Proposal of probability risk Evaluation for System Development Project Based on Requirements Analysis and Bayesian estimation

Shinji Mochida
Faculty of Commerce
University of Marketing and Distribution Sciences
mochida@red.umds.ac.jp

Abstract

All companies employ business systems during the development of new software, mechanical equipment or other services. Additionally In most business system development projects are planned in order to reduce amount of running cost or increase benefit. Thus More efficient project management is needed in order to meet schedules and reduce cost. Although there are several stages in business system development projects. There are some factors that cause over cost or schedule delay of projects. Especially finesse in estimating for requirement from customer is the most essential in project managements. Although there are subjective factors to evaluate requirements. Thus evaluating scientifically is needed using requirements analysis and Bayesian estimation in project management.

1 Introduction

Business System development projects are challenging in that are many requirements demanded from customers even while these requirements are proposed with the same priority. Thus it is important to narrow down and prioritize requirements according to their essentiality and criticality to finish on schedule. Although system developers estimate according to the complexity of projects [2], but customers expect the cost to be based on the number of requirements they propose. Then customers and system developers estimate differently, there are often **conflicting estimates**. Thus, this paper proposes cost share rate for business system development projects based on requirements analysis in order to estimate accurately. Cost share rate is defined as the percentage of total cost assigned to each requirement. Cost share rate could distinguish essential requirements. And requirements

that has large cost share rate must have large risk, also should be under strict control. Because big change or modification for essential requirement give large impact to costs or schedule. Then this paper aim to propose a method to identify essential requirements in order to estimate accurately. Then risks are assigned to each requirements with cost share rate and probability. This research show methods in following steps; (see Figure 1). First, this research predict cost for requirement version one of past project. Next this research compares with predicted estimate(version four) and total cost at completion of the past project. Additionally there is essential point estimating is subjective. Thus this paper propose the method how to predict cost share rate accurately using Bayesian estimation.

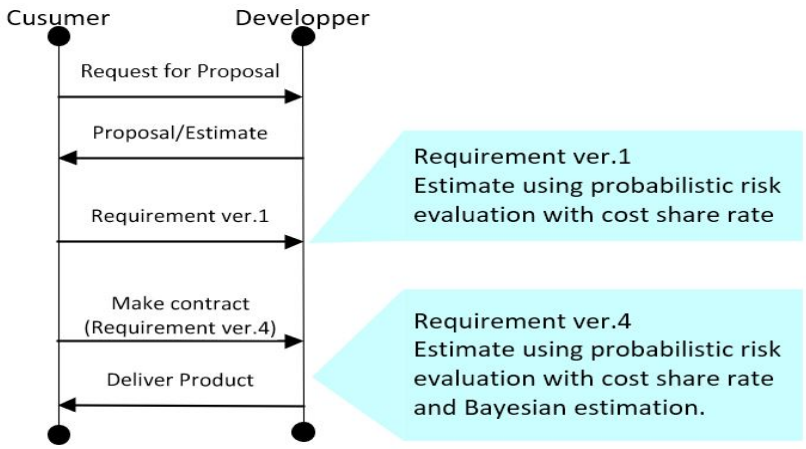

Figure 1: Steps for probabilistic estimation with cost share rate

2 Previous Studies

One purpose of this paper is showing that risk management could contribute reducing cost of business system development project. Additionally this paper demonstrates the potential to evaluate risk by requirements analysis for business system development project management with Bayesian estimation. This research is not focused on estimating costs based on the method to measure amount of source code of system, but rather allocating costs to each requirement. Previous research typically focused on either schedule, cost estimation or productivity. Improving productivity contribute to finish project fine[3]. About changing requirement in mechanical engineering design, one example of requirements analysis research explored the ability of predicting requirements change through graphical models of the requirements documents and historical change trends[4]. There are plus and minus risks in system development project. although there are no research that

refers subjective factor. This propose the method take into subjective factor in account with Bayesian estimation.

