

# A Study of Probabilistic Risk Evaluation for System Development Project Based on Requirements Analysis and Bayesian Estimation

Shinji MOCHIDA<sup>1</sup>

*Faculty of Commerce University of Marketing and Distribution Sciences, Japan*

**Abstract.** For building a business system quickly, more efficient project management is needed. However, uncertain and subjective factors in requirements from customer cause cost overruns or schedule delays in the project. Furthermore, uncertain and subjective factors can lead to misunderstandings and false estimates when converting requirements into specifications or scheduling within a project. Thus this paper proposes a probabilistic risk evaluation method with Requirements Analysis and Bayesian estimation for project management to accurately assess the project risk. If this method works well, efficient project management will be realized.

**Keywords.** Risk Evaluation, System Development Project, Requirements Analysis, Bayesian Estimation

## Introduction

Business System development projects are challenging, because there are many requirements demanded from customers even while these requirements are proposed with the same priority. Additionally, requirements have subjective factors. Thus it is important to correct accuracy and prioritize requirements according to their essentiality and criticality to finish by schedule under budget. Although system developers estimate according to the complexity of projects [1] [2], on the other hand customers expect the cost to be based on the number of requirements they demand. As a result, there is a difference between the customer and the system developer in the evaluation of estimate. Over cost or schedule delay is caused by missing estimate. Additionally, evaluating requirements properly is important to finish project successfully. There is no research on the accuracy, uncertainty or subjective factors of customer requests. Previous paper proposed cost share rate for projects based on requirements analysis in order to estimate and evaluate requirements accurately [3]. But the attention must be paid to the fact that estimation has subjective factors, for example productivity and skills of the programmer. This paper proposes the cost share rate and Bayesian estimation to predict costs. Cost share rate is defined as the percentage of total cost assigned to each requirement. This paper shows the potential for project risk management.

---

<sup>1</sup> Corresponding Author, Email: Shinji\_Mochida@red.umds.ac.jp.

## 1. Cost Prediction Methods for Projects

### 1.1 Previous research of project management

This paper demonstrates the potential to evaluate risk of requirements using Bayesian estimation based on requirement analysis in system development project management. Previous research of project management focused by schedule, cost estimation or productivity [1]-[4]. Improving productivity contribute to finish project successfully [1]. There was also a previous research that deals with changing requirement in mechanical engineering design. This research showed one example of requirements analysis by exploring the possibility of predicting requirements change with graphical models of the requirement documents and historical change trends [4]-[8]. There is no research that refers plus risks and subjective factors in estimating based on requirement analysis. Therefore this paper proposes the method taking plus risk and subjective factors into account by the means of Bayesian estimation based on requirement analysis.

### 1.2 Typical Cost Prediction Methods

Usually the amount of program source code is predicted by some prediction method in order to assess projects [7]-[9]. Then the amount of program source code is converted into basic monetary cost. Next total cost is made by adding contingency budget to basic monetary cost. It is set as a budget for a project at first. In case of COCOMO method [8]. The amount of program source code acquired by this way would be converted into monetary cost using a parameter (PM: Person-Months) [8]. On the other hand, in case of Function Point method, the point are accumulated according to the complexity of system, for example the number of DB tables, dialog boxes, print forms and interfaces. Next, the acquired points would be converted into monetary cost using a parameter (PM: Person-Months) as those of COCOMO methods [8]. In the Function point method, in addition, it is necessary to count the number of internal and external tables, and internal and external interfaces.

## 2. Requirement Analysis and Evaluation with Cost Share Rate

### 2.1 Cost Share Rate

Cost overrun or schedule delay is caused by lots of changes in requirements. In particular, changing and uncertainty in essential requirement could bring a high risk. It is vital to distinguish the essential requirement which gives significant impacts to system specification or budget of the project. Thus, this research proposes the method to distinguish influential requirements that has high risks for cost share rate [3]. Cost share rates indicate the impact of requirements based on requirements analysis. Costs for projects are usually estimated by grouping costs with the number of dialog boxes, interfaces or print forms and complexity based on requirements. Alternatively, costs are estimated by associating amount of cost with logic design, development, test, adjustment and documents of requirements, not according to the essentiality of requirements. System developers estimate according to the complexity of specifications, but customers expect the cost according to the number of requirements they demand. 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 for mutual understanding of the developer and customers. Cost share rate also indicates the importance and distinguishes essential requirements to prioritize properly. Requirements that have a high cost share rate must have a high risk, and should be also under strict control, because change or modification for essential requirement give impact to costs or the schedule.

