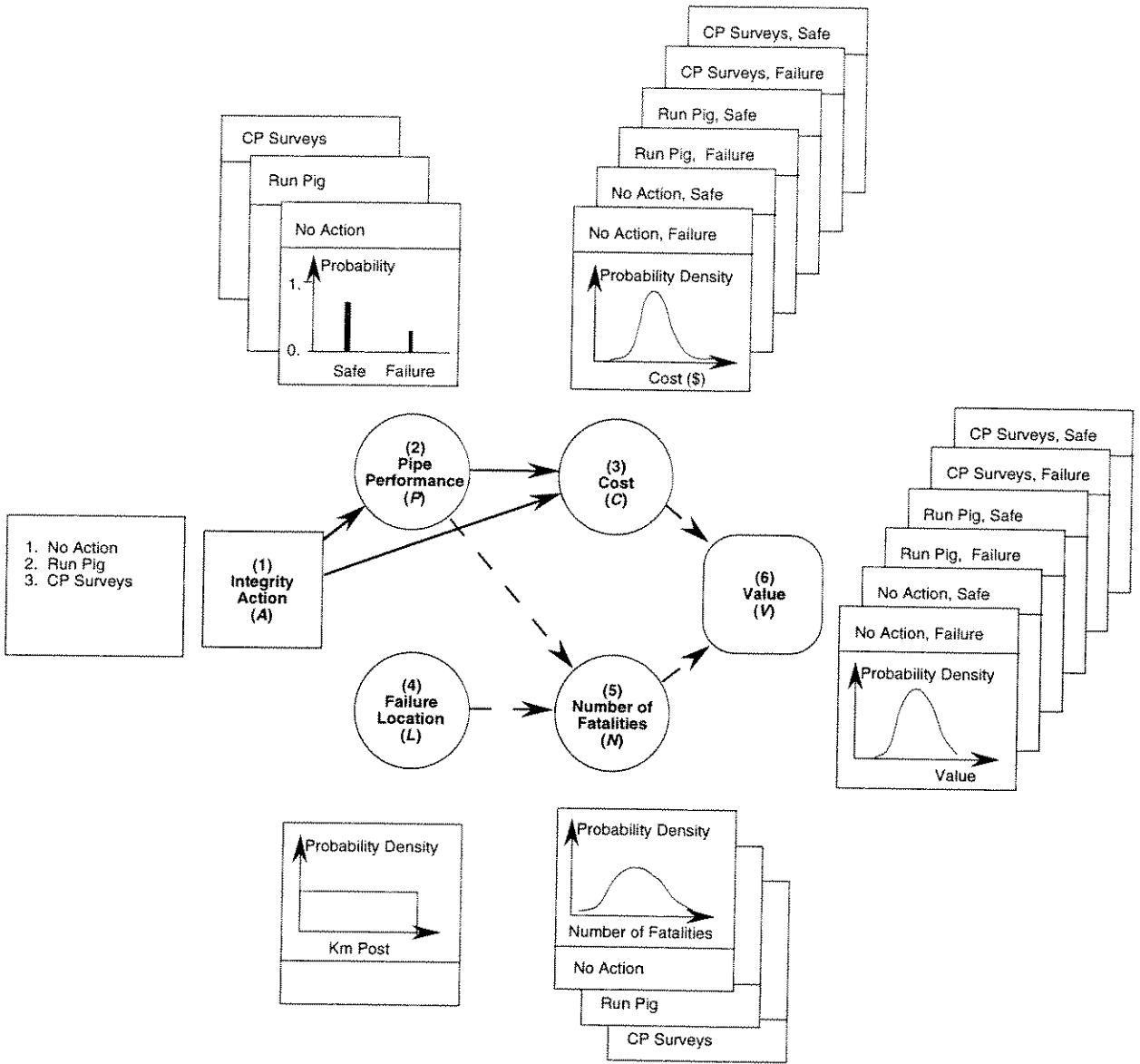
Figure 3.2 Illustration of the distribution combinations for different diagram nodes
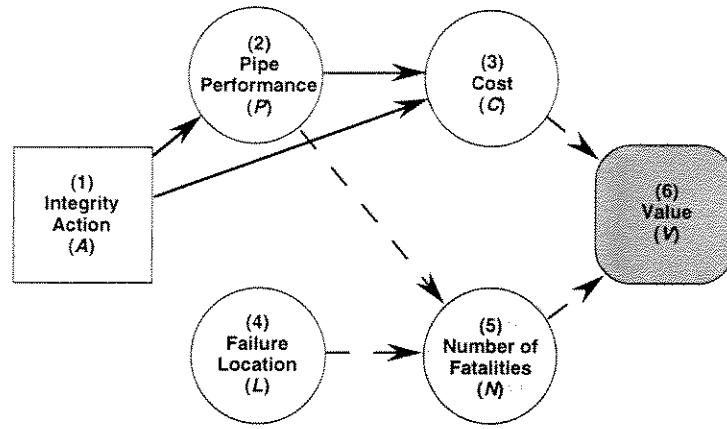
a) Original Diagram
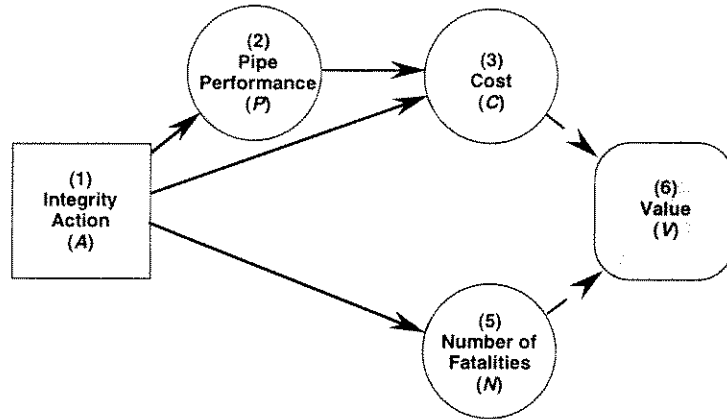
b) After Calculation of Node 5
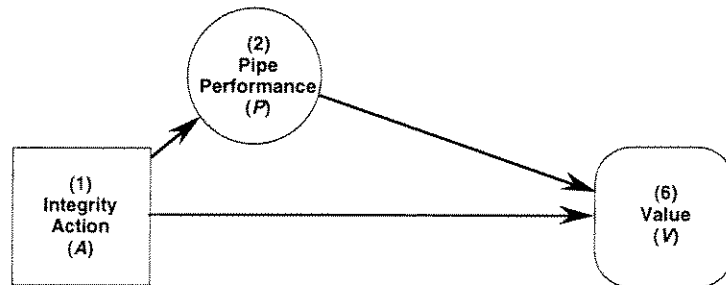
c) After Calculation of Node 6

