

Schätzung der Erfolgsaussichten, der Dauer und
der Kosten von Softwareprojekten mit strengen
wahrscheinlichkeitstheoretischen Methoden

Frank Padberg

1998

Schätzung der Erfolgsaussichten, der Dauer und
der Kosten von Softwareprojekten mit strengen
wahrscheinlichkeitstheoretischen Methoden

Dissertation zur Erlangung des Grades Doktor der
Ingenieurwissenschaften (Dr.-Ing.) der Technischen
Fakultät der Universität des Saarlandes

Frank Padberg
Fachbereich Informatik
Universität Saarbrücken
1998

Diese Arbeit wurde von der Deutschen Forschungsgemeinschaft durch ein Promotionsstipendium am Graduiertenkolleg des Fachbereichs Informatik in Saarbrücken gefördert.

Kolloquium : 3. 11. 1998

Dekan : Prof. Dr. Wolfgang Paul

Gutachter : Prof. Dr. Wolfgang Paul

Prof. Dr. Reinhard Wilhelm

Inhalt

1. Einleitung	1
2. Softwareprojekte	4
3. Takte und Phasen	6
4. Phasenverläufe	7
5. Langzeitverteilungen	11
6. Verschobene Verteilungen	14
7. Zustandsvektoren und Projektverlauf	17
8. Rückschritte und Abhängigkeiten	20
9. Übergangswahrscheinlichkeiten	23
10. Beweis zu den Übergangswahrscheinlichkeiten	26
11. Beweis zu den Phasenverläufen	29
12. Schätzungen	40
13. Beispiele und Rechenaufwand	45
Tabellen	56
Literatur	61
Summary	63

1. Einleitung*

Um die Durchführbarkeit und die Wirtschaftlichkeit eines Softwareprojekts beurteilen zu können, ist es nötig, zu einem frühen Zeitpunkt die Dauer und die Kosten des Projekts möglichst gut zu schätzen. Das gilt als schwierig. So meint Fairley [3] auf Seite 64:

“Estimating the cost of a software product is one of the most difficult and error-prone tasks in software engineering.“

Die Schätzwerte allein genügen jedoch nicht. Um das Risiko beurteilen zu können, das in einem Projekt steckt, ist es wichtig, die Schwankungsbreiten für die Dauer und die Kosten abzuschätzen. Vor allem muß man die Erfolgsaussichten des Projekts abschätzen, also die Chancen dafür, das Projekt innerhalb eines vorgegebenen Zeitraumes und mit einer vorgegebenen Kostengrenze erfolgreich abzuschließen.

Es gibt bereits eine Reihe von Modellen zur Schätzung der Kosten und der Dauer von Softwareprojekten, die in den Lehrbüchern beschrieben sind [13] [16] [12] [9] [3]. Ein typisches Beispiel ist das Modell von Boehm. Er gibt eine einfache Formel an, mit der die Kosten eines Projekts aus der Größe des zu entwickelnden Softwareprodukts berechnet werden. Mit Größe ist dabei die Anzahl der Quellcode-Instruktionen gemeint. Die Größe muß vorab geschätzt werden. Die Formel enthält einige Konstanten, mit denen sie an Besonderheiten des Projekts und der Entwicklungsumgebung angepaßt wird. Die Werte der Konstanten sind teilweise voneinander abhängig und müssen für jedes Projekt neu bestimmt werden. Die Dauer des Projekts wird ähnlich behandelt. Eine Abschätzung der Schwankungsbreiten oder der Erfolgsaussichten ist nicht möglich. Verschiedene Studien zeigen, daß die Ergebnisse, die dieses und andere Modelle liefern, nicht brauchbar sind [8] [10] [6] [7]. Es gibt daher immer wieder neue Beiträge, etwa vor zwei Jahren auf der International Conference on Software Engineering [14]. Auch ungewöhnliche Methoden werden versucht [18] [17] [5]. Ich denke, daß Sallis, Tate und Mac Donell [13] mit ihrem Kommentar auf Seite 66 den Punkt treffen:

“It seems likely that such simplistic models reflect too few of the complexities of actual software development.“

* mit Zusammenfassung

Die Dauer und die Kosten eines Projekts hängen unmittelbar von dem Verlauf ab, den das Projekt nehmen wird. Der Verlauf läßt sich aber kaum vorab bestimmen, denn es ist kennzeichnend für Softwareprojekte, daß Entwicklungsabläufe unerwartet wiederholt werden müssen [2] . Ich möchte in meiner Arbeit zeigen, daß und wie man dieses Problem mit strengen wahrscheinlichkeitstheoretischen Methoden angehen kann. Ich konstruiere ein Modell, das es erlaubt, die *Wahrscheinlichkeiten von Projektverläufen* zu berechnen. Die Schätzungen für die Dauer und die Kosten eines Projekts ergeben sich dann als die Erwartungswerte von geeigneten Funktionen, die jedem Projektverlauf seine Dauer und seine Kosten zuordnen. Die Wahrscheinlichkeiten dafür, daß die Dauer und die Kosten des Projekts innerhalb bestimmter Schwankungsbreiten um die Schätzwerte bleiben, lassen sich mit meinem Modell ebenso berechnen wie die Erfolgsaussichten.

In das Modell gehen *statistische Daten* über den Verlauf bereits abgeschlossener Projekte als Parameter ein. Die Zahlen, die das Modell liefert, können also nur so zuverlässig sein wie die verfügbare Datenbasis. Die benötigten Rohdaten wären leicht bei laufenden Projekten aufzuzeichnen, aber das ist vermutlich bisher nicht gemacht worden. Ferner geht in das Modell ein, in welchem Umfang die einzelnen Teile des Softwareprodukts voneinander abhängen. Diese Parameter müssen aus dem *Entwurf* abgeleitet werden.

Das Modell ist wie folgt konstruiert. Der zeitliche Verlauf eines Projekts wird in Phasen unterteilt. Der Zustand des Projekts ändert sich von Phase zu Phase. Ein Projektverlauf wird formal als eine Folge von Zuständen beschrieben. Die Übergangswahrscheinlichkeiten für einen Zustand geben an, mit welchen Wahrscheinlichkeiten das Projekt von diesem Zustand in die anderen wechselt. Die Formeln für die Übergangswahrscheinlichkeiten sind komplex. Sie werden nach und nach aus den Eingangsdaten konstruiert. Dann wird bewiesen, daß sich die Formeln für jeden Zustand zu Eins aufsummieren, so daß man tatsächlich von Übergangswahrscheinlichkeiten sprechen kann. Der Beweis ist aufwendig, aber konstruktiv. Die Phasen und die Zustände sind so festgelegt, daß die Übergangswahrscheinlichkeiten für einen Zustand nur von diesem Zustand abhängen und nicht davon, wie das Projekt in den Zustand gekommen ist. Die Wahrscheinlichkeit für einen Projektverlauf wird dann einfach berechnet als das Produkt der Übergangswahrscheinlichkeiten für die Zustandsübergänge, aus denen sich der Projektverlauf zusammensetzt. Technisch gesehen ist das Modell eine Markovkette.

Ich benutze für das Modell eine einfache Sichtweise von Softwareprojekten. Ein Projekt besteht aus einem Projektmanager und mehreren Entwicklerteams, die an der Umsetzung des gemeinsamen Entwurfs arbeiten. Der Fortschritt jedes Teams hängt von den Fortschritten der anderen Teams ab. Um überhaupt einmal bis zu konkreten Formeln zu kommen, habe ich einige Annahmen gemacht, die nachfolgend aufgeführt sind. Es wird angenommen, daß sich die Anzahl der Teams nicht ändert. Die Teams beginnen ihre Arbeit gleichzeitig mit dem Projektbeginn. Die Auftragstellung durch den Kunden ändert sich nicht. Das Ausmaß der Abhängigkeiten zwischen den Teilaufgaben und die Komplexität der einzelnen Teilaufgaben ändern sich im Verlauf des Projekts praktisch nicht. Ich bin überzeugt, daß sich diese vorläufigen Annahmen in zukünftigen Fassungen des Modells vermeiden lassen, nachdem hier der Anfang gemacht ist.

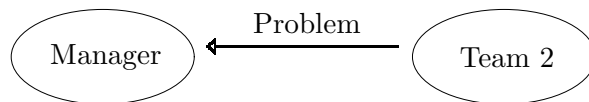
Man kann interessante Beobachtungen machen, wenn man Zahlen anstelle der Parameter in die Formeln einsetzt und dann mit dem Modell rechnet. So kann man zum Beispiel für jede Zeitspanne die Wahrscheinlichkeit dafür berechnen, daß das Projekt nach Ablauf dieser Zeitspanne erfolgreich abgeschlossen sein wird, aber nicht früher. Trägt man die Werte monatsweise in einem Balkendiagramm auf, dann erhält man eine Kurve, die bei allen Beispielen dieselbe Grundform hat. Das ist erstaunlich. Auch zeigt sich, daß sich das Modell hinsichtlich der Schätzwerte, ihrer Schwankungsbreiten und der Erfolgsaussichten wie zu erwarten verhält, wenn die Eingangsdaten verändert werden. Nebenbei bemerkt hat sich an den Beispielen bestätigt, daß sich die Übergangswahrscheinlichkeiten für einen Zustand zu Eins aufsummieren. An den Beispielen läßt sich ferner erkennen, daß im Modell viele erfolgreiche Projekte so verlaufen, daß jedes Team in jeder der Phasen arbeitet. Ich denke, diese Beobachtungen sind Belege dafür, daß es mit meinem wahrscheinlichkeitstheoretischen Ansatz möglich ist, über die Schätzwerte hinaus Einsichten in die Dynamik von Softwareprojekten zu gewinnen.

An Vorkenntnissen aus der Mathematik werden lediglich die Grundzüge der klassischen, diskreten Wahrscheinlichkeitstheorie vorausgesetzt [4] [11] [15] [1] .

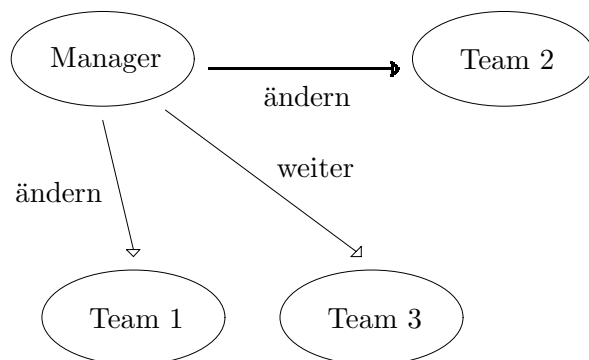
2. Softwareprojekte

An einem Softwareprojekt arbeiten mehrere Entwicklerteams und ein Projektmanager. Aus dem ersten *Systementwurf* für das Softwareprodukt ergibt sich eine Aufteilung der Arbeit in Teilaufgaben. Jede Teilaufgabe wird von einem Team bearbeitet. Die Teams beginnen ihre Arbeit gleichzeitig mit dem Projektbeginn. Jedes Team arbeitet, bis seine Teilaufgabe fertiggestellt ist.

Die Teams arbeiten gleichzeitig und die meiste Zeit auch unabhängig voneinander. Es gibt keine direkte Kommunikation zwischen den Teams. Aber es kommt zu einer gegenseitigen Beeinflussung, wenn eines der Teams dem Manager ein weitreichendes Problem meldet, das sich möglicherweise auf andere Teams auswirken wird.

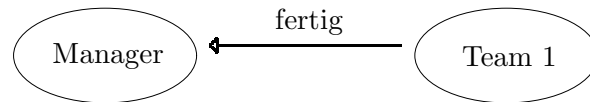


Der Systementwurf gilt dann als „offen“ und der Manager sorgt dafür, daß der Entwurf überarbeitet wird. In der Zwischenzeit wartet das Team, welches das Problem gemeldet hat. Die anderen Teams arbeiten weiter. Es ist möglich, daß von ihnen noch weitere Probleme gemeldet werden, während der Systementwurf offen ist. Diese werden dann auch bei dem neuen Entwurf berücksichtigt. Sobald der neue Systementwurf fertig ist, teilt der Manager jedem Team mit, ob es von den Änderungen betroffen ist oder nicht.



Der Systementwurf ist nun wieder gültig. Die betroffenen Teams arbeiten die Änderungen ein. Es kann auch sein, daß Teams, die vorher schon fertig geworden waren, deshalb ihre Arbeit wieder aufnehmen müssen. *Der Fortschritt eines Teams hängt also indirekt vom Fortschritt der übrigen Teams ab.*

Wenn ein Team mit seiner Teilaufgabe fertig wird, meldet es das beim Manager.

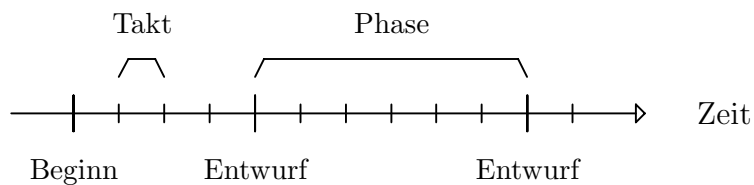


Der *Entwicklungszyklus* endet, wenn alle Teams mit ihren Teilaufgaben fertig sind. Das System ist programmiert und kann getestet werden. Enthält es noch Fehler, wird ein neuer Entwicklungszyklus begonnen.

3. Takte und Phasen

Im Modell wird die Zeitachse in Abschnitte gleicher Dauer unterteilt, die *Takte*. Die Dauer eines Taktes kann später je nach Anwendung des Modells gewählt werden. Einstweilen kann man sich vorstellen, daß ein Takt einen Monat dauert. Ob die Teams Fortschritte erzielt haben oder nicht, wird immer erst am Ende eines Taktes festgestellt. Der genaue Zeitpunkt der Fortschritte innerhalb eines Taktes wird dabei vernachlässigt. Das heißt, *die Zeit ist im Modell diskret*. Das ist eine übliche Abstraktion.

Im Verlauf eines Projekts gibt es in unregelmäßigen Abständen neue Systementwürfe. Die Zeitspanne zwischen zwei aufeinanderfolgenden Systementwürfen heißt eine *Phase*.



Da die Zeit diskret modelliert ist, besteht eine Phase aus einer gewissen Anzahl von Takten, die von Phase zu Phase unterschiedlich sein kann. Ein Projekt erscheint als eine Folge von Phasen.

Das Ziel ist nun, sowohl die möglichen Verläufe der einzelnen Phasen als auch die möglichen Folgen von Phasen wahrscheinlichkeitstheoretisch zu modellieren. Für die Phasenverläufe werden in den Kapiteln 4 bis 6 Formeln entwickelt, für die Folgen von Phasen in den Kapiteln 7 und 8. Die Ergebnisse werden dann in Kapitel 9 zusammengesetzt.

4. Phasenverläufe

In einer Phase zwischen zwei Systementwürfen wartet der Manager auf Meldungen von den Teams. Ein Team meldet sich, wenn es einen von zwei Zwischenzuständen erreicht. Entweder es ist mit seiner Teilaufgabe fertig geworden, erster Zwischenzustand. Oder es hat ein Problem entdeckt, zweiter Zwischenzustand. Die Fortschritte des i -ten Teams im Verlauf einer bestimmten Phase sind also wahrscheinlichkeitstheoretisch beschrieben, wenn für jeden Zwischenzustand $x = 1, 2$ und jede Zeitspanne k die Wahrscheinlichkeit

$$P_i(x, k)$$

dafür angegeben ist, daß das Team genau k Takte nach Beginn der Phase den Zwischenzustand x erreicht. Diese Wahrscheinlichkeiten hängen unter anderem davon ab, wie weit das Team zu Beginn der Phase war.

Die Dauer einer Phase hängt davon ab, ob und wann die Teams während der Phase Probleme melden. Umgekehrt hängt es von der Dauer der Phase ab, wie weit die Teams vor dem nächsten Systementwurf überhaupt kommen können. Hält man die in Takten ausgedrückte Dauer d fest, dann ist der genaue *Verlauf v einer Phase* gegeben, wenn neben der Anzahl N der Teams noch die beiden Szenarien

$$\mathcal{K} = \left\{ (K_\nu, k_\nu) \right\}_{\nu = 1, \dots, n} \quad \text{und} \quad \mathcal{L} = \left\{ (L_\mu, l_\mu) \right\}_{\mu = 1, \dots, m}$$

gegeben sind. Dabei sind $1 \leq k_1 < \dots < k_n \leq d - 1$ die verschiedenen Zeitpunkte in der Phase, zu denen Probleme gemeldet werden, und $1 \leq l_1 < \dots < l_m \leq d$ die verschiedenen Zeitpunkte, zu denen Teams fertig werden. Diese Angaben sind relativ zum Beginn der Phase. Falls im letzten Takt d eine Problemmeldung auftreten würde, müßte sie noch beim neuen Systementwurf berücksichtigt werden und die Phase würde dann wenigstens einen Takt länger dauern. Die Menge K_ν enthält die Nummern der Teams, die zum Zeitpunkt k_ν ein Problem melden. Entsprechend ist die Menge L_μ für den Zeitpunkt l_μ gebildet. Ferner enthalte L_0 die Nummern derjenigen Teams, die während der ganzen Phase nichts melden. Darunter sind auch die Teams, die schon in einer früheren

Phase fertig geworden sind. Die Mengen K_ν , L_μ , L_0 sind allesamt disjunkt. Ihre Vereinigung ist die Menge $\{1, \dots, N\}$.

Die Problemmeldungen einzelner Teams wirken sich erst am Ende der Phase auf die übrigen Teams aus. Erst dann teilt der Manager jedem Team mit, ob sich aus dem neuen Systementwurf Änderungen für das Team ergeben. Auch gibt es keine direkte Kommunikation zwischen den Teams. Das bedeutet, *bei gegebener Dauer der Phase können die Fortschritte der Teams innerhalb der Phase als voneinander unabhängig betrachtet werden*. Generell ist die Wahrscheinlichkeit für das gleichzeitige Eintreten unabhängiger Ereignisse gleich dem Produkt der Wahrscheinlichkeiten der einzelnen Ereignisse. Daher enthält die Wahrscheinlichkeit

$$P(v)$$

dafür, daß die Phase bei gegebener Dauer d den beschriebenen Verlauf v nimmt, die folgenden Faktoren.