2.2 Calculation Method of Cost Share Rate

This section shows calculation procedure for improved cost share rate. In this calculation procedure, start with counting overlapping keywords which appear both requirements and technical term dictionary. The technical term dictionary has typical keywords categorized in the main design, dialog design, programming, and other. Other categories have keywords in print, test, interface and document categories. These categories are derived from the experience of engineers from project management. Overlapping keywords indicate relationships between one requirement and other requirements or categories. The number of relationships indicates the essentiality and importance of requirements. Steps of procedure for calculating cost share rate are as follows:

- (1) Extracting keywords from each requirement (see Figure 1).
- (2) Counting overlapping key words, first order and complexity of each requirement from extracted keywords (see Figure 2). First order is the number of paths that is directly from one requirement to other requirements. Complexity is the number o keyword including each requirement. And categorized each requirement into categories according to the technical term dictionary.
- (3) Calculating cost share rates of each requirement by regression analysis with parameters; First order, Complexity, Category weight(see Table 1). And Evaluate essential requirements (see Figure 3).by

In Figure 3, cost share rates of staffs for each requirements are obtained subjectively from three engineers who worked on the system development project. On the other hand ,in Figure 3 cost share rates are obtained by steps as above.

keyword	Req01	Req02	Req03	Req04	Req05
name	0	0	1		1
equipment	1	1	1		1
user	0	0	1		1
screen	0	0	1		1
tag	1	1	0		0
displayed	0	0	1		1
gate	0	0	1		1
IC	1	0	0		0

Figure 1. Part of Extracted keywords from requirements.

Table 1. Category weight for requirements.

Type	Other	Screen	System	Main
Category	Other	Dialog design	Programming	Main design
Weight	2.3	9.1	4.5	6.8

	Req01	Req02	Req03	Req04	Req05
Req01	0	6	3	2	
Req02	6	0	3	2	
Req03	3	3	0	20	
Req04	2	2	20	0	
Req05	2	3	10	11	
Req06	3	4	8	6	
Req07	0	0	0	0	
Req08	1	1	1	1	

Figure 2. Overlapping keywords from requirement.

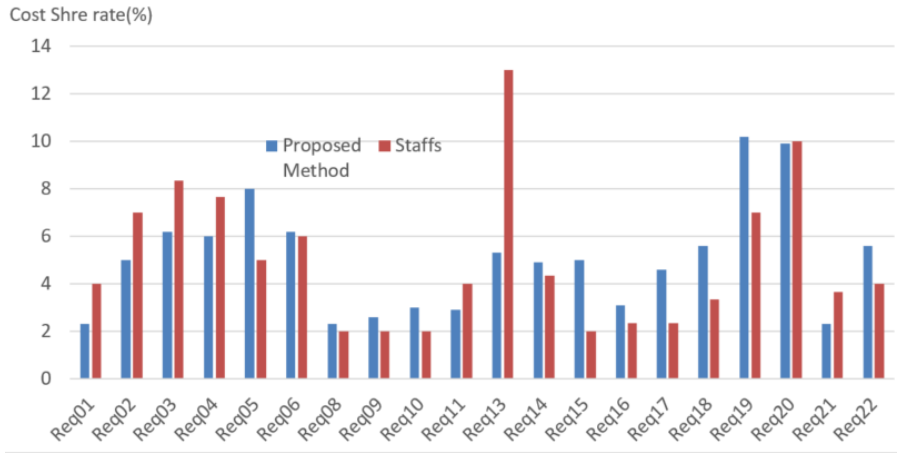


Figure 3. Cost share rate.