3 Project management and Risk management

3.1 Project management and Risk management

In project management there are two important methods[5]. One is schedule management, and another one is risk management. On the other hand Equation 1 shows there are controllable factors or uncontrollable factors in business. Sale is uncontrollable factor, because sale come under the influence of markets, customers. Although cost is controllable factor, cost includes payment, material costs, for example, payment for staff could be cut off by manager. It is important to identify which factor is controllable and uncontrollable. And how appropriately controllable factor could be controlled. Also there is possibility uncontrollable factor could be controlled with Bayesian analyses. Usually extra budget is settled aside for refinement or fixing trouble in project management. This extra budget is called for contingency budget or only contingency. If risk management would work well, contingency budget would not be used, then contingency budget would come to benefit. At the result prospect of profit would increase. Thus risk management could contribute reducing contingency cost,And risk management has potential to increase contingency profit.

$$Gaining = Sale - Cost \tag{1}$$

3.2 Risk analysis

Risk is defined as factors that make uncertain when they will achieve their objectives under ISO31000. Usually risk Analysis is started from risk identification in risk management. Then, risk evaluation is considered by qualitative evaluation and quantitative evaluation. Thus quantitative risk analysis is calculated by possibility $\times$ cost. This calculated risk(cost) should be spend ,if risks comes up. It is called expected monetary value. Risks is evaluated by expected monetary value Equation 2. And risks are prioritized by the order of expected monetary value. In quantitative evaluation for risk management usually probability is given subjectively by staffs subjectively. Or probability is given by experts, Delphi method or questionnaire for skilled staffs. Thus getting accurate probability is very essential to cucullate expected monetary value correctly. Thus this paper takes into account β distribution to calculate probability, and expect monetary value. This paper propose conditional possibility in order to cucullate risk correctly. It is very essential point in risk management there are plus and minus risks, in addition probability is subjective; (see Figure 2).In this research conditional possibility is given to essential

risks. In many project risk comes up by misunderstanding requirements or defects in requirements from clients in business system development project. It is vital to get certain requirements and predict risk in requirements properly.

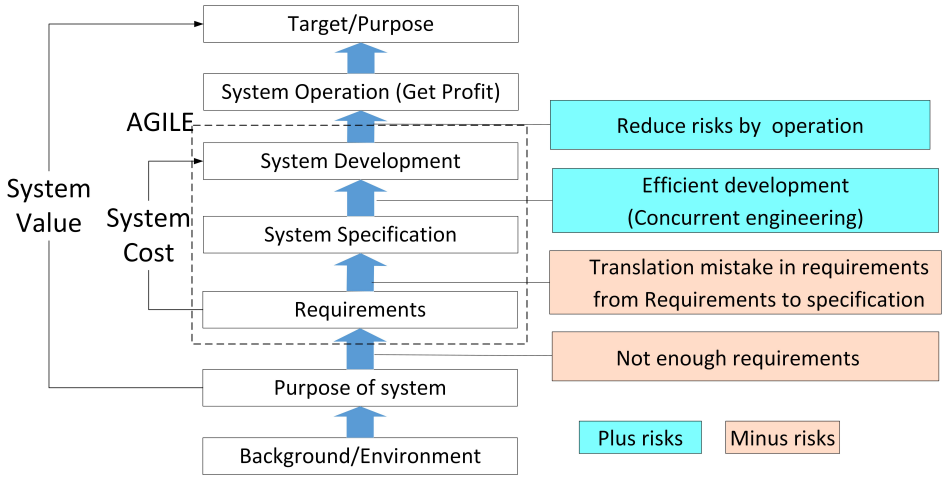

Figure 2: Plus and Minus risks in system development project

$$Risk(Expected\ Monetary\ Cost) = Probability \times Cost \quad (2)$$

4 Cost Prediction Methods for Business System Development

4.1 Cost Prediction Methods

Proper estimate is essential to finish projects on schedule and under budget. Over cost or schedule delay is caused by missing estimate. Usually amount of program source code is predicted by some prediction method in order to estimate in business system development project. Then amount of program source code is converted into base monetary cost. Next total cost is made by adding contingency cost to base monetary cost. It is finally budget for project. Then this section explains some current methods to predict costs of business system development projects. There is typical methods to estimate for business system, typical methods;COCOMO method and Function Point method. Both methods predict costs for business

system. In case of COCOMO method, it estimate by amount of program source code. On the other hand In case of Function Point method it accumulates points according to the complexity of system;the number of db tables, dialog boxes, print forms and interfaces. Acquired points could be converted into cost.