- Für ein Team, das zum Zeitpunkt k_ν ein Problem meldet: $P_i(2, k_\nu)$.
- Für ein Team, das zum Zeitpunkt l_μ fertig wird: $P_i(1, l_\mu)$.
- Für ein Team, das die Phase über nichts meldet: $1 - P_i(1) - P_i(2)$.

Dabei ist vorübergehend

$$P_i(x) = \sum_{k=1}^d P_i(x, k).$$

Die Überlegung führt also auf den Ausdruck

$$\begin{aligned} P(\mathcal{K}, \mathcal{L}; d) &= \prod_{\nu=1}^n \prod_{i \in K_\nu} P_i(2, k_\nu) \\ &\times \prod_{\mu=1}^m \prod_{i \in L_\mu} P_i(1, l_\mu) \\ &\times \prod_{i \in L_0} (1 - P_i(1) - P_i(2)) \end{aligned} \tag{4.1}$$

als Faktor in der Wahrscheinlichkeit dafür, daß die Phase den beschriebenen Verlauf nimmt. Für ein Team, das vor Beginn der Phase schon mit seiner Arbeit fertig war, gilt $P_i(1) = P_i(2) = 0$.

Sobald im Takt k_1 die ersten Probleme gemeldet werden, beginnt die Überarbeitung des Systementwurfs. Zu weiteren Problemmeldungen im Takt k_2 der Phase kann es nur kommen, wenn es in den $k_2 - k_1 - 1$ dazwischen liegenden Takten nicht gelingt, den neuen Entwurf fertigzustellen. Andernfalls wäre die Phase ja schon in weniger als k_2 Takten beendet. Auch die nachfolgenden Problemmeldungen konkurrieren zeitlich mit der Überarbeitung des Entwurfs. Zur Vereinfachung wird nun angenommen, daß die Entwurfsüberarbeitung mit jedem weiteren Problem von vorne beginnt. Die Wahrscheinlichkeit dafür, daß die Phase wie beschrieben verläuft, enthält dann die Faktoren

$$1 - \Gamma(k_{\nu+1} - k_{\nu} - 1).$$

Dabei ist

$$\Gamma(k) = \sum_{j=0}^k \gamma(j)$$

die kumulierte Verteilung zur Verteilung $\gamma(j)$. Letztere beschreibt, mit welcher Wahrscheinlichkeit es genau j Takte dauert, bis nach einer Problemmeldung ein neuer Systementwurf erstellt ist, sofern kein weiteres Problem dazwischen kommt. *Diese Verteilung muß aus den Daten abgeschlossener Projekte statistisch ermittelt werden.* Der Faktor $1 - \Gamma(k_{\nu+1} - k_{\nu} - 1)$ ist also die Wahrscheinlichkeit dafür, daß in den Takten zwischen den Zeitpunkten k_{ν} und $k_{\nu+1}$ kein neuer Entwurf fertig wird. Es gilt

$$\sum_j \gamma(j) = 1,$$

da γ eine Wahrscheinlichkeitsverteilung ist.

Der neue Entwurf ist schließlich genau $d - k_n$ Takte nach der letzten Problemmeldung fertig. Die Wahrscheinlichkeit dafür gibt den weiteren Faktor

$$\gamma(d - k_n).$$

Die Überlegungen zur Verteilung γ führen also auf den Ausdruck

$$P(\gamma, \mathcal{K}; d) = \gamma(d - k_n) \cdot \prod_{\nu=1}^{n-1} (1 - \Gamma(k_{\nu+1} - k_{\nu} - 1)) \quad (4.2)$$

als Korrekturfaktor zum Ausdruck $P(\mathcal{K}, \mathcal{L}; d)$. Insgesamt ist die Wahrscheinlichkeit dafür, daß die Phase bei gegebener Dauer d den beschriebenen Verlauf v nimmt, gleich

$$P(v) = P(\mathcal{K}, \mathcal{L}; d) \cdot P(\gamma, \mathcal{K}; d), \quad (4.3)$$

sofern die Phase Problemmeldungen enthält.

Es bleibt noch der besondere Fall, daß die Phase beendet wird, ohne daß ein Team ein Problem gemeldet hat. Das geht nur, wenn alle Teams im Lauf der Zeit fertig geworden sind, die letzten in dieser Phase. Das System ist dann programmiert und kann getestet werden. Die Phase endet im Takt $d = l_m$. Die Wahrscheinlichkeit für einen solchen Verlauf enthält nur die Faktoren $P_i(1, l_{\mu})$, ist also gleich

$$P(v) = P(\mathcal{L}; d) = \prod_{\mu=1}^m \prod_{i \in L_{\mu}} P_i(1, l_{\mu}). \quad (4.4)$$

5. Langzeitverteilungen

Wenn ein Team lange genug arbeiten kann, ohne vom Manager wegen einer das Team betreffenden Änderung am Systementwurf unterbrochen zu werden, wird es letztlich ein Problem melden oder fertig werden. Zur Modellierung ist es nötig zu wissen, mit welcher Wahrscheinlichkeit der eine oder der andere Fall zu diesem oder zu jenem Zeitpunkt auftreten wird, vorausgesetzt, das Team bleibt ungestört. *Diese Wahrscheinlichkeiten müssen aus den Daten abgeschlossener Projekte statistisch ermittelt werden.* Die Wahrscheinlichkeiten für ein Team hängen von verschiedenen Faktoren ab, etwa von der Erfahrung der Teammitglieder und dem Schwierigkeitsgrad der zu bearbeitenden Teilaufgabe. Wenn die statistische Datenbasis groß genug und detailliert genug ist, wird man versuchen, das bei der Ermittlung der Wahrscheinlichkeiten zu berücksichtigen. Man wird die Zahlen für jedes Team gesondert und dann nur für Teilaufgaben vergleichbarer Komplexität ermitteln. Dabei ist man frei in der Wahl des verwendeten Komplexitätsmaßes. Man ermittelt also die Wahrscheinlichkeiten der Ereignisse

$$E_{x,k}^i$$

dafür, daß das i -te Team von sich aus genau k Takte nach Beginn seiner Arbeit den Zwischenzustand x erreicht, wenn es ungestört bleibt.

Die Wahrscheinlichkeit $P(E_{x,k}^i)$ läßt sich als ein Produkt zweier anderer Wahrscheinlichkeiten schreiben. Nach der Definition bedingter Wahrscheinlichkeiten gilt

$$P(E_{x,k}^i) = P(E_x^i) \cdot P(k | E_x^i).$$

Dabei bezeichnet E_x^i das Ereignis, daß das ungestörte Team den Zwischenzustand x erreicht, und $P(k | E_x^i)$ ist die bedingte Wahrscheinlichkeit dafür, daß das genau k Takte dauert. Anstelle der Wahrscheinlichkeiten $P(E_{x,k}^i)$ könnte man also auch die beiden *Teamergebniswahrscheinlichkeiten*

$$q_i \quad \text{und} \quad p_i$$

und die beiden zugehörigen *Langzeitverteilungen*

$$\left\{ h_i(k) \right\}_{k=1, 2, 3, \dots} \quad \text{und} \quad \left\{ g_i(k) \right\}_{k=1, 2, 3, \dots}$$

angeben. Dabei entsprechen $q_i = P(E_1^i)$ und die $h_i(k) = P(k | E_1^i)$ dem Fall, daß das Team fertig werden wird, sowie $p_i = P(E_2^i)$ und die $g_i(k) = P(k | E_2^i)$ dem Fall, daß es ein Problem melden wird. Insbesondere gilt

$$q_i + p_i = 1.$$

Da die Langzeitverteilungen Wahrscheinlichkeitsverteilungen sind, gilt

$$\sum_k g_i(k) = 1 \quad \text{und} \quad \sum_k h_i(k) = 1.$$

Die Darstellung mit den Langzeitverteilungen läßt sich wegen $E_x^i = \bigcup_k E_{x,k}^i$ aus den $P(E_{x,k}^i)$ berechnen.

Die Wahrscheinlichkeiten für die Fortschritte der Teams in der ersten Phase eines Projekts lassen sich unmittelbar an den statistischen Daten der Teams ablesen. Für den Rest des Kapitels wird der Index i weggelassen. Dann ist für die erste Phase

$$P(1, k) = q \cdot h(k) \quad \text{und} \quad P(2, k) = p \cdot g(k). \quad (5.1)$$

Wenn nun ein Team am Ende der ersten Phase nicht fertig war und von den Änderungen des Systementwurfs nicht betroffen ist, dann sind seine Fortschritte in der zweiten Phase immer noch von seinen Langzeitverteilungen bestimmt. Bei der Berechnung der genauen Wahrscheinlichkeiten muß berücksichtigt werden, daß zu Beginn der zweiten Phase bereits d Takte verstrichen sind, wobei d die Dauer der ersten Phase ist. In dieser Zeit hat das Team ungestört gearbeitet. Die gesuchten Wahrscheinlichkeiten $P(x, k)$ für die zweite Phase sind dann genau die bedingten Wahrscheinlichkeiten

$$P(E_{x,d+k} | B).$$

Dabei wird auf das vorübergehend mit B bezeichnete Ereignis bedingt, daß das ungestörte Team nach frühestens $d + 1$ Takten einen der beiden Zwischenzustände erreicht. Die Wahrscheinlichkeit dafür ist

$$1 - W(d).$$

Dabei ist

$$W(k) = \sum_{j=1}^k w(j)$$

die kumulierte Wartezeitverteilung des Teams, und

$$w(j) = q \cdot h(j) + p \cdot g(j)$$

seine *Wartezeitverteilung*. Letztere beschreibt die Wahrscheinlichkeit dafür, daß sich das ungestörte Team genau j Takte nach Beginn seiner Arbeit beim Manager meldet. Die Wartezeitverteilung ist ebenfalls eine Wahrscheinlichkeitsverteilung. Für die zweite Phase erhält man dann

$$P(1, k) = \frac{q \cdot h(d+k)}{1 - W(d)} \quad \text{und} \quad P(2, k) = \frac{p \cdot g(d+k)}{1 - W(d)}. \quad (5.2)$$

Wenn das Team am Ende der zweiten Phase immer noch nicht fertig ist und von den neuerlichen Änderungen des Systementwurfs wieder nicht betroffen ist, dann lassen sich auf die gleiche Weise die Wahrscheinlichkeiten für die dritte Phase berechnen. Bedingt wird dann auf das Ereignis B , daß das ungestörte Team nach frühestens $d + e + 1$ Takten einen der Zwischenzustände erreicht, wobei e die Dauer der zweiten Phase ist.

Formal ist noch der Fall $W(d) = 1$ zu beachten. Das bedeutet, daß das Team schon innerhalb der bereits verstrichenen d Takte fertig geworden sein oder sich mit einem Problem beim Manager gemeldet haben muß. Es folgt $h(d+k) = g(d+k) = 0$. Da Zähler und Nenner des Bruches dann gleich Null sind, wird die Vereinbarung getroffen:

$$\frac{0}{0} = 0.$$

6. Verschobene Verteilungen

Die Wahrscheinlichkeiten für die Fortschritte eines ungestörten Teams in der zweiten Phase lassen sich auch anders berechnen. Zunächst bildet man

$$\Theta_d(q) = q \cdot \frac{1 - H(d)}{1 - W(d)} \quad \text{und} \quad \Theta_d(p) = p \cdot \frac{1 - G(d)}{1 - W(d)}$$

sowie

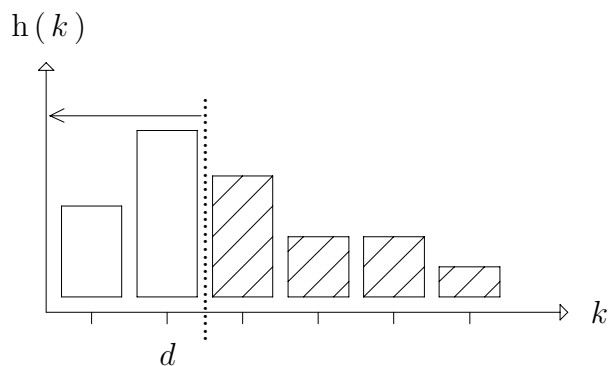
$$(\Theta_d h)(k) = \frac{h(d+k)}{1 - H(d)} \quad \text{und} \quad (\Theta_d g)(k) = \frac{g(d+k)}{1 - G(d)}.$$

Das sind die um die Dauer d der ersten Phase „verschobenen“ Langzeitverteilungen des Teams und die zugehörigen Ergebniswahrscheinlichkeiten. Der Index i wurde wieder weggelassen. Der Operator Θ_d bezeichnet eine Verschiebung um d Takte. Ferner sind

$$H(k) = \sum_{j=1}^k h(j) \quad \text{und} \quad G(k) = \sum_{j=1}^k g(j)$$

die kumulierten Langzeitverteilungen des Teams.

Die verschobenen Verteilungen ergeben sich anschaulich dadurch, daß die ursprünglichen Langzeitverteilungen vor $k = d + 1$ abgeschnitten werden und jeweils der gesamte rechte, im Bild schraffierte Teil normiert wird (Teilen durch seine „Fläche“) und dann an den Ursprung verschoben wird.



Durch das Normieren sind die verschobenen Verteilungen tatsächlich Wahrscheinlichkeitsverteilungen, sofern man nicht „zu weit rechts“ abgeschnitten hat, also im Nenner die kumulierte Verteilung nicht schon gleich Eins ist. Sonst gilt wieder die Vereinbarung über Brüche der Form $\frac{0}{0}$. Die verschobene Wartezeitverteilung ist gleich

$$(\Theta_d w)(k) = \frac{w(d+k)}{1 - W(d)}.$$

Für die verschobenen Verteilungen gilt ähnlich wie für die ursprünglichen Verteilungen

$$(\Theta_d w)(j) = \Theta_d(q) \cdot (\Theta_d h)(j) + \Theta_d(p) \cdot (\Theta_d g)(j). \quad (6.1)$$

Sofern d nicht zu groß ist, so daß die verschobenen Verteilungen noch Wahrscheinlichkeitsverteilungen sind und somit Summe Eins haben, folgt daraus nebenbei bemerkt

$$\begin{aligned} \Theta_d(q) + \Theta_d(p) &= \Theta_d(q) \cdot \sum_j (\Theta_d h)(j) + \Theta_d(p) \cdot \sum_j (\Theta_d g)(j) \\ &= \sum_j (\Theta_d w)(j) = 1, \end{aligned}$$

ähnlich wie für die ursprünglichen Ergebniswahrscheinlichkeiten.

Nun werden an den verschobenen Langzeitverteilungen die Wahrscheinlichkeiten für die zweite Phase genau so abgelesen, wie die Wahrscheinlichkeiten für die erste Phase an den ursprünglichen Langzeitverteilungen, siehe Gleichung (5.1). Für die zweite Phase gilt in Übereinstimmung mit Gleichung (5.2)

$$P(1, k) = \Theta_d(q) \cdot (\Theta_d h)(k) \quad (6.2)$$

und

$$P(2, k) = \Theta_d(p) \cdot (\Theta_d g)(k).$$

Wenn das Team in die dritte Phase hinein ungestört bleibt, dann müßten zur Berechnung der Wahrscheinlichkeiten für die dritte Phase die um d Takte verschobenen Langzeitverteilungen nochmals um e Takte verschoben werden. Das geht auch einfacher wegen

$$\Theta_e \circ \Theta_d = \Theta_{d+e}. \quad (6.3)$$

Man kann also die ursprüngliche Langzeitverteilung gleich um $d + e$ Takte verschieben. Zum Beweis sei φ eine beliebige Wahrscheinlichkeitsverteilung und Φ die kumulierte Verteilung zu φ . Ferner sei $\psi(j) = (\Theta_d \varphi)(j)$ gesetzt und Ψ die kumulierte Verteilung zu ψ . Dann folgt

$$\begin{aligned}
 (\Theta_e \psi)(k) &= \frac{\psi(e+k)}{1 - \Psi(e)} = \frac{\varphi(d+e+k)}{1 - \Phi(d)} \cdot \frac{1}{1 - \sum_{j=1}^e \frac{\varphi(d+j)}{1 - \Phi(d)}} \\
 &= \frac{\varphi(d+e+k)}{1 - \Phi(d) - \sum_{j=1}^e \varphi(d+j)} = (\Theta_{d+e} \varphi)(k).
 \end{aligned}$$

7. Zustandsvektoren und Projektverlauf

Ein Projekt verläuft als eine Folge von Phasen. Sofern nicht die letzten Teams gerade fertig geworden sind, haben am Ende einer Phase einige der Teams ein Problem gemeldet und es kommt zu einem neuen Systementwurf. Betrachten wir für einen Moment ein Team, etwa das i -te, das zu Beginn einer Phase bereits ζ_i Takte lang ungestört gearbeitet hatte. Die Wahrscheinlichkeiten für seine Fortschritte in dieser Phase können dann an seinen Langzeitverteilungen durch Verschieben um ζ_i Takte abgelesen werden. Wenn das Team nicht von den Änderungen des Entwurfs betroffen ist, dann wird es zu Beginn der nächsten Phase entweder fertig sein oder $\zeta_i + d$ Takte weit fortgeschritten sein, wobei d die Dauer der Phase ist. Wenn es betroffen ist, dann wird es in seiner Arbeit zurückgeworfen. Nun liegt die Idee nahe, diesen Rückschritt dadurch zu modellieren, daß für den Beginn der nächsten Phase das Team auf einen noch näher festzulegenden, kleineren Wert zurückgesetzt wird. Die Wahrscheinlichkeiten für die Fortschritte des Teams in der nächsten Phase ließen sich dann wiederum an seinen Langzeitverteilungen ablesen, diesmal durch Verschieben um den neuen, kleineren Wert.