Figure 3.3 Simplification of influence diagram due to evaluation of functional nodes
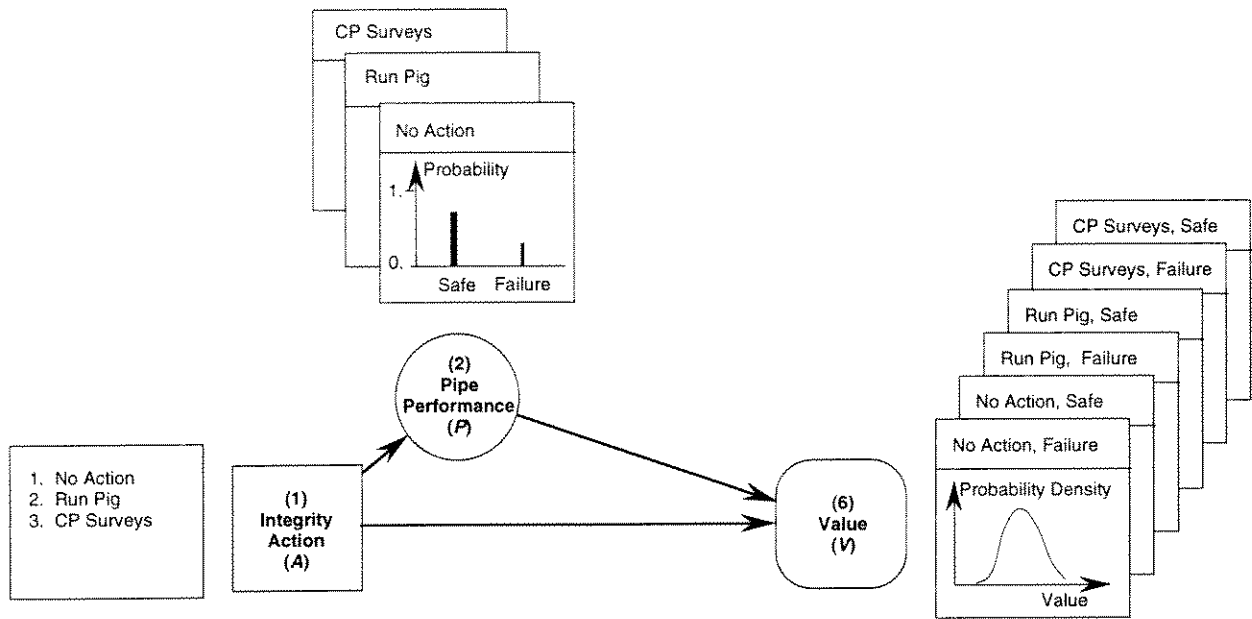
a) Diagram before elimination
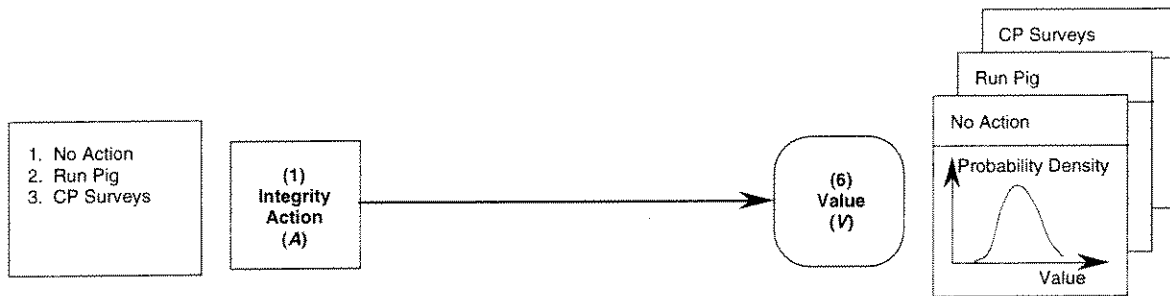
b) Diagram after elimination

Figure 3.4 Elimination of conditional nodes to find the expectation of the value

4.0   USE OF THE RESULTS

4.1   Decision Making and Sensitivity Analysis

The main outcome of the influence diagram solution is the expected value, which is used as a basis for decision making as illustrated throughout this document. In addition to this, influence diagrams are well suited to sensitivity analyses. Examples of such analyses include:

1. *Sensitivity to a given parameter* can be assessed by changing the value of the parameter and re-analyzing the problem to determine the effect of the change on the outcome. This may be done for input parameters that cannot be quantified with accuracy. If, for example, the cost node is difficult to evaluate in a given situation, the decision maker may define bounding values representing the maximum possible and minimum possible costs for each combination of the Cost predecessors. The analysis could then be performed for the two scenarios (and possibly other in-between scenarios) to find out the impact of this parameter on the final decision. Such analyses may be used to determine ranges of the Cost for which certain choices are optimal. This information is very valuable because the optimal decision in this case can be obtained by placing the cost in a certain range, which may be much easier than defining its probability distribution.

2. *Sensitivity to randomness regarding a given parameter* can be assessed by setting this parameter at its most likely value and examining the effect on the results. This type of analysis can be performed if additional information that eliminates the randomness associated with a given parameter can be obtained (*e.g.*, a request for quotations from repair vendors to define the cost of failure). There is a cost associated with obtaining the information and the sensitivity analysis can be used to determine whether obtaining the information is worthwhile. This is often referred to as assessing the value of information. A similar type of analysis can be carried out to determine the value of control, where control is used to indicate an action (with an associated cost) that reduces uncertainty regarding a given parameter. An example is the installation of pressure relief devices to ensure that the pressure does not exceed a specified value.

4.2   Intermediate Node Outcomes

In the course of solving the diagram it was shown that the conditional probability distributions for all node parameters must be defined. This information in itself can be very valuable. For example,

16

Use of the Results

the probability distribution of the Number of Fatalities was calculated for each choice. This information can be used as a basis to estimate the individual risk associated with each choice, which in itself is useful in regulatory submissions and public forums. For more realistic diagrams, many other useful nodal outcomes will be calculated including information on hazard scenarios, cost and its different components, and environmental impact. These distributions can be obtained for any combination of the relevant conditional nodes.

In addition, the probability distribution and expectation of any node can be obtained conditional on the decision node only. This can be achieved by treating the node in question as the final node in the diagram, and isolating the relevant portion from the remaining nodes. This is illustrated in Figure 4.1 for the Cost node and the Number of Fatalities node. Solving the reduced diagram in Figure 4.1a gives the probability distributions and expectations of the Cost conditional only on the choices. This information in itself is useful in understanding the consequences of a given decision.

As such, influence diagrams do not have to include a Value node. They can in fact be seen as a general probabilistic analysis tool that can be used to represent the relationships between a number of random variables and then derive the dependent variables from the basic variables.

4.3   Other Comments

Realistic influence diagrams may include a large number of conditional and functional nodes. For example, the consequence analysis influence diagram includes approximately 30 nodes (Stephens *et al.* 1995b). This results in compounding the conditions affecting a given node and may lead to a large number of distributions (up to several hundred) for a given node. In such cases it is not practical to calculate all of these distributions. An alternate approach is to calculate only the mean and standard deviation corresponding to each conditional distribution. This drastically reduces the computational effort and does not affect the decision making process sine the decision is based on the expectation or mean value in both cases. Coupled with this, specific probability distributions of particular interest can be calculated separately if desired. The algorithm used to calculate the mean and standard deviation is described in Appendix B.

Finally, it must be mentioned that the example influence diagram used in this report is a simple version that does not use all features of the approach. It does however utilize the essential features required for the consequence analysis model. One major influence diagram feature not discussed here is that of multiple decision nodes. A diagram including a number of decision nodes can be

17

Use of the Results

solved using the same methodology. In such cases, the decision and uncertain nodes for which the outcomes will be known before any decision is made must be defined and reflected in the diagram. The outcome in this case would be a sequence of optimal strategies defining the optimal response at each decision node. Multiple decision nodes are not implemented in PIRAMID.

Figures
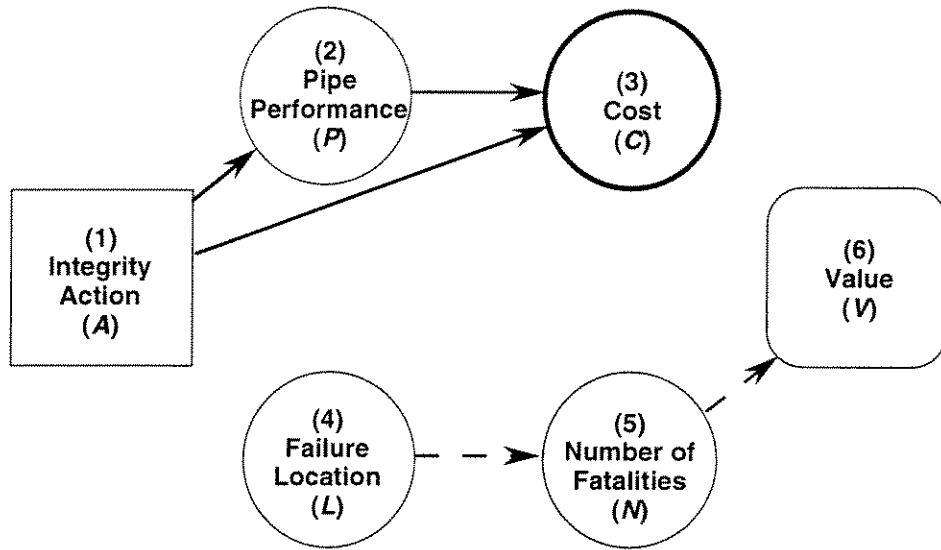
a) Cost node
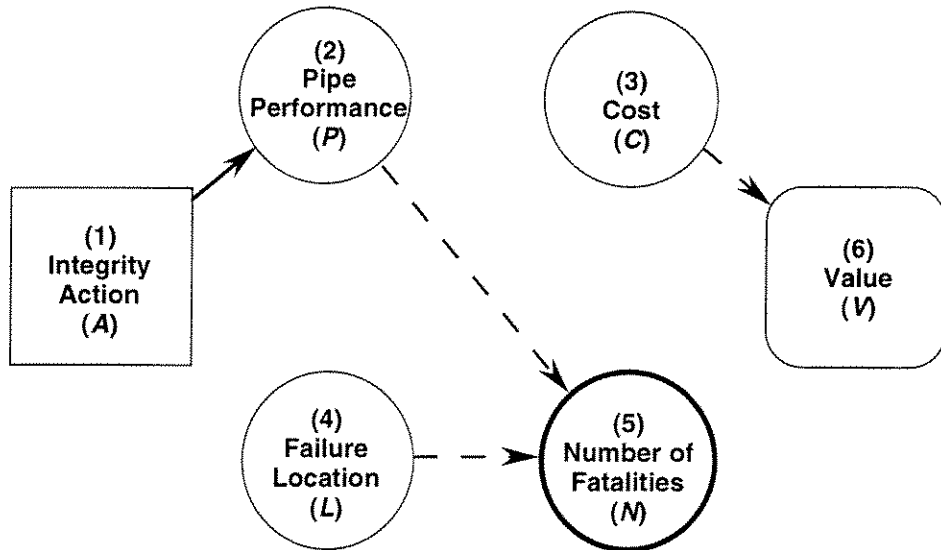
b) Number of Fatalities node

