Probabilistic Budgeting and Time-Planning

J.K. Vrijling and P.H.A.J.M. van Gelder
Delft University of Technology

Abstract: In this paper an overview is given of the principles of probabilistic budgeting and time planning. Uncertainties related to normal – and special events are described. Analytical expressions are presented. To deal with correlations between special events, an alternative for the classical product moment correlation coefficient is proposed.

1 Introduction

The management and control of the costs and the duration of civil engineering projects have been studied in literature since the early 1970’s [5, 3, 7, 8]. Apart from uncertainties concerning quantities and production times (such as unit prices and wages), uncertainties involving economics and “influences from the outside world” (essentially changes in the design of the project), the difficulties in modelling the budgeting and time-planning are caused by the long period over which a civil engineering project is stretched. This period starts when the social demand for change is felt and the first plan for the project is outlined. The end of the period can be set at the delivery of the final product and the settlement of the bill.

The estimate of the budget is an approximation of the real costs. If all knowledge and facts have to be expressed in one single number, as is often required, discrepancies between the estimate and the finally realised costs cannot be avoided. Normally, in comparisons between the original estimate of the building costs and the total expenses at the end of the project, no correction is incorporated for the overall increase in prices. In Table 1, the exceedence of the budgets for the hydraulic engineering projects of the reclamation of some IJsselmeerpolders is given [1]:

Table 1:

<table>
<thead>
<tr>
<th>Polder:</th>
<th>Exceedence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wieringermeerpolder</td>
<td>12%</td>
</tr>
<tr>
<td>Northeast polder</td>
<td>16%</td>
</tr>
<tr>
<td>Easterly Flevoland</td>
<td>-3%</td>
</tr>
</tbody>
</table>

The difference in accuracy between an estimate of the budget in an early stage of the project (the study-of-plan phase) and the final estimate (builder’s specifications at the start of the engineering) is illustrated in the next table [3]:

Table 2:

<table>
<thead>
<tr>
<th>Project</th>
<th>Difference in % of the final costs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate in study-of-plan phase</td>
</tr>
<tr>
<td>Haringvliet locks</td>
<td>77%</td>
</tr>
<tr>
<td>Grevelingen dam</td>
<td>-19%</td>
</tr>
<tr>
<td>Volkerak dam</td>
<td>56%</td>
</tr>
<tr>
<td>Brouwers dam</td>
<td>-39%</td>
</tr>
</tbody>
</table>

Often final project costs exceed their estimate. Historical estimates of the budget at the Ministry of Public Works in The Netherlands (Rijkswaterstaat) [4] clearly show a pattern of increase of costs and delay in work (which, due to inflation and loss of interest, also increases the costs):

Table 3:

<table>
<thead>
<tr>
<th>Project</th>
<th>Start</th>
<th>Planned</th>
<th>Reality</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mf</td>
<td>years</td>
</tr>
<tr>
<td>Noordhollands Canal</td>
<td>1818</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Haarlemmer lake</td>
<td>1837</td>
<td>8.4</td>
<td>5 (?)</td>
</tr>
<tr>
<td>Nieuwe Waterweg</td>
<td>1858</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Maas &amp; Waal</td>
<td>1864</td>
<td>6.5</td>
<td>10</td>
</tr>
</tbody>
</table>

In road-construction projects there were (and there are) large fluctuations in the differences between the estimated budgets and the real costs as well. For road-construction projects in The Netherlands from 1980 up to 1985, differences between estimates in the pre-design phase and the real costs as a percentage of the real costs are given in the following histogram:
Publications on exceedence of the estimates of budgets in other countries show the same tendency [2] and [5].

2 The classical approach to budget estimates and time-planning

From early days, the calculation of estimates of the budget and time-planning schemes are based on the most likely values. Uncertainty of the decision maker is expressed as an increase of the deterministic final amount by a certain percentage of that amount.