Der Zustand des i -ten Teams zu Beginn einer bestimmten Phase wird also im Modell ausgedrückt als die Zahl ζ_i der Takte, um welche die ursprünglichen Langzeitverteilungen des Teams verschoben werden müssen, wenn die Wahrscheinlichkeiten für die weiteren Fortschritte des Teams berechnet werden sollen. Die Zahl ζ_i summiert alle bisherigen Fortschritte und Rückschritte des Teams. Zu Beginn des Projekts ist $\zeta_i = 0$. Der Wert $\zeta_i = \infty$ soll bedeuten, daß das Team bereits fertig ist. Der Vektor

$$\zeta = (\zeta_1, \zeta_2, \dots, \zeta_N)$$

beschreibt im Modell den *Zustand des Projekts* zu Beginn einer Phase, wobei N die Anzahl der Teams ist. Der Vektor $\underline{\infty} = (\infty, \infty, \dots, \infty)$ bedeutet, daß alle Teams mit ihren Teilaufgaben fertig sind.

Nach Abschluß einer Phase liegt im Modell ein neuer Vektor vor, der beschreibt, mit welchem neuen Zustand die Teams in die nächste Phase gehen. Für die folgenden Kapitel

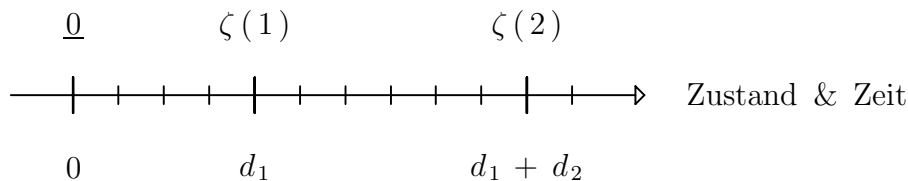
wird vereinfachend angenommen, daß die Anzahl N der Teams fest ist. Der gesamte *Projektverlauf* kann dann modelliert werden als eine Folge

$$\{ \zeta(t) \}_{t=0, 1, 2, \dots}$$

von Zuständen des Projekts. Der Index t ist die Nummer der Phase, und $\zeta(t)$ ist der Zustandsvektor am Ende der Phase. Die Zahlen $\zeta_i(t)$ ($t = 1, 2, \dots$) geben wieder, wie sich das i -te Team im Verlauf des Projekts entwickelt. Zum Projektverlauf gehört ferner die Zahlenfolge

$$d_1, d_2, \dots$$

welche die Dauer der einzelnen Phasen in Takten angibt.



Für den Projektbeginn ist $\zeta(0) = \underline{0} = (0, 0, \dots, 0)$ gesetzt.

Ist das Projekt zu Beginn einer Phase im Zustand ζ , so läßt sich für jeden Verlauf v berechnen, mit welcher Wahrscheinlichkeit

$$P_{\zeta}(v)$$

die begonnene Phase diesen Verlauf nimmt. Benötigt werden dazu lediglich die gemäß dem Vektor ζ verschobenen Langzeitverteilungen der Teams sowie die Verteilung γ , die das Erstellen neuer Systementwürfe beschreibt. Der Verlauf v führt mit einer gewissen bedingten Wahrscheinlichkeit

$$P_{\zeta}(\eta | v)$$

auf den neuen Zustand η für die nächste Phase. Die Wahrscheinlichkeit dafür, daß der Zustand ζ nach einer Phase der Dauer d auf den Zustand η führt, ist dann gleich

$$P_{\zeta}(d, \eta) = \sum_{v = (d, \mathcal{K}, \mathcal{L})} P_{\zeta}(v) \cdot P_{\zeta}(\eta | v). \quad (7.1)$$

Summiert wird über alle Phasenverläufe v der Dauer d , letztlich also über die möglichen Szenarien \mathcal{K} und \mathcal{L} mit dieser Dauer. Die Wahrscheinlichkeiten $P_{\zeta}(\eta | v)$ müssen noch bestimmt werden.

Für die Berechnung der Wahrscheinlichkeit $P_{\zeta}(d, \eta)$ spielt es keine Rolle, über welche Zustände sich das Projekt zuvor bis zum Zustand ζ entwickelt hat. Sind die Langzeitverteilungen der Teams und die Verteilung γ fest vorgegeben, so hängt die Wahrscheinlichkeit nur von den beiden Zuständen ζ und η sowie der Dauer d ab. Das bedeutet, *das Modell ist eine Markovkette*. Mathematisch streng genommen müßte ein Zustand der Kette dabei als Paar (e, ζ) definiert werden, wobei e die Dauer der zurückliegenden Phase ist. Die $P_{\zeta}(d, \eta)$ sind die *Übergangswahrscheinlichkeiten* der Kette. Das Produkt

$$P(\omega) = \prod_{t=1}^n P_{\zeta(t-1)}(d_t, \zeta(t)) \quad (7.2)$$

ist die Wahrscheinlichkeit dafür, daß das Projekt nach seinem Beginn mit den zeitlichen Abständen d_1, \dots, d_n die Zustände $\zeta(1), \dots, \zeta(n)$ durchläuft, das heißt, den Verlauf ω nimmt.

Die Anzahl der Zustandsvektoren, die in einem Projekt mit positiver Wahrscheinlichkeit auftreten können, ist endlich, wenn es eine Zahl k_0 gibt, so daß $g_i(k) = h_i(k) = 0$ für alle $k \geq k_0$ und für alle i gilt. Das bedeutet, daß ungestörte Teams spätestens nach k_0 Takten fertig sind oder sich bis dahin mit einem Problem beim Manager melden. Gilt das entsprechend auch für die Verteilung γ , dann existiert eine Obergrenze für die Dauer von Phasen. Es genügt dann ein endlicher Zustandsraum für die Markovkette.

8. Rückschritte und Abhängigkeiten

Am Ende einer Phase werden sich die Änderungen des Systementwurfs auf einige der Teams als Rückschritte auswirken. Die Menge der Nummern derjenigen Teams, die von Änderungen betroffen sind, sei mit X bezeichnet. Da ein Team, das selbst ein Problem gemeldet hat, sicher betroffen ist, gilt

$$K = \bigcup_{\nu=1}^n K_{\nu} \subset X.$$

Für den Rest des Kapitels sind der Zustand $\zeta \neq \infty$ zu Beginn der betrachteten Phase und der Verlauf v der Phase fest vorgegeben. Da ein Team, das schon seit einer früheren Phase fertig ist, sich nicht in dieser Phase melden wird, soll neben $K \neq \emptyset$ auch noch

$$L_{\infty} = \{ i \mid \zeta_i = \infty \} \subset L_0$$

für den Phasenverlauf gelten.

Nehmen wir für einen Moment an, daß die Menge X der betroffenen Teams bekannt ist. Ist das i -te Team nicht betroffen, gilt also $i \notin X$, dann läßt sich am Zustand ζ und am Verlauf v ablesen, was der mit χ_i bezeichnete Zustand des Teams zu Beginn der nächsten Phase sein muß.

- Das Team ist in dieser Phase fertig geworden: $i \in L_{\mu} \setminus X$, dann $\chi_i = \infty$.
- Das Team hat die ganze Phase hindurch gearbeitet: $i \in L_0 \setminus X$ und $\zeta_i < \infty$, dann $\chi_i = \zeta_i + d$.
- Das Team ist seit einer früheren Phase fertig: $i \in L_{\infty} \setminus X$, dann $\chi_i = \infty$.

In dieser Fassung des Modells wird nun angenommen, daß der Zustand eines Teams, das von Änderungen betroffen ist, auf Null zurückgesetzt wird. Formal ausgedrückt,

$$i \in X \implies \chi_i = 0.$$

Der *Zustand eines Teams* ist dann nichts anderes als die Anzahl der Takte, die das Team seit der letzten Unterbrechung durch den Manager ungestört gearbeitet hat. Das muß beim

Ermitteln der Langzeitverteilungen aus den Daten abgeschlossener Projekte berücksichtigt werden. Unter dieser Annahme ist durch die Angabe der Menge X der Folgezustand

$$\chi = \chi(X) = \chi(\zeta, v, X)$$

für die nächste Phase bereits vollständig festgelegt.

Es sei nun $\eta \neq \underline{\infty}$ ein beliebiger Zustandsvektor. Wenn η nach dem Verlauf v der Phase der Folgezustand von ζ wäre, dann ließe sich im nachhinein anhand von η feststellen, auf welche Teams sich die Änderungen des Systementwurfs ausgewirkt haben. Dann wäre

$$X = R(\eta)$$

mit

$$R(\eta) = \{ i \mid \eta_i = 0 \}.$$

Die Menge X legt aber den Zustand $\chi(X)$ als den einzig möglichen Folgezustand fest. Also kann η nur dann ein Folgezustand von ζ sein, wenn gilt :

$$(*) \quad K \subset R(\eta) \quad \text{und} \quad \eta = \chi(R(\eta)).$$

Erfüllt η diese Bedingungen nicht, wird die bedingte Wahrscheinlichkeit dafür, daß der Verlauf v auf den Zustand η führt, auf

$$P_{\zeta}(\eta \mid v) = 0 \tag{8.1}$$

gesetzt. Erfüllt η diese Bedingungen, dann gilt

$$P_{\zeta}(\eta \mid v) = P_{\zeta}(R(\eta); v). \tag{8.2}$$

Dabei bezeichnet allgemein

$$P_{\zeta}(X; v)$$

die bedingte Wahrscheinlichkeit dafür, daß bei gegebenem Verlauf v der Phase genau die Teams aus X von den Änderungen betroffen sind.

Die Wahrscheinlichkeit $P_{\zeta}(X; v)$ ist letztlich dadurch bestimmt, in welchem Umfang die Teilaufgaben, die von den Teams aus X bearbeitet werden, von den Teilaufgaben derjenigen Teams abhängen, die in der Phase ein Problem gemeldet haben. Die Wahrscheinlichkeit ist daher gleich dem *Abhängigkeitsgrad*

$$\alpha(K, X)$$

der Teilaufgaben mit den Nummern X von den Teilaufgaben mit den Nummern K . *Diese Abhängigkeitsgrade müssen aus dem Systementwurf abgeleitet werden.* Zählt man bei festem K die Abhängigkeitsgrade für alle möglichen Teilmengen $X \subset \{1, \dots, N\}$ zusammen, dann ist

$$\sum_X \alpha(K, X) = 1.$$

Eine der Mengen X muß ja die Nummern der betroffenen Teams enthalten. Ferner gilt

$$K \not\subset X \implies \alpha(K, X) = 0.$$

Die indirekten Abhängigkeiten zwischen den Teilaufgaben sollen schon in den Abhängigkeitsgraden berücksichtigt sein.

Für einen beliebig gewählten Zustandsvektor $\eta \neq \infty$ ergibt sich insgesamt

$$P_{\zeta}(\eta | v) = \alpha(K, R(\eta)) \cdot I_{\zeta, v}^{(1)}(\eta). \quad (8.3)$$

Dabei ist der Faktor $I_{\zeta, v}^{(1)}(\eta)$ gleich 1 oder gleich 0, je nachdem, ob η die obigen Bedingungen (*) erfüllt oder nicht.

Das Modell bleibt nur dann eine Markovkette, wenn die Abhängigkeitsgrade $\alpha(K, X)$ nicht vom Projektverlauf vor der betrachteten Phase abhängen. Dazu muß vorausgesetzt werden, daß bei allen Änderungen des Systementwurfs die Abhängigkeitsgrade praktisch gleich bleiben.

9. Übergangswahrscheinlichkeiten

Die Ergebnisse der bisherigen Kapitel werden nun zu Formeln für die Übergangswahrscheinlichkeiten $P_\zeta(d, \eta)$ zusammengesetzt. Ausgangspunkt ist die Gleichung (7.1).

Für $\eta \neq \underline{\infty}$ erhält man mit den Gleichungen (4.3) und (8.3) die Formel

$$P_\zeta(d, \eta) = \sum_{\substack{v = (d, \mathcal{K}, \mathcal{L}) \\ K \neq \emptyset}} I_\zeta^{(0)}(v) \cdot P_\zeta(\mathcal{K}, \mathcal{L}; d) \cdot P_\zeta(\gamma, \mathcal{K}; d) \cdot \alpha(K, R(\eta)) \cdot I_{\zeta, v}^{(1)}(\eta). \quad (9.1)$$

Summiert wird über alle Phasenverläufe der Dauer d mit Problemmeldungen. Der Faktor $I_\zeta^{(0)}(v)$ ist gleich 1 oder gleich 0, je nachdem, ob der Verlauf v die Bedingung

$$L_\infty \subset L_0$$

erfüllt oder nicht. Aus den Gleichungen (4.1) (6.1) und (6.2) folgt

$$\begin{aligned} P_\zeta(\mathcal{K}, \mathcal{L}; d) &= \prod_{\nu=1}^n \prod_{i \in K_\nu} \Theta_{\zeta_i}(p_i) \cdot (\Theta_{\zeta_i} g_i)(k_\nu) \\ &\times \prod_{\mu=1}^m \prod_{i \in L_\mu} \Theta_{\zeta_i}(q_i) \cdot (\Theta_{\zeta_i} h_i)(l_\mu) \\ &\times \prod_{\substack{i \in L_0 \\ \setminus L_\infty}} \left(1 - \sum_{j=1}^d (\Theta_{\zeta_i} w_i)(j) \right). \end{aligned} \quad (9.2)$$

Der Faktor $P_\zeta(\gamma, \mathcal{K}; d)$ ist in Gleichung (4.2) angegeben. Der Faktor $I_{\zeta, v}^{(1)}(\eta)$ ist gleich 1 oder gleich 0, je nachdem, ob der Zustand η die Bedingungen

$$K \subset R(\eta) \quad \text{und} \quad \eta = \chi(R(\eta))$$

erfüllt oder nicht.

Der Zustand $\underline{\infty}$ wird nur erreicht, wenn in der Phase die letzten Teams fertig werden. In diesem Fall erhält man mit Gleichung (4.4) die Formel

$$\boxed{P_{\zeta}(d, \underline{\infty}) = \sum_{v = (d, \emptyset, \mathcal{L})} I_{\zeta}^{(2)}(v) \cdot P_{\zeta}(\mathcal{L}; d).} \quad (9.3)$$

Summiert wird über alle Phasenverläufe der Dauer d , in denen keine Probleme gemeldet werden. Aus den Gleichungen (4.4) und (6.2) folgt

$$P_{\zeta}(\mathcal{L}; d) = \prod_{\mu=1}^m \prod_{i \in L_{\mu}} \Theta_{\zeta_i}(q_i) \cdot (\Theta_{\zeta_i} h_i)(l_{\mu}). \quad (9.4)$$

Da alle Teams mit dem letzten Takt der Phase fertig sein müssen, ist $I_{\zeta}^{(2)}(v)$ gleich 1 oder gleich 0, je nachdem, ob der Verlauf v die Bedingungen

$$l_m = d \quad \text{und} \quad \bigcup_{\mu=1}^m L_{\mu} = \{i \mid \zeta_i \neq \infty\}$$

erfüllt oder nicht.

Die Übergangswahrscheinlichkeiten sind bisher nur für den Fall $\zeta \neq \underline{\infty}$ angegeben, das heißt, nur für Zustandsübergänge innerhalb eines Entwicklungszyklus des Projekts. Ein Entwicklungszyklus endet mit einem Systemtest. Die Übergangswahrscheinlichkeit

$$P_{\underline{\infty}}(d, \eta)$$

ist die Wahrscheinlichkeit dafür, daß nach einem d Takte langen Systemtest ein neuer Entwicklungszyklus mit $\eta \neq \underline{\infty}$ als Anfangszustand begonnen wird. Der Verlauf eines Sytemtests ist durch die Wahrscheinlichkeiten

$$\beta(d, K)$$

dafür beschrieben, daß der Test d Takte dauert und dabei in den Teilaufgaben mit den Nummern K Fehler auftreten. Da β eine Wahrscheinlichkeitsverteilung ist, gilt

$$\sum_{d, K} \beta(d, K) = 1.$$

Die Verteilung β muß statistisch ermittelt werden. Die Fehler führen mit der Wahrscheinlichkeit $\alpha(K, X)$ dazu, daß die Teams mit den Nummern X ihre Teilaufgaben überarbeiten müssen. Wieder ist der Anfangszustand $\chi = \chi(X) = \chi(\underline{\infty}, X)$ für den nächsten Zyklus bereits durch die Menge X festgelegt. Für ein betroffenes Team ist $\chi_i = 0$, für die anderen ist $\chi_i = \infty$. Ein Zustandsvektor $\eta \neq \underline{\infty}$ kann daher nur dann ein Anfangszustand für den nächsten Zyklus sein, wenn gilt

$$\eta = \chi(R(\eta)).$$

Die gesuchte Übergangswahrscheinlichkeit ist für $\eta \neq \underline{\infty}$ somit gleich

$$\boxed{P_{\underline{\infty}}(d, \eta) = \sum_{K \neq \emptyset} \beta(d, K) \cdot \alpha(K, R(\eta)) \cdot I_{\underline{\infty}}^{(3)}(\eta).} \quad (9.5)$$

Summiert wird über alle nichtleeren Teilmengen $K \subset \{1, \dots, N\}$. Der Faktor $I_{\underline{\infty}}^{(3)}(\eta)$ ist gleich 1 oder gleich 0, je nachdem, ob η die angegebene Bedingung erfüllt oder nicht. Wenn der Zustandsraum der Markovkette endlich bleiben soll, muß es auch eine Obergrenze für die Dauer des Systemtests geben.