4.2 COCOMO Method

COCOMO Method[6] estimate the duration (Person-Months) in system development projects. In the COCOMO method, volume of source code is estimated by Equation 3. Duration (Person-Months) could be calculated with dividing volume of source code by the number of staffs. COCOMO Method propose Equation 3. And it uses the parameters as follows: C_0,C_e, P_1, P_2, and P_3.

C_e: estimate duration (PM:Person-Months) for expectation
C_0: estimate volume of source code
P_1: parameter for estimated productivity
P_2: exponent parameter for software development
P_3: calibration parameter
A challenge with this method is the parameters used for the cost estimation method are empirically derived and contextually dependent on many different factors, such as team size, project complexity, cultural environment, and others.

$$C_e = (C_0 \times P_1)^{P_2} \times P_3 \tag{3}$$

4.3 Function Point Method

Function Point Method estimate duration (PM:Person-Months) as those of COCOMO methods [5][6]. In the Function point method it is necessary to count the number of internal and external files, tables and internal and external interfaces. Function Point Method propose equation 4.And it uses the parameters as follows: C_e,F_p, F_1, P_1, and P_2.

C_e: estimate duration (PM:Person-Months) for expectation
F_p: estimate function points
F_1: function points
P_1: parameter for estimated productivity
P_2: calibration parameter

$$C_e = (F_p \times P_1), \quad F_p = F_1 \times P_2 \tag{4}$$

4.4 Other Methods to Estimate

Additionally, there is another cost estimate method as Experience method. In the experience method, total cost is estimated based on previous experiences. In these methods, there are often gaps between system developers' cost estimates and customer expectations. This results from differences in how developers and customers group costs. Estimating correctly is important to finish building system on schedule and under budget. Thus, translating requirements into factors to estimate is essential. Certain requirements are needed to estimate properly. But there is no cost estimate method taking certainty of requirements into account. This paper considers taking certainty of requirements into account to estimate in order to gain customer agreement.

5 Requirement Analysis using Linguistic analysis

	Req01	Req02	Req03	Req04	Req05	Req13	Req14	Req15	Req16	Req17	Req18	Req19	Req20	Req21	Req22	Sum
Req01	0	6	3	2	2	1	3	3	1	0	2	4	1	0	0	38
Req02	6	0	3	2	3	1	2	1	1	0	3	3	1	0	0	35
Req03	3	3	0	20	10	0	4	0	1	1	3	8	15	0	1	81
Req04	2	2	20	0	11	0	4	0	0	1	4	8	15	0	1	76
Req05	2	3	10	11	0	1	1	0	0	2	7	16	18	0	1	78
Req06	3	4	8	6	3	0	1	0	1	1	2	2	5	0	0	40
Req07	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Req08	1	1	1	1	1	1	0	0	0	0	1	2	2	2	0	18
Req09	2	2	2	0	0	1	1	0	1	0	0	1	1	2	0	19
Req10	4	2	0	0	0	1	1	4	0	0	0	3	0	0	0	15
Req17	0	0	1	1	2	0	0	0	0	0	7	7	1	0	0	20
Req18	2	3	3	4	7	0	1	0	0	7	0	13	5	0	1	50
Req19	4	3	8	8	16	2	1	2	1	7	13	0	10	0	1	84
Req20	1	1	15	15	18	1	1	0	0	1	5	10	0	1	1	79
Req21	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	6
Req22	0	0	1	1	1	0	0	0	0	0	1	1	1	0	0	6
Sum.	38	35	81	76	78	10	22	11	7	20	50	84	79	6	6	702

Figure 3: Count over lapping keywords from requirement version four

Taking correct requirements is essential to estimate properly. And over cost or schedule delay is caused by missing evaluation of requirements. Otherwise over cost or schedule delay is caused by many remediation of requirements. Remediation for essential requirement and uncertain requirement has large risk.thus It is vital to distinguish the requirement which gives large impacts to specification or budget of project. This research propose the method to distinguish influential requirements that has large risks. This paper shows a method to distinguish an influential requirements with with linguistic analysis and cost share rate. Cost share rate is defined as the percentage of total cost assigned to each requirement. This paper analyzes the requirements that were requested in past small system development projects by linguistic analysis. This small project is building a knowledge collecting system. In this project requirements were revised four times. Thus this paper analyzes requirement version one and version four. This paper predict risks from

the result from analyzing requirement version one. Risk is considered as cost in risk management. This results; acquired cost is compared with the actual cost at

