

Risk Analysis of Development Programs

Peter P. Pandolfini

A set-theory methodology for risk analysis was first proposed by Benjamin Blanchard. The methodology is explained here in terms of unfavorable events in technical, schedule, and cost performances. This is a comprehensive way to evaluate the risks contained in development programs from a program management perspective. By using program state tables and combining them with the assignment of Bayesian probabilities based on a maximum entropy principle, the overall risk of a program issue is evaluated using only the information at hand. The determination of iso-risk boundaries implied by this methodology is illustrated, and sample presentations of results are shown.

BASIC THEORY

The analysis of risk can be based on an adaptation of a method introduced by Blanchard¹ in the early 1980s that uses set theory to establish the probability of unfavorable events—the union of failures and impacts. In a program management environment, system development can be viewed in terms of favorable and unfavorable events, which contain all the events that can happen during program development (Fig. 1).

An event is the basic unit of the space. The sets considered may contain technical, schedule, or cost performance events (or combinations thereof). Only the unfavorable events (i.e., failures and impacts) are used for determining risk. A failure is an unfavorable event, and an impact is an unfavorable event that follows a failure;

an impact may also be a failure. The unfavorable events, comprising several sets, are contained in Fig. 1.

In this diagram, T, S, and C denote technical, schedule, and cost performance events, respectively. L is the set of failures that can contain these events. The sets I_T , I_S , and I_C contain the (unfavorable) impact events. Figure 1 contains 13 regions in its union of unfavorable events, as summarized in Table 1. For example, the intersection of the I_T and I_S sets contains those unfavorable events that have technical and schedule impacts; at the intersection of all these sets are the events that are simultaneously technical, cost, and schedule failures.

Risk is defined as the probability of occurrence of the unfavorable events (i.e., the probability of the union of

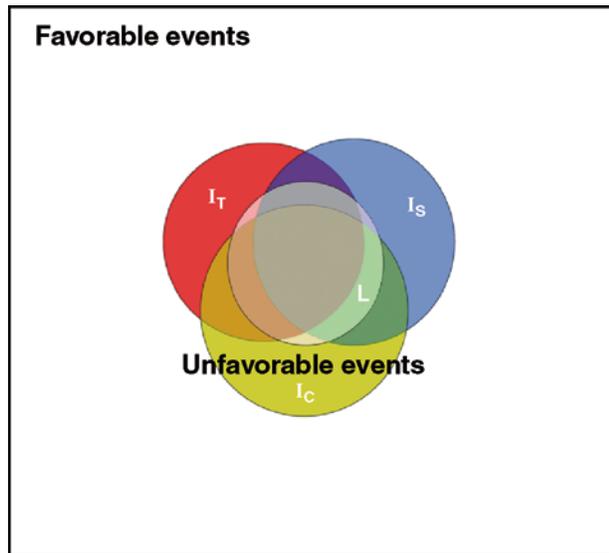


Figure 1. Event space for program development. All program events are contained within the box. The unfavorable events are represented in a Venn diagram.

Table 1. Descriptions of regions containing unfavorable events.

Region	Description
1	Technical impact events
2	Schedule impact events
3	Cost impact events
4	Technical and schedule impact events
5	Schedule and cost impact events
6	Technical and cost impact events
7	Technical FAILURE events
8	Technical and schedule FAILURE events
9	Schedule FAILURE events
10	Technical and cost FAILURE events
11	Cost FAILURE events
12	Schedule and cost FAILURE events
13	Technical, schedule, and cost FAILURE events

all sets containing unfavorable events). In terms of the notation for the failure and impact sets,

$$RISK = P(L \cup I_T \cup I_S \cup I_C) = P(L \cup I),$$

where $I = I_T \cup I_S \cup I_C$, $P()$ is the probability of the arguments (in this case, the union of several sets), and “ \cup ” is the symbol used to denote the union. Thus, $L \cup I_T \cup I_S \cup I_C$ (or $L \cup I$ in the shorthand introduced) is a compound event comprising the union of the four sets

of unfavorable events. The probability of occurrence of this compound event is derived from set theory and is

$$RISK = P(L) + P(I) - P(I|L)P(L),$$

where $P(I|L)$ denotes a conditional probability (i.e., the probability of I given that L has occurred). (A similar expression can be derived with the analysis of the RISK complement, $RISK' = P(L')P(I')$, where $P(L')$ is the probability of the complement of a failure (i.e., a success) equal to $(1 - P(L))$, and $P(I')$ is the probability of the complement of unfavorable impacts (i.e., favorable impacts) equal to $(1 - P(I))$. Substitution of the complementary relationships, noting $RISK = 1 - RISK'$, yields $RISK = P(L) + P(I) - P(I)P(L)$.)