Figure 4.1 Illustration of using the influence diagram to analyze arbitrary uncertain nodes

5.0 REFERENCES

Call, H. J. and Miller W. A. 1990. A Comparison of Approaches and Implementations for Automating Decision Analysis. Reliability Engineering and system Safety, pp. 115 - 162.

Keeney, R. L. and Raiffa, H. 1976. Decisions with Multiple Objectives: Preferences and Value Tradeoffs. John Wiley & Sons, Inc.

Shachter, R. D. 1986. Evaluating Influence Diagrams. Operations research, Vol. 34, No. 6, November - December, pp. 871 - 882.

Stephens, M. J., Nessim, M. A. and Chen, Q. 1995a. A Methodology for Risk-Based Optimization of Pipeline Integrity Maintenance Activities - PIRAMID Technical Reference Manual No. 1.2. Confidential to C-FER's Pipeline Program Participants, C-FER Project 94006, October.

Stephens, M. J., Nessim, M. A., and Chen, Q. 1995b. Probabilistic Assessment of Onshore Pipeline Failure Consequences - PIRAMID Technical Reference Manual No. 3.1. Confidential to C-FER's Pipeline Program Participants, C-FER Project 94035, October.

APPENDIX A

INFLUENCE DIAGRAM SOLUTION ALGORITHMS

A.1 Introduction

This appendix gives the algorithms used to solve the influence diagram according to the methodology described in the main body of the report. The terminology and basic characteristics of the influence diagrams used are the same as given in Figure 2.2 of the report. Notation used is defined in Section A.2, and the algorithms are given in the following sections.

The algorithms described in Section A.3 of this Appendix correspond to the *extensions* developed by C-FER an added to the original influence diagram methodology. These extensions consist of including nodes that describe continuous random variables and allowing functional dependencies between nodes. Execution of these algorithms reduces the influence diagram to a standard diagram that can be solved using the original algorithm developed by Shachter (1986).

The algorithm described in Section A.4 is a simplified version of Shachter's algorithm that contains all the features needed to solve the influence diagrams developed in this project. The algorithm is used to calculate the probability distribution of any functional node. Since the utility node is a special case of a functional node, the same algorithm is used to define the expected utility for each decision, thus giving the final solution of the influence diagram.

An example illustrating these algorithms is included in the main body of the report (Section 3.3.3).

A.2 Notation

X_j Random variable corresponding to the jth diagram node.

$P(x_j|E)$ Probability distribution function of discrete random variable X_j conditioned on E.

$f(x_j|E)$ Probability density function of continuous random variable X_j conditioned on E.

$typ(j)$ Type of node of jth diagram node: D is a decision node; C is a conditional node; F is a functional node; and V is the value node.

Appendix A

pre(j)   Set of nodes that are direct predecessors to node j.

fcpre(j)   Set of direct conditional predecessors of node j.

A.3   Evaluating the Conditional Probability of a Functional Node

A.3.1   Finding the Direct Conditional Predecessors of a Functional Node

Consider node j with typ(j) = F. Let pre(j) = $\{i1,...,in_{prej}\}$ denote its direct predecessors. The procedure for finding the direct conditional predecessors of node j, fcpre(j) is as follows:

1. For node j, let fpre(j) equal to pre(j) = $\{i1,...,in_{prej}\}$, where fpre(j) is a temporary set.

2. Check each node $k \in$ fpre(j), if typ(k) is not equal to "C" or "D", erase node k from fpre(j) and add pre(k,1) = $\{i1,...,in_{prek}\}$ to fpre(j).

3. Let fcpre(j) = $(\bigcup_{k \in fpre(j)} pre(k)) \bigcup$ (all decision nodes $\in$ fpre(j)); and let fcpre'(j) = fcpre(j).

4. For each node $i \in$ fcpre(j), if a path from node i to node j contains $k \in$ fcpre(j) and $k \neq i$, then i is deleted from fcpre'(j).

5. Set fcpre(j) = fcpre'(j).

A.3.2   Calculating the Conditional Probability Distribution of a Functional Node j

1. Using the direct conditional predecessors of j, define the combination of conditions for which the distribution of X_j is to be calculated.

2. The set of conditions identified in (1) can be ordered using a permutation algorithm.

3. For each ordered set of conditions in (2) return to the direct predecessors of node j and identify their corresponding information (probability distribution).

4. If all the random variables are continuous:

    4.1   Use the node function to calculate an approximation of the mean and standard deviation of X_j for each ordered set of conditions given in (2). This uses the approach described in Appendix B.

    4.2   Call FORM (see Appendix B) to generate a numerical probability distribution function in a specified range around the mean value (*e.g.*, 3 standard deviations on either sides of the

Appendix A

mean, but always greater than 0.

5. If some of the random variables are discrete, carry out step (4) for each value of the discrete variable, which is treated as a deterministic parameter. The final distribution is the sum of the distributions at each point of the discrete parameter, each weighted by the corresponding probability.

6. For a specified set of conditions, if all the predecessors are discrete random variables, the distribution of X_j is also discrete. In this case, the value of X_j is evaluated at each set of conditions. The probability of each value is equal to the probability of the corresponding set of conditions, calculated as the product of the probabilities of the individual conditions in the set.

Note: The algorithm implemented in PIRAMID as of the date of this document, calculates only the mean and standard deviation of each node parameter by eliminating the call to FORM (step 4.2) and using a higher order approximation for the mean and standard deviation in step 4.1 (see Appendix B).

A.4 Evaluating the Unconditional Probability distribution of a Functional Node

Note: This algorithm is based on the original influence diagram solution except that it assumes that reversal of arcs and no forgetting arcs are not required (see Shachter 1986). It assumes that the diagram is oriented and regular. An oriented influence diagram has a value node. An influence diagram is said to be a regular if it satisfies the conditions that it has no cycles; the value node, if present has no successors; and there is a direct path to the value node that contains all the decision nodes.

1. Using the algorithm in Section A.3, calculate the conditional probability density function, $f(x_j|$ an event of fcpre(j)).

2. Define a set of direct predecessors of node j, which is identical to the direct conditional predecessor set fcpre(j). This results in eliminating all direct predecessors of the functional node being solved for.

3. If there are only decision nodes in fcpre(j) then stop calculation.

4. Set new fpre(j) = fcpre(j) and use algorithm in Section A.3 to find a new set, fcpre(j).

Appendix A

5. Use the total probability rule to calculate a probability density function or distribution function conditional on an event of fcpre(j) using:

$$f(x_j| \text{ an event, } E_c, \text{ of fcpre(j)}) = \sum_{\text{all possible events of fpre(j)}} f(x_j| \text{ an event, } E, \text{ of fpre(j)}) \; P(E|E_c)$$

or

$$P(x_j| \text{ an event, } E_c, \text{ of fcpre(j)}) = \sum_{\text{all possible events of fpre(j)}} P(x_j| \text{ an event, } E, \text{ of fpre(j)}) \; P(E|E_c)$$

The probability density function or distribution function now is conditioned on an event of fcpre(j). The calculation is performed for all events of fcpre(j).

6. Repeat (3) - (5) until the probability density function or distribution function of the node is conditional only on the decision node and the evaluation of the functional node is complete.

Note: The algorithm implemented in PIRAMID as of the date of this document calculates only the mean and standard deviation of the distribution by replacing the conditional distribution functions in (1) by the conditional means and standard deviations.