3. Risks Evaluation of Projects using Schedule Delay and Cost share rate

3.1 Prediction of probability of the schedule delay for each task

In quantitative evaluation for risk management, risks are evaluated by the monetary loss[9]. The monetary loss is defined as the monetary damage multiplied by the probability of risk. Thus this section explains the trial for calculating risks by multiplying probability and cost share rate of each requirements. Table. 2 shows the schedule delay of these processes: investigation, design, programming, making document, test process from the past system development project which is building RFID system. Table 2. shows that the test and document making process have no risk for the schedule delay. In this research, risk is considered as the probability of the schedule delay. Thus this research suppose that the probability of the schedule delay follows  $\beta$  distribution (see Equation 1). Parameters in Table 3. are obtained by fitting the sample data from Table 2 into  $\beta$  distribution. Thus the parameters (see Table 3) and  $\beta$  distribution curve are obtained (see Figure 4). In Figure 4 horizontal axis(x) shows the normalized start date of tasks. The vertical axis(y) shows the probability for the schedule delay against planned days for each task. Each task is categorized into process category(see Table 2).

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

From these results the average of the probability for the schedule delay is considered to be 0.225. Thus this probability of 0.225 is assigned to standard requirements. Standard requirements are the requirements that belong to the design or programming, because they have average complexity and average risks for cost overrun and schedule delay.

Table 2. sample data from past project.

Process Category	Schedule day(day)	Actual work day	Schedule delay(day)	Normalized Start day
Investigation/Main design	11	35	24	0.00
Dialog design	11	16	5	0.55
Function design	11	11	0	0.52
Detailed Dialog design	18	18	0	0.66
Detailed Function design	16	25	9	0.60
Detailed DB design	27	27	0	0.59
Dialog programming	9	9	0	0.71
Function programming	32	31	-1	0.74
Interface programming	26	24	-2	0.59
DB programming	39	62	23	0.59
Utility programming	3	3	0	0.91
Document	1	1	0	0.93
Test	3	2	-1	0.963

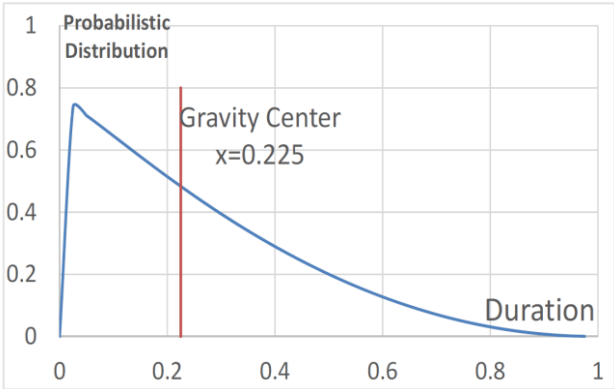
Table 3. Obtained Parameters for  $\beta$  distribution.

Coefficient	c	$\alpha$	$\beta$
Value	0.840	1.021	3.048

Project risk is the defined possibility of cost overrun after this. On the other hand, requirement1 is about investigation and the main design(see Table 2). It indicates the probability of cost overrun risk as 0.1, because the uncertainty of the main requirement is the main reason for arranging an extra budget called contingency budget as 10% of total cost in addition to the basic budget. Table 4 shows the proportion of breakdown of cost in the past actual projects: building information management system, medical record system, knowledge management system. These past projects needed extra cost of 10% of total cost, because there are uncertainties at the beginning of the project. Because tasks are planned on the basis of requirements, it is important to evaluate requirements accurately in order to avoid the cost overrun. If requirements would be accurate, tasks would be set up appropriately, as a result the project could be finished successfully. Therefore, in this research, in order to avoid the cost overrun, it propose a method to categorize appropriately and evaluate requirements according to the degree of risk. Requirements should be categorized as the main design, design, development, and others according to their risk by the linguistic analysis and regression analysis mentioned in section 2.

Table 4. Proportion breakdown of cot in past projects.

Project	Design	Progra mming	Testing	Docum ent	Refine ment	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



**Figure 4.** Curve for possibility of schedule delay.

3.2 Applying probability for another project

This section shows a trial to apply the probability of cost overrun to another small project which is building document management system. In this project, there are requirements2, requirements4, requirement7 and requirement8 which concern the design. These four requirements are standard requirements. Standard requirement is the requirement which has normal complexity and risks for the cost overrun. Thus, it has been assigned the probability of the schedule delay from  $\beta$  distribution as 0.225. The requirement 1 is concerning the main design and it has been assigned the probability of 0.1, because usually the schedule delay is caused by the uncertainty of the main design (see Table 2). Table 5 shows the estimated cost adapting the above mentioned probability of the schedule delay to each requirement. At the result, estimated total cost that includes expected monetary risk is about 110 ; on the other hand, actual total cost at completion is about 127(see Table 5).

