Uncertainty calculations for estimates of project durations

David Meisch

DTU Informatics
Technical University of Denmark

1-3.11.2012 YEQT VI in Eindhoven

outline

- Project management
- The Successive Principle
- Different modeling approach
- Extension to a bivariate model
- Discussion

Construction Planning

Construction Planning

construction start 2007, estimated finish time 2010

Construction Planning

construction start 2007, estimated finish time 2010
not finished yet, estimated finish time 2014/2015

General problem

- Elbphilharmonie in Hamburg

General problem

- Elbphilharmonie in Hamburg
- The delay makes it impossible to book an opening act

General problem

- Elbphilharmonie in Hamburg
- The delay makes it impossible to book an opening act
- Not a unique case

"'Forecasts for planned projects have been constantly and remarkably inaccurate"'

Kahnemann (2003),Wachs (1990)

Project Management

Part of project management is

- organizing
- planning
- leading
- ...

Goal: Ensure quality of the project, keep the budget and finish the project on time.

General properties of a project

- A project consist of several subtask
- The duration of a subtask is either random or deterministic
- Subtask can run parallel, have more then one predecessor, ...

Classical duration estimates

Let us consider a project with the following structure

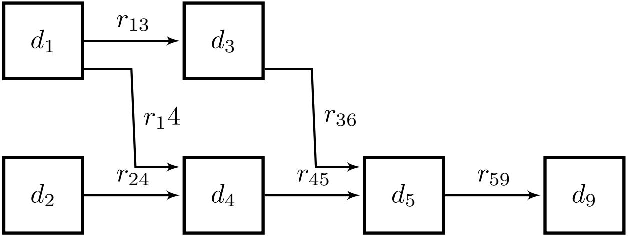

Figure: Main activities

Let D be the total duration of the project, d_i be the duration of subtask i and r_{ij} the specific relations between subtask i and j.

Classical duration estimates

Let us consider a project with the following structure

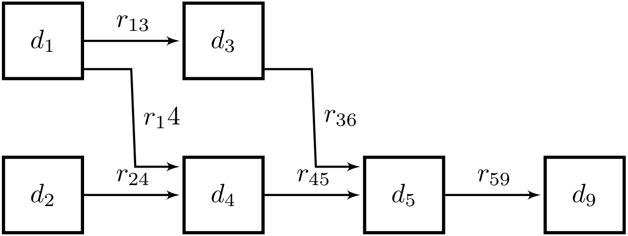

Figure: Main activities

If we assume $r_{ij} = 0$ then we can calculate $E[D] = max\{E[d_1] + E[d_3], max\{E[d_1], E[d_2]\} + E[d_4]\} + E[d_5] + E[d_9]$

Merge event bias

Let us consider a project with the following structure

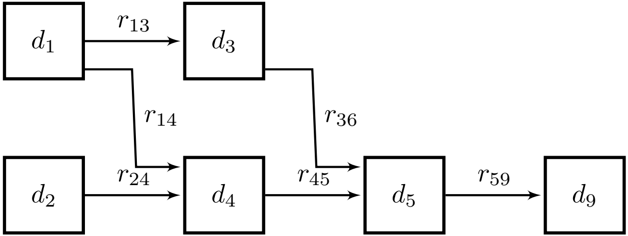

Figure: Main activities

$$E[\max\{d_1, d_2\} \geq \max\{E[d_1], E[d_2]\}$$

Merge event bias

Let us consider a project with the following structure

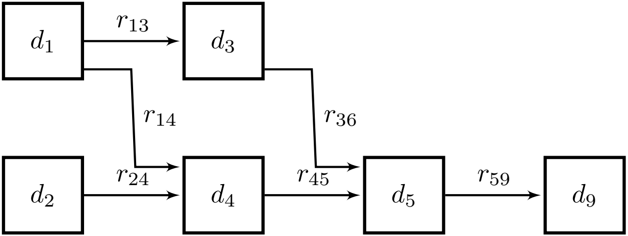

Figure: Main activities

$E[\max\{d_1, d_2\} \geq \max\{E[d_1], E[d_2]\}$ assume a distribution for the merge event bias and correct the term

Group Estimates

How many breweries are owned by Heineken? How high is the annual beer production of Heineken?

Group Estimates

Heineken owns over 125 breweries and produces over 139 million hectoliter ($3.67199153 \cdot 10^9$ Gallons) beer per year.

Group Estimates

Heineken owns over 125 breweries and produces over 139 million hectoliter ($3.67199153 \cdot 10^9$ Gallons) beer per year.