	Req01	Req02	Req03	Req20	Req21	Req22
1	system	system	When	following	simultaneou	Schedule
2	achieves	reads	equipment	screen	connected	screen
3	taking	tag	passes	installed	number	schedule
4	out	pasted	gate	as	clients	function
5	equipment	taking	direction	display	PC	
6	efficiency	out	taking	online	assumed	
7	improvemer	worker	out	monitor	50	
8	return	equipment	image	history		
9	managemer	noncontact	vicinity	inspection		
37				assumed		
38				one		
39				renew		
40				still		
41				picture		
42				regularly		

Figure 4: Extract keywords from requirement version four

the completion of project. If risk for requirement could be predicted properly, it would contribute project management. In this research, overlapping keywords are extracted form each requirements with linguistic analysis. Overlapping keywords are words that appear in one requirement and groups of keywords that appear in each category or phase in system development project. Categories or phases are Design, Development, Print, Test, Interface and Document. Overlapping keywords indicates relationships between one requirement and each other. The number of relationships that each requirement has with other requirements indicate essentiality and importance. Steps of linguistic analysis are as follows:
(1) Extract keywords from each requirement in version four. (see Figure 4).

(2) Count overlapping key words from extracted keywords, also count overlapping key words from extracted keywords in version four (see Figure 3).

(3)Distinguish essential requirements by counting the number of over lapping keywords and subjective cost share rate from three engineers, and measure the distance from most essential requirement to each requirement in version four (see Figure 5).

(4)Distinguish essentiality of each requirements from these results, and distinguish categories that each requirements belong in version one.

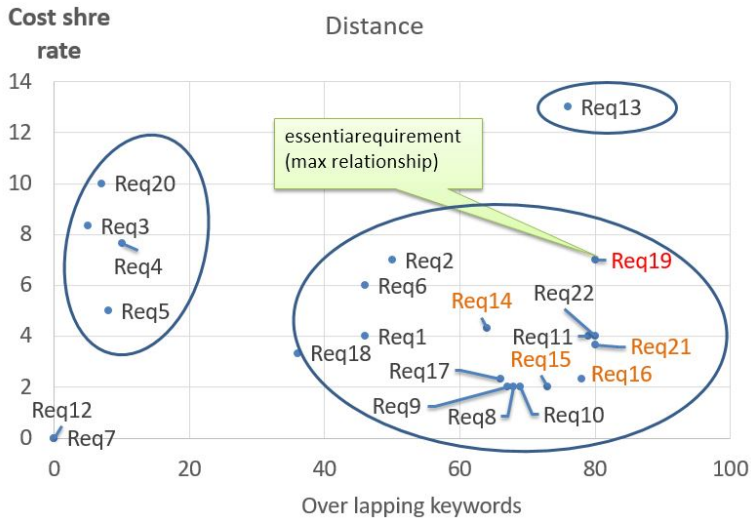

Figure 5: Specify important requirement from requirement version four

6 Prediction of Cost Share Rate

Usually costs for system development projects are estimated by grouping costs with the number of dialog boxes, interfaces or print forms . Alternatively, costs may be estimated by associating cost to logic design, development, test, adjustment and documents; but, not according to the requirements .System developers estimate according to the complexity of projects, but customers expect the cost according to the number of requirements. Thus, customers could not understand the estimates provided by system developers. This paper shows a method to calculate cost share rate for each requirement in order to evaluate requirements accurately with mutual understanding of the developer and customers. Also cost share rate indicates importance of each requirement. Cost share rate is defined as the percentage of total cost assigned to each requirement. Figure 6 sows cost share rate of version our of past project. Although in this research cost rate(%) for each requirements are subjective figure from the two staffs that worked on this system development project. And cost share rates are gained by multiply cost rate by cost under the estimate. This cost rate and estimated cost are the value that according to the category.