In the above expression three probabilities— $P(L)$, $P(I)$, and $P(I|L)$ —must be evaluated to determine the iso-risk curves. It can be shown that the assumption that a value of $P(I)$ equal to the value of $P(I|L)$ yields a conservative set of iso-risk lines that are compatible with program management control (otherwise, risks may be underestimated). A discussion of other assumptions for determining $P(I|L)$ is beyond the scope of this article but would yield some interesting insights into the way iso-risk boundaries have been arbitrarily drawn in other analyses.

EVALUATION PROCESS

To apply this basic approach to a development program risk analysis, a statement of potential risk is analyzed with respect to four state tables. These tables contain language suggesting the state of the program with respect to a given issue (the issue for which risk is being evaluated). The “likelihood” table permits evaluation of probability of the failure associated with the issue, and the other three tables evaluate the extent of the impact of the failure in the technical, schedule, and cost dimensions. Figure 2 illustrates the flow.

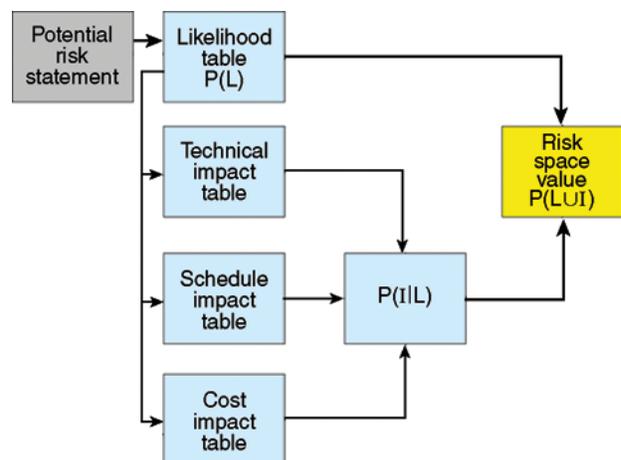


Figure 2. Risk evaluation process flow.

Considering the potential risk statement, the state of the program is chosen from a likelihood table that lists program states from best to worst. This state will result in the assignment of $P(L)$. Next, the state of the program with respect to the conditional realization of the situation contained in the risk statement is chosen for technical, schedule, and cost impact using the “impact” tables. These states will result in the assignments of $P(I_T|L)$, $P(I_S|L)$, and $P(I_C|L)$, which are then combined into $P(I|L)$.

The coordinates $P(L)$ and $P(I|L)$ are plotted in the risk space to illustrate where in the risk region (i.e., low, medium, or high) the potential risk is located. Figure 3 is an example taken from a recently completed study; the data are evaluations from several subject matter experts of a particular risk statement.

The two curves in Fig. 3 are the iso-risk lines dividing the space into the low-, medium-, and high-risk regions. These lines are projections onto the $P(L) - P(I|L)$ plane of curves that divide a complex surface into three equal areas. The blue triangles are the evaluations of individual subject matter experts. The red diamond is the result of means of the responses to the state tables for all respondents; it does not represent a consensus.