A.5   Reference

Shachter, R. D. 1986. Evaluating Influence Diagrams. Operations research, Vol. 34, No. 6, November - December, pp. 871 - 882.

Appendix B

the cumulative distribution function. The derivation of $F_X(x)$ from these inputs is illustrated in Figure B.1.

The function g will be referred to as the *response function* because it defines the *response parameter* X given the input parameters Y_1, Y_2, ..., Y_n. The input parameters Y_i, i = 1, 2, ..., n are referred to as the *basic parameters*. For simplicity, we will assume throughout this section that Y_i are independent parameters. However, all methods described in this Appendix can be extended to address correlated parameters.

In an influence diagram analysis, X can be seen as the parameter of a functional node and the Y's as the parameters of its immediate predecessors. For the influence diagram in Figure 2.1 of the main report, the methods discussed here can be used, for instance, to calculate the probability distribution of the Number of Fatalities from the probability distributions off the Pipe Performance and Failure Location.

B.3 Estimation of the Mean and Standard Deviation of X - Simplified Methods

B.3.1 Introduction

This section describes the methods used to estimate the mean and standard deviation of the response parameter from the moments (*e.g.*, mean standard deviation, skewness and kurtosis) of the basic parameters. These methods are useful for cases where the mean and standard deviation are sufficient for the decision analysis.

B.3.2 Calculation Procedure for a Linear Response Function

If the relationship between X and Y_i, i = 1, 2, ..., n is linear, then Eq. [B.1] becomes

$$X = a_0 + a_1 \, Y_1 + a_2 \, Y_2 + \text{.......} + a_n \, Y_n \qquad \text{[B.2]}$$

where a_0, a_1, a_2, ... , a_n are constants. In this case (recall that we assume Y_i's to be independent throughout) the mean value and standard deviation of X, μ_X and σ_X are given by:

$$\mu_X = a_0 + a_1 \, \mu_{Y1} + a_2 \, \mu_{Y2} + \text{.......} + a_n \, \mu_{Y2} \qquad \text{[B.3]}$$

B.2

Appendix B

$$\sigma_X = \sqrt{a_1^2 \sigma_{Y1}^2 + a_2^2 \sigma_{Y2}^2 + \ldots + a_n^2 \sigma_{Yn}^2} \qquad [B.4]$$

This is given without proof but can be verified by the following example. Details and proofs can be found in Ang and Tang (1984).

Example B.1: Columns 1 and 2 of Table B.1 contain data samples Y_1 and Y_2 with means μ_{Y1} and μ_{Y2} of 1.86 and 3.01 and standard deviations σ_{Y1} and σ_{Y2} of 0.59 and 0.86. Column 3 is created using:

$$X = 0.4 \ Y_1 + 1.2 \ Y_2$$

which is identical to Eq. [B.1] with $a_1 = 0.4$ and $a_2 = 1.2$. The mean of X calculated using Eq. [B.3] is $\mu_X = 0.4 \times 1.86 + 1.2 \times 3.01 = 4.36$, and similarly the standard deviation from Eq. [B.4] is $\sigma_X = 1.06$. These are almost identical to the values calculated from the data at the bottom of column 3 in Table B.1 and this illustrates the validity of Eqs. [B.3] and [B.4].

B.3.3 Calculation Procedure for a Non-Linear Response Function

A non-linear response function can be approximated by a linear function using the tangent to the original function at any given point. This is illustrated in Figure B.2 for a response variable X defined as a function g(Y) of one basic random variable (Y). The tangent at the mean value of Y is a straight line given by:

$$X = g(\mu_Y) + (Y - \mu_Y) \ \partial g/\partial Y \qquad [B.5]$$

where $\partial g/\partial Y$ is the derivative of g with respect to Y evaluated at μ_Y (*i.e.*, the slope of the tangent to g(Y) at $Y = \mu_Y$). This equation is a first order Taylor series expansion of the original function around the mean of Y. It has the same format as Eq. [B.2] with $a_0 = g(\mu_Y) - \mu_Y \ (\partial g/\partial Y)$ and $a_1 = \partial g/\partial Y$. Substituting in Eqs. [B.3] and [B.4] and generalizing to n basic variables (Y_i, $i = 1,2, \ldots, n$) gives:

$$\mu_X = g(\mu_{Y1}, \mu_{Y1}, \ldots, \mu_{Yn}) \qquad [B.6]$$

$$\sigma_X = \sqrt{(\partial g/\partial Y_1)^2 \ \sigma_{Y2}^2 + (\partial g/\partial Y_1)^2 \ \sigma_{Y2}^2 + \ldots + (\partial g/\partial Y_n)^2 \ \sigma_{Yn}^2} \qquad [B.7]$$

Appendix B

Example B.2: This example is similar to Example B.1 except that the function considered is non-linear as follows:

$$X = Y_1 Y_2$$

The data samples of Y_1 and Y_2 are given in Table B.2. Their means μ_{Y1} and μ_{Y2} are 1.87 and 2.99 and the standard deviations σ_{Y1} and σ_{Y2} are 0.46 and 0.97. Column 3 is created using the above equation. The mean of X from Eq. [B.6] is given by $\mu_X = 1.87 \times 2.99 = 5.60$. The derivative $\partial g/\partial Y_1$ and $\partial g/\partial Y_2$ are given by Y_2 and Y_1 and when evaluated at the mean value they equal μ_{Y2} and μ_{Y1}. The standard deviation calculated by using these values in Eq. [B.7] is $\sigma_X = 2.29$. These are close to the values calculated from the data at the bottom of column 3 in Table B.2 ($\mu_X = 5.48$, $\sigma_X = 2.03$).

B.3.4 Extensions Using Higher Order Terms

The accuracy of the linear approximation used in Section B.3.3 for the response function, can be improved by adding higher order terms in the Taylor series expansion. These terms mean that the approximate function is quadratic or third order and can therefore represent the original function more accurately resulting in better approximations of the required mean and standard deviation.

In this case, the following equations result (Hahn and Shapiro 1967):

$$\mu_X = g(\mu_{Y1}, \mu_{Y2}, ..., \mu_{Yn}) + \frac{1}{2} \sum_{i=1}^{n} \frac{\partial^2 g}{\partial Y_i^2} \sigma_{Yi}^2 \qquad [B.8]$$

$$\sigma_X^2 = \sum_{i=1}^{n} \left(\frac{\partial g}{\partial Y_i} \right)^2 \sigma_{Yi}^2 + \sum_{i=1}^{n} \left(\frac{\partial g}{\partial Y_i} \right) \left(\frac{\partial^2 g}{\partial Y_i^2} \right) \mu_{3_{Yi}} \qquad [B.9]$$

where μ_3 and μ_4 denote the third and fourth order moments about the mean. These parameters can be calculated from the probability distributions of the basic variables (Christensen 1989). For the response variable, they can be calculated from:

$$\mu_{3_X} = \sum_{i=1}^{n} \left(\frac{\partial g}{\partial Y_i} \right)^3 \mu_{3_{Yi}} \qquad [B.10]$$

$$\mu_{4_X} = \sum_{i=1}^{n} \left(\frac{\partial g}{\partial Y_i} \right)^4 \mu_{4_{Yi}} + 6 \sum_{i, i<j} \sum_{i} \left(\frac{\partial g}{\partial Y_i} \right)^2 \left(\frac{\partial g}{\partial Y_j} \right)^2 \sigma_{Yi}^2 \sigma_{Yj}^2$$

B.4

Appendix B

B.4 Estimation of the Probability distribution of X Using Reliability Methods

B.4.1 Introduction

Reliability Methods are approximate techniques for estimating the probabilities of events that are defined on the basis of functions of basic random parameters. They have been developed over the past two decades in the area of structural reliability. These methods were developed as an efficient alternative to Monte Carlo simulations, because the latter can be computationally prohibitive for the estimation of low probability events.

Reliability methods concentrate on estimating the probability $p\ (X \leq 0)$, where X is defined as a function $g(Y_1, Y_2, ..., Y_n)$ as in Eq. [B.1]. This is given by:

$$p = p\ [g(Y_1, Y_2, ..., Y_n) \leq 0] \qquad [B.12]$$

The probability distribution of X can be calculated by evaluating a series of problems defined as p $(X-a_i \leq 0)$, $i = 1, 2, ..., n$. This probability is equal to $p\ (X \leq a_i)$, $i=1, 2, ..., n$, which is in fact the cumulative probability distribution of X.