- Two heads are better then one
- An interval estimate is easier then a point estimate

Group Estimates

Heineken owns over 125 breweries and produces over 139 million hectoliter ($3.67199153 \cdot 10^9$ Gallons) beer per year.

- Two heads are better then one
- An interval estimate is easier then a point estimate
- PERT
- Successive Principle

Group Estimates

Heineken owns over 125 breweries and produces over 139 million hectoliter ($3.67199153 \cdot 10^9$ Gallons) beer per year.

- Two heads are better then one
- An interval estimate is easier then a point estimate
- PERT
- Successive Principle

Not only sharp estimates but also "‘soft"’ values, e.g. awareness of uncertainties, problems,...

The Successive Principle

- Developed in the 1970s
- Assumes a project can be split in independent subtask
- Subtask being analyzed using group estimates
- The analyzing group should be consistent of different kind of personality, e.g. positive and negative opinion about the project
- Used mainly for cost and duration estimates
- Uses subjective probabilities
- Assumes the cost/duration of a subtask is a random variable following an Erlang seven distribution

The Successive Principle

The estimated total cost of the project are the cumulative estimates of the cost estimates for all subtasks

The Successive Principle

The estimated total cost of the project are the cumulative estimates of the cost estimates for all subtasks

For the duration estimate of a project

1. Critical path
2. Near critical path
3. Merge even bias (MEB)
4. Sum of the durations of the subtask on the critical path plus MEB

Assumptions

There is space to improve the mathematical modeling in the Successive Principle

Assumptions

There is space to improve the mathematical modeling in the Successive Principle

Temporary assumptions

- Subtasks follow Erlang 2 distributions

Assumptions

There is space to improve the mathematical modeling in the Successive Principle

Temporary assumptions

- Subtasks follow Erlang 2 distributions
  Can easily be extended to Erlang 7 distributions

Assumptions

There is space to improve the mathematical modeling in the Successive Principle

Temporary assumptions

- Subtasks follow Erlang 2 distributions
  Can easily be extended to Erlang 7 distributions
- $r_{ij} = 0$

Assumptions

There is space to improve the mathematical modeling in the Successive Principle

Temporary assumptions

- Subtasks follow Erlang 2 distributions
  Can easily be extended to Erlang 7 distributions
- $r_{ij} = 0$
  later: r_{ij} follow an bilateral distribution

An international IT development project

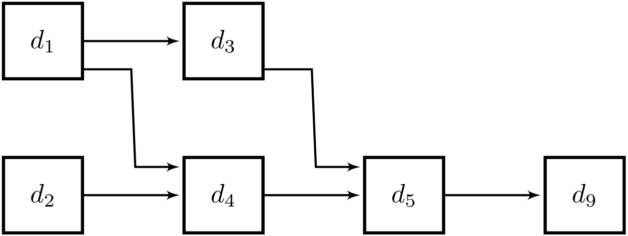

Figure: Project flow

With $E[d_1] = 12,4$, $E[d_2] = 15,8$, $E[d_3] = 7$, $E[d_4] = 6,4$, $E[d_5] = 4,2$ and $E[d_9] = 2$

An international IT development project

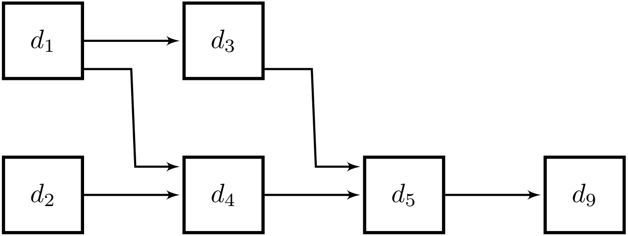

Figure: Project flow

simple approach: $E[D_s] = E[d_2] + E[d_4] + E[d_5] + E[d_9] = 28,4$

An international IT development project

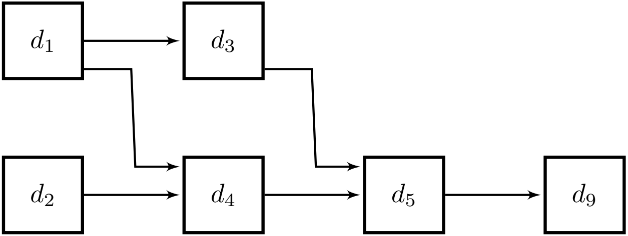

Figure: Project flow

simple approach: $E[D_s] = E[d_2] + E[d_4] + E[d_5] + E[d_9] = 28,4$
Successive Pinciple:
$E[D_{SP}] = E[d_2] + MEB_{12} + E[d_4] + E[d_5] + E[d_9] = 29,15$