An alternative way to present the combined results in terms of box plots is illustrated in Fig. 4. The analysis of Fig. 3 is contained in the left-most box plot. The values of the ordinate scale correspond to the values for risk, and the iso-risk lines divide the green-yellow and

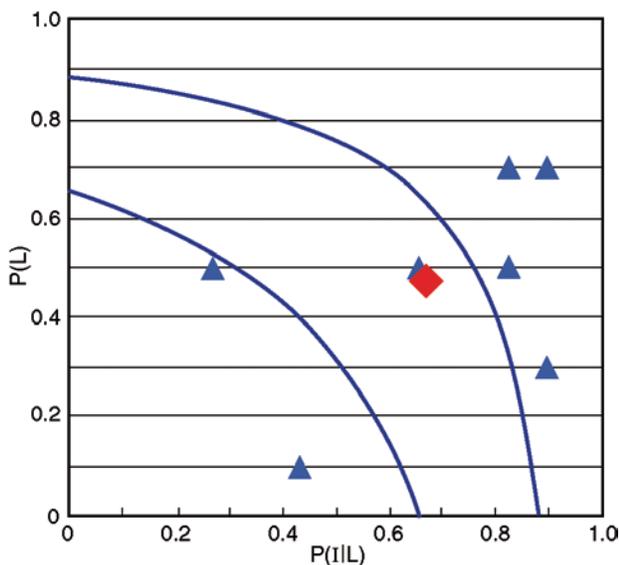


Figure 3. Plotted results in risk space.

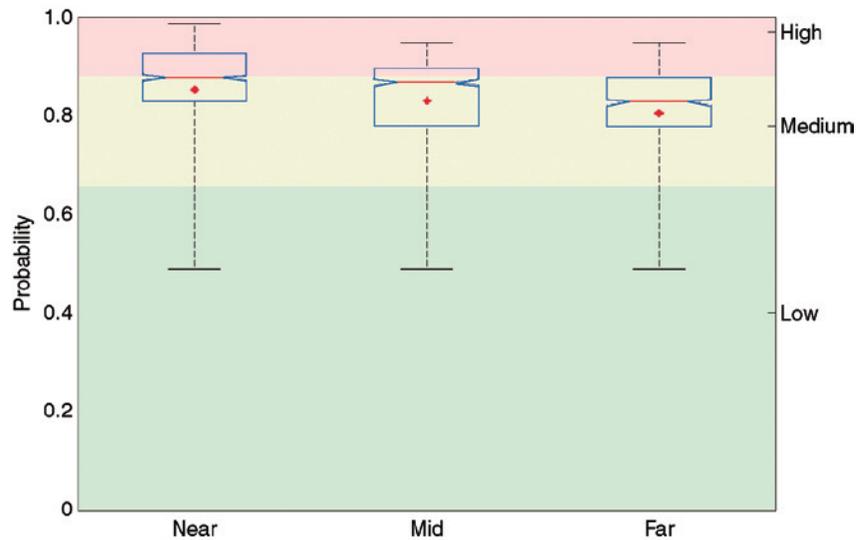


Figure 4. Box plot for the results of the example illustrated in Fig. 3.

yellow-red regions (see Fig. 7). The red diamonds are the result of using the medians of state-table responses, the tick marks indicate the spread in the results of all the responses, and the tops and bottoms of the boxes define the middle 50 percentile. The determination of the low-to-medium and medium-to-high boundaries is explained below in a separate section.

This box-plot method of illustrating the results permits the side-by-side comparison of several risk statements. In the example shown in the Fig. 4, a progression of a risk over three separate periods (i.e., “Near,” “Mid,” and “Far”) is analyzed as a design is changed and mitigation projects are completed. If the notches of two box plots do not overlap, then the results are deemed to be different to a 95% confidence level.

STATE TABLES

The state tables for evaluation of $P(L)$, $P(I_T)$, $P(I_S)$, and $P(I_C)$ follow the general pattern of best-to-worst state. In a recent analysis five levels per table were used. Table 2 is a generic version of likelihood states. Over the course of the study, several variations of this generic table were used depending on the nature of the issue for which risk was to be evaluated. Any number of states ≥ 1 can be used; however, five results is a reasonably fine detail without oversimplifying the choices or adding unnecessarily finer precision incompatible with the accuracy of the measurement. Table 3 illustrates the generic versions of impact states used for all risk statements. For each of these tables, a sixth choice (shown in italics) is added for those respondents who judge that they do not have enough information to choose a state. Depending on the content of justification statements that are required for each response to a program state, the treatment of the “opt out” response varies.

Table 2. Program likelihood states.