Es bleibt noch der Fall, daß der Systemtest nach d Takten erfolgreich bestanden wird. Das Projekt ist nun abgeschlossen, wofür formal ein ganz neuer Zustand ε eingeführt wird, der sich insbesondere vom Zustand $\underline{\infty}$ unterscheidet. Die zugehörigen Übergangswahrscheinlichkeiten ergeben sich aus der Verteilung β und sind gleich

$$\boxed{P_{\underline{\infty}}(d, \varepsilon) = \beta(d, \emptyset).} \quad (9.6)$$

Da der Zustand ε nicht mehr verlassen wird, setzt man zuletzt noch formal

$$\boxed{P_{\varepsilon}(0, \varepsilon) = 1.} \quad (9.7)$$

Alle nicht explizit angegebenen Übergangswahrscheinlichkeiten, wie etwa $P_{\underline{\infty}}(d, \underline{\infty})$, sind auf Null gesetzt.

10. Beweis zu den Übergangswahrscheinlichkeiten

Die Formeln $P_\zeta(d, \eta)$ bilden die Übergangswahrscheinlichkeiten einer Markovkette, denn für jeden Zustandsvektor ζ gilt der

SATZ:
$$\sum_{d, \eta} P_\zeta(d, \eta) = 1.$$

Das muß noch formal bewiesen werden. Die Fälle $\zeta = \varepsilon$ und $\zeta = \underline{\infty}$ werden vorneweg behandelt. Der Fall $\zeta \neq \varepsilon, \underline{\infty}$ wird auf zwei andere Aussagen zurückgeführt. Die erste Aussage wird in diesem, die zweite im nächsten Kapitel bewiesen. Es wird angenommen, daß es eine Obergrenze für die Dauer von Phasen sowie die Dauer des Systemtests gibt. Die Anzahl der zu berücksichtigenden Zustandsvektoren und somit alle im Folgenden auftretenden Summen sind dann endlich.

BEWEIS FÜR $\zeta = \varepsilon$. Der Zustand ε wird erreicht, nachdem der Systemtest bestanden ist. Er wird nicht mehr verlassen, deshalb ist in Gleichung (9.7) formal $P_\varepsilon(0, \varepsilon) = 1$ gesetzt. Für die übrigen Zustände $(d, \eta) \neq (0, \varepsilon)$ ist $P_\varepsilon(d, \eta) = 0$. Daher ist

$$\sum_{d, \eta} P_\varepsilon(d, \eta) = P_\varepsilon(0, \varepsilon) = 1.$$

BEWEIS FÜR $\zeta = \underline{\infty}$. Der Zustand $\underline{\infty}$ wird mit dem Ende eines Entwicklungszyklus erreicht. Der anschließende Systemtest führt entweder zum Projektabschluß (Zustand ε) oder zu einem neuen Zyklus. Deshalb ist $P_{\underline{\infty}}(d, \underline{\infty}) = 0$. Treten beim Testen Fehler auf, $K \neq \emptyset$, dann gibt es zu jeder Menge X von betroffenen Teams genau einen Anfangszustand $\chi(X)$ für den nachfolgenden Zyklus. Nur für $\eta = \chi(X)$ ist dann $I_{\underline{\infty}}^{(3)}(\eta) = 1$. Mit Gleichung (9.5) folgt

$$\begin{aligned} \sum_{\eta \neq \varepsilon, \underline{\infty}} P_{\underline{\infty}}(d, \eta) &= \sum_X \sum_{K \neq \emptyset} \beta(d, K) \cdot \alpha(K, X) = \\ &= \sum_{K \neq \emptyset} \beta(d, K) \cdot \left(\sum_X \alpha(K, X) \right) = \sum_{K \neq \emptyset} \beta(d, K). \end{aligned}$$

Daraus folgt mit Gleichung (9.6)

$$\begin{aligned} \sum_{d,\eta} P_{\underline{\infty}}(d, \eta) &= \sum_d P_{\underline{\infty}}(d, \varepsilon) + \sum_d \sum_{\eta \neq \varepsilon, \underline{\infty}} P_{\underline{\infty}}(d, \eta) \\ &= \sum_d \beta(d, \emptyset) + \sum_d \sum_{K \neq \emptyset} \beta(d, K) = 1. \end{aligned}$$

BEWEIS FÜR $\zeta \neq \varepsilon, \underline{\infty}$. Innerhalb eines Entwicklungszyklus ist der Zustand ε nicht erreichbar, also ist $P_{\zeta}(d, \varepsilon) = 0$. Daraus folgt

$$\sum_{d,\eta} P_{\zeta}(d, \eta) = \sum_d P_{\zeta}(d, \underline{\infty}) + \sum_d \sum_{\eta \neq \varepsilon, \underline{\infty}} P_{\zeta}(d, \eta).$$

Geht man auf den ursprünglichen Ansatz in Gleichung (7.1) für die Wahrscheinlichkeiten $P_{\zeta}(d, \eta)$ zurück, so erhält man die Gleichung

$$\begin{aligned} \sum_d \sum_{\eta \neq \varepsilon, \underline{\infty}} P_{\zeta}(d, \eta) &= \sum_{\eta \neq \varepsilon, \underline{\infty}} \sum_{\{v \mid K \neq \emptyset\}} P_{\zeta}(v) \cdot P_{\zeta}(\eta \mid v) \\ &= \sum_{\{v \mid K \neq \emptyset\}} P_{\zeta}(v) \cdot \left(\sum_{\eta \neq \varepsilon, \underline{\infty}} P_{\zeta}(\eta \mid v) \right). \end{aligned}$$

Es genügt daher, die folgenden beiden Aussagen zu zeigen.

(I) Für jeden Zustand $\zeta \neq \varepsilon, \underline{\infty}$ und für jeden Phasenverlauf v mit $K \neq \emptyset$ und mit $P_{\zeta}(v) \neq 0$ gilt:

$$\sum_{\eta \neq \varepsilon, \underline{\infty}} P_{\zeta}(\eta \mid v) = 1.$$

(II) Für jeden Zustand $\zeta \neq \varepsilon, \underline{\infty}$ gilt:

$$\sum_{\{v \mid K = \emptyset\}} P_{\zeta}(v) + \sum_{\{v \mid K \neq \emptyset\}} P_{\zeta}(v) = 1.$$

Die Aussage (II) bedeutet, daß die Formeln für die $P_\zeta(v)$ tatsächlich ein Wahrscheinlichkeitstheoretisches Modell für eine Phase mit dem Anfangszustand ζ bilden. Das wird im nächsten Kapitel bewiesen. Um Aussage (I) zu beweisen, genügt es, nur solche Verläufe v zu betrachten, die ausgehend vom Zustand $\zeta \neq \varepsilon, \underline{\infty}$ möglich sind. Aus $I_\zeta^{(0)}(v) = 0$ folgt ja entgegen der Voraussetzungen $P_\zeta(v) = 0$. Zu jeder Menge X von betroffenen Teams (nach Voraussetzung ist $K \neq \emptyset$) gibt es genau einen Anfangszustand $\chi(X)$ für den nachfolgenden Zyklus. Nur für $\eta = \chi(X)$ ist dann $I_{\zeta,v}^{(1)}(\eta) = 1$. Mit Gleichung (8.3) für die Wahrscheinlichkeiten $P_\zeta(\eta | v)$ folgt daraus

$$\sum_{\eta \neq \varepsilon, \underline{\infty}} P_\zeta(\eta | v) = \sum_X \alpha(K, X) = 1.$$

□

11. Beweis zu den Phasenverläufen

Die Formeln $P_\zeta(v)$ bilden ein wahrscheinlichkeitstheoretisches Modell für eine Phase mit dem Anfangszustand ζ , denn es gilt der

$$\text{SATZ:} \quad \sum_{\{v \mid K=\emptyset\}} P_\zeta(v) \cdot I_\zeta^{(2)}(v) + \sum_{\{v \mid K \neq \emptyset\}} P_\zeta(v) \cdot I_\zeta^{(0)}(v) = 1.$$

Das muß noch formal bewiesen werden. Die Faktoren $I_\zeta^{(0)}(v)$ und $I_\zeta^{(2)}(v)$ wurden explizit herausgezogen, um zu verdeutlichen, daß letztlich nur über solche Phasenverläufe summiert wird, die ausgehend vom Zustand ζ möglich sind. Für den formalen Beweis wird eine geeignete Darstellung der Gesamtheit der möglichen Phasenverläufe als ein endlicher Baum entwickelt. Dabei wird angenommen, daß es eine Obergrenze für die Dauer von Phasen gibt. Dann wird zu diesem Baum mit Hilfe eines allgemeinen, bekannten Schemas ein diskreter Wahrscheinlichkeitsraum konstruiert. Dessen Elementarereignisse entsprechen genau den möglichen Phasenverläufen. Die einzelnen Schritte der Konstruktion werden so gestaltet, daß die Wahrscheinlichkeiten der Elementarereignisse genau durch die Formeln $P_\zeta(v)$ beschrieben sind. Da sich generell in einem diskreten Wahrscheinlichkeitsraum die Wahrscheinlichkeiten der Elementarereignisse zu Eins aufsummieren, ist der Satz dann bewiesen.

Darstellung als Baum

Eine Phase verläuft in *Schritten*. Ein Schritt ist dadurch festgelegt, daß sich eines oder mehrere Teams gleichzeitig beim Manager melden. Je nachdem, ob im Verlauf der Phase Probleme aufgetreten sind oder nicht, besteht ihr letzter Schritt entweder darin, daß ein neuer Systementwurf gültig wird, oder darin, daß das letzte Team mit seiner Arbeit fertig wird (oder die letzten Teams gleichzeitig). Jeder Schritt kann als ein Tupel

$$(m, A, B, \delta)$$

beschrieben werden. Dabei ist m die Anzahl der Takte, die seit dem vorigen Schritt ereignislos verstrichen sind. Die Menge A enthält die Nummern der Teams, die genau

zu dem betrachteten Zeitpunkt Probleme melden, und die Menge B die Nummern der Teams, die genau zu diesem Zeitpunkt fertig werden. Eine der beiden Mengen kann auch leer sein. Die Variable δ ist gleich 1 oder gleich 0, je nachdem, ob zu dem Zeitpunkt ein neuer Systementwurf gültig wird oder nicht. Falls Probleme gemeldet werden, muß $\delta = 0$ sein. Nur wenn ein neuer Systementwurf gültig wird, dürfen beide Mengen leer sein.

LEMMA 1. Zu jedem Phasenverlauf $v = (d, \mathcal{K}, \mathcal{L})$ gehört in eindeutiger Weise eine Folge von Schritten. Zu verschiedenen Phasenverläufen gehören auch verschiedene Folgen von Schritten.

BEWEIS. Um die Folge zu bestimmen, werden alle Zeitpunkte k_ν und l_μ gemeinsam aufsteigend angeordnet. Zu jedem in dieser Reihe einzelnen k_ν beziehungsweise l_μ gehört ein Schritt mit den Mengen $A = K_\nu$ und $B = \emptyset$ beziehungsweise mit den Mengen $A = \emptyset$ und $B = L_\mu$. Fallen in der Reihe ein k_ν und ein l_μ zusammen, ergibt das einen Schritt mit den Mengen $A = K_\nu$ und $B = L_\mu$. Die Zahlen m der Schritte sind die Abstände zwischen den angeordneten Zeitpunkten, wobei die Reihe um die Dauer d ergänzt wird. Unterscheiden sich zwei Phasenverläufe in einem K_ν, L_μ, k_ν oder l_μ , so unterscheiden sich auch die zugehörigen Schrittfolgen. Haben zwei Phasenverläufe nur unterschiedliche Dauer, dann unterscheiden sich die zugehörigen Schrittfolgen in der Zahl m für den letzten Schritt. \square

Da zu Beginn einer Phase höchstens N Teams arbeiten, die sich dann zu unterschiedlichen Zeitpunkten melden könnten, gibt es zusammen mit dem letzten Schritt höchstens $N + 1$ Schritte in einer Phase.

LEMMA 2. Eine Schrittfolge mit der Länge $r \leq N + 1$ gehört genau dann zu einem möglichen Phasenverlauf mit dem Anfangszustand ζ , wenn sie folgende Eigenschaften hat. Der Laufindex der Schrittfolge ist mit ϱ bezeichnet.

- Nur die Nummern solcher Teams kommen vor, die noch nicht fertig waren :

$$A_\varrho, B_\varrho \subset M_0 = \{i \mid \zeta_i \neq \infty\}.$$

- Die Mengen A_ϱ, B_ϱ sind allesamt disjunkt.
- Mit der möglichen Ausnahme des letzten Schrittes ist für alle Schritte $\delta_\varrho = 0$.

- Für den letzten Schritt ist $A_r = \emptyset$.
- Ist für den letzten Schritt $\delta_r = 1$, dann ist wenigstens ein $A_\varrho \neq \emptyset$.
- Ist auch für den letzten Schritt $\delta_r = 0$, dann sind alle Teams fertig geworden:

$$\delta_r = 0 \implies \bigcup_{\varrho} B_{\varrho} = M_0.$$

BEWEIS. Die Schrittfolge, die zu einem gegebenen Phasenverlauf gehört, wurde schon in Lemma 1 bestimmt. Die zweite bis vorletzte Eigenschaft wird von jeder solchen Schrittfolge erfüllt, siehe den Beweis von Lemma 1. Umgekehrt wird der Phasenverlauf zu einer gegebenen Schrittfolge wie folgt bestimmt. Die Mengen K_ν sind die nichtleeren Mengen A_ϱ . Die Mengen L_μ sind die nichtleeren Mengen B_ϱ . Die Zeitpunkte k_ν und l_μ bestimmt man durch schrittweises Zusammenzählen der m_ϱ . Die Dauer der Phase ist $d = \sum_{\varrho} m_\varrho$. Die erste Eigenschaft verwirft genau diejenigen Schrittfolgen, für deren Phasenverlauf $I_\zeta^{(0)}(v) = 0$ wäre. Die letzte Eigenschaft verwirft genau diejenigen, für deren Phasenverlauf $I_\zeta^{(2)}(v) = 0$ wäre. \square

Die Gesamtheit aller Schrittfolgen, die nach Lemma 2 zu den möglichen Phasenverläufen mit dem Anfangszustand ζ gehören, bildet einen Baum

$$\mathbb{B}(\zeta)$$

der Tiefe $|M_0| + 1$. Die Knoten und Kanten des Baums werden induktiv festgelegt. Die Knoten der ersten Stufe sind die Anfangsschritte der Folgen, wobei die doppelten Schritte weggelassen werden. Der Weg von der Wurzel bis zu einem Knoten der Stufe ϱ ist das Anfangsstück einer oder mehrerer Schrittfolgen. Die Nachfolger des Knotens sind dann die Schritte Nummer $\varrho + 1$ aus eben diesen Schrittfolgen, wobei doppelte Schritte wieder weggelassen werden. Die Beschriftungen der Knoten wiederholen sich also. Da eine Obergrenze für die Dauer von Phasen angenommen wird, ist der Baum endlich. Die Knoten auf einem Weg durch den Baum von der Wurzel bis zu einem Endknoten sind alle verschieden, da die Mengen A_ϱ , B_ϱ der entsprechenden Schrittfolge alle disjunkt sind. Der Baum ist nicht symmetrisch.

Schema

Man stelle sich einen endlichen Baum vor. An jede Kante wird eine Wahrscheinlichkeit geschrieben. Dabei soll für jeden Knoten gelten, daß sich die Wahrscheinlichkeiten der Kanten, die von diesem Knoten ausgehen, zu Eins aufsummieren. Man erhält dann wie folgt einen diskreten Wahrscheinlichkeitsraum.

- Die Elementarereignisse sind die Wege von der Wurzel bis zu den Endknoten.
- Die Wahrscheinlichkeit eines Elementarereignisses ist das Produkt der Wahrscheinlichkeiten, die an den Kanten des entsprechenden Weges stehen.

Dabei ist mit „Weg“ immer der kürzestmögliche gemeint.

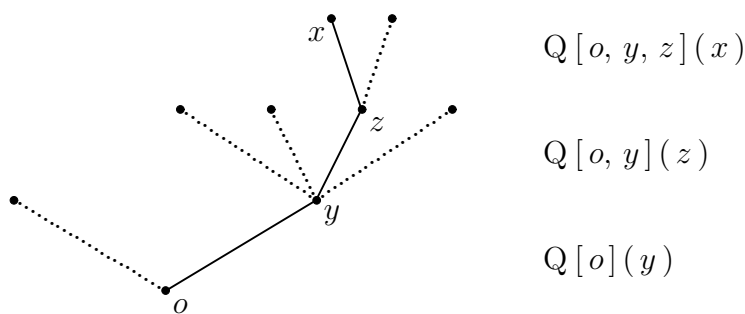
Präziser wird die Wahrscheinlichkeit einer bestimmten Kante mit

$$Q[S](x)$$

bezeichnet. Hier ist S die Folge von Knoten, die ein Weg von der Wurzel o bis zum Anfangspunkt der Kante durchläuft, und x ist der Endpunkt der Kante. Die Bedingung an die Kantenwahrscheinlichkeiten lautet formal, daß für jeden solchen Weg S gilt

$$\sum_x Q[S](x) = 1.$$

Die Wahrscheinlichkeit $Q[S](x)$ kann aufgefaßt werden als die bedingte Wahrscheinlichkeit für das Auftreten von x unter der Bedingung, daß die „Vorgeschichte“ S ist.



Die Wahrscheinlichkeit des Weges (o, s_1, \dots, s_j) von der Wurzel bis zu einem beliebigen Knoten ist festgelegt als

$$Q(o, s_1, \dots, s_j) = \prod_{\varrho=1}^j Q[o, \dots, s_{\varrho-1}](s_{\varrho}).$$

Es ist bekannt und kann induktiv bewiesen werden, daß die Summe dieser Ausdrücke für die Wege von der Wurzel bis zu den Endknoten, also die Summe der Wahrscheinlichkeiten der Elementarereignisse, gleich Eins ist. Das gilt auch dann noch, wenn die Bedingung an die Kantenwahrscheinlichkeiten nur für die Wege S mit $Q(S) \neq 0$ gefordert wird.