The whole project as a PH distribution

Markov process $J(t)$ with states $(iijj)$, (ijj), (ij), (ii), i, and generator matrix

$$Q = \begin{pmatrix} T & -Te' \\ 0 & 0 \end{pmatrix}, \text{ and } R = r_{ij} \; i \; \in \{1, \dots, n\} \text{ and } j \in \{1, \dots, m\}, \; r_{ij} \geq 0$$

and initial distribution $e_1 = (1, 0, \dots, 0)$

The whole project as a PH distribution

Markov process $J(t)$ with states $(iijj)$, (ijj), (ij), (ii), i, and generator matrix

$$Q = \begin{pmatrix} T & -Te' \\ 0 & 0 \end{pmatrix}, \text{ and } R = r_{ij} \ i \ \in \{1, \ldots, n\} \text{ and } j \in \{1, \ldots, m\}, \ r_{ij} \geq 0$$

and initial distribution $e_1 = (1, 0, \ldots, 0)$
Transition rates $\lambda_1 = 0,1613$, $\lambda_2 = 0,1266$, $\lambda_3 = 0,2857$, $\lambda_4 = 0,3125$, $\lambda_2 = 0,4762$ and $\lambda_9 = 1$

The whole project as a PH distribution

PH	1122	122	112	2233	12	11	223	233	1	22	23	3344	2	344	334	34	44	33	3	4	55	5	99	9
1122	-λ_{12}	λ_1	λ_2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
122	0	-λ_{12}	0	λ_1	λ_2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
112	0	0	-λ_{12}	0	λ_1	λ_2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2233	0	0	0	-λ_{23}	0	0	λ_3	λ_2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	0	0	0	0	-λ_{12}	0	0	λ_1	λ_2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	-λ_1	0	0	λ_1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
223	0	0	0	0	0	0	-λ_{23}	0	0	λ_3	λ_2	0	0	0	0	0	0	0	0	0	0	0	0	0
233	0	0	0	0	0	0	0	-λ_{23}	0	0	λ_3	λ_2	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	-λ_1	0	0	λ_1	0	0	0	0	0	0	0	0	0	0	0	0
22	0	0	0	0	0	0	0	0	0	-λ_2	0	0	λ_2	0	0	0	0	0	0	0	0	0	0	0
23	0	0	0	0	0	0	0	0	0	0	-λ_{23}	0	λ_3	λ_2	0	0	0	0	0	0	0	0	0	0
3344	0	0	0	0	0	0	0	0	0	0	0	-λ_{34}	0	λ_3	λ_4	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0	-λ_2	0	0	0	λ_2	0	0	0	0	0	0	0
344	0	0	0	0	0	0	0	0	0	0	0	0	0	-λ_{34}	0	λ_4	λ_3	0	0	0	0	0	0	0
334	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-λ_{34}	λ_3	0	λ_4	0	0	0	0	0	0
34	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-λ_{34}	0	0	λ_4	λ_3	0	0	0	0
44	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-λ_4	0	0	λ_4	0	0	0	0
33	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-λ_3	λ_3	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-λ_3	0	λ_3	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-λ_4	λ_4	0	0	0
55	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-λ_5	λ_5	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-λ_5	λ_5	0
99	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-λ_9	λ_9
9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-λ_9

With $\lambda_{ij} = \lambda_i + \lambda_j$

The different results

- simple approach: $E[D_s] = E[d_2] + E[d_4] + E[d_5] + E[d_9] = 28,4$
- Successive Pinciple:
  $E[D_{SP}] = E[d_2] + MEB_{12} + E[d_4] + E[d_5] + E[d_9] = 29,15$

The different results

- simple approach: $E[D_s] = E[d_2] + E[d_4] + E[d_5] + E[d_9] = 28,4$
- Successive Pinciple:
  $E[D_{SP}] = E[d_2] + MEB_{12} + E[d_4] + E[d_5] + E[d_9] = 29,15$
- PH approach: $E[D_{ph}] = -e_1 T^{-1} e = 33,47$ with e being a vector of ones and $e_1 = (1, 0, \ldots, 0)$

The different results

- simple approach: $E[D_s] = E[d_2] + E[d_4] + E[d_5] + E[d_9] = 28,4$
- Successive Pinciple:
  $E[D_{SP}] = E[d_2] + MEB_{12} + E[d_4] + E[d_5] + E[d_9] = 29,15$
- PH approach: $E[D_{ph}] = -e_1 T^{-1} e = 33,47$ with e being a vector of ones and $e_1 = (1,0,\ldots,0)$
- We get the density, cumulative distribution, moments,...