L	Program state
A	Best
B	Good
C	Average
D	Fair
E	Worst
F	<i>Opt out</i>

ASSIGNMENT OF PROBABILITY VALUES

A respondent (i.e., a subject matter expert) is asked to choose the state of the program with respect to the risk statement, but not assign a value to the probability that the program will experience the failure or impact. The assignment of probabilities to each state is developed by applying the maximum entropy principle.²

The probability of any event occurring in a given state is constrained by $0 < P < 1$ (note that the end values indicating certainty, 0 and 1, are not included in this range). For determining a value for $P(L)$, given a state, it is asserted that the probability of a given failure happening is such that $P(L)_A < P(L)_B < P(L)_C < P(L)_D < P(L)_E$. Consider Fig. 5: the worse the state, the greater the probability of failure. That is the extent of our prior information with respect to the probabilities of failure assigned to the states. (If additional information is known, e.g., the specific probability of a particular state, then it is incorporated directly into the state tables. The remaining unknown probabilities are determined by using the maximum information entropy method.) Assigning the intervals between the discrete probabilities requires another constraint. This results in $P(L)_A = 0.1$, $P(L)_B = 0.3$, $P(L)_C = 0.5$, $P(L)_D = 0.7$, and $P(L)_E = 0.9$.

This is how we arrive at these values: Referring to Fig. 5,

Table 3. Program impact states.

I_T, L	I_S, L	I_C, L	Program state
F1	G1	H1	Little impact
F2	G2	H2	Low impact
F3	G3	H3	Moderate impact
F4	G4	H4	Large impact
F5	G5	H5	Very significant impact
F6	G6	H6	<i>Opt out</i>

the sizes of the intervals (a , b , c , d , and e) dividing up the probability interval (0 to 1) can take on many different values and still satisfy $P(L)_A < P(L)_B < P(L)_C < P(L)_D < P(L)_E$. However, if the additional maximum information entropy constraint is imposed, then the quantity $[-a(\log_2 a) - b(\log_2 b) - c(\log_2 c) - d(\log_2 d) - e(\log_2 e)]$ must be maximized. The maximum information entropy occurs when $a = b = c = d = e = 0.2$. Thus, $P(L)_A$ is greater than 0 and less than or equal to 0.2, and is set equal to 0.1, the mid-point of the interval. Similarly, the other probabilities $P(L)_B$, $P(L)_C$, $P(L)_D$, and $P(L)_E$ are set to 0.3, 0.5, 0.7, and 0.9, respectively. A similar argument is made for $P(I_x | L)_A$, $P(I_x | L)_B$, $P(I_x | L)_C$, $P(I_x | L)_D$, and $P(I_x | L)_E$, where $x = T, S$, or C , to establish these values as 0.1, 0.3, 0.5, 0.7, and 0.9, respectively.

The two equations required to evaluate RISK,

$$\text{RISK} = P(L \cup I) = P(L) + P(I) - P(I | L)P(L)$$

and

$$\begin{aligned} P(I | L) &= P(I_T | L \cup I_S | L \cup I_C | L) \\ &= P(I_T | L) + P(I_S | L) + P(I_C | L) - P(I_S | I_T, L)P(I_T | L) \\ &\quad - P(I_C | I_T, L)P(I_T | L) - P(I_C | I_S, L)P(I_S | L) \\ &\quad + P(I_T | L)P(I_S | L)P(I_C | L), \end{aligned}$$

contain eight independent probabilities: $P(L)$, $P(I)$, $P(I_T | L)$, $P(I_S | L)$, $P(I_C | L)$, $P(I_S | I_T, L)$, $P(I_C | I_T, L)$, and $P(I_C | I_S, L)$. During a survey, four probabilities are measured, viz., $P(L)$, $P(I_T | L)$, $P(I_S | L)$, and $P(I_C | L)$. For purposes of a practical evaluation, the values of $P(I_S | I_T, L)$, $P(I_C | I_T, L)$, and $P(I_C | I_S, L)$ are assumed to be equal to $P(I_S | L)$, $P(I_C | L)$, and $P(I_C | L)$, respectively. This assumption simplifies the second equation; sample measurements from a completed study showed no significant differences from the $P(I_T | L)$, $P(I_S | L)$, and $P(I_C | L)$ measurements. The additional assumption that the value of $P(I)$ is equal to $P(I | L)$, as previously discussed, adds an additional level of conservatism in drawing the iso-risk boundaries.