Example B.3: Consider a simple concrete beam with a random mid-section bending moment capacity of R and a random applied mid-section bending moment of S. Failure of the beam in bending at the mid-section occurs if $R \leq S$, or if $R - S \leq 0$. This means that the probability of failure $p_f = p\ (R - S \leq 0)$, which is the same format as Eq. [B.12] with $g(R,S) = R - S$. If the resistance R and S are defined as functions of a number of basic random parameters $Y_1, Y_2, ..., Y_n$ representing the material strength, beam dimensions and load intensity, then $p_f = p\ [g(Y_1, Y_2, ..., Y_n) \leq 0]$.

In the theory of structural reliability, a failure condition is usually referred to as a limit state and consequently, g is referred to as the limit state function. Because the method is presented here in a generalized sense, g will be referred to as the response function.

Appendix B

B.4.2 The Basic Principles of Reliability Methods

B.4.2.1   General Probability Integrals

The calculation of the probability in Eq. [B.12] is a multidimensional integration problem. To explain this, consider the one-dimensional case in which X is a function of only one random variable, *i.e.*, $X = g(Y_1)$. The calculation of $p(X \leq 0) = p[g(Y_1) \leq 0]$ is illustrated in Figure B.3. The procedure consists of finding the range of Y_1 values for which $g(Y_1) \leq 0$ and then calculating the probability that Y_1 is in this range. This probability is equal to the area under the PDF of Y_1 for all values of Y_1 leading to $g(Y_1) \leq 0$, which can be calculated by the integral

$$p[g(Y_1) \leq 0] = \int_{g(Y_1) \leq 0} f_{Y1}(y_1) \, dy_1 \qquad\qquad [B.13]$$

Now, consider the two-dimensional case in which $X = g(Y_1,Y_2)$. In this case, $g(Y_1,Y_2) = 0$ can be plotted in the $\{Y_1,Y_2\}$ plane as shown in Figure B.4. The function $g(Y_1,Y_2) = 0$ divides the $\{Y_1,Y_2\}$ plane into two regions: a region in which $g(Y_1,Y_2) \leq 0$ and a region in which $g(Y_1,Y_2) > 0$. If a perpendicular axis to the $\{Y_1,Y_2\}$ plane is drawn, on which a bivariate probability density function of Y_1 and Y_2 is defined (see Figure B.5), then $p[g(Y_1,Y_2) \leq 0]$ is given by the volume under this density function and over the corresponding region in the $\{Y_1,Y_2\}$ plane. This volume can be calculated by integration of the joint (bivariate) probability density function over the appropriate domain. This is given by:

$$p(g \leq 0) = \int \int_{g(Y1,Y2) \leq 0} f_{Y1,Y2}(y_1,y_2) \, dy_1 \, dy_2 \qquad\qquad [B.14]$$

In the general case where there are n Y's, a visual representation is not possible but the probability $p[g(Y_1,Y_2, ...,Y_n) \leq 0)]$ is calculated by generalizing Eq. [B.14] to an n-dimensional integral. These integrals are often referred to as *general probability integrals*.

B.4.2.2   The Evaluation of Probability Integrals Using Reliability Methods

The evaluation of this type of probability integral is a well known problem that appears repeatedly in probabilistic analysis. Reliability methods focus on estimating general probability integrals using approximate efficient algorithms. These algorithms are based on the special case of the problem: if the response function is linear (as in Eq. [B.2]), and if the Y_i variables are normally distributed and independent, then X is also normally distributed with a mean and standard deviation

Appendix B

given by Eqs. [B.3] and [B.4]. Knowledge of the distribution of X can be used to estimate the required probability directly from normal distribution tables. The essence of reliability methods is to exploit this special case by transforming any general case with nonlinear g and non-normal Y's into an analogous case with a linear g and normal Y's. Two transformations are needed:

1. Variables Y_i that do not have normal distributions are transformed into variables U_i with independent normal distributions. (See Madsen *et al.* 1986 for details).

2. The response surface is approximated by a tangent (*i.e.*, linear function) at the point closest to the origin (see Figure B.6). This point is called the design point and can be found by solving an optimization problem in which the distance between the origin and the g surface is minimized subject to the constraint $g = 0$. Efficient algorithms exist for this purpose and some of these are available in commercial computer programs (*e.g.*, Gollwitzer *et al.* 1988).

These transformations create a new problem for which the special case discussed above can be used directly, and the required probabilities can be calculated from normal distribution tables. Because a linear approximation is used this method is referred to as the *First Order Reliability Method (FORM)*. Extensions of this method that use a second order (parabolic) function to approximate the response surface have been developed and are referred to as *Second Order Reliability Methods (SORM)*. Detailed description of the methods can be found in Madsen *et al.* (1986).

B.4.2.3 Comments on FORM and SORM

* *Limitations on convergence.* Algorithms used in finding the design point are not guaranteed to converge. They assume that the function g is continuous and smooth in the solution zone. Functions that include discontinuities (*e.g.*, IF branches in computer algorithms) may cause problems in convergence. Also the algorithm may converge to a local minimum if one exists, leading to erroneous solutions (see Figure B.7). FORM and SORM work best for smooth well behaved g functions.

* *Solution speed.* The number of calls to the response function is approximately equal to the number of random variables multiplied by the number of iterations required to find the design point. The latter depends on the shape of the g functions. Experience shows that many problems can be solved in 5 to 10 iterations. Unlike simulation methods, this is not affected by the level of probability being estimated. A problem with 10 random variables may be solved

Appendix B

within 100 calls to the g function. This is much less than the number of simulations required for a Monte Carlo simulation.

- *Accuracy of FORM and SORM results.* Because the design point has the highest probability density, the contribution of the region close to the design point to the volume expressed by the probability integral in Eq. [B.14] is relatively high. As can be seen from Figure B.6, this is also the area where the straight line approximation is closest to the original function. For these reasons FORM and SORM give results that are in most cases very close to the exact solution.

B.4.3 Application

Example B.4 (from Ang and Tang 1984): This example deals with the calculation of the probability of failure of a steel beam in flexure. The flexural capacity of the beam can be estimated as $Y_1 Y_2$, where

$Y_1 = $ the yield strength of steel.

$Y_2 = $ the section modulus

If the applied moment on the beam is Y_3, then the margin of safety X can be estimated as

$$X = g(Y_1, Y_2, Y_3) = Y_1 Y_2 - Y_3$$

Failure occurs if the margin of safety is less than zero (i.e. the load exceeds the capacity). The probability of failure can be calculated as

$$pf = p(X \leq 0) = p [(Y_1 Y_2 - Y_3) \leq 0]$$

The input variables are assumed to be independent with the probability distributions given in Table B.3.

Details of the FORM solution are not given here since we have not developed the necessary mathematical background. This problem can be solved by hand, but computer programs are available for more complex g functions. These programs require programming the response function in a separate subroutine, linking it to the program and defining the input random variable distributions. The calculated value of the probability of failure is 0.003. The program used to

B.8

Appendix B

calculate this result (and also used in the decision analysis software for this JIP) is produced by RCP GmbH of Munich, Germany (Gollwitzer *et al*. 1988).

B.5   References

Ang, A. H-S. and Tang, W. H. 1984.   Probability Concepts in Engineering Planning and Design Volume II: Decision, Risk, and Reliability.   John Wiley & Sons, Inc.

Christensen, R. 1989.   Data Distributions: A Statistical Handbook.   Entropy Limited.

Gollwitzer, S., Abdo, T. and Rackwitz, R. 1988.   First Order Reliability Method.   RCP GmbH, Munich, Germany.

Hahn, G. J. and Shapiro, S. S. 1967.   Statistical Models in Engineering.   John Wiley & Sons, Inc.

Madsen, H. O., Krenk, S. and Lind, N. C. 1986.   Methods of Structural Safety.   Prentice-Hall, Inc., Englewood Cliffs, NJ.