**Table 5.** Estimated cost using beta distribution.

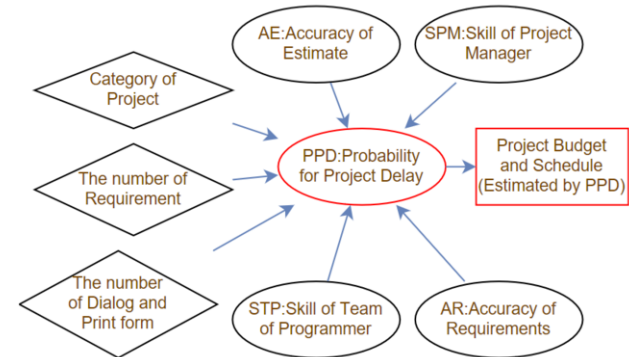
Requir ement No	Category	Cost share Rate(%) (Estimated)	Probability for Schedule Delay	Expected Monetary Risk	Estimated Cost (Normalized)	Actual cost (Normalized)
1	Main	6.4	0.100	0.64	7.02	10.6
2	Design	10.6	0.225	2.39	13.03	10.6
3	Program	2.1	0.000	0.00	2.13	2.1
4	Design	4.3	0.225	0.96	5.21	8.5
5	Program	2.1	0.000	0.00	2.13	2.1
6	Program	12.8	0.000	0.00	12.77	17.0
7	Design	6.4	0.225	1.44	7.82	8.5
8	Design	21.3	0.225	4.79	26.06	12.8
9	Program	12.8	0.000	0.00	12.77	25.5
10	Other	12.8	0.000	0.00	12.77	17.0
11	Program	8.5	0.000	0.00	8.51	12.77
		100.0		10.21	110.21	127.7

\*Program: Programming

4 Risks Evaluaton of Projects using Cost Share Rate and Bayesian Estimation

4.1 Influence diagram for system development project

This research considers an influence diagram that shows the relationship between factors in a project as shown in Figure 5[10]. This influence model includes type of chance nodes as follows: AE (Accuracy of Estimate), SPM (Skill of Project Manager), STP (Skill of Team of Programmers), AR (Accuracy of Requirements). Additionally, this influence model includes decided nodes as follows: Category of Project, The number of Requirements, The number of Dialog and Print form, and decision node as Project Budget and Schedule (see Figure 11). In addition, this is expanded, by adding Decided nodes( $\diamond$ ). Decided nodes has only information about projects, Decided nodes give no infrequence for calculating risks of projects.



Chance nodes( $\circ$ ), Decision nodes( $\square$ ), Decided nodes( $\diamond$ ).

Figure 5. Expanded Influence Diagram for project.

4.2 Risk Evaluation with Bayesian network

Figure 6 shows the Bayesian network using Weka for the influence diagram shown in Figure 6 before given evidence, and it shows the conditional probability given by Table 6 [11][12]. Numbers in Table 6 are obtained from two project manager subjectively. Thus Figure 7 called case1 shows the conditional probability for project delay is 0.125 when engineers' skill is high and accuracy of requirement is standard. On the other hand, Figure 8 called case 2 shows the conditional probability for project delay is 0.2631 when engineers' skill is high and accuracy of requirement is low. For predicting risks of the project using Bayesian estimation and cost share rate, in this method, the expected monetary value for risk is obtained by cost share rate multiplied by the probability of schedule delay of whole project using Bayesian estimation with Weka. Table 7 shows the result of expected monetary value for risk is 0.125 when engineers' skill is high with standard accuracy of the requirements (Case1). And the expected monetary value for risk is 0.263 when engineers' skill is high with low accuracy of the requirements (Case2). This result fit to sence of project manager. For example, actual cost is 127 (normalized) in Table 5 showed the case when engineers' skill is high with low accuracy of the requirements.