Uncertainty regarding the budget estimate or the time-planning scheme is not constant during the project. The later on in the project an estimate of the budget or a time-planning scheme is made, the more about the project is known and the decision maker’s uncertainty concerning the estimated amounts of money and duration of the activities will be less than in the early stages of the project. A classification of project phases in order of time is given in Table 4. The project parts in the phase at hand are estimated or planned in detail, for other parts in other phases the estimates are determined roughly.

Table 4:

<table>
<thead>
<tr>
<th>Class</th>
<th>Project phase</th>
</tr>
</thead>
<tbody>
<tr>
<td>D</td>
<td>study-of-plan</td>
</tr>
<tr>
<td>C</td>
<td>pre-design</td>
</tr>
<tr>
<td>B</td>
<td>pre-builder’s specifications</td>
</tr>
<tr>
<td>A</td>
<td>builder’s specifications</td>
</tr>
</tbody>
</table>

Because of the decision maker’s greater uncertainty in the early phases of the project, it is of no use to make detailed estimates and time-planning schemes for phases to come. Making an
estimate of the budget in detail for builder’s specifications when the project is still in the study-of-plan phase will turn out to be a waste of time, although in the early stages more detailed estimates and time-planning schemes (or parts of those) are made.

**Budget-estimates**

First, some examples of specifications of budget estimates in several phases of a project are given.

**Estimate of the budget, Class D (Study-of-plan phase)**

\[
\begin{align*}
1 \text{ viaduct} \times \text{price of 1 viaduct} &= \text{item viaduct} \\
5 \text{ km of road} \times \text{price per km of road} &= \text{item road} \\
1 \text{ tunnel} \times \text{price per tunnel} &= \text{item tunnel}
\end{align*}
\]

\[-=\]

Total of Direct costs

Indirect costs

\[-=\]

Primary costs

Additional costs

Miscellaneous\(^1\)

\[-=\]

Basic estimate

Unforeseen \(^2\)

\[-=\]

Estimate (ex. VAT)

VAT

\[-=\]

**Study-of-plan phase estimate**

In the successive phases the items of the class D- estimate are worked out in more detail. An example is given in the following class B- estimate.

**Estimate of the budget, class B (Pre-builders specifications phase):**

\[
\begin{align*}
800 \text{ m}^3 \text{ soil} \times \text{price per m}^3 &= \text{item "soil"} \\
800 \text{ m}^3 \text{ concrete} \times \text{price per m}^3 &= \text{item "concrete"} \\
1 \text{ ton reinforcement} \times \text{price per ton} &= \text{item "reinforcement"} \\
800 \text{ m}^2 \text{ formwork} \times \text{price per m}^2 &= \text{item "formwork"}
\end{align*}
\]

\[-=\]

Subtotal viaduct

The other items are detailed analogously.

In a class A-estimate (estimate for builders specifications) the prices per m\(^3\), m\(^2\) or ton are determined in more detail, based on working methods and quantities. In an estimate in this

\(^1\) In “Miscellaneous” those costs are categorized which are known but which are not specified. For a study-of-plan phase estimate these could be: land (has to be bought) preparation, deflection of conduit-pipes and water courses, temporary diversion of traffic, etc..

\(^2\) “Unforeseen” is taken as an additional percentage on the Basic estimate here. If there is little insight in the character of the item Unforeseen then this way of calculation is applicable.
phase, time and equipment are taken into consideration. Contractors prefer this method of estimating.

A sub-item of the item SOIL of the partial project ROAD from the class D-estimate (5 km of road) is chosen as an example of specification of an item in a class A-estimate:

For the delivery and processing of 80,000 m$^3$ of soil for the partial project "road" the costs of the following means of production are estimated:

\[
\begin{align*}
\text{Delivery at the quay by ship} & \quad 80000 \text{ m}^3 \times \text{ price per m}^3 = \text{ partial item 1} \\
\text{Lease of an unloading plant} & \quad 80 \text{ days} \times \text{ day tariff} = \text{ partial item 2} \\
\text{Transport to location (by cars)} & \quad 800 \text{ days} \times \text{ day tariff} = \text{ partial item 3} \\
\text{Equipment for processing and compaction} & \quad 85 \text{ days} \times \text{ day tariff} = \text{ partial item 4}
\end{align*}
\]

---------- +

Subtotal soil

The price per m$^3$ of processed soil is calculated by division by the volume in m$^3$ (here: 80,000).