Nessim, M. A. 1992.   Introduction to Risk Analysis of Engineering Systems - Course Notes. Centre for Frontier Engineering Research, October.
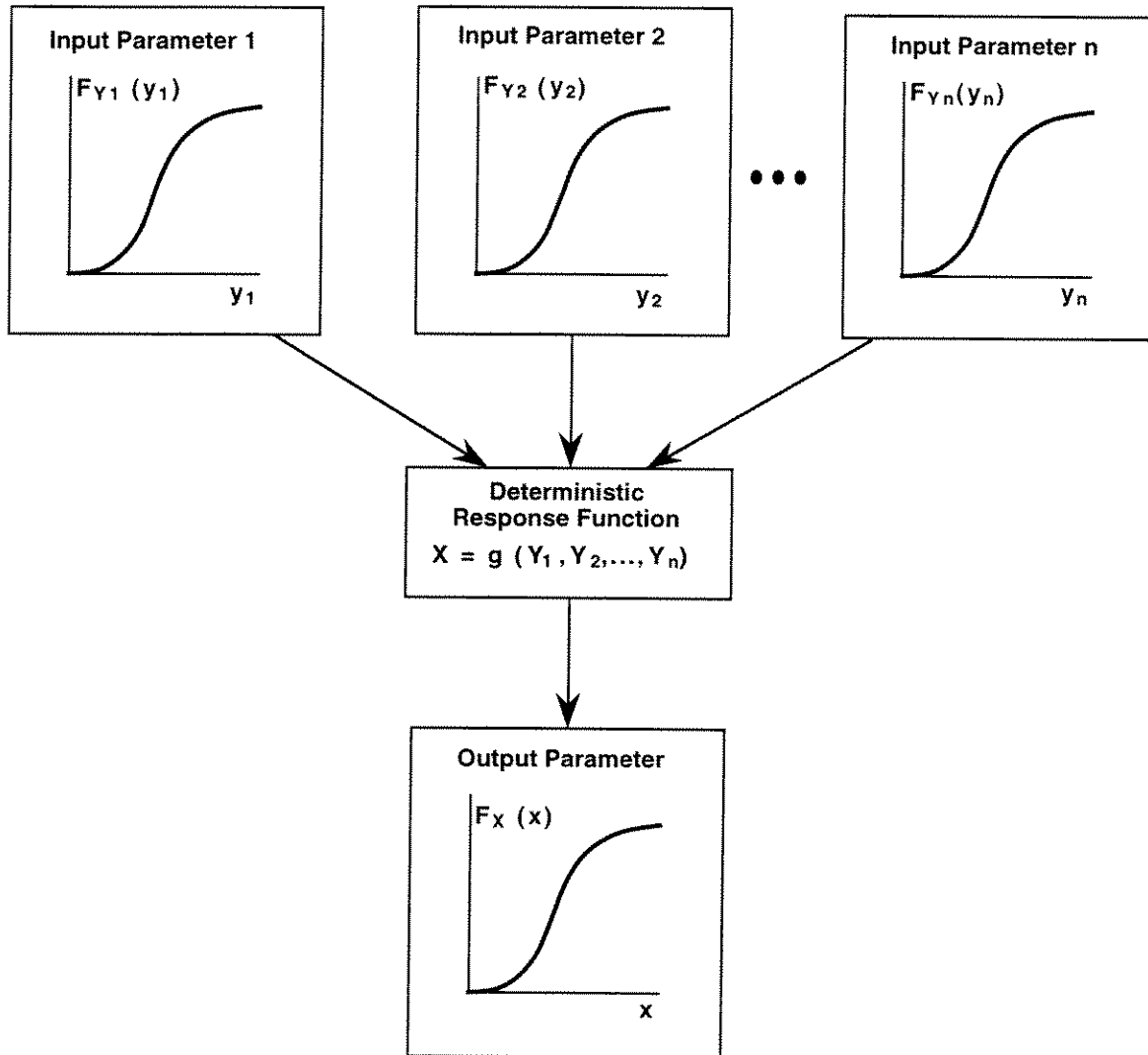
Figures and Tables