Cost and Time

Cost and Time

construction delayed by ca. 5 years

Cost and Time

- construction delayed by ca. 5 years
- original cost estimate 114 million euro, now 476 million euro due to extra cost and delays

Bivariate Extension

Cost and duration of a project are closely correlated

Bivariate Extension

Cost and duration of a project are closely correlated

Total cost of a project=general cost + duration dependent cost

$$TC = GC + TDC$$

Bivariate Extension

Cost and duration of a project are closely correlated

Total cost of a project=general cost + duration dependent cost

$$TC = GC + TDC$$

Define: r(l) = time dependent cost of state l

$$\begin{aligned} TDC &= \int_0^\tau r(J(t))dt \\ D_{ph} &= \int_0^\tau J(t)dt \end{aligned}$$

where τ is the finish time of the project. i.e. the absorption time of $J(t)$.

Bivariate Extension

Cost and duration of a project are closely correlated

Total cost of a project=general cost + duration dependent cost

$$TC = GC + TDC$$

Define: r(l) = time dependent cost of state l

$$\begin{aligned} TDC &= \int_0^\tau r(J(t))dt \\ D_{ph} &= \int_0^\tau J(t)dt \end{aligned}$$

where τ is the finish time of the project. i.e. the absorption time of $J(t)$.

$$Y = (D_{ph}, TDC) \sim \mathrm{MPH}^\star$$

where MPH* is multivariate phase type after Kulkarni

Multivariate PH

As before, let $J(t)$ be a continuous Markov chain (CTMC) with state space $\{1, 2, \cdots, m, m+1\}$, initial distribution α, generator matrix Q and rewards r_{ij}

$$Q = \begin{pmatrix} T & -Te' \\ 0 & 0 \end{pmatrix}, \text{ and } R = r_{ij} \; i \; \in \{1, \ldots, n\} \text{ and } j \in \{1, \ldots, m\}, \; r_{ij} \geq 0$$

Multivariate PH

As before, let $J(t)$ be a continuous Markov chain (CTMC) with state space $\{1, 2, \cdots, m, m+1\}$, initial distribution α, generator matrix Q and rewards r_{ij}

$$Q = \begin{pmatrix} T & -Te' \\ 0 & 0 \end{pmatrix}, \text{ and } R = r_{ij} \; i \; \in \{1, \ldots, n\} \text{ and } j \in \{1, \ldots, m\}, \; r_{ij} \geq 0$$

$R_i(j) = r_{ij}$
$Y_i = \int_0^\tau r_i\left(X(t)\right) dt$

Multivariate PH

As before, let $J(t)$ be a continuous Markov chain (CTMC) with state space $\{1, 2, \cdots, m, m+1\}$, initial distribution α, generator matrix Q and rewards r_{ij}

$$Q = \begin{pmatrix} T & -Te' \\ 0 & 0 \end{pmatrix}, \text{ and } R = r_{ij} \; i \; \in \{1, \ldots, n\} \text{ and } j \in \{1, \ldots, m\}, \; r_{ij} \geq 0$$

$R_i(j) = r_{ij}$
$Y_i = \int_0^\tau r_i \left(X(t)\right) dt$
$Y = (Y_1, \ldots, Y_n)$ is multivariate phase type distributed after Kulkarni,
$Y \sim MPH^\star(\alpha, T, R)$

Properties of MPH*

- The survival function follows a set of PDEs
- $E[Y_i] = \alpha \cdot (-T)^{-1} r_i$
- $E[Y] = (E[Y_1], \ldots, E[Y_n])$
- $E[Y_i^2] = 2 \cdot \alpha(-T)^{-1} \cdot \Delta(r_i)(-T)^{-1} \cdot r_i$,
  with $\Delta(r_i)$ being a matrix with the elements of r_i on the diagonal
- $E[Y_i Y_j] = \alpha(-T)^{-1} \cdot \Delta(r_i)(-T)^{-1} \cdot r_j + \alpha(-T)^{-1} \cdot \Delta(r_j)(-T)^{-1} \cdot r_i$

Discussion

- Successive Principle is widely use in Scandinavia
- Subtasks are PH distributed that allows us to model the entire project instead of handling only the expectation and variance of the subtask
- More natural then dealing only with the mean and the variance of each subtask
- More complex, higher modeling cost
- Cost/Benefit ratio might not be favorable
- Density, moments,... vs expectation, variance of the subtasks $\longrightarrow$ law of large numbers
- extension to the bivariate case allows to deal with correlation between cost and duration