In principle, the estimate of Direct costs (an example of which was given for a class D-estimate at the bottom of the previous page) follows from an addition over all N partial items of the multiplication of quantity, $h_i$, and the prices per unit, $p_i$ (see the calculation scheme of the budget estimate on the next page). Indirect costs can be calculated by an additional percentage, $\%_1$, which is a fixed percentage of the Direct costs.

Additional costs and Miscellaneous can both be expressed as a percentage $\%_2$ of the Primary costs. As stated above, additional costs are established as a percentage of the preceding part of the estimate. The Total estimate can thus be expressed as a function of the total of the Direct costs. A percentage of (the part of) the afore calculated estimate is called an additional percentage $^1$).

In any phase of the project such percentages can differ. Generally, the Total estimate, in which the additional percentages are included, is calculated from the total Direct costs:

\[
\text{Estimate} = \left( \prod_{j=1}^{M} (1 + \%_j) \right) \left( \sum_{i=1}^{N} h_i p_i \right)
\]

in which:

$M = \text{ the number of additional percentages}$

$\%_j = \text{ the } j^{th} \text{ addition over the foregoing subtotal}$

$N = \text{ the number of cost items in the Direct costs}$

---

1) An additional percentage (for example the percentage Unforeseen) can be seen as a random variable or as a deterministic constant. For example the percentage VAT was fixed at 17.5% for a long time. It can be regarded as a fixed constant, unless it is expected to be changed in the future. Then, introducing the percentage VAT as a random variable is an option. In 2009, VAT is 19%.
\[ h_i = \text{the quantity of the } i\text{th item in the Direct costs} \]

\[ p_i = \text{the unit price in the } i\text{th item in the Direct costs}. \]

**Time-planning**

In an estimate of the budget the unit prices have to be multiplied by their quantity and then added. In a time-planning only the duration or the lengths of time of all activities have to be added and no multiplication is needed:

\[
D = \sum_{i=1}^{N} T_i
\]

where:

- \( i \) = rotation number of the activity
- \( T_i \) = duration of the activity \( i \)
- \( N \) = number of activities (in Figure 2: \( N = 4 \)).

If the various activities succeed each other in time the time-planning is simple. Example: In a small building pit, pile driving cannot be started before digging of the pit is completed. The digging hinders pile driving too much.

Usually not all activities succeed each other in time. Various activities are (partially or totally) executed simultaneously. A consecutive activity can often only be started when more than one preceding activity have been completed.

In the figure below it is shown that the pre-builder’s specifications phase can only be started when the pre-design phase has been completed and the licences have been granted. Both the
activities have to be completed before the pre-builder’s specifications phase can be started. If two activities are executed in parallel (in Figure 3: the pre-design phase and the granting of licences) the time-planning can be sketched as follows:

![Time-planning Diagram](image)

Figure 3.

In this example there are two time-paths:

\[
D_a = T_1 + T_2 + T_4 + T_5 \\
D_b = T_1 + T_3 + T_4 + T_5
\]

The total duration of the project, \( D_{tot} \), is the maximum amount of time according to the duration of the various time-paths: \( D_{tot} = \max(D_a, D_b) \).

In the example in Figure 3, the duration of the activities 1, 3, 4 and 5 determine the duration of the total project, \( D_{tot} \). It is said that these activities form the critical time-path (critical path for short).