The protocol of assigning a risk value for *opt out* responses is to rely on the maximum entropy principle. Usually, a value of 0.5 is used if the respondent indicates an unbiased ignorance of the state. If a collection of responses is being combined, the *opt out* response is usually not included because there are other responses to process, unless all the responses are *opt out*.

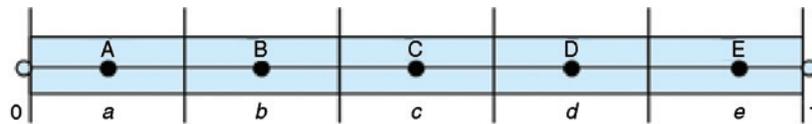


Figure 5. Division of the probability space.

ISO-RISK BOUNDARIES

The equation $RISK = P(L) + P(I) - P(I|L)P(L)$, with its simplifying assumption $P(I) = P(I|L)$, yields the complex surface illustrated in Fig. 6. An iso-risk is defined as the locus of points on this surface where RISK is constant.

To establish the classification boundaries among, say, low, medium, and high risks, the iso-risk curves that divide the three-dimensional surface area into equal parts must be found. The analysis was conducted using *Mathematica*, and the result is that $R = 0.657917$ and 0.880472 divides the surface bounded by $0 < P, C < 1$ into thirds. The projection of these boundaries onto the P-C plane (corresponding to the $P(L) - P(I|L)$ plane in the current analysis) is shown in Fig. 7. Figure 7 represents a projection of the three-dimensional surface onto a plane yielding a “distorted” look to the low-medium-high risk areas when they are viewed in two dimensions and is analogous to Fig. 3, showing the source of the iso-risk boundaries. The shape of these iso-risk boundaries depends on the relationship between $P(I|L)$ and $P(I)$ —assumed to have equal values in the current analysis. Distortion of these boundaries by other assumptions is a subject for a separate paper.

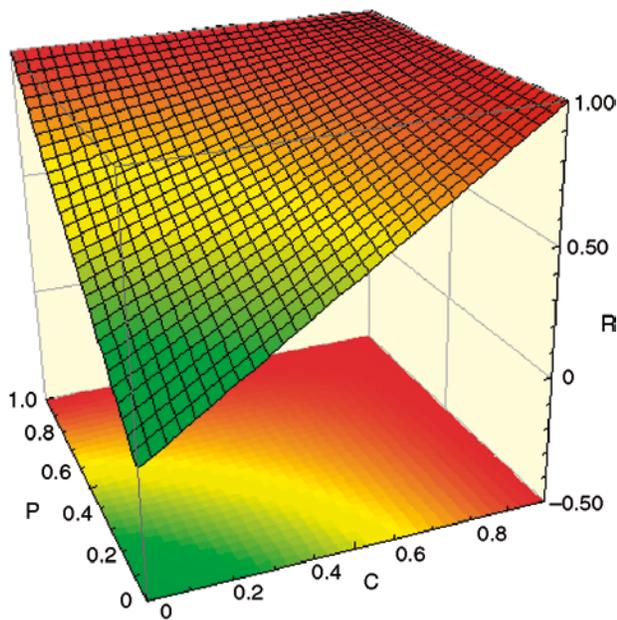


Figure 6. Risk surface for $R = P + C - P \cdot C$.

RISK PHASE SPACE

As a result of the foregoing analysis, a large but finite phase space for the discrete values of RISK is established. For the risk equation and its simplifying assumptions that reduce its evaluation to four probability measurements (i.e., $P(L)$, $P(I_T|L)$, $P(I_S|L)$, and $P(I_C|L)$), and the five values (i.e., 0.1, 0.3, 0.5, 0.7, and 0.9) per the maximum information entropy argument, to which they are limited, there are 625 ways to compute a RISK value (5^4). However, many of these values occupy the same phase (i.e., equal values of RISK, although the location may be different in the plane); there are only 55 discrete phases, 4 of which fall into the “green” region (7.3%), 12 into the “yellow” region (21.8%), and 39 into the “red” region (70.9%).