Konstruktion

Es sei nun eine Folge S von Schritten fest vorgegeben, die das Anfangsstück eines oder mehrerer Phasenverläufe ist. Ferner sei (m, A, B, δ) ein zulässiger nächster Schritt, dessen Wahrscheinlichkeit festgelegt werden soll. Anschaulich befinden wir uns also an einem Knoten im Baum $\mathbb{B}(\zeta)$, der über den Weg S erreicht wird, und möchten die Wahrscheinlichkeit der zum Knoten (m, A, B, δ) ausgehenden Kante festlegen. Aus dem vorgegebenen Anfangsstück S ergibt sich die Zahl e der Takte, die die Phase bis einschließlich des vorigen Schrittes schon gedauert hat. Die Teams aus A und B haben sich also bis dahin e Takte lang nicht beim Manager gemeldet. Die Wahrscheinlichkeit dafür, daß sich das i -te Team nun m Takte später meldet, wird von seinen um $\zeta_i + e$ verschobenen Langzeitverteilungen bestimmt. Das legt nahe, die Wahrscheinlichkeit der Kante festzulegen als

$$\begin{aligned} Q[S](m, A, B, \delta) &= \prod_{i \in A} \Theta_{\zeta_i + e}(p_i) \cdot (\Theta_{\zeta_i + e} g_i)(m) \\ &\times \prod_{i \in B} \Theta_{\zeta_i + e}(q_i) \cdot (\Theta_{\zeta_i + e} h_i)(m) \\ &\times \prod_{i \in M} \sum_{j > m} (\Theta_{\zeta_i + e} w_i)(j) \\ &\times Q'. \end{aligned}$$

Ist $A = \emptyset$ oder $B = \emptyset$, dann fallen die entsprechenden Faktoren weg. Die Menge

$$M = M_r \setminus (A \cup B) = \left(M_0 \setminus \left(\bigcup_{\varrho=1}^r A_\varrho \cup B_\varrho \right) \right) \setminus (A \cup B)$$

enthält die Nummern aller Teams, die sich in diesem Schritt auch noch hätten melden können, wobei r die Anzahl der Schritte im Anfangsstück und $M_0 = \{i \mid \zeta_i \neq \infty\}$ ist. Der Faktor Q' hängt davon ab, ob zu dem betrachteten Zeitpunkt der Systementwurf noch gültig oder bereits offen ist. Letzteres läßt sich daran erkennen, daß bei einem vorangegangenen Schritt $A_\varrho \neq \emptyset$ war. Ist der Systementwurf noch gültig, dann ist notwendig $\delta = 0$, und man setzt

$$Q' = 1.$$

Ist der Entwurf bereits offen, dann wurde schon mit seiner Überarbeitung begonnen. Bezeichnet c die Zahl der Takte, die die letzte Problemmeldung im Anfangsstück zurückliegt, dann ist die Wahrscheinlichkeit für die Fertigstellung des neuen Systementwurfs von der um c verschobenen Verteilung γ bestimmt. Entsprechend sind für Q' die Fälle

$$A = \emptyset \text{ und } \delta = 0: \text{ setze } Q' = \sum_{j>m} (\Theta_c \gamma)(j);$$

$$A = \emptyset \text{ und } \delta = 1: \text{ setze } Q' = (\Theta_c \gamma)(m);$$

$$A \neq \emptyset: \text{ setze } Q' = \sum_{j \geq m} (\Theta_c \gamma)(j)$$

zu unterscheiden, je nachdem, ob es nun zu einem neuen Entwurf kommt oder nicht.

Um später das Schema anwenden zu können, müssen sich die Wahrscheinlichkeiten der zulässigen nächsten Schritte zumindest für $Q(S) \neq 0$ zu Eins summieren. Die Schrittfolge S und somit die Zahlen e, r, c sind immer noch vorgegeben.

LEMMA 3. Ist der Systementwurf offen und eine der verschobenen Verteilungen $\Theta_c \gamma$ oder $\Theta_{\zeta_i+e} w_i$ überall gleich Null, oder ist der Systementwurf gültig und eine der verschobenen Verteilungen $\Theta_{\zeta_i+e} w_i$ überall gleich Null, dann gilt bereits $Q(S) = 0$.

BEWEIS. Die Aussage wird für den Fall gezeigt, daß der Systementwurf offen und die Verteilung $\Theta_c \gamma$ überall gleich Null ist. Die Beweise für die anderen Fälle laufen fast

genauso. Da γ selbst nicht überall gleich Null ist, kann $c \neq 0$ angenommen werden. Daraus folgt $r \geq 2$. Es bezeichne S_0 die Schrittfolge, die aus S durch Weglassen des letzten Schrittes $x = (m_r, A_r, B_r, 0)$ entsteht. Aus $c \neq 0$ folgt, daß es im letzten Schritt x keine Problemmeldungen gab, also ist $A_r = \emptyset$. Der Faktor Q' in der Formel für $Q[S_0](x)$ war daher gleich

$$\begin{aligned} Q' &= \sum_{j > m_r} (\Theta_{c-m_r} \gamma)(j) = \sum_{j > m_r} \frac{\gamma(c - m_r + j)}{1 - \Gamma(c - m_r)} \\ &= \frac{1 - \Gamma(c)}{1 - \Gamma(c - m_r)} \cdot \sum_{j \geq 1} (\Theta_c \gamma)(j) = 0. \end{aligned}$$

Daraus folgt $Q(S) = 0$. □

LEMMA 4. Für $Q(S) \neq 0$ summieren sich die Wahrscheinlichkeiten der zulässigen nächsten Schritte zu Eins.

BEWEIS. Das wird für den Fall gezeigt, daß der Systementwurf bereits offen ist. Der andere Fall wird fast genauso gezeigt. Aus Gleichung (6.1) folgt für $A \cup B \neq \emptyset$ die Gleichung

$$\begin{aligned} Q[S](m, \emptyset, A \cup B, 0) + Q[S](m, \emptyset, A \cup B, 1) \\ + \sum_{\substack{X_1 \cup X_2 = A \cup B \\ X_1 \neq \emptyset, X_1 \cap X_2 = \emptyset}} Q[S](m, X_1, X_2, 0) &= \prod_{i \in A \cup B} (\Theta_{\zeta_i + e} w_i)(m) \\ &\times \prod_{i \in M} \sum_{j > m} (\Theta_{\zeta_i + e} w_i)(j) \\ &\times \sum_{j \geq m} (\Theta_c \gamma)(j). \end{aligned}$$

Summiert wird auf der linken Seite über alle Zerlegungen der nichtleeren Menge $A \cup B$ in zwei disjunkte Teilmengen X_1 und X_2 . Für $A \cup B = \emptyset$ hat man die Gleichung

$$\begin{aligned} Q[S](m, \emptyset, \emptyset, 1) &= \prod_{i \in M_r} \sum_{j > m} (\Theta_{\zeta_i + e} w_i)(j) \\ &\times (\Theta_c \gamma)(m). \end{aligned}$$

Nun betrachtet man den Produkt-Wahrscheinlichkeitsraum

$$(Y_0, Q_0) \times \prod_{i \in M_r} (Y_i, Q_i)$$

mit

$$Y_0 = Y_i = \{0, 1, \dots, D\}$$

$$Q_0(j) = (\Theta_c \gamma)(j) \quad \text{und} \quad Q_i(j) = (\Theta_{\zeta_i + e} w_i)(j) \quad (j \geq 1)$$

$$Q_0(0) = 1 - \sum_{j=1}^D (\Theta_c \gamma)(j) \quad \text{und} \quad Q_i(0) = 1 - \sum_{j=1}^D (\Theta_{\zeta_i + e} w_i)(j).$$

Da verschobene Verteilungen entweder wieder Wahrscheinlichkeitsverteilungen sind oder überall gleich Null sind, nimmt $Q_0(0)$ beziehungsweise $Q_i(0)$ nur den Wert 0 oder 1 an. Dabei ist D die Obergrenze für die Dauer von Phasen. Die rechte Seite der Gleichung für $A \cup B \neq \emptyset$ ist im Produktraum genau die Wahrscheinlichkeit des Ereignisses

$$,, y_i = m \text{ für } i \in A \cup B, \quad y_i > m \text{ für } i \in M, \quad y_0 \geq m "$$

und die rechte Seite der Gleichung für $A \cup B = \emptyset$ ist im Produktraum genau die Wahrscheinlichkeit des Ereignisses

$$,, y_i > m \text{ für } i \in M_r, \quad y_0 = m "$$

Durchlaufen m und $A \cup B$ alle Möglichkeiten, so bilden diese Ereignisse zusammen eine Zerlegung der Teilmenge

$$,, y_i \geq 1 \text{ für } i \in M_r, \quad y_0 \geq 1 "$$

Daher sind die Wahrscheinlichkeiten der zulässigen nächsten Schritte genau dann allesamt gleich Null, wenn die Wahrscheinlichkeit

$$\left(1 - Q_0(0)\right) \cdot \prod_{i \in M_r} \left(1 - Q_i(0)\right)$$

dieser Teilmenge gleich Null ist. Das ist genau dann der Fall, wenn wenigstens eine der verschobenen Verteilungen überall gleich Null ist. Nach Lemma 3 und der Voraussetzung $Q(S) \neq 0$ ist aber keine der verschobenen Verteilungen überall gleich Null. Daher ist $Q_0(0) = Q_i(0) = 0$, und die Wahrscheinlichkeiten der zulässigen nächsten Schritte summieren sich zu Eins. \square

Wahrscheinlichkeiten der Elementarereignisse

Im Baum $IB(\zeta)$ ist nun für jede Kante eine Wahrscheinlichkeit festgelegt. Wendet man auf diesen Baum das Schema an, so erhält man einen Wahrscheinlichkeitsraum, dessen Elementarereignisse die Schrittfolgen sind, die den Phasenverläufen mit dem Anfangszustand ζ entsprechen. Ihre Wahrscheinlichkeiten sind die Produkte der Kantenwahrscheinlichkeiten $Q[o, \dots, s_{\varrho-1}](s_{\varrho})$ für die einzelnen Schritte s_{ϱ} . Abschließend muß nachgewiesen werden, daß gilt:

LEMMA 5. Die Wahrscheinlichkeit eines Elementarereignisses, das dem Phasenverlauf v entspricht, stimmt genau mit der Formel $P_{\zeta}(v)$ überein.

BEWEIS. Betrachten wir einen Schritt (m, A, B, δ) in dem Elementarereignis. Hat sich ein Team, etwa das i -te, nach dem Schritt noch nicht gemeldet, dann trägt die Kantenwahrscheinlichkeit dieses Schrittes den Faktor

$$\sum_{j>m} (\Theta_{\zeta_i+e} w_i)(j) = \frac{1 - W_i(\zeta_i + e + m)}{1 - W_i(\zeta_i + e)}$$

für das i -te Team zur Wahrscheinlichkeit des Elementarereignisses bei. Dabei ist e die Dauer der Phase bis zum vorigen Schritt. Meldet sich das Team auch im nächsten Schritt (m', A', B', δ') nicht, kommt der Faktor

$$\frac{1 - W_i(\zeta_i + e + m + m')}{1 - W_i(\zeta_i + e + m)}$$

dazu. Meldet das Team hingegen im nächsten Schritt ein Problem, dann kommt der Faktor

$$\Theta_{\zeta_i+e+m}(p_i) \cdot (\Theta_{\zeta_i+e+m} g_i)(m') = \frac{p_i \cdot g_i(\zeta_i + e + m + m')}{1 - W_i(\zeta_i + e + m)}$$

dazu. Man sieht, daß sich beide Male der Zähler aus dem ersten Faktor und der Nenner aus dem zweiten Faktor kürzen. Die Faktoren, welche die Kantenwahrscheinlichkeiten der einzelnen Schritte für das i -te Team beitragen, bilden also ein „Teleskopprodukt“. Von einem Teleskopprodukt bleiben nur der Nenner des ersten Faktors und der Zähler des letzten Faktors übrig. Meldet sich das i -te Team die gesamte Schrittfolge hindurch nicht, ergibt das Teleskopprodukt gerade

$$\frac{1}{1 - W_i(\zeta_i)} \cdot \frac{1 - W_i(\zeta_i + d)}{1} = 1 - \sum_{j=1}^d (\Theta_{\zeta_i} w_i)(j).$$

Dabei ist d die Summe der zeitlichen Abstände zwischen den Schritten des betrachteten Elementarereignisses, also die Dauer des zugehörigen Phasenverlaufs. Meldet sich das Team im j -ten Schritt mit einem Problem, ergibt das Teleskopprodukt gerade

$$\frac{1}{1 - W_i(\zeta_i)} \cdot \frac{p_i \cdot g_i(\zeta_i + k_\nu)}{1} = \Theta_{\zeta_i}(p_i) \cdot (\Theta_{\zeta_i} g_i)(k_\nu).$$

Dabei ist der Zeitpunkt k_ν nach Phasenbeginn die Summe der zeitlichen Abstände zwischen den ersten j Schritten des betrachteten Elementarereignisses. Entsprechendes gilt, wenn das Team zum Zeitpunkt l_μ nach Phasenbeginn fertig wird. Es bleibt dann

$$\Theta_{\zeta_i}(q_i) \cdot (\Theta_{\zeta_i} h_i)(l_\mu)$$

vom Teleskopprodukt übrig. Hat sich ein Team in einem bestimmten Schritt gemeldet, dann tragen die darauffolgenden Schritte keine Faktoren mehr für das Team zur Wahrscheinlichkeit des Elementarereignisses bei.

Ähnlich verfährt man mit den Faktoren Q' , die in den Kantenwahrscheinlichkeiten der Schritte enthalten sind. Ein Schritt $(m, \emptyset, B, 0)$ bei offenem Entwurf, in dem weder

weitere Probleme gemeldet werden, noch der neue Entwurf fertig wird, ergibt den Faktor

$$\sum_{j>m} (\Theta_c \gamma)(j) = \frac{1 - \Gamma(c + m)}{1 - \Gamma(c)}.$$

Wieder kürzen sich der Zähler dieses Faktors und der Nenner des Faktors für den nächsten Schritt, und zwar für alle drei Arten (siehe Q') möglicher nächster Schritte. Für den Schritt unmittelbar nach einer Problemmeldung ist $c = 0$. Von den Schritten nach den Problemmeldungen im Zeitpunkt k_ν bis einschließlich der Problemmeldungen im Zeitpunkt $k_{\nu+1}$ bleibt nur

$$\frac{1}{1 - \Gamma(0)} \cdot \frac{1 - \Gamma(k_{\nu+1} - k_\nu - 1)}{1} = 1 - \Gamma(k_{\nu+1} - k_\nu - 1).$$

Von den Schritten nach den letzten Problemmeldungen im Zeitpunkt k_n bis zur Fertigstellung des neuen Systementwurfs bleibt nur

$$\frac{1}{1 - \Gamma(0)} \cdot \frac{\gamma(d - k_n)}{1} = \gamma(d - k_n).$$

□

12. Schätzungen

Das Modell wird nun eingesetzt, um die Erfolgsaussichten, die Dauer und die Kosten von Softwareprojekten zu schätzen. Die Idee ist, jedem Projektverlauf seine Dauer und seine Kosten zuzuordnen und dann die Wahrscheinlichkeiten zu berechnen, mit denen diese beiden Funktionen ihre jeweiligen Werte annehmen. Die Erfolgsaussichten berechnet man unmittelbar aus diesen Wahrscheinlichkeiten. Als Schätzungen für die Dauer und die Kosten berechnet man die Erwartungswerte der beiden Funktionen.

Ein Projektverlauf wurde formal als die Folge

$$\omega = \left\{ (d_j, \zeta(j)) \right\}_{j=0, 1, 2, \dots}$$

der durchlaufenen Zustände beschrieben. Ein abgeschlossenes Projekt entspricht einer endlichen Folge, die mit einem Übergang vom Zustand $\underline{\infty}$ in den Zustand ε endet. Da es möglich ist, daß ein Projekt nie zum Ende kommt, müssen im Modell auch unendliche Folgen berücksichtigt werden. Der Raum Ω aller möglichen Projektverläufe ist daher überabzählbar. Streng genommen müßte dieser Raum mit den Methoden der höheren Wahrscheinlichkeitstheorie behandelt werden. Insbesondere läßt sich der Erwartungswert einer Funktion meist nicht mehr einfach berechnen. Für die praktischen Zwecke dieser Arbeit genügen jedoch folgende Näherungen.

Man wählt eine „absolute Höchstdauer“ von x_0 Takten, nach deren Ablauf ein Projekt endgültig als Mißerfolg abgebrochen wird. Eine Höchstdauer gibt es ganz natürlich für jedes Projekt. Für die Näherungen werden dann nur die Wahrscheinlichkeiten der erfolgreichen Projektverläufe betrachtet, also solcher Verläufe, deren Dauer die Höchstdauer nicht überschreitet, sowie die verbleibende Wahrscheinlichkeit dafür, daß das Projekt abgebrochen wird. Genauer betrachtet man anstelle des Raumes Ω den Raum Ω^+ aller endlichen Folgen, deren letzter Zustand, und nur dieser, gleich ε ist. Darunter sind auch Folgen, die nicht wirklich als Projektverlauf auftreten, etwa solche, die einen Übergang von $\underline{\infty}$ nach $\underline{\infty}$ enthalten. Diese Folgen werden später durch ihre Wahrscheinlichkeit, die gleich Null ist, aussortiert. Die Wahrscheinlichkeit eines Projektverlaufs ω wird mit Gleichung (7.2) berechnet. Ferner betrachtet man anstelle der auf Ω definierten Funktionen lediglich ihre Einschränkungen auf Ω^+ .