### 3 Uncertainty concerning budget estimates and time-planning

In order to express one’s uncertainty about the estimate or the time-planning, probabilistic techniques are employed. If the random character of the cost items or the duration of the various activities are taken into account, the budget estimate or time-planning of the project is said to be statistically controlled.

The estimated amount of money or the planned duration of a project can be interpreted in various ways, depending on what the person who estimates or plans has in mind. Does he or she focus on an estimate of the mean costs or duration (the expected values) or on the amount that is most likely (the mode of the costs or the duration)? In the first case it is commonly assumed that the costs or the duration are normally distributed. The mean and the mode then
coincide. In the second case other (skewed) probability density functions are possible. Triangular probability density functions are often used for this purpose.

In addition to the mode or the mean of the estimate of the budget or the time-planning, the deviation from it (or the spread around it) is important. The size of the margin depends on the phase the project is in and on the required reliability with which the budget or the planned duration (quantification of the estimate or time-planning plus margin) will not be exceeded.

Considering the first point (size of the margin depends on the phase the project is in), in an early stage of the project one is much more uncertain about the budget estimate or the planned duration than in a later stage. Estimates and time-plans are often classified according to the phase of the project they were made in. Characteristic numbers for the magnitude and for the spreading around costs items depend on the project phase. For time-planning such characteristic numbers are not (yet) available.

4 Classification of uncertainty

Uncertainty can be classified in three categories:

♦ uncertainty related to Normal events;
♦ uncertainty related to Special events;
♦ project uncertainty.

Uncertainty related to Normal events

Although the costs items in a Basic budget estimate of a project become increasingly clear in the course of time, and the estimate becomes more accurate, many causes of uncertainty will remain as long as the project is not finished. With the necessary changes made, this can be applied to time-planning. The degree of uncertainty can be classified as follows:

1. There is no cause of uncertainty. The item concerned is deterministic. This concerns costs items or activities that are known exactly in size or duration. If, for example, the contract settling the purchase of land has been signed, this amount of money is known exactly. An “activity” with deterministic duration is the tide. The duration (along the North Sea coasts) is “exactly” 12 hours 25 minutes.
Often the costs are not so uniquely determined and one is uncertain about the duration of an activity. When the negotiations are still in progress, there is a notion about how much money the land (meant in point 1) will cost, but one cannot be certain. An example of uncertainty about the duration of an activity is a barge with heavy draught that has to be towed over a sill. Suppose this can only be done at high tide. (The keel clearance has to be sufficient). Usually the final costs or the spreading around the point in time of arrival at the sill will be within a band width. The probability density can then be as indicated in Figure 6.

![Figure 5.](image)

![Figure 6.](image)

**Uncertainty related to Special events**

Often another type of uncertainty plays a role with the evaluation of costs of a project or its duration, namely uncertainty caused by the Unforeseen or by Special events (mainly calamities). Two criteria characterize a Special event: the first is that it is not meant to occur and the second is that occurrence is not likely. The probability of occurrence, p, is small (less than 0.05), but if the event occurs, the consequence (damage or exceedence of the duration, B) is large. The probability of no occurrence (and accordingly: no damage or exceedence of the duration) is 1 - p. In a “classical” estimate of the budget or time-planning such events are seldom taken into account. Contractors insure against such events, associated with small probabilities but with large consequences. In a statistically controlled estimate or time-planning the probabilities and the consequences can be indicated as follows:
3. Figure 7 shows the mass density of a Special event as a function of the damage or exceedence of the duration.

![Figure 7](image1)

Figure 7.

4. The probability density function of a “Special event”, of which the consequences (the damage or the duration) are subject to uncertainty, could be as is illustrated in Figure 8.

![Figure 8](image2)

Figure 8.

