



HAL
open science

Fiabilité des Systèmes : Inference pour des modèles de défaillance de cause commune dans des contextes de données manquantes

Huu Du Nguyen

► **To cite this version:**

Huu Du Nguyen. Fiabilité des Systèmes : Inference pour des modèles de défaillance de cause commune dans des contextes de données manquantes. Statistics [math.ST]. Université de Bretagne Sud, 2019. English. NNT : 2019LORIS530 . tel-02875396

HAL Id: tel-02875396

<https://theses.hal.science/tel-02875396>

Submitted on 19 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE DE DOCTORAT DE

L'UNIVERSITE BRETAGNE SUD

COMUE UNIVERSITÉ BRETAGNE LOIRE

ECOLE DOCTORALE N° 601

*Mathématiques et Sciences et Technologies
de l'Information et de la Communication*

Spécialité : Mathématiques et leurs Interactions

Par

Huu Du NGUYEN

System Reliability : Inference for Common Cause Failure Model in Contexts of Missing Information

Thèse présentée et soutenue à Vannes, le 12 Juillet 2019

Unité de recherche : LMBA CNRS UMR 6205

Thèse N° : 530

Rapporteurs avant soutenance :

Anne Barros Professeure, Norwegian University of Science and Technology

Massimiliano Giorgio, Professeur, Università degli Studi di Napoli Federico II

Composition du Jury :

Président :

Examineurs : **Mitra Fouladirad**, Professeure, Université de Technologie de Troyes

Bruno Castanier, Professeur, Université d'Angers

Emmanuel Frenod, Professeur, Université de Bretagne Sud

Dir. de thèse : **Evans Gouno**, MCF-HDR, Université de Bretagne Sud

Invité : **Gilles Deleuze**, Senior Researcher, EDF R&D

Titre : Fiabilité des Systèmes : Inference pour des modèles de défaillance de cause commune dans des contextes de données manquantes.

Mots clés : Fiabilité des systèmes, défaillance de cause commune, données manquantes, estimateur de Bayes, algorithme EM, algorithme de Metropolis Hasting, Inverse Bayes formula.

Résumé: Le bon fonctionnement de l'ensemble d'un système industriel est parfois fortement dépendant de la fiabilité de certains éléments qui le composent. Une défaillance de l'un de ces éléments peut conduire à une défaillance totale du système avec des conséquences qui peuvent être catastrophiques en particulier dans le secteur de l'industrie nucléaire ou dans le secteur de l'industrie aéronautique. Pour réduire ce risque de panne catastrophique, une stratégie consiste à dupliquer les éléments sensibles dans le dispositif. Ainsi, si l'un de ces éléments tombe en panne, un autre pourra prendre le relais et le bon fonctionnement du système pourra être maintenu. Cependant, on observe couramment des situations qui conduisent à des défaillances simultanées d'éléments du système : on parle de **défaillance de cause commune**.

Analyser, modéliser, prédire ce type d'événement revêt donc une importance capitale et sont l'objet des travaux présentés dans cette thèse. Il existe de nombreux modèles pour les défaillances de cause commune. Des méthodes d'inférence pour étudier les paramètres de ces modèles ont été proposées.

Dans cette thèse, nous considérons la situation où l'inférence est menée sur la base de données manquantes. L'actualité récente a mis en évidence l'importance de la fiabilité des systèmes redondants et nous espérons que nos travaux contribueront à une meilleure compréhension et prédiction des risques de catastrophes majeures.

Title : System Reliability: Inference for Common Cause Failure Models in Contexts of Missing Information

Keywords: System reliability, common cause failure, missing data, Bayes estimator, EM algorithm, Metropolis-Hasting algorithm, inverse Bayes formula.

Abstract: The effective operation of an entire industrial system is sometimes strongly dependent on the reliability of its components. A failure of one of these components can lead to the failure of the system with consequences that can be catastrophic, especially in the nuclear industry or in the aeronautics industry. To reduce this risk of catastrophic failures, a redundancy policy, consisting in duplicating the sensitive components in the system, is often applied. When one of these components fails, another will take over and the normal operation of the system can be maintained.