Für einen Projektverlauf ω aus Ω^+ bezeichne $|\omega|$ die Anzahl der Phasen in ω , wobei Systemtests als Phasen gezählt werden. Dann wird durch

$$f(\omega) = \sum_{j=1}^{|\omega|} d_j$$

eine Funktion

$$f : \Omega^+ \longrightarrow \mathbb{N}$$

definiert, die jedem endlichen Projektverlauf seine Dauer zuordnet. Die Wahrscheinlichkeit dafür, daß ein Projekt nach genau x Takten erfolgreich abgeschlossen sein wird, ist gleich

$$\varphi(x) = \mathbb{P}(f^{-1}(x)) = \sum_{\{\omega: f(\omega) = x\}} \mathbb{P}(\omega). \quad (12.1)$$

Summiert wird über alle endlichen Projektverläufe von x Takten Dauer. Die Wahrscheinlichkeit dafür, das Projekt innerhalb der Höchstdauer abzuschließen, ist gleich

$$\Phi(x_0) = \sum_{x \leq x_0} \varphi(x).$$

Da die Anzahl der Projektzustände endlich ist, sind auch die Summen endlich.

Als Ersatz für den Erwartungswert auf Ω eignet sich für auf Ω^+ definierte, nichtnegative Funktionen h der „bei x_0 abgeschnittene“ Erwartungswert

$$\mathbb{E}_0[h] = \sum_{\{\omega: f(\omega) \leq x_0\}} h(\omega) \cdot \mathbb{P}(\omega) + m_0[h] \cdot (1 - \Phi(x_0)). \quad (12.2)$$

Die Summanden $h(\omega) \cdot \mathbb{P}(\omega)$ bilden das „gewichtete Mittel“ der Funktionswerte für die erfolgreichen Projektverläufe, wobei das Gewicht eines Projektverlaufs die Wahrscheinlichkeit seines Auftretens ist. Die übrigen Projektverläufe, die zusammen das Gewicht $1 - \Phi(x_0)$ haben, gehen mit dem Faktor

$$m_0[h] = \max \{ h(\omega) : \omega \in \Omega^+, f(\omega) \leq x_0 \} \quad (12.3)$$

in den abgeschnittenen Erwartungswert ein. Die Summanden $h(\omega) \cdot P(\omega)$ sind die Anteile der erfolgreichen Projektverläufe am tatsächlichen Erwartungswert der geeignet auf ganz Ω ausgedehnten Funktion h . Der Faktor $m_0[h]$ ist so gewählt, daß die Zahlen $E_0[h]$ mit wachsendem x_0 gegen den tatsächlichen Erwartungswert konvergieren, sofern die Funktion h die Eigenschaft hat, daß ihre Werte $h(\omega)$ größer werden, wenn die Anzahl $|\omega|$ der Phasen in ihren Argumenten zunimmt. Das ist der Fall für die hier interessierenden Funktionen wie die Dauer oder die Kosten von Projektverläufen. Der Beweis dieser Aussagen erfordert Begriffe und Schlußweisen aus der höheren Wahrscheinlichkeitstheorie und wird daher nicht hier angegeben.

Eine *Schätzung für die Projektdauer* erhält man, wenn $h = f$ gesetzt wird, das heißt als abgeschnittenen Erwartungswert für die Dauer :

$$E_{dauer} = E_0[f].$$

Die Formel läßt sich schreiben als

$$E_{dauer} = \sum_{x \leq x_0} x \cdot \varphi(x) + x_0 \cdot (1 - \Phi(x_0)) \quad (12.4)$$

wegen $m_0[f] = x_0$.

Auf ähnliche Weise lassen sich die Projektkosten schätzen. Dazu baut man eine Funktion

$$g : \Omega^+ \longrightarrow \mathbb{R}^+,$$

die jedem endlichen Projektverlauf die mit ihm verbundenen Kosten zuordnet. Die Kosten setzen sich hauptsächlich aus den Arbeitskosten für die Teams zusammen. Eine obere Schranke dafür ist

$$g(\omega) \leq f(\omega) \cdot \sum_{i=1}^N a_i,$$

wobei a_i der Betrag ist, den das i -te Team pro Takt kostet. Die Schranke entspricht also einem Projekt, bei dem alle Teams durcharbeiten. Bezeichnet c_j die Kosten für die j -te Phase, dann ist allgemein

$$g(\omega) = \sum_{j=1}^{|\omega|} c_j.$$

Nutzt man die Information, die der Projektverlauf über den Verlauf der einzelnen Phasen enthält, dann lassen sich die Kosten besser ansetzen. Für eine normale Phase setzt man

$$c_j = \sum_{i=1}^N d_j \cdot a_i \cdot \delta_{i,j}$$

wobei $\delta_{i,j}$ gleich Eins ist, wenn das i -te Team in der j -ten Phase aktiv ist. Das läßt sich am Zustand $\zeta_i(j-1)$ des Teams zu Beginn der Phase erkennen. Andernfalls ist $\delta_{i,j}$ gleich Null gesetzt. Die Kosten für die Phase umfassen hier nur die Kosten für die aktiven Teams. Für einen Systemtest setzt man

$$c_j = d_j \cdot a_0.$$

Dabei ist a_0 der Betrag, den ein Systemtest pro Takt kostet. Die Wahrscheinlichkeit dafür, daß ein erfolgreiches Projekt genau den Betrag z kostet, ist gleich

$$\psi(z) = P(g^{-1}(z) \cap f^{-1}[1, x_0]) = \sum_{\{\omega: f(\omega) \leq x_0, g(\omega) = z\}} P(\omega). \quad (12.5)$$

Die Wahrscheinlichkeit dafür, das Projekt innerhalb der Höchstdauer x_0 und innerhalb einer vorgegebenen Kostengrenze y_0 abzuschließen, ist gleich

$$\boxed{\Psi(y_0) = \sum_{z \leq y_0} \psi(z)}. \quad (12.6)$$

Das sind die *Erfolgsaussichten* für das Projekt.

Eine *Kostenschätzung* für das Projekt erhält man, wenn $h = g$ gesetzt wird, das heißt als abgeschnittenen Erwartungswert für die Kosten:

$$E_{kosten} = E_0[g].$$

Sind die Kosten pro Takt für einen Systemtest nicht größer als die Summe der Kosten pro Takt für alle Teams zusammen, dann ist

$$m_0[g] = (x_0 - 1) \cdot \sum_{i=1}^N a_i + a_0.$$

Das entspricht einem Projekt, bei dem alle Teams $x_0 - 1$ Takte lang bis zum Systemtest durcharbeiten, der dann nur einen Takt lang dauert. Setzt man $z_0 = m_0[g]$, dann läßt sich die Formel für die Kostenschätzung wegen $\Psi(z_0) = \Phi(x_0)$ schreiben als

$$E_{kosten} = \sum_{z \leq z_0} z \cdot \psi(z) + z_0 \cdot (1 - \Psi(z_0)).$$

(12.7)

Summiert wird letztlich über die endlich vielen Werte, die als Kosten für die erfolgreichen Projektverläufe auftreten.

Die Wahrscheinlichkeiten dafür, daß die Dauer oder die Kosten eines erfolgreichen Projekts um nicht mehr als eine vorgegebene Zahl von der Schätzung abweichen, lassen sich unmittelbar aus den Funktionen φ und ψ berechnen. Für die Dauer erhält man

$$P(|f - E_{dauer}| \leq y) = \Phi(\lfloor E_{dauer} + y \rfloor) - \Phi(\lceil E_{dauer} - y - 1 \rceil). \quad (12.8)$$

Dabei ist $\lfloor x \rfloor$ die größte ganze Zahl, die kleiner oder gleich x ist, und $\lceil x \rceil$ die kleinste ganze Zahl, die größer oder gleich x ist. Für die Kosten erhält man

$$P(|g - E_{kosten}| \leq y, f \leq x_0) = \Psi(E_{kosten} + y) - \Psi(E_{kosten} - y) + \psi(E_{kosten} - y). \quad (12.9)$$

Beispielsweise kann man für y die *Streuung* einsetzen, also die Wurzel aus der passend zu den abgeschnittenen Erwartungswerten definierten *abgeschnittenen Varianz*

$$V_0[h] = E_0[h^2] - (E_0[h])^2.$$

13. Beispiele und Rechenaufwand

Die folgenden Beispiele geben einen Eindruck davon, wie die Ergebnisse aussehen, die das Modell liefert. Neben den Schätzwerten für die Dauer und die Kosten werden vor allem die Diagramme für die *Erfolgskurve* φ und die *Kostenkurve* ψ betrachtet, aus denen ja die Schätzungen und die Schwankungsbreiten abgeleitet werden. Die einzelnen Werte sind in Tabellen im Anhang aufgeführt. Die Beispiele zeigen auch, daß sich das Modell wie zu erwarten verhält, wenn die Eingangsdaten verändert werden. Alle Wahrscheinlichkeiten sind in Prozent angegeben.

Erstes Beispiel

Am ersten Projekt sind drei Teams beteiligt. Ein Takt entspricht einem Monat. Die Teams melden sich nach spätestens sechs Monaten beim Manager, sofern sie nicht unterbrochen werden. Die Ergebniswahrscheinlichkeiten der Teams sind

q_1	p_1	q_2	p_2	q_3	p_3
67.4	32.6	47.2	52.8	62.3	37.7

Die Langzeitverteilungen der Teams sind

k	$h_1(k)$	$g_1(k)$	$h_2(k)$	$g_2(k)$	$h_3(k)$	$g_3(k)$
1	0.0	30.7	21.2	56.8	16.0	13.3
2	13.4	27.6	38.2	22.7	13.6	22.6
3	32.0	22.1	31.8	11.4	32.7	54.2
4	32.0	13.3	7.6	6.8	24.0	7.2
5	18.0	5.3	1.1	2.0	11.5	2.5
6	4.6	1.0	0.1	0.3	2.2	0.2

Neue Systementwürfe werden nach einem Monat fertig,

$$\gamma(1) = 100.$$

Die Abhängigkeitsgrade $\alpha(K, X)$ sind in der Tabelle auf der nächsten Seite angegeben. Die Tabelle enthält eine Zeile für jedes K und eine Spalte für jedes X . Ein erheblicher

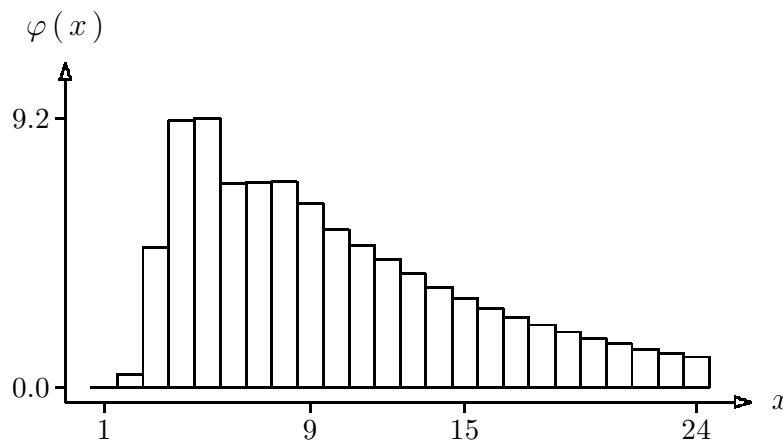
Teil der Einträge ist gleich Null, da dies im Modell für $K \nsubseteq X$ verlangt wird. Die positiven Einträge sind so gewählt, daß sie für jedes K gleichverteilt sind.

$\alpha(K, X)$	1	2	3	1,2	1,3	2,3	1,2,3
1	25	0	0	25	25	0	25
2	0	25	0	25	0	25	25
3	0	0	25	0	25	25	25
1,2	0	0	0	50	0	0	50
1,3	0	0	0	0	50	0	50
2,3	0	0	0	0	0	50	50
1,2,3	0	0	0	0	0	0	100

Bei der Berechnung der Ergebnisse wird nicht das gesamte Projekt, sondern lediglich ein Entwicklungszyklus mit dem Anfangszustand $\underline{0}$ betrachtet. Als Höchstdauer wird

$$x_0 = 24$$

Monate gewählt. Es wird keine Kostengrenze vorgegeben. Für einen Entwicklungszyklus mit diesen Eingangsdaten sieht dann die Erfolgskurve so aus :



Die Kurve steigt schnell bis zu ihrem höchsten Wert an und fällt dann langsam gegen Null ab. Dabei steigt sie noch einmal kurz an, fällt aber ansonsten monoton ab. Was der höchste Wert ist, wann er erreicht wird und wie langsam die Kurve abfällt, hängt von den jeweiligen Eingangsdaten ab.

Die Erfolgswahrscheinlichkeit dafür, den Entwicklungszyklus innerhalb von zwei Jahren abzuschließen, beträgt in diesem Beispiel etwa 92%. Die 50%-Grenze ist nach 9 Monaten erreicht, die 75%-Grenze nach 15 Monaten. Als Schätzung für die Dauer erhält man

$$E_{dauer} = 11.1$$

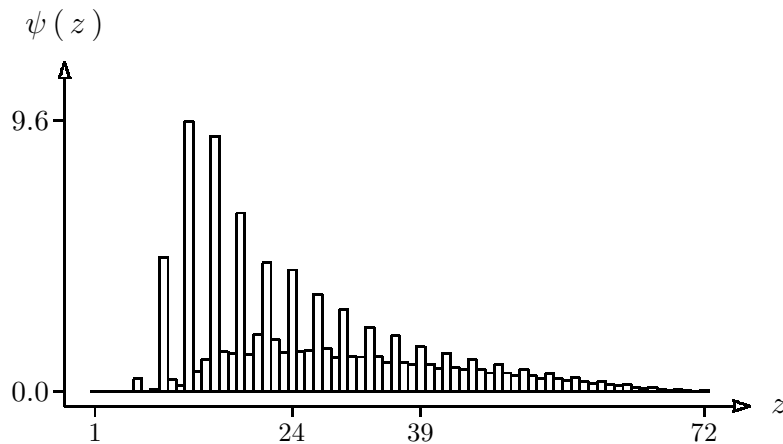
Monate. Die Streuung beträgt 6.5 Monate. Die Wahrscheinlichkeit dafür, daß die Dauer um nicht mehr als die Streuung von der Schätzung abweicht, ist gleich

$$P(|f - 11.1| \leq 6.5) = \Phi(17) - \Phi(4) \approx 67\%.$$

Im Beispiel kostet jedes Team denselben Betrag pro Monat,

$$a_i = 1.$$

Ein Takt kostet also drei „Einheiten“, wenn alle Teams arbeiten. Die Kostenkurve dazu sieht dann so aus:



Die z -Achse ist so skaliert, daß das Diagramm dieselbe Breite hat wie das vorherige. Die Kurve, welche die dritten Balken $3 \cdot x$ bilden, ähnelt der Erfolgskurve. Zur Erklärung muß man sich überlegen, wie die Kostenkurve zustande kommt. Unter den erfolgreichen Verläufen sind auch solche, bei denen nicht alle Teams die ganze Zeit gearbeitet haben. Die Verläufe der Dauer x tragen also nicht nur zu den Balken $3 \cdot x$, sondern auch zu den Balken $z < 3 \cdot x$ bei. So entstehen die zwischen den dritten Balken gelegenen Balken,

aber auch ein Anteil an den dritten Balken selbst. Ausschlaggebend ist nun, daß der Hauptanteil an den Balken $3 \cdot x$ von den Verläufen der Dauer x stammt. Das geht für das Beispiel aus folgender Tabelle hervor :

x	2	3	4	5	6	7	8	9	10	11	12
$\psi(3 \cdot x)$	0.5	4.8	9.6	9.1	6.3	4.6	4.3	3.4	2.9	2.3	2.0
$P(f = x, g = 3 \cdot x)$	0.5	4.7	8.6	7.6	4.3	3.1	2.7	2.1	1.6	1.2	0.9

Im Modell verlaufen also viele erfolgreiche Entwicklungszyklen so, daß jedes Team in jeder der Phasen arbeitet.