**Project uncertainty or plan uncertainty: variants**

5. In the study-of-plan phase several variants have to be considered and estimated (and sometimes planned). Beforehand, one is not certain which variant will be chosen (for example a tunnel, a ferry or a bridge for a road crossing of a river). Only at the end of the pre-design phase a decision is made. Awaiting the choice, elaborating and estimating (and eventually time-planning) several variants mainly meet the uncertainty. Sometimes, the decision between the variants is so unpredictable that all variants are considered equally likely. Sometimes one variant is preferential and it is unlikely that another one will be chosen.
Figure 9.

If more than one variant is estimated or planned, the problem could be that the estimate of the budget is required to be one total amount of money or the time-planning should be one total duration. One estimate of the budget or one time-planning (possibly with a margin or expressed as a probability density function) for presentational purposes is then acquired rationally by weighing each variant by its (estimated) probability of selection. The following figure presents the result for two variants.

Figure 10.

The disadvantage is that the result is not recognized as a reasonable estimate or time-planning of each of the individual variants.

The classified uncertainty and the formulae to calculate the associated means and the standard deviations of the probability-weighed consequence (the related risks) are summarized in the following table. Mutatis mutandis the formulae hold for a time-planning with only one time-path.

<table>
<thead>
<tr>
<th>Case</th>
<th>mean</th>
<th>standard deviation</th>
<th>Description</th>
</tr>
</thead>
</table>

Table 5:
A deterministic amount of money, \( B \), expressed in units of money.

A stochastic amount of money, with mean \( B \), and some spreading, \( \sigma_B \).

An event with probability of occurrence \( p \) that has a deterministic consequence \( B \).

An event with probability of occurrence \( p \) that has a statistic consequence with mean \( B \) and some spreading, expressed by \( \sigma_B \).

There are two (or more) variants, each associated with a probability of realization, \( p_i \). Their probabilities add up to one as it is certain that one of the variants will be selected.

The spreading for an item of the estimate increases with the related uncertainty. In the first case in Table 5, one is absolutely certain about the size of the sum, \( B \). The standard deviation equals zero. In the second case there is some uncertainty. The spreading, \( \sigma_B \), is smaller than the expected value, \( B \). (If this were not so, the estimate of the item was of no significance. It then suggests that there is not the vaguest idea of the size of \( B \).) In case of Special events (cases 3 and 4), one is not certain if there will be costs (damage) at all. The probability that there will be costs is \( p \) (\( p \ll 1 \)). There is a larger probability (\( 1 - p \)) that there will be no costs. In fact the risk \( 1 \) (probability \( \times \) consequence = \( p \times B \)) is estimated. If the Special event occurs, the estimated amount of money (\( p \times B \)) is not nearly enough to cover the costs (\( B \)). According to this, the third case is associated with a larger spread (in the order of \( B \times \sqrt{p} \)) than in the case of a Normal event. So the spreading for Special events is approximately \( \frac{1}{\sqrt{p}} \) times the expected value \( p \times B \).

---

1 From a mathematical point of view the estimates of Normal events are estimates of the risks as well. The probability of occurrence of Normal events is 1 (or 100% certainty).
5 Calculation formulae for probabilistic budgeting

Monte Carlo simulations can be easily used in budgeting and time planning problems, but many expressions in budgeting and time planning problems can also be derived analytically.

Assume that \( X \) and \( Y \) are random variables (of for instance prices and amounts) with PDF’s \( f \) and \( g \) respectively. The following four functional operators are often encountered in budgeting and time planning problems:

\[
Z = X + Y, \quad U = X - Y, \quad V = XY \quad \text{and} \quad W = X/Y
\]

Then the PDF’s of \( Z, U, V \) and \( W \) are, respectively, given by:

\[
f_Z(z) = \int f(x)g(z-x) \, dx
\]

\[
f_U(u) = \int f(u+y)g(y) \, dy
\]

\[
f_V(v) = \int f(x)g(v/x) \, |x|^{-1} \, dx
\]

\[
f_W(w) = \int f(xw)g(x) \, |x| \, dx
\]