7 Probability Prediction of schedule delay

This paper aim to predict risk of system development project based on requirements analysis. This paper consider one of risk of system development project is

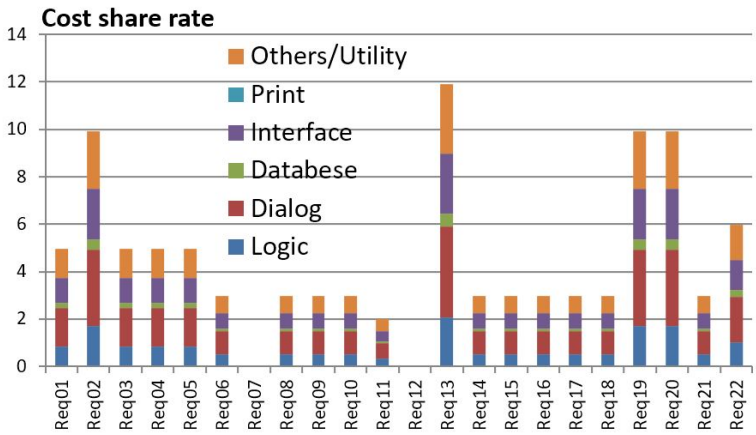

Figure 6: Cost share rate of requirement version four

schedule delay. Thus this research suppose probability of schedule delay follows β distribution (see Equation 5), Sample data in Table 1 are surveyed in past system development project; Rdf system for tool tracking in machine factory. Then parameters (see Table 2) are gained by curve fitting sample data into β distribution. In this analyzing process x is probability parameter that indicates start day for each task, and y is ratio for schedule delay against actual days. Table 1 shows that survey/preparation, design and programming process have risk of schedule delay. Although test and writing document process have no risk of schedule delay. Figure 7 shows β distribution curve in this case from parameters (see Table 2).

$$f(x) = c \times x^{\alpha\ -1}(1-x)^{\beta\ -1} \tag{5}$$

Table 1: Sample data from past project

Item	Survey/	Design			Programing			Test		
(days)	Prepare	Dia.	Fun	DB	Dia.	Fun	DB	Test	Doc.	Sum
Schejule	11	29	27	27	9	58	39	3	13	216
Actual	35	34	36	27	9	55	62	2	7	267
Delay	24	5	9	0	0	-3	23	-1	-6	51
Prog	0	0.52	0.6	0.59	0.7	0.74	0.59	0.95	0.93	

Prog.:Progress Rate Fun:function Doc.:Document Dia.:Dialog

Table 2: Acquire Parameter for β distribution

C	α	β
0.839835	1.020625	3.047617

8 Risk Prediction for business system development project

Figure 7 shows that probability of schedule delay is 0.225. Thus this research

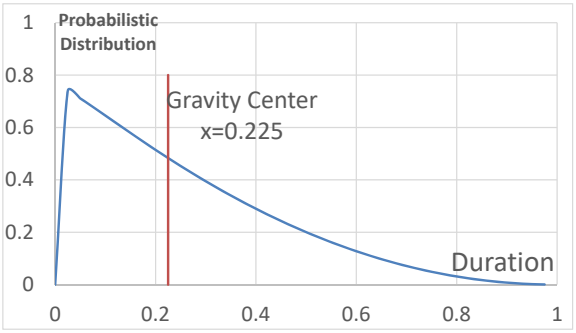

Figure 7: Beta distribution from past project

assign probability of risk;0.225 to essential requirements that distinguished by requirements analysis. Actually Table 3 shows that requirement3 and requirement4 are about design and development, requirement6 is about development. These requirements are essential, and assigned probability of risk as 0.225. Requirement1 is about Main Design, and assigned probability of risk as 0.1,because in past three project contingency is set as10% (see Table 4). Contingency is set for refinement. Refinment contains deleting, adding or refactoring. Table 4 shows occupying cost rate for each work process from surveying of past system development project; medical record system, knowledge management system. Thus risk is calculated probability of risk $\times$ cost (see equation 2),cost is analyzed by cost share rate. In estimated cost of requirement version one estimated total cost is 109 (see Table 3).And at completion actual total cost is 109.1 (see Table 3). At the result these results are equal.

9 Bayesian estimation

First of all, costs is estimated in the project management. Also accuracy of estimation is needed. However, estimate isn't accurate usually.