Figure B.1 Illustration of the problem of deriving random variables based on mathematical models
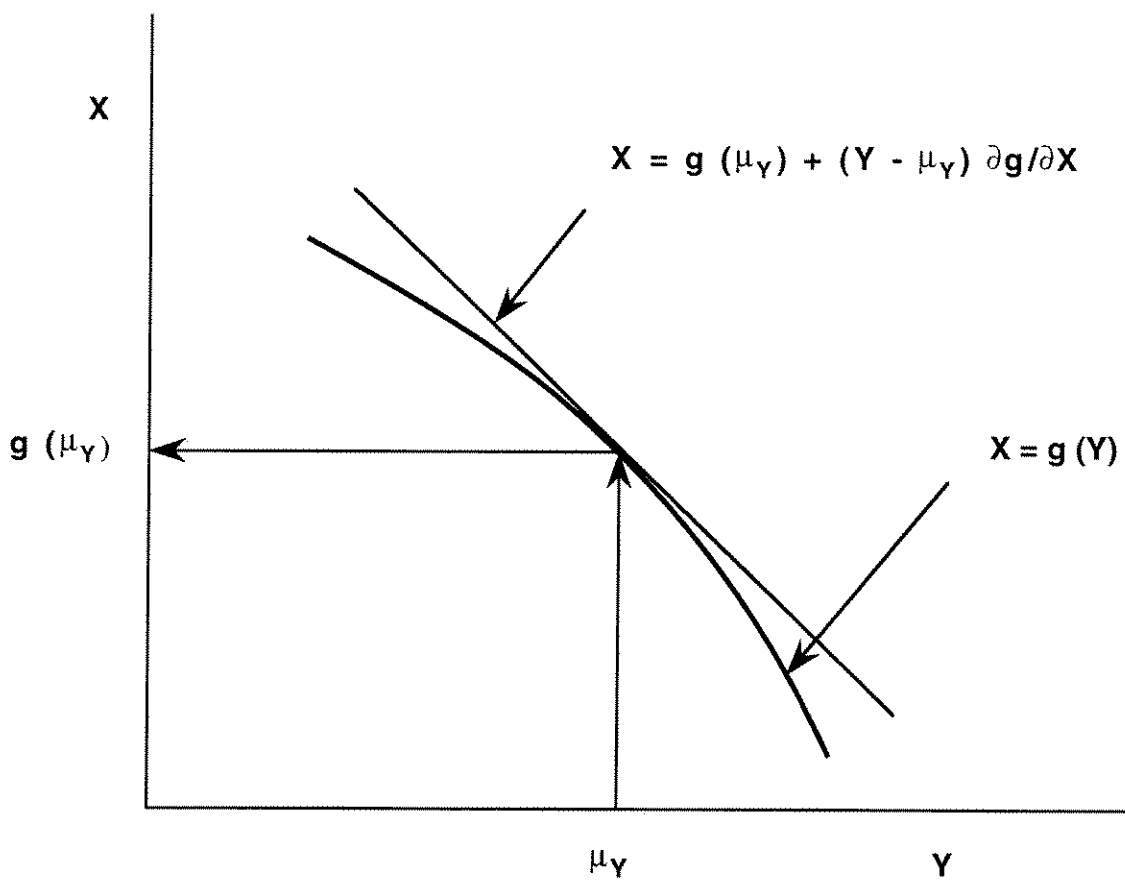
Figure B.2 Illustration of the linearization procedure using a Taylor series expansion about the mean of y
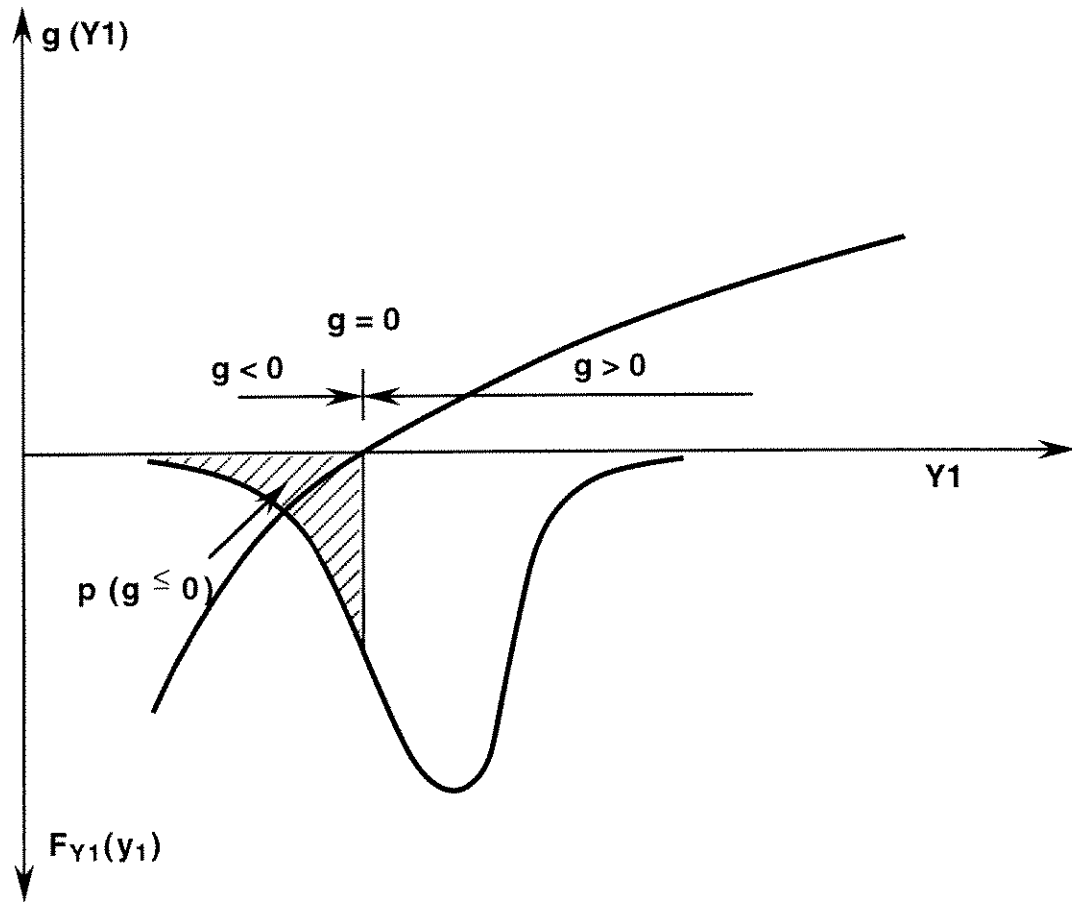
Figure B.3 Illustration of reliability calculation in one dimension
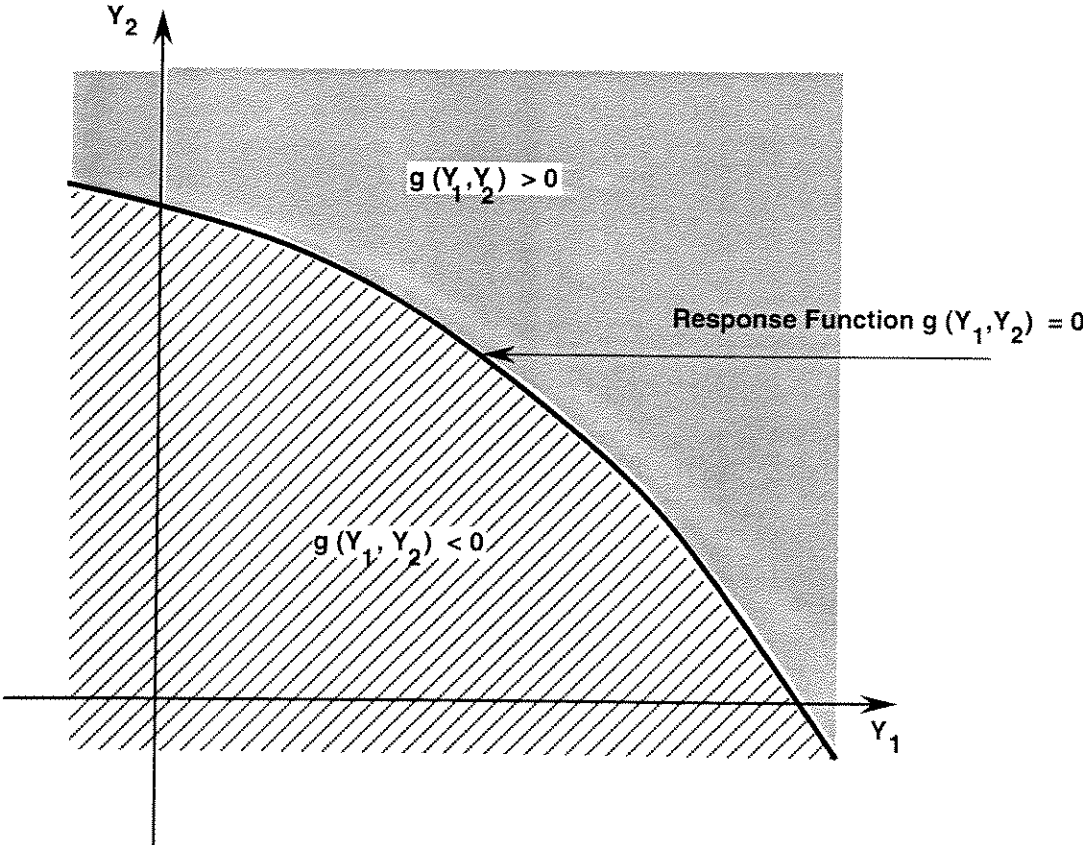
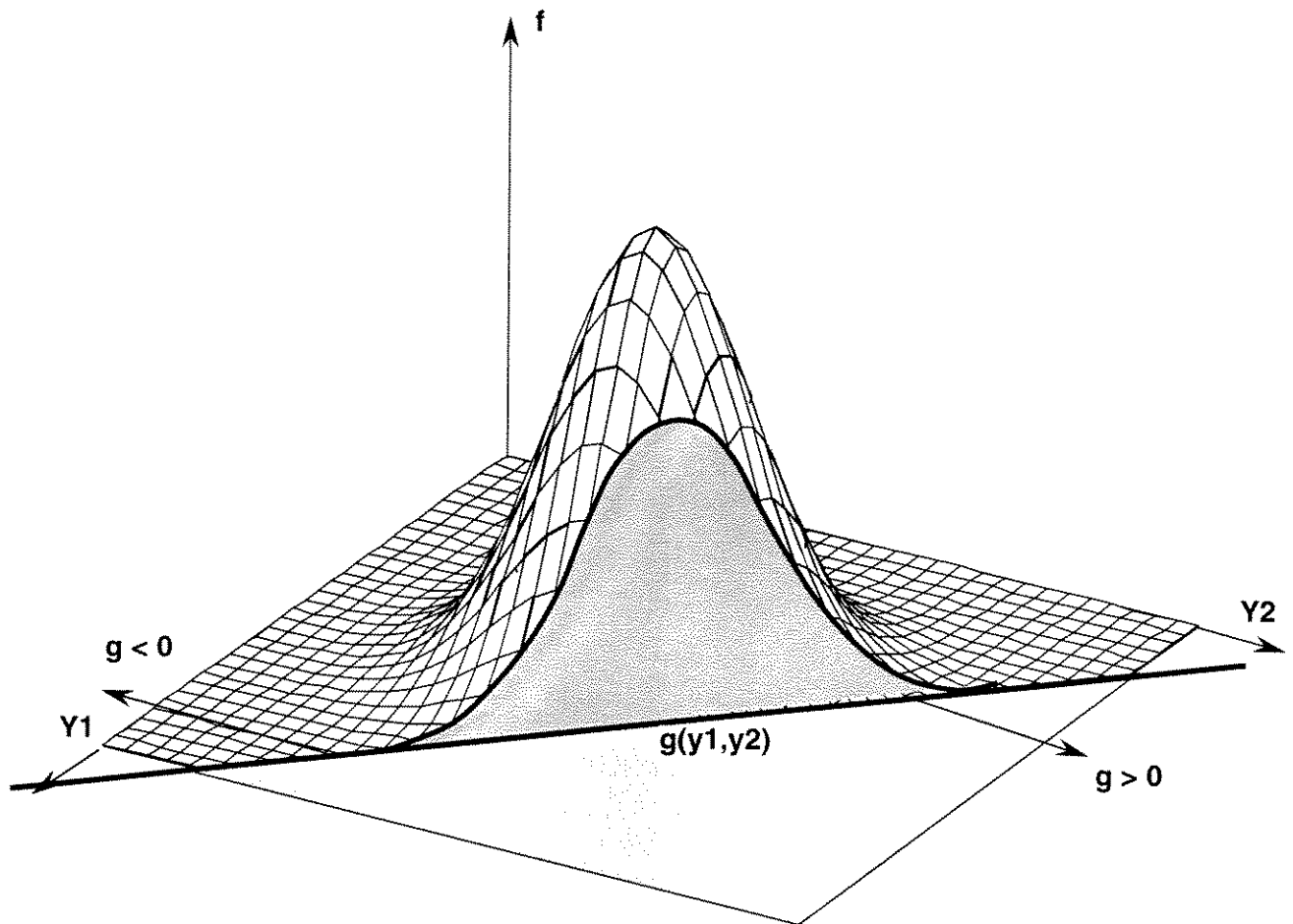Figure B.4 Illustration of reliability calculation in two dimensions