In textbooks on statistics the following relation is proven:

\[
E(XY) = E(X)E(Y)
\]

\[
\text{Var}\left( \sum a_i X_i \right) = \sum a_i^2 \text{Var}(X_i) + 2 \sum \sum a_i a_j \text{Cov}(X_i, X_j)
\]

Furthermore it is possible to derive the following property for the product of random variables:

\[
\text{Var}(V) = \text{Var}(X)\text{Var}(Y) + E^2(X)\text{Var}(Y) + E^2(Y)\text{Var}(X)
\]

If exact calculations are not possible, the following approximation rules can be used (using Taylor’s formula):

\[
g(X) = g(m_X) + (X - m_X) \frac{dg(x)}{dx} \bigg|_{x=m_X} + \frac{(X - m_X)^2}{2} \frac{d^2g(x)}{dx^2} \bigg|_{x=m_X} + ...
\]

From this, we can derive:
E(g(X)) . g(E(X))

Var(g(X)) . Var(X) [g’(mX)]^2

If the coefficient of variation of X is less than c, the error involved in these approximations is less than c^2. In particular, the following useful approximations in budgeting and time planning models can be used:

\[
E(\sqrt{X}) \approx \sqrt{E(X)}, \quad \text{Var}(\sqrt{X}) \approx \frac{\text{Var}(X)}{4E(X)}
\]

\[
E(X^{-1}) \approx \frac{1}{E(X)}, \quad \text{Var}(X^{-1}) \approx \frac{\text{Var}(X)}{E^4(X)}
\]

6 Dependent Bernoulli distributed random variables for special events

The notation X ~ BBG(p1,G1) is used for a random variable X which has a probability p1 of the occurrence of an event with consequences G1.

Mean and standard deviation of X are:

\[
\mu = p1 \times G1
\]

\[
\sigma = G1 \times \sqrt{p1(1-p1)}
\]

Extensions to this univariate random variable by allowing ‘horizontal and vertical uncertainties’ in p1 and G1 are described by Van Gelder [10].

Dependence

The classical product-moment correlation coefficient of Pearson is usually used as a measure for dependence between two random variables. This correlation coefficient however, can only be applied for normal distributed random variables. For non-normal distributions, the correlation structure should be described differently. In the remainder of this section, a suggestion for such structure is proposed.

Bivariate BBG’s

The following 4 situations can be distinguished:
Situation 1

Assume $X \sim BBG(p_1,G_1)$ and $Y \sim BBG(p_2,G_2)$ and independence, then the following probability table can be derived:

Table 6.

<table>
<thead>
<tr>
<th>X \ Y</th>
<th>0</th>
<th>G2</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>$(1-p_1)(1-p_2)$</td>
<td>$p_2(1-p_1)$</td>
<td>$1-p_1$</td>
</tr>
<tr>
<td>G1</td>
<td>$p_1(1-p_2)$</td>
<td>$p_1p_2$</td>
<td>$p_1$</td>
</tr>
<tr>
<td>Sum</td>
<td>$1-p_2$</td>
<td>$p_2$</td>
<td>$1$</td>
</tr>
</tbody>
</table>

Situation 2

If $X$ and $Y$ are completely positive dependent, then it follows,
- if $X=0$ then $Y=0$
- if $X=G_1$ then $Y=G_2$.

Using conditional probabilities: $P(Y=0|X=0) = 1$ and $P(Y=G_2|X=G_1) = 1$.

From which it follows: $P(X=0 \text{ and } Y=0) = P(X=0)P(Y=0|X=0)=1-p_1$
and

$P(X=G_1 \text{ and } Y=G_2) = P(X=G_1)P(Y=G_2|X=G_1)=p_1$

and

$P(X=0 \text{ and } Y=G_2) = P(X=0)P(Y=G_2|X=0)=0$

and

$P(X=G_1 \text{ and } Y=0) = P(X=G_1)P(Y=0|X=G_1)=1-p_1$.

Which can be summarized in the following probability table:
Table 7.