Table 3: Total cost at completion

Aitem	Attr	Cost Share Rate(%)	Schdule delay Probability	Conditional Probability	Monitary Risk	Estimated Cost
Req1	M	22.5	0	0.1	2.25	24.75
Req2		25	0	0	0	25
Req3	D/De	5	0.225	0	1.125	6.125
Req4	D/De	7.5	0.225	0	1.6875	9.1875
Req5		22.5	0	0	0	22.5
Req6	De	17.5	0.225	0	3.9375	21.4375
Toatal		100			9	109

Req:Requirement M:Main, D:design, De:Development Attr:Attribute

The reason is that there are subjective factors, and the accuracy of the requirements is low. Thus, this paper propose the method to estimate of the project-properly using Bayesian estimation(expression 6) and cost share rate proposed in before section. The example of the adjustment to the cost estimate of past actual projects is shown (see Table 4) This table shows there are over cost as 10% in almost

Table 4: Data from past project

Project	Design	Development	Testing	Document	Refinement	Total
P1	3	4	1	1	1	10
P2	1	6	1	1	1	10
P3	1	6	1	1	1	10
Average	1.7	5.3	1	1	1	10

project. There is subjective factor to estimate, and there is plus and minaus lisks. Also there is inaccuracy in requirements. Figure 8 shows there is valiance of cost share rate according to subjective factors and inaccuracy in requirements .Thus data from past project(Table 5) are inputted into Bayesian estimation(expression 7)[6]. Thus The result was 1.07 is gained. It is important point plus risks are estimated as 1.0. Plus risk are risks that have minus difference. This result shows only minus risks influence cost or estimate. Also this results shows possibility estimate would be gained accurately with Bayesian estimation.

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)} \tag{6}$$

$$P(Obud|Est) = \frac{P(Est|Obud)P(Obud)}{P(Obud)P(Est|Obud) + P(Obud)P(Est|Obud)} \quad (7)$$

$Obud : Proceedon\ budget\ Actually$

$Est : Estimation$

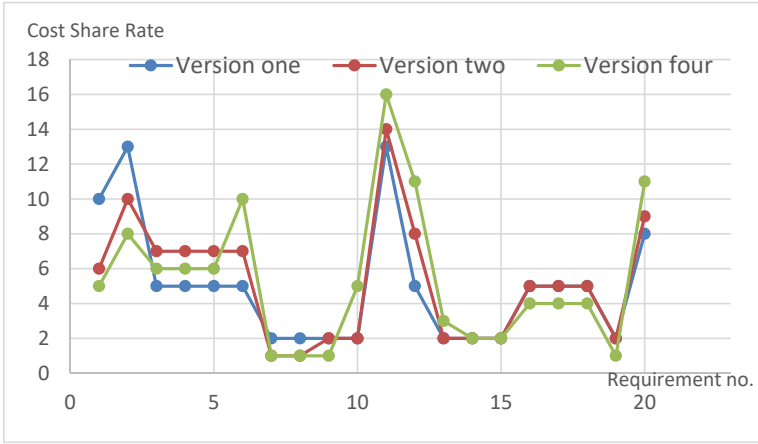

Figure 8: Variation of each cost share rate and revise

10 Conclusion

Meeting budget, finishing on schedule, and maintaining high quality are all important in project management. If cost and duration would be gained accurately, meeting budget and finishing on schedule would be achieved and managing project well. Estimate miss by developer causes finally over cost and schedule delay. Subjective factor and inaccuracy in requirements are reasons why estimate miss is caused. Generally estimate is measured by amount of source code of system or complexity of system. Also it is not estimated based on essentiality or risk of requirements. Misunderstanding of requirements and subjective factors cause misses in estimate or schedule. Then, this paper proposes cost share rate to measure essentiality of requirements in order to estimate accurately. Cost share rate is defined as the percentage of total cost assigned to each requirement Usually on risk management risks are evaluated according to staffs' experience. Also risks are not evaluated based requirement analysis. Evaluating risk properly is needed in order to manage project well. Thus this paper show better results by using cost share

rate from requirement analysis. Additionally, this method helps prioritize requirements and narrow down specifications of the project. Prioritizing requirements and narrowing down specifications accurately help ensure it meets budget and duration targets. Additionally, this paper shows possibility to obtaining probability for risks accurately with cost share rate and Bayesian estimation. However, this result was obtained by small case. Thus further research and study is needed to refine and improve this method to obtain cost share rate and risk more accurately.