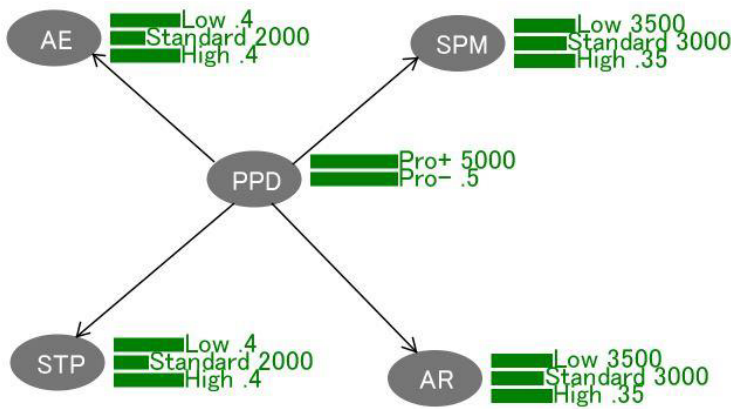


Figure 6. Bayesian network for the project (before given evidence).

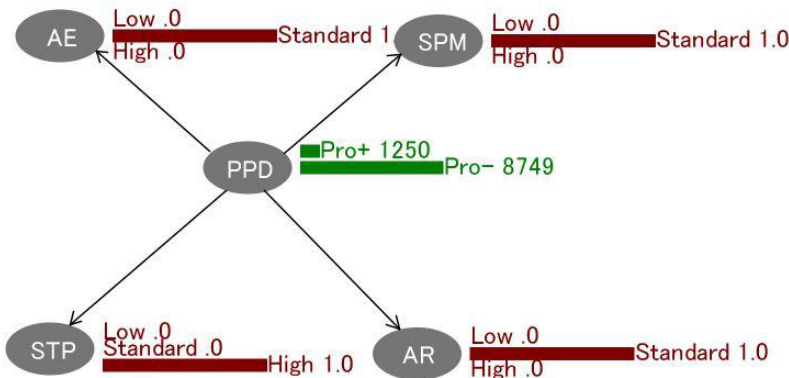


Figure 7. Bayesian network for the project (Case1)

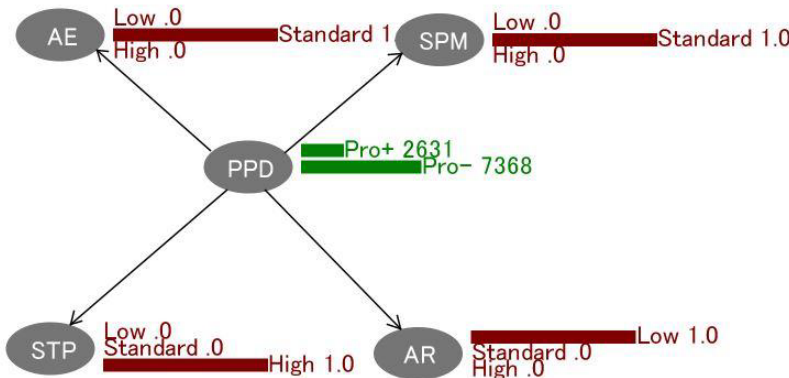


Figure 8. Bayesian network for the project (Case2).



**Table 6.** Conditional probability for project.

Probability Distribution Table for AE				
	AE:Accuracy of Estimate			
Project Status	Low	Standard	High	Sum
Delay	0.7	0.2	0.1	1.0
On Schedule	0.1	0.2	0.7	1.0
Probability Distribution Table for SPM				
	SPM:Skill of Project Manager			
Project Status	Low	Standard	High	Sum
Delay	0.6	0.3	0.1	1.0
On Schedule	0.1	0.3	0.6	1.0
Probability Distribution Table for STP				
	STP:Skill of Team of Programmer			
Project Status	Low	Standard	High	Sum
Delay	0.6	0.3	0.1	1.0
On Schedule	0.1	0.3	0.6	1.0
Probability Distribution Table for AR				
	AR:Accuracy of Requirements			
Project Status	Low	Standard	High	Sum
Delay	0.5	0.3	0.2	1.0
On Schedule	0.2	0.3	0.5	1.0