As part of an analysis, the distribution of the risk levels may be compared to the distribution of levels in the phase space (by, for example, employing a chi-squared test on the distribution of green-yellow-red results). Such an analysis may provide insight into whether the results are random or a consequence of deliberation by subject matter experts.

CONCLUDING REMARKS

This article presents a derivation of the methodology first introduced by Blanchard and combines his approach with a way of assigning probabilities that are determined by a maximum entropy constraint. In this approach, program issues are analyzed using state tables. Both failures and impacts are treated as unfavorable events and

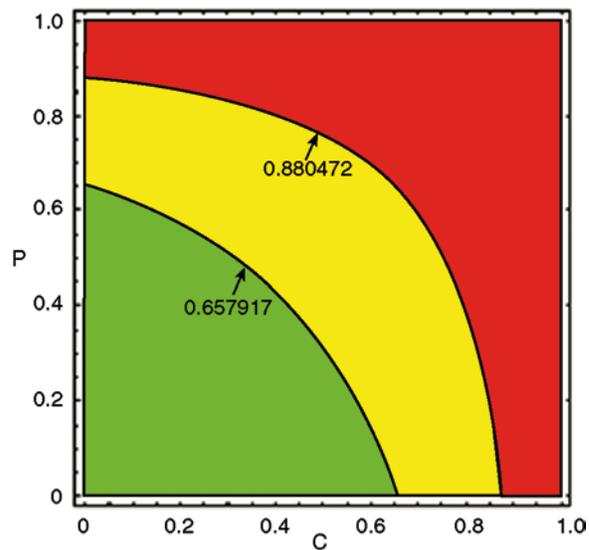


Figure 7. Projection of iso-risk boundaries.

are linked in their union. The assigned probabilities to failures and impacts are used in the risk equation, which computes the probability of the union of unfavorable events. The use of the probability of the union ensures that program risks are not underestimated.

The separation of the levels of risk (into, e.g., three categories—low, medium, or high) is accomplished by dividing the area of the complex surface generated by the risk equation and projecting the divisions, the iso-risks, onto a plane.

The use of set theory and the maximum entropy principle leads to a straightforward and elegant way to analyze program risks on a sound mathematical basis. This method, with the judicious use of state tables, analyzes a program issue comprehensively in its technical, schedule, and cost performances—the types of events of interest in the management of development programs.

ACKNOWLEDGMENTS. The author wishes to thank Eric Greenberg and Jack Calman for their assistance in preparing several illustrations. Eric's analysis supplied the box plot illustrated in Fig. 4, and Jack's analysis prepared Fig. 6.

REFERENCES AND NOTES

¹Blanchard, B. S., *System Engineering Management*, 3rd Ed., John Wiley & Sons, Inc., Hoboken, NJ (2004).

²Jaynes, E. T., *Probability Theory, The Logic of Science*, Cambridge University Press, Cambridge, UK (2004). See, especially, Chap. 11, which discusses the assignment of discrete prior probabilities and application of the entropy principle (i.e., application of Shannon's information entropy). Jaynes published many other works employing this principle, but this text summarizes his lifework in this field.

The Author

Peter P. Pandolfini is a member of APL's Principal Professional Staff. He graduated from the City College of New York (CCNY) with a civil engineering degree and from Rutgers University with M.S. and Ph.D. degrees in mechanical engineering. In 1984, he was a member of the first technical management class to graduate from the JHU G. W. C. Whiting School evening program. As a 35-year member of the APL staff, Dr. Pandolfini has performed research on air-breathing engines and on alternate energy conversion power systems. During the 1990s, he led APL's participation in several aircraft development programs. Currently, he is a member of the National Security Analysis Department and leads a section that specializes in modeling and simulation assessment, including tasks for the Assessment Division of the Navy. Dr. Pandolfini is a Registered Professional Engineer in Maryland. His e-mail address is peter.pandolfini@jhuapl.edu.



Peter P. Pandolfini