<table>
<thead>
<tr>
<th>X \ Y</th>
<th>0</th>
<th>G2</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1-p1</td>
<td>0</td>
<td>1-p1</td>
</tr>
<tr>
<td>G1</td>
<td>0</td>
<td>p1</td>
<td>p1</td>
</tr>
<tr>
<td>Som</td>
<td>1-p2</td>
<td>p2</td>
<td>1</td>
</tr>
</tbody>
</table>

From this table, it follows that $1-p_1=1-p_2$.

Therefore, BBG distributed X and Y can only be 100% positively correlated if $p_1=p_2$.

**Situation 3**

If X and Y are completely negatively dependent, then it follows that:

- if $X=0$ then $Y=G_2$
- if $X=G_1$ then $Y=0$.

The following probability table can be derived:

Table 9.

<table>
<thead>
<tr>
<th>X \ Y</th>
<th>0</th>
<th>G2</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>1-p1</td>
<td>1-p1</td>
</tr>
<tr>
<td>G1</td>
<td>p1</td>
<td>0</td>
<td>p1</td>
</tr>
<tr>
<td>Som</td>
<td>1-p2</td>
<td>p2</td>
<td>1</td>
</tr>
</tbody>
</table>

Requirement: $p_1=1-p_2$.

BBG distributed X and Y can only be 100% negatively correlated if $p_1+p_2=1$.

**Situation 4**

X and Y are partially dependent from each other:

- if $X=0$ then $Y=0$ with probability $a_1$ and $Y=G_2$ with probability $1-a_1$
- if $X=G_1$ then $Y=0$ with probability $b_1$ and $Y=G_2$ with probability $1-b_1$.

Table 10.

<table>
<thead>
<tr>
<th>X \ Y</th>
<th>0</th>
<th>G2</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>(1-p1)a1</td>
<td>(1-p1)(1-a1)</td>
<td>1-p1</td>
</tr>
<tr>
<td>G1</td>
<td>p1b1</td>
<td>p1(1-b1)</td>
<td>p1</td>
</tr>
<tr>
<td>Som</td>
<td>1-p2</td>
<td>p2</td>
<td>1</td>
</tr>
</tbody>
</table>
Requirement: \((1-p1)a1+p1b1=1-p2\)

BBG distributed X and Y can only be partially correlated if:

\[a1+p1(b1-a1)=1-p2.\]

Situation 4 is the most general situation. Situations 1, 2 and 3 can be derived directly from situation 4 with the following choices for \(a1\) and \(b1\):

Table 11.

<table>
<thead>
<tr>
<th>Situation</th>
<th>Choice for (a1) and (b1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Situation 1 (independent)</td>
<td>(a1=1-p2, b1=1-p2)</td>
</tr>
<tr>
<td>Situation 2 (100% positively dependent)</td>
<td>(a1=1, b1=0)</td>
</tr>
<tr>
<td>Situation 3 (100% negatively dependent)</td>
<td>(a1=0, b1=1)</td>
</tr>
</tbody>
</table>

We conclude that the correlation structure between 2 BBG distributed \(X \sim BBG(p1,G1)\) and \(Y \sim BBG(p2,G2)\), needs to be described by 2 constants \(a1\) and \(b1\) which satisfy the following 3 boundary conditions:

\[
0 \leq a1 \leq 1 \\
0 \leq b1 \leq 1 \\
a1+p1(b1-a1)=1-p2.
\]

Complete positive dependence is reached when \(a1 \to 1\) and \(b1 \to 0\) and \(p1 \to p2\).

Complete negative dependence is reached when \(a1 \to 0\) and \(b1 \to 1\) and \(p1 \to 1-p2\).

7 Conclusions

Budgeting and time planning should be handled by probabilistic methods in order to deal with uncertainties prior to the start of the project. Correlations between events need special attention, in particular the correlation structure between special events. Analytical solutions are presented in this paper and a correlation structure is proposed.

8 Literature