Als Kostenschätzung für das Beispiel erhält man

$$E_{kosten} = 29.9$$

Einheiten. Die Streuung beträgt 18.1 Einheiten. Die Wahrscheinlichkeit dafür, daß die Kosten um nicht mehr als die Streuung von der Schätzung abweichen, ist gleich

$$P(|g - 29.9| \leq 18.1, f \leq 24) = \Psi(48.0) - \Psi(11.8) + \psi(11.8) \approx 78\%.$$

Einfluß der Abhängigkeitsgrade

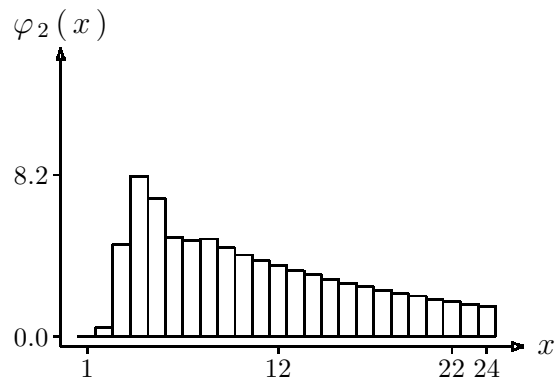
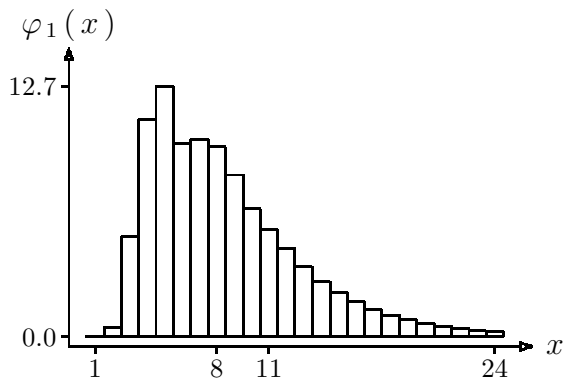
Das Modell verhält sich wie zu erwarten, wenn die Eingangsdaten verändert werden. Die nächsten beiden Beispiele zeigen den Einfluß der Abhängigkeitsgrade auf die Ergebnisse. Für das eine Beispiel wird angenommen, daß die Abhängigkeiten minimal sind,

$$\alpha(K, K) = 100.$$

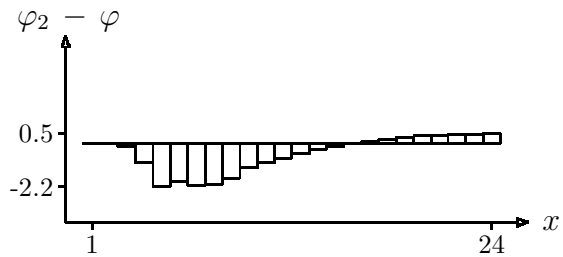
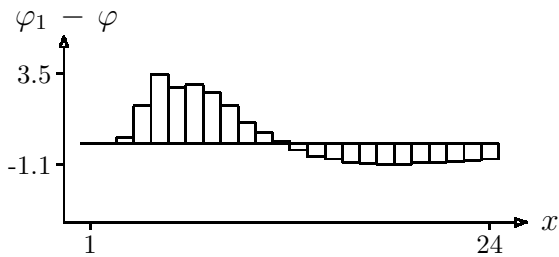
Das bedeutet, daß immer nur diejenigen Teams von Änderungen am Systementwurf betroffen sind, die selbst ein Problem gemeldet haben. Für das andere Beispiel wird angenommen, daß die Abhängigkeiten maximal sind,

$$\alpha(K, \{1, 2, 3\}) = 100.$$

Das bedeutet, daß immer alle Teams von Änderungen betroffen sind, gleichgültig, welche der Teams ein Problem gemeldet haben. Alle anderen Eingangsdaten werden festgehalten. Das ergibt diese Erfolgskurven :

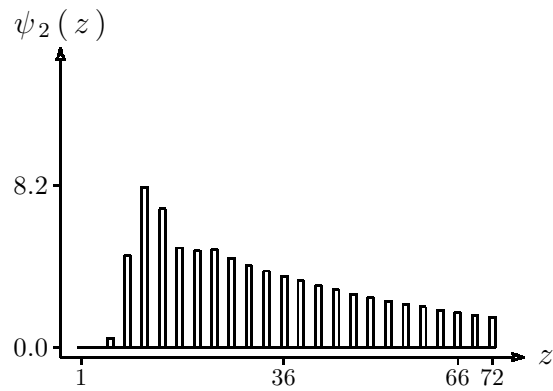
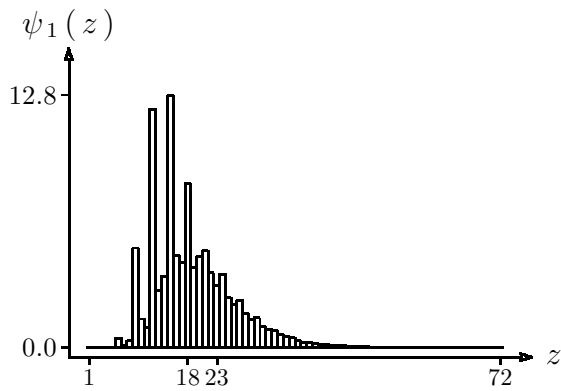


Die erste Kurve steigt auf einen höheren Wert an als die Kurve für das ursprüngliche Beispiel und fällt dann schneller gegen Null ab. Die zweite Kurve hingegen steigt weniger hoch an und fällt langsamer ab. Die Erfolgswahrscheinlichkeiten sind 99 % und 79 %. Die Unterschiede im Kurvenverlauf werden besonders deutlich, wenn man die Differenzen zum ursprünglichen Beispiel betrachtet :



Beide Male gibt es einen Wendepunkt, da die Summe der Werte einer Erfolgskurve höchstens 100 % beträgt. Das höhere Gewicht der ersten paar Balken bei minimalen Abhängigkeiten reicht aus, um eine günstigere Schätzung von 8.6 Monaten für die Dauer zu erhalten. Bei maximalen Abhängigkeiten ist die Schätzung durch das höhere Gewicht der hinteren Balken mit 13.5 Monaten ungünstiger als im ursprünglichen Beispiel. Die Streuungen sind 4.5 Monate und 7.6 Monate. Die Wahrscheinlichkeiten dafür, um nicht mehr als die Streuungen von den Schätzungen abzuweichen, betragen etwa 70 % und etwa 54 %.

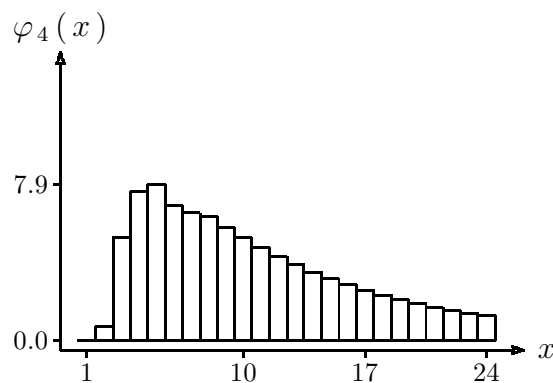
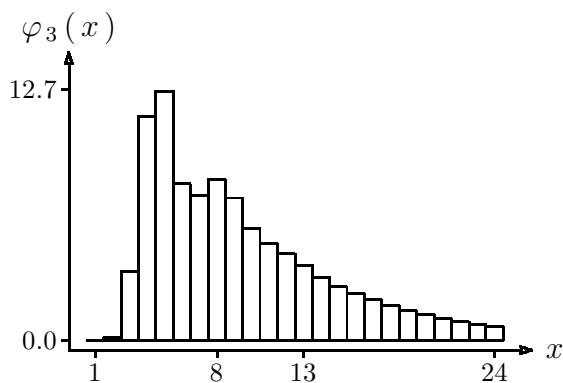
Kostet jedes der Teams unverändert denselben Betrag pro Monat, dann ergibt sich bei minimalen Abhängigkeiten eine Kostenschätzung von 19.2 Einheiten, bei maximalen Abhängigkeiten von 40.4 Einheiten. Das sind die Kostenkurven für diese Beispiele :

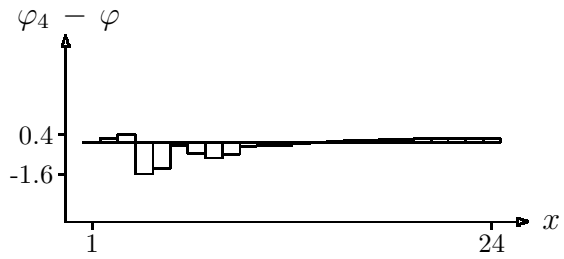
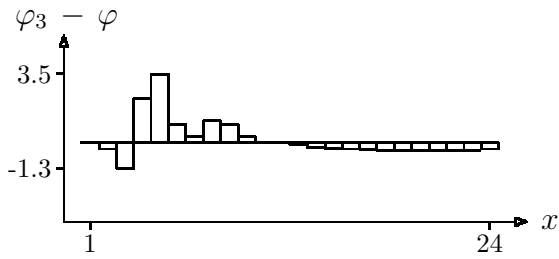


Im Fall minimaler Abhängigkeiten ähnelt die von den jeweils dritten Balken gebildete Kurve der Erfolgskurve. Für das Beispiel mit maximalen Abhängigkeiten ist sie sogar gleich der Erfolgskurve. In diesem Fall sind immer alle Teams von Änderungen betroffen und somit in jeder Phase aktiv, so daß nach Konstruktion der Kostenfunktion die Kosten in jedem Takt maximal sind. Daher treten keine Zwischenwerte in der Kostenkurve auf. Weitere Zahlen zu diesen Beispielen sind in der Tabelle weiter unten angegeben.

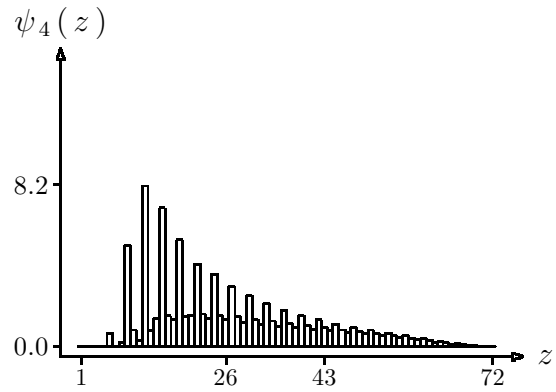
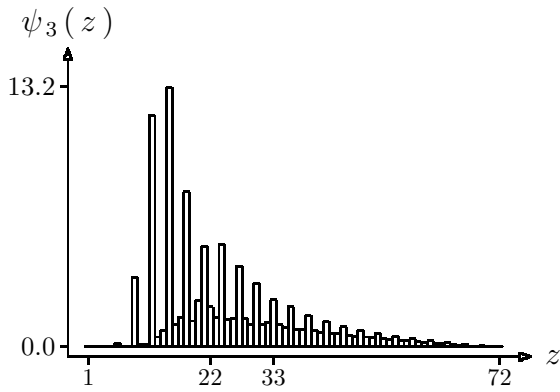
Einfluß der Langzeitverteilungen

Das Modell verhält sich wie zu erwarten, wenn die statistischen Daten der Teams verändert werden. Im nächsten Beispiel werden die Ergebniswahrscheinlichkeiten und Langzeitverteilungen des zweiten Teams gegen die aussichtsreicheren des ersten Teams ausgetauscht. Im darauffolgenden Beispiel werden stattdessen die des dritten Teams gegen die weniger aussichtsreichen des zweiten Teams ausgetauscht. Die anderen Eingangsdaten werden unverändert vom ursprünglichen Beispiel übernommen. Das ergibt diese Erfolgskurven und Differenzen:





Im ersten Fall haben sich die Aussichten verbessert, im zweiten Fall verschlechtert. Die Erfolgswahrscheinlichkeiten sind 96 % und 89 %. Die Schätzungen für die Dauer sind 9.9 und 11.8 Monate. Das sind die zugehörigen Kostenkurven :



Als Kostenschätzungen erhält man 26.8 und 32.0 Einheiten.

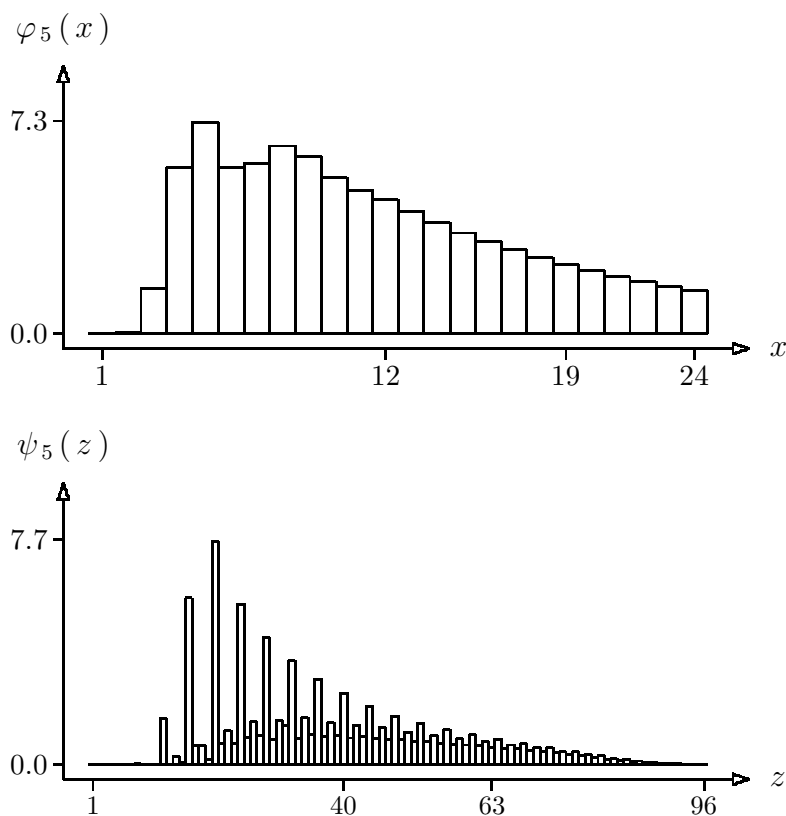
Hier sind die Zahlen für die Erfolgskurven und für die Kostenkurven der bisherigen Beispiele im Überblick :

	φ	φ_1	φ_2	φ_3	φ_4
Erfolgschance	92 %	99 %	79 %	96 %	89 %
Schätzung der Dauer	11.1	8.6	13.5	9.9	11.8
- Streuung	6.5	4.5	7.6	5.7	6.8
- Fehlerschranke	67 %	70 %	54 %	68 %	65 %
Kostenschätzung	29.9	19.2	40.4	26.8	32.0
- Streuung	18.1	8.5	22.7	15.5	19.6
- Fehlerschranke	78 %	81 %	54 %	82 %	67 %

Einfluß der Anzahl der Teams

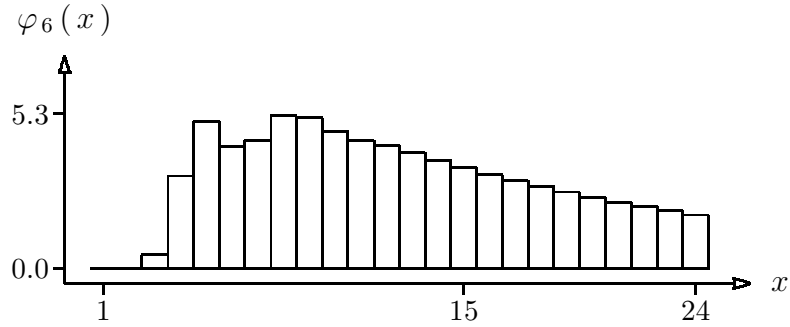
In den folgenden Beispielen wird die Anzahl der Teams schrittweise erhöht. Der Einfachheit halber haben die zusätzlichen Teams dieselben Ergebniswahrscheinlichkeiten und Langzeitverteilungen wie das erste Team. Die Tabellen mit den Abhängigkeitsgraden haben zunehmend mehr Zeilen und Spalten. Es wird angenommen, daß für jedes K die positiven Einträge in etwa gleichverteilt sind. Die übrigen Eingangsdaten sind unverändert. Insbesondere kostet jedes Team denselben Betrag pro Monat.

Für einen Entwicklungszyklus eines Projekts mit vier Teams erhält man diese Kurven:

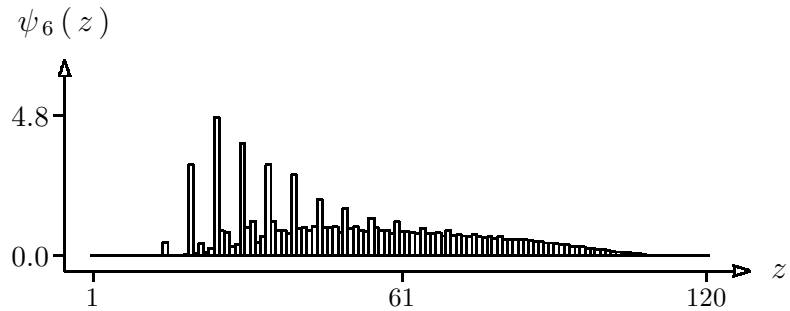


Die Form der Erfolgskurve hat sich im Vergleich zum ursprünglichen Beispiel mit drei Teams kaum geändert. Die von den Balken $4 \cdot x$ der Kostenkurve gebildete Kurve ähnelt wieder der Erfolgskurve. Die Erfolgswahrscheinlichkeit beträgt 85%. Die zu erwartende Dauer ist 13.2 Monate, die geschätzten Kosten sind 46.8 Einheiten.

Für ein Projekt mit fünf Teams erhält man diese Erfolgskurve :

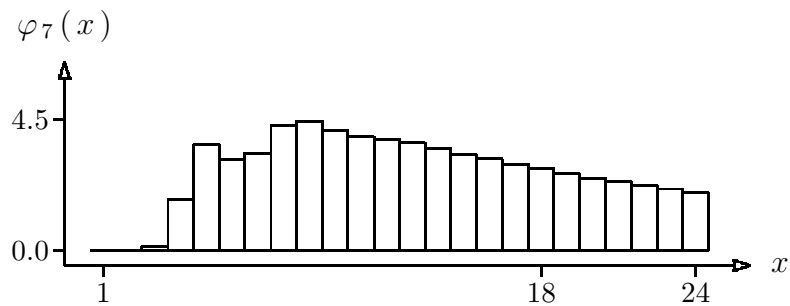


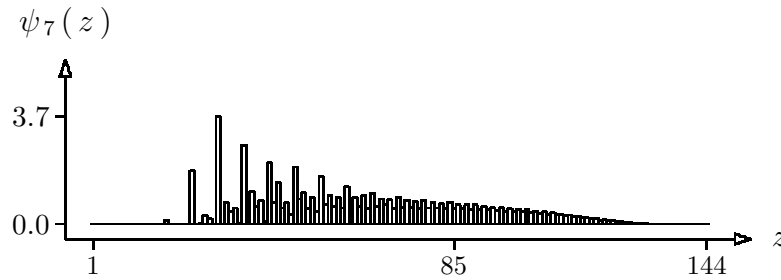
Hier ist eine Veränderung aufgetreten. Während im Beispiel mit drei Teams das zweite lokale Maximum wenig ausgeprägt war, hat sich hier der höchste Wert der Kurve vom ersten auf das zweite lokale Maximum verlagert. Der höchste Wert der Kostenkurve liegt dagegen immer noch beim ersten lokalen Maximum der Erfolgskurve :



Die Erfolgswahrscheinlichkeit beträgt knapp 75%. Die Dauer wird auf 15.3 Monate geschätzt, die Kosten auf 69.1 Einheiten.