Table 5: Bayesian estimation and actual cost by cost share rate

Aitem	First	Middle	Final	Difference	Accuracy	Bayesian estimation	Correction value
Req1	10	6	5	5	0.95	0.99	5.03
Req2	13	10	8	5	0.95	0.99	8.05
Req3	5	7	6	-1	1	1.00	6.04
Req4	5	7	6	-1	1	1.00	6.04
Req5	5	7	6	-1	1	1.00	6.04
Req6	5	7	10	-5	1	1.00	10.06
Req8	2	1	1	1	0.99	1.00	1.01
Req9	2	1	1	1	0.99	1.00	1.01
Req10	2	2	1	1	0.99	1.00	1.01
Req11	2	2	5	-3	1	1.00	5.03
Req13	13	14	16	-3	1	1.00	16.09
Req14	5	8	11	-6	1	1.00	11.06
Req15	2	2	3	-1	1	1.00	3.02
Req16	2	2	2	0	1	1.00	2.01
Req17	2	2	2	0	1	1.00	2.01
Req18	5	5	4	1	0.99	1.00	4.02
Req19	5	5	4	1	0.99	1.00	4.02
Req20	5	5	4	1	0.99	1.00	4.02
Req21	2	2	1	1	0.99	1.00	1.01
Req22	8	9	11	-3	1	1.00	11.06
Total	100	104	107	-7	21.83	21.98	107.63

11 Discussion

Customers and developers estimate costs differently, resulting in differing expectations for project cost. Because in business system development projects there are many ways to implement requirements, there are large variability in translating user's requirement into system specification. It differs greatly according to staff's skill. Additionally subjective factors is another reasons why estimate does not meet final cost. Thus proper estimate by requirement analysis is needed in order to finish business system development project fine. Also correct probability for risk is needed to build proper estimate. Requirements have invisible risks. There are plus risk and minus risk. Minus risk gets prospect of profit worse, but plus risk gets prospect of profit well. But plus risk is not visible, plus risk is only in mind of staffs individually. This is one reason why estimated cost:109 match total cost:109 (see Table 3).Additionally, this paper shows potential accurate estimation would be gained by ruling out plus risks in estimating with Bayesian estimation. Therefore there is potential to predict risk accurately using conditional probability or Bayesian estimation [7][8][9]in order to predict cost accurately.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Number 17K00354.
The author would like to thank INTERLINK Inc. and TECHNO SOLUTION Inc. for data and advices.

References

[1] Shinji Mochida,James Righter, Joshua D. Summers: *A Study of Project Cost Management Based on Requirements Analysis, International Journal of BMFSA Vol. 21, No. 1* ,jp, 2016. 21-27

[2] Parviz F. Rad, Kou Itou, japanese translation: *PROJECT ESTIMATING AND COST MANAGEMENT* ,Seisansei Syutupan, jp, 2004

[3] Shinji Mochida: *A Study on Method of Measuring Performance for Project Management, 20th ISPE International Conference on Concurrent Engineering* ,IOS Press Ebooks,US, 2013. 264-273

[4] Beshoy Morkos,James Mathieson,Joshua D. Summers: *Comparative analysis of requirements change pre-diction models: manual, linguistic, and neural network, Research in Engineering Design Vol. 25* ,Springer,US, 2014.139-156

[5] Project Management Institute: *A Guide to the Project Management Body of Knowledge Fourth Ed. (JAPANESE TRANSLATION*,Project Management Institute,US, 2008

[6] Kazuya Fujita: *That is Bayes*,OHMSHA,jp, 2015

[7] Christian Robert : *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation* ,Springer Texts in Statistics,US, 2007

[8] Paul G Hoel,Akira Asai,Masayasu Murakami japanese translation : *ELEMENTARY STATISTICS* ,Springer Texts in Statistics, jp, 2011

[9] David Blackwell,Kazuyoshi Mikuni japanese translation : *BASIC STATISTICS* ,Kogakusha, jp, 1972