**Table 7.** Conditional probability for project.

NO	Requirement No	Category	A:Cost Share Rate(%) (from Table 7)	B:Probability for Schedule Delay (Case1)	Estimated Cost $A*(1+B)$ (Case1)	C:Probability for Schedule Delay (Case2)	Estimated Cost $A*(1+C)$ (Case2)	Actual Cost
1	Req1	Main	6.4	0.125	7.18	0.26	8.06	10.6
2	Req2	Design	10.6	0.125	11.97	0.26	13.44	10.6
3	Req3	Program	2.1	0.125	2.39	0.26	2.69	2.1
4	Req4	Design	4.3	0.125	4.79	0.26	5.37	8.5
5	Req5	Program	2.1	0.125	2.39	0.26	2.69	2.1
6	Req6	Program	12.8	0.125	14.36	0.26	16.12	17.0
7	Req7	Design	6.4	0.125	7.18	0.26	8.06	8.5
8	Req8	Design	21.3	0.125	23.94	0.26	26.87	12.8
9	Req9	Program	12.8	0.125	14.36	0.26	16.12	25.5
10	Req10	Other	12.8	0.125	14.36	0.26	16.12	17.0
11	Req11	Program	8.5	0.125	9.57	0.26	10.75	12.77
			100.00	1.38	112.50	2.89	126.31	127.7

## 5. Conclusion

More efficient project management is needed in order to meet budget, finish by schedule, and maintain high quality in projects. Though it is difficult to evaluate risks accurately, because there are uncertain and subjective factors in projects, especially requirements have much uncertainty. Thus this paper proposes the cost share rate in order to distinguish essentiality of each requirements. In section 3, it shows it is difficult to predict risk of projects with cost share rate and probability for schedule delay is obtained from  $\beta$  distribution. Next this paper shows a trial to predict risk of projects accuracy using the cost share rate and Bayesian estimation. As a result, this paper shows the potential to evaluate the risk of projects accuracy using Bayesian estimation and cost share rate, in addition this method provides the opportunity to break down the cause of risks into risk factors [8][9]. The conditional probability showed Table 6 includes subjective factor since it was obtained by a questionnaire filled in by project engineers, thus obtaining the conditional probability more logically is needed in future works. In addition, in order to manage projects better, further research is needed on risk evaluation with requirements analysis that takes into account uncertain factors and subjective factors in requirements.

## Acknowledgement

This work was supported by JSPS KAKENHI Grant Number 17K00354.

The author would like to thank Professor Radim Jiroušek, and Professor Václav Kratochvíl for advices. And the author would like to thank Professor Miloslava Hirsova for advices. Finally the author would like to thank INTERLINK Inc. and TECHNO SOLUTION Inc. for data.

## References

- [1] S. Mochida, A Study on Method of Measuring Performance for Project Management, *20<sup>th</sup> ISPE International Conference on Concurrent Engineering*, IOS Press, Amsterdam, 2013, pp. 264-273.
- [2] Parviz F. Rad, Kou Itou, Japanese translation: Project Estimating and Cost management, Seisansei Syutupan Japan, 2004.
- [3] S. Mochida, J. Richter, J.D. Summers, A Study of Project Cost Management Based on Requirements Analysis, *International Journal of BMFSA*, 2016, Vol. 21, No. 1, pp. 21-27.
- [4] R. Young, *Requirements Engineering Handbook*, Artech House, Boston, 2003.
- [5] B. Morkos, J. Mathieson, J.D. Summers, Comparative analysis of requirements change prediction models: manual, linguistic, and neural network, *Research in Engineering Design*, 2014, Vol. 25, pp. 139-156.
- [6] S. Kato, *The science of the plan-we can use everywhere PERT • CPM*, KODANSHA, 1965.
- [7] N.N., *Project Management Institute: A Guide to the Project Management Body of Knowledge*, Fourth Ed. (Japanese translation), Project Management Institute, US, 2008.
- [8] N. Kaneko, *Project Manager text*, Nikkei Publishing Inc., 2014.
- [9] Q.W. Fleming, J.M. Koffleman, *Earned Value Project Management Japanese translation*, JMA Management Center, 2004.
- [10] R.A. Howard and J.E. Matheson, Influence Diagrams, *Decision Analysis*, 2005, Vol. 2, No. 3, September 2005, pp. 127-143.
- [11] K. Fujita, *That is Bayes*, Ohmsha, Japan, 2015.
- [12] C. Robert, *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation*, Springer, New York, 2007.