Die Veränderung ist deutlicher zu sehen bei einem Projekt mit sechs Teams :





Die Erfolgswahrscheinlichkeit beträgt nur mehr 66%. Die geschätzte Dauer ist 16.9 Monate, die Kostenschätzung 91.7 Einheiten.

Insgesamt werden die Erfolgskurven mit zunehmender Anzahl der Teams flacher, und sie fallen langsamer gegen Null ab. Entsprechend sinken die Erfolgsaussichten und die Schätzwerte werden größer.

Rechenaufwand

Die Anzahl der Zustandsvektoren ζ steigt exponentiell mit der Anzahl N der Teams an. Daher ist zu erwarten, daß die Rechenzeiten und der Speicherbedarf für die Berechnung der Übergangswahrscheinlichkeiten und der Kurven sehr stark zunehmen.

Man kann sich vorstellen, daß die Übergangswahrscheinlichkeiten $P_\zeta(d, \eta)$ zu einer Matrix angeordnet sind. Dabei ist ζ der Zeilenindex und (d, η) der Spaltenindex. Für die Berechnung der Wahrscheinlichkeiten $P(\omega)$ genügt es, nur solche Zustände ζ zu betrachten, die nach einer oder mehreren Phasen vom gegebenen Anfangszustand $\underline{0}$ des Entwicklungszyklus aus erreichbar sind. Das spart einigen Rechenaufwand. Für das ursprüngliche Beispiel mit drei Teams muß nur etwa ein Fünftel aller Zeilen berechnet werden, für das Beispiel mit sechs Teams nur etwa ein Zwölftel. Ferner stellt sich heraus, daß für jeden Zustand ζ nur ein Teil der Übergänge (d, η) tatsächlich auftritt, so daß in den betrachteten Zeilen auch nur ein Teil der Einträge berechnet werden muß. Für das Beispiel mit sechs Teams sind das trotzdem für alle betrachteten Zeilen zusammen noch über sechs Millionen Einträge.

Die Verläufe ω , die ein Entwicklungszyklus nehmen kann, lassen sich als Wege in einem Baum auffassen. Jeder Knoten dieses Baumes ist mit einem Paar (d, η) beschriftet,

wobei sich die Beschriftungen wiederholen. Die Wurzel ist mit $(0, \underline{0})$ beschriftet. Die Kante zwischen zwei Knoten (e, ζ) und (d, η) ist mit der Übergangswahrscheinlichkeit $P_{\zeta}(d, \eta)$ und den Kosten c des Übergangs beschriftet. Für die Berechnung der Erfolgskurve und der Kostenkurve muß ein Weg in diesem Baum nur solange verfolgt werden, bis ein Knoten $(d, \underline{\infty})$ erreicht ist oder bis die Gesamtdauer des Weges die Höchstdauer x_0 überschreitet. Da eine Phase mit Problemmeldungen mindestens zwei Takte dauert, wird ein Weg nach spätestens $\lfloor \frac{x_0+1}{2} \rfloor$ Phasen abgebrochen. Je näher man dieser Grenze kommt, desto weniger Kanten im Baum müssen noch weiter verfolgt werden. Das verringert den Rechenaufwand beim Durchlaufen des Baumes erheblich.

Die folgende Tabelle enthält einige gemessene Daten über den Rechenaufwand. Ausgewertet wurden das ursprüngliche und die letzten Beispiele. Gemessen wurden die Anzahl der berechneten Zeilen der Matrix, die Anzahl der insgesamt für diese Zeilen berechneten Einträge, sowie die Rechenzeiten. Die Zeiten sind als Minuten: Sekunden angegeben.

Teams	3	4	5	6
Zeilen	74	396	867	8772
Einträge	3768	48 120	259 086	6 225 408
Rechenzeit – Matrix	0:01	0:06	2:30	119:10
Rechenzeit – Kurven	0:02	0:21	2:27	117:28

Die Zeiten wurden auf einem Intel Pentium II 266 gemessen. Alle Wahrscheinlichkeiten wurden als `long double` abgespeichert. Der Rechner hatte so viel Hauptspeicher, daß darin alle Daten gehalten werden konnten, so daß die Werte nicht durch Paging beeinflußt sind. Für das Beispiel mit sechs Teams wurde fast ein halbes Gigabyte Hauptspeicher benötigt.

Tabellen

x	φ	φ_1	φ_2	φ_3	φ_4	φ_5	φ_6	φ_7
1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	0.47	0.47	0.47	0.15	0.71	0.04	0.00	0.00
3	4.78	5.08	4.65	3.50	5.19	1.54	0.48	0.15
4	9.13	11.02	8.14	11.36	7.55	5.68	3.17	1.76
5	9.18	12.66	7.00	12.64	7.89	7.22	5.05	3.62
6	6.97	9.78	5.04	7.91	6.82	5.72	4.19	3.11
7	7.02	9.98	4.89	7.35	6.47	5.84	4.38	3.36
8	7.04	9.61	4.95	8.17	6.26	6.43	5.22	4.28
9	6.30	8.19	4.54	7.20	5.73	6.07	5.17	4.45
10	5.40	6.46	4.14	5.70	5.20	5.37	4.71	4.14
11	4.86	5.44	3.87	4.91	4.69	4.92	4.40	3.92
12	4.39	4.46	3.62	4.41	4.25	4.59	4.22	3.83
13	3.90	3.56	3.37	3.81	3.84	4.20	3.98	3.69
14	3.44	2.79	3.13	3.21	3.47	3.81	3.71	3.49
15	3.06	2.25	2.91	2.76	3.13	3.47	3.45	3.29
16	2.72	1.78	2.71	2.39	2.82	3.16	3.23	3.13
17	2.41	1.39	2.53	2.06	2.55	2.88	3.02	2.96
18	2.14	1.09	2.35	1.76	2.30	2.62	2.81	2.80
19	1.90	0.86	2.19	1.51	2.07	2.38	2.62	2.65
20	1.68	0.68	2.04	1.30	1.87	2.17	2.45	2.50
21	1.49	0.53	1.90	1.12	1.69	1.97	2.28	2.36
22	1.32	0.41	1.77	0.96	1.52	1.80	2.13	2.23
23	1.18	0.32	1.64	0.82	1.37	1.63	1.98	2.11
24	1.04	0.25	1.53	0.71	1.24	1.49	1.85	1.99

z	ψ	ψ_1	ψ_2	ψ_3	ψ_4	ψ_5	ψ_6	ψ_7
1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
6	0.47	0.47	0.47	0.15	0.71	0.00	0.00	0.00
7	0.02	0.06	0.00	0.00	0.04	0.00	0.00	0.00
8	0.10	0.36	0.00	0.01	0.21	0.04	0.00	0.00
9	4.76	5.03	4.65	3.52	5.13	0.00	0.00	0.00
10	0.43	1.44	0.00	0.12	0.84	0.01	0.00	0.00
11	0.23	1.01	0.00	0.11	0.32	0.01	0.00	0.00
12	9.57	12.08	8.14	11.71	8.15	1.57	0.00	0.00
13	0.71	2.87	0.00	0.49	0.82	0.01	0.00	0.00
14	1.14	3.63	0.00	0.83	1.44	0.28	0.00	0.00
15	9.05	12.74	7.00	13.13	7.04	0.06	0.48	0.00
16	1.42	4.66	0.00	1.13	1.56	5.70	0.01	0.00
17	1.38	4.30	0.00	1.46	1.37	0.65	0.00	0.00
18	6.33	8.31	5.04	7.84	5.43	0.65	0.04	0.15
19	1.33	4.04	0.00	1.31	1.52	0.18	0.07	0.00
20	2.01	4.58	0.00	2.35	1.57	7.65	3.13	0.01
21	4.56	4.92	4.89	5.08	4.20	0.74	0.09	0.01
22	1.85	3.77	0.00	2.01	1.66	1.15	0.43	0.02
23	1.40	3.16	0.00	1.47	1.42	0.74	0.12	0.00
24	4.32	3.72	4.95	5.18	3.68	5.50	0.28	1.83

z	ψ	ψ_1	ψ_2	ψ_3	ψ_4	ψ_5	ψ_6	ψ_7
25	1.44	2.52	0.00	1.41	1.57	0.94	4.74	0.00
26	1.45	2.19	0.00	1.42	1.35	1.49	0.88	0.05
27	3.42	2.39	4.54	4.07	3.06	1.01	0.83	0.29
28	1.53	1.71	0.00	1.45	1.51	4.35	0.35	0.18
29	1.20	1.41	0.00	1.15	1.23	0.86	0.42	0.01
30	2.90	1.50	4.14	3.19	2.61	1.49	3.86	3.69
31	1.28	1.06	0.00	1.15	1.39	1.33	0.99	0.02
32	1.23	0.92	0.00	1.20	1.15	3.56	1.21	0.74
33	2.29	0.84	3.87	2.37	2.19	0.90	0.49	0.46
34	1.25	0.64	0.00	1.11	1.29	1.62	0.67	0.55
35	1.04	0.51	0.00	0.97	1.05	1.05	3.11	0.04
36	1.97	0.48	3.62	2.00	1.86	2.93	1.21	2.72
37	1.07	0.34	0.00	0.90	1.18	0.98	0.88	0.54
38	0.98	0.28	0.00	0.88	0.97	1.45	0.88	1.14
39	1.61	0.25	3.37	1.58	1.58	0.99	0.80	0.63
40	0.99	0.18	0.00	0.82	1.07	2.44	2.78	0.81
41	0.84	0.14	0.00	0.73	0.88	0.92	0.95	0.10
42	1.37	0.13	3.13	1.29	1.35	1.35	0.99	2.11
43	0.86	0.09	0.00	0.68	0.96	0.96	0.89	0.74
44	0.77	0.07	0.00	0.65	0.80	1.98	1.02	1.44
45	1.13	0.06	2.91	1.02	1.16	0.87	1.95	0.55
46	0.77	0.04	0.00	0.60	0.87	1.26	0.99	0.74
47	0.67	0.03	0.00	0.54	0.72	0.86	0.99	0.35
48	0.96	0.03	2.71	0.83	0.99	1.67	1.03	1.97

z	ψ	ψ_1	ψ_2	ψ_3	ψ_4	ψ_5	ψ_6	ψ_7
49	0.66	0.02	0.00	0.49	0.77	0.84	0.82	0.87
50	0.59	0.02	0.00	0.47	0.64	1.12	1.62	1.09
51	0.79	0.01	2.53	0.66	0.84	0.79	0.95	0.54
52	0.58	0.01	0.00	0.42	0.67	1.42	1.03	0.92
53	0.50	0.01	0.00	0.38	0.56	0.77	0.90	0.46
54	0.65	0.01	2.35	0.52	0.70	1.00	0.84	1.63
55	0.47	0.00	0.00	0.33	0.56	0.73	1.31	0.68
56	0.41	0.00	0.00	0.30	0.46	1.20	0.97	0.99
57	0.51	0.00	2.19	0.40	0.57	0.71	0.87	0.61
58	0.38	0.00	0.00	0.25	0.45	0.90	0.90	0.92
59	0.31	0.00	0.00	0.22	0.36	0.66	0.79	0.58
60	0.38	0.00	2.04	0.29	0.43	1.02	1.17	1.29
61	0.27	0.00	0.00	0.18	0.33	0.65	0.84	0.60
62	0.22	0.00	0.00	0.15	0.25	0.79	0.84	0.92
63	0.26	0.00	1.90	0.19	0.29	0.60	0.83	0.63
64	0.17	0.00	0.00	0.11	0.21	0.87	0.79	0.97
65	0.12	0.00	0.00	0.08	0.14	0.57	0.96	0.54
66	0.16	0.00	1.77	0.12	0.17	0.69	0.80	1.05
67	0.08	0.00	0.00	0.05	0.10	0.53	0.77	0.60
68	0.05	0.00	0.00	0.04	0.05	0.72	0.80	0.88
69	0.08	0.00	1.64	0.06	0.08	0.50	0.70	0.65
70	0.03	0.00	0.00	0.02	0.04	0.58	0.86	0.84
71	0.00	0.00	0.00	0.00	0.00	0.44	0.73	0.53
72	0.04	0.00	1.53	0.03	0.03	0.58	0.73	0.94

z	ψ_5	ψ_6	ψ_7	z	ψ_6	ψ_7	z	ψ_7
73	0.41	0.71	0.59	97	0.28	0.47	121	0.13
74	0.46	0.67	0.82	98	0.27	0.54	122	0.13
75	0.35	0.75	0.60	99	0.23	0.46	123	0.10
76	0.45	0.68	0.77	100	0.22	0.52	124	0.10
77	0.30	0.65	0.54	101	0.19	0.43	125	0.07
78	0.34	0.65	0.82	102	0.17	0.50	126	0.08
79	0.24	0.60	0.58	103	0.15	0.41	127	0.05
80	0.31	0.67	0.74	104	0.13	0.46	128	0.05
81	0.20	0.60	0.56	105	0.11	0.39	129	0.04
82	0.22	0.59	0.71	106	0.09	0.43	130	0.04
83	0.14	0.58	0.53	107	0.07	0.36	131	0.02
84	0.18	0.55	0.75	108	0.07	0.40	132	0.02
85	0.11	0.57	0.54	109	0.05	0.33	133	0.01
86	0.11	0.53	0.67	110	0.04	0.36	134	0.01
87	0.07	0.50	0.54	111	0.03	0.30	135	0.01
88	0.09	0.50	0.65	112	0.02	0.32	136	0.01
89	0.04	0.45	0.51	113	0.02	0.26	137	0.00
90	0.04	0.47	0.66	114	0.01	0.29	138	0.00
91	0.02	0.43	0.50	115	0.01	0.23	139	0.00
92	0.03	0.41	0.61	116	0.01	0.24	140	0.00
93	0.01	0.38	0.50	117	0.00	0.20	141	0.00
94	0.01	0.35	0.59	118	0.00	0.21	142	0.00
95	0.00	0.34	0.47	119	0.00	0.16	143	0.00
96	0.01	0.32	0.58	120	0.00	0.17	144	0.00

Literatur

- [1] Billingsley: *Probability and Measure*, Wiley 1995
- [2] Dowson (editor): *Iteration in the Software Process*, IEEE Computer Press 1987, 3rd International Software Process Workshop (1986)
- [3] Fairley: *Software Engineering Concepts*, McGraw-Hill 1985
- [4] Feller: *An Introduction to Probability Theory and Its Applications*, Wiley 1968
- [5] Gray, MacDonell: “ A Comparison of Techniques for Developing Predictive Models of Software Metrics “,
Information and Software Technology 39 (1997) 425 - 437
- [6] Kemerer: “ An Empirical Validation of Software Cost Estimation Models “,
Communications of the ACM 30 (1987) 416 - 429
- [7] Kemerer: “ Reliability of Function Point Measurements: A Field Experiment “,
Communications of the ACM 36 (1993) 85 - 97
- [8] Kitchenham, Taylor: “ Software Project Development Cost Estimation “,
Journal of Systems and Software 5 (1985) 267 - 278
- [9] Londeix: *Cost Estimation for Software Development*, Addison-Wesley 1987
- [10] Miyazaki, Mori: “ Cocomo Evaluation and Tailoring “,
8th International Conference on Software Engineering (1985) 292 - 299
- [11] Parzen: *Modern Probability Theory and Its Applications*, Wiley 1960
- [12] Pressman: *Software Engineering*, McGraw-Hill 1992
- [13] Sallis, Tate, MacDonell: *Software Engineering*, Addison-Wesley 1995
- [14] Shepperd, Schofield, Kitchenham: “ Effort Estimation Using Analogy “,
18th International Conference on Software Engineering (1996) 170 - 178

- [15] Sirjaev: *Wahrscheinlichkeit*, Deutscher Verlag der Wissenschaften 1988
- [16] Sommerville: *Software Engineering*, Addison-Wesley 1996
- [17] Srinivasan, Fisher: “ Machine Learning Approaches to Estimating Software Development Effort “,
IEEE Transactions on Software Engineering 21 (1995) 126 - 137
- [18] Wittig, Finnie: “ Using Artificial Neural Networks and Function Points to Estimate 4GL Software Development Effort “,
Australian Journal of Information Systems 1 (1994) 87 - 94

Summary

A probabilistic model for software development projects is constructed. The model provides formulas for computing the probabilities of the courses that a project could take. Using these probabilities, the chances of succeeding with a given amount of time and money can be computed. In addition, formulas are provided to compute estimates for the development time and for the cost of a project. Error bounds for the estimates can be obtained, too. The input data required for the model are statistical data on the courses of past projects and data about the high-level design of the software product being developed. Numerical examples are provided to show that the model behaves as expected with respect to changes of the input data.

Ich konstruiere ein wahrscheinlichkeitstheoretisches Modell für Softwareprojekte. Das Modell enthält Formeln, mit denen sich die Wahrscheinlichkeiten der Projektverläufe berechnen lassen. Davon ausgehend können die Erfolgsaussichten für ein Projekt berechnet werden, also die Wahrscheinlichkeit dafür, das Projekt in einem vorgegebenen Zeitraum und mit einer vorgegebenen Kostengrenze abzuschließen. Außerdem gebe ich Formeln an, mit denen sich Schätzwerte für die Dauer und für die Kosten eines Projekts berechnen lassen. Die Schwankungsbreiten für die Schätzungen lassen sich ebenfalls berechnen. Für das Modell werden statistische Daten über den Verlauf bereits abgeschlossener Projekte und Daten über den Entwurf des zu erstellenden Softwareprodukts als Parameter benötigt. An Zahlenbeispielen zeigt sich, daß sich das Modell wie zu erwarten verhält, wenn die Eingangsdaten verändert werden.