Figure B.5 Illustration of a response function in two dimensions with bivariate PDF
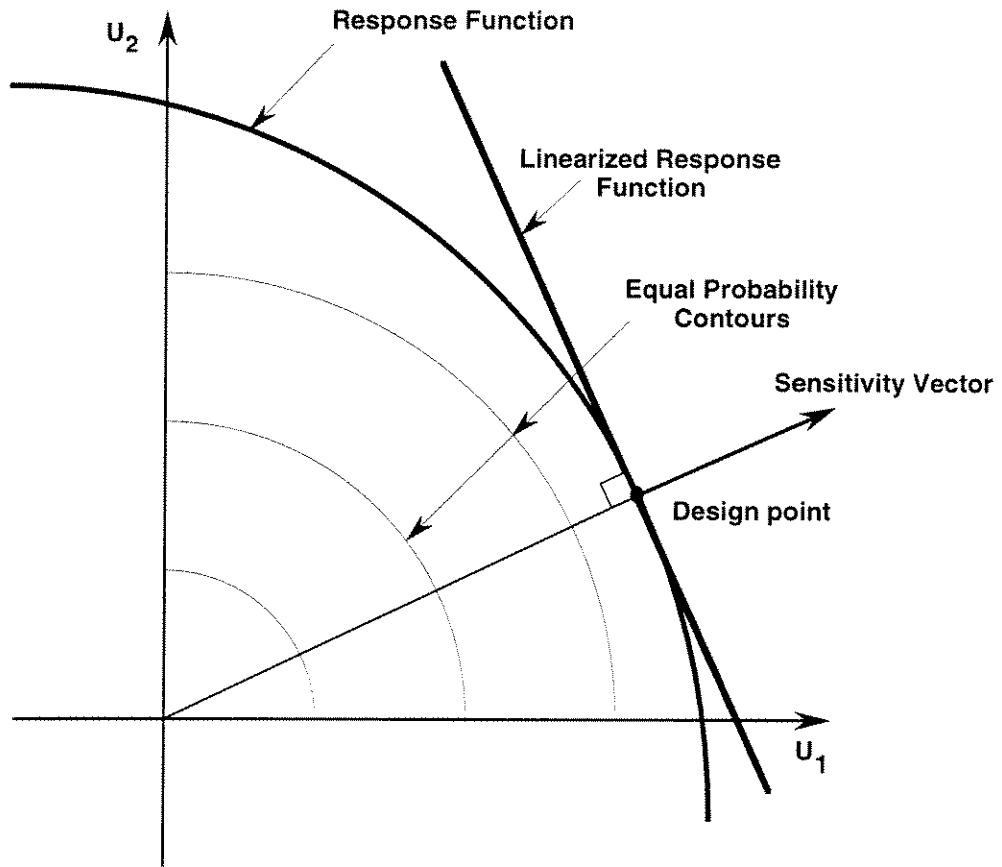
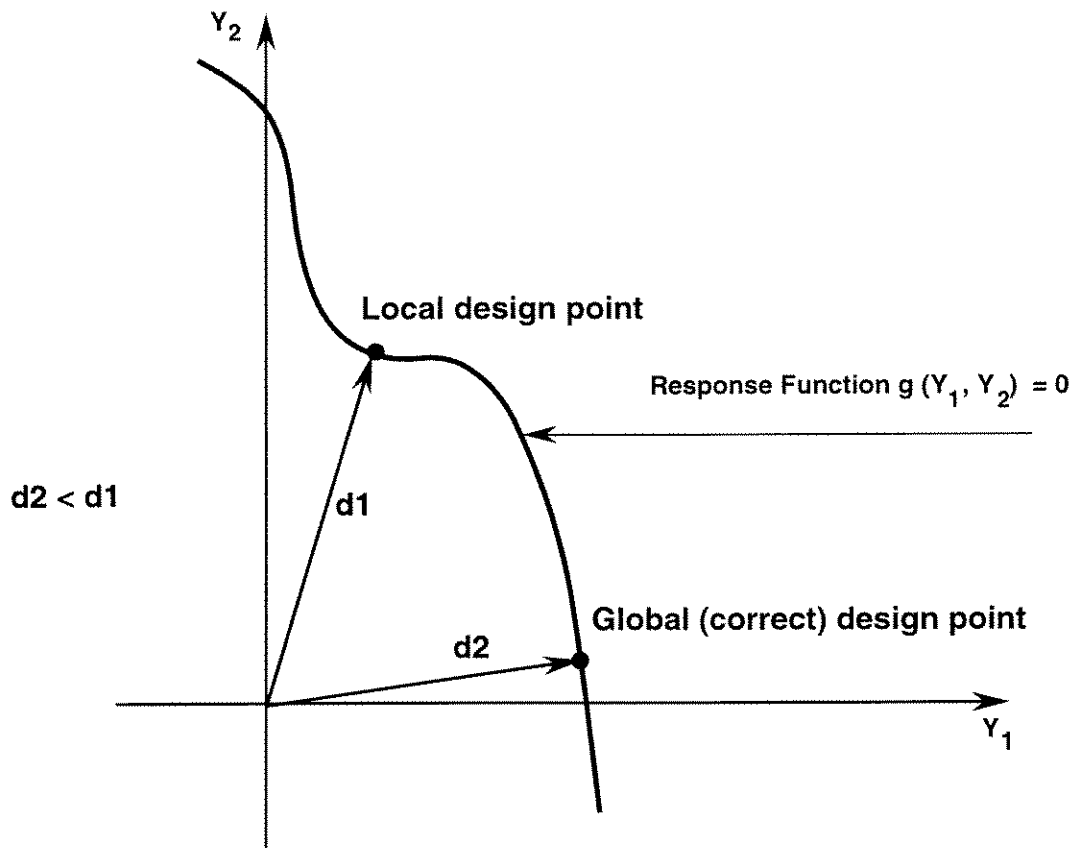Figure B.6 Illustration of the linearized response function

Figure B.7 Illustration of an erroneous local design point

No.	Y1	Y2	X = 0.4 Y1 + 1.2 Y2
1	2.86	3.56	5.42
2	1.84	2.82	4.12
3	2.41	3.40	5.05
4	0.57	3.30	4.19
5	1.55	2.31	3.39
6	1.54	2.53	3.66
7	1.20	3.13	4.23
8	2.34	2.83	4.33
9	1.41	3.72	5.03
10	1.62	4.34	5.86
11	2.74	3.62	5.44
12	1.75	2.89	4.17
13	2.19	2.00	3.27
14	1.59	4.17	5.64
15	2.59	3.76	5.55
16	2.30	4.90	6.79
17	1.59	3.41	4.72
18	1.68	0.72	1.54
19	2.16	2.83	4.26
20	1.92	1.83	2.97
21	0.72	3.01	3.90
22	1.41	3.25	4.46
23	1.81	1.77	2.85
24	2.10	3.26	4.76
25	2.22	2.02	3.32
26	1.03	3.19	4.24
27	2.34	3.56	5.21
28	2.74	3.16	4.89
29	1.72	1.96	3.04
30	2.68	3.17	4.88
Mean	1.86	3.01	4.37
Std. Dev.	0.59	0.86	1.08

Table B.1 Data from normal distributions combined linearly

Variable	Definition	Units	Mean	COV	Distribution Type
Y_1	yield strength	ksi	40	0.125	lognormal
Y_2	section modulus	in^3	50	0.05	lognormal
Y_3	applied moment	in-kips	1000	0.20	Gumbel

Table B.3 Probability distributions on input random variables used in FORM analysis

No.	Y1	Y2	X = Y1 . Y2
1	2.03	2.05	4.16
2	1.89	2.75	5.20
3	2.34	3.19	7.46
4	1.47	3.31	4.86
5	1.97	3.63	7.14
6	1.43	5.26	7.53
7	1.94	3.72	7.22
8	2.10	2.58	5.42
9	1.23	4.44	5.48
10	2.79	1.02	2.84
11	2.43	4.02	9.79
12	1.78	3.10	5.52
13	1.43	2.83	4.04
14	2.10	1.94	4.09
15	2.18	3.31	7.23
16	2.37	2.20	5.20
17	1.36	3.50	4.75
18	1.23	3.49	4.29
19	0.97	2.49	2.41
20	2.02	1.73	3.49
21	1.47	4.16	6.11
22	1.25	3.10	3.89
23	2.59	2.16	5.59
24	2.51	4.52	11.34
25	2.15	1.60	3.43
26	1.84	1.77	3.27
27	2.17	3.18	6.91
28	1.56	3.10	4.84
29	1.63	2.17	3.55
30	2.07	3.50	7.24
Mean	1.87	2.99	5.48
Std. Dev.	0.46	0.97	2.03

Table B.2 Data from normal distributions combined non-linearly