However, some situations that lead to simultaneous failures of components in the system could be observed. They are called **common cause failure (CCF)**. Analyzing, modeling, and predicting this type of failure event are therefore an important issue and are the subject of the work presented in this thesis.

Recent events have highlighted the importance of reliability redundant systems and we hope that our work will contribute to a better understanding and prediction of the risks of major CCF events.

**System Reliability: Inference for Common Cause
Failure Models in Contexts of Missing Information**

Huu Du Nguyen

Laboratoire de Mathématiques de Bretagne Atlantique
Université de Bretagne Sud

Acknowledgements

This is to express my sincere gratitude to all the people who have supported me during my stay in France and assisted me in completing this work.

I would like to especially thank my true mentor, Prof. Evans Gouno for his kindness and enthusiasm. Honestly, I cannot imagine doing my PhD without his instructions. Moreover, I have also learned from him many simple but valuable lessons in life. I will always remember his stories and sharing during the conference at Palermo, the meal he prepared for me when I came to his home and the time we worked together. I wish to have opportunities to welcome him in my hometown.

I would also like to thank the referees of this thesis, Prof. Anne Barros and Prof. Massimiliano Giorgio, for agreeing to review this manuscript and for their constructive comments. I am grateful to the members of the jury, Prof. Mitra Fouladirad, Prof. Bruno Castanier, Prof. Emmanuel Frenod and Dr. Gilles Deleuze for taking the time to evaluate my work.

I render my sincere thank to LMBA members for their cheerfulness. Prof. Bertrand is a very nice man and we had wonderful meals together with his family. Mrs. Véronique and Mrs. Sandrine helped me a lot with the administrative tasks. Prof. Salim shared a lot of time with me and other students in the Lab. Prof. Sylvain, the French captain, showed me a lot of interesting things. I thank my fellow doctoral students: to Éric, a perfect friend who is always willing to help others with a smile never off on the lips; to Ronan, a true French friend who taught me the kindness and perseverance; to Kevin, a great French guy who gave me a great French friendship and made me drunk; to Phú, a brother and a best Vietnamese friend of mine in France; to Thuy, a kind and gentle sister; to Hui, an interesting bookworm who told me that I have a very good Chinese friend; to Augustin, a funny guy from Burkina Faso; to Jamila, a lovely Moroccan friend; to Cuong, who introduces me to our mentor and helps me in the early days to France; to Jonathan; to Hélène; to Anne-Charlotte; to Erwan and to Tarik.

I am very much thankful to all the friends who helped me in the time of living away from home. Thanks to Phúc and his wife for motivating and encouraging me in doing research. Thanks to Thong for his enthusiastic help in all the problems of studying and daily life. Thanks to Tan and his wife for their attentions and hospitality. Thanks to Thao, Phuong, Trang, Trung beo, Hien tro, Hien beo, Quang, for the time we spent together in the Vietnamese days. Thanks to Dominique, Lucien and his family for their kindness. I would also like to express my sincere thanks to some Vietnamese families in Vannes: ba Noi, Mrs. Le Tho and her family, Mr Hung and his family, Mrs Vinh and her family and Mr Dong and his family; they truly treated me like a family member.

Finally, I would like to convey thanks to my family for their support throughout this four years stay far away from home, especially my mother, my beloved wife and my two children. Their trust is my motivation to overcome difficulties in life.

Contents

1	Introduction	17
2	Inference on BFR model from confounded data	25
2.1	Introduction	25
2.2	The data and the model	26
2.3	Inference	29
2.3.1	Maximum likelihood method	29
2.3.2	Bayesian approach	33
2.4	Concluding remarks	39
3	Inference for the α-factor model from uncertain data	41
3.1	Introduction	41
3.2	The α -factor model	43
3.3	Inference on complete data	45
3.3.1	Maximum likelihood method	45
3.3.2	Bayesian approach	45
3.4	Inference on data with uncertainties	47
3.4.1	Impact vector method	47
3.4.1.1	The definition of impact vector	47
3.4.1.2	Inference for CCF with impact vector	48
3.4.2	Confounded data method	52
3.4.2.1	Inference for scheme 1	54
3.4.2.2	Inference for scheme 2	59
3.4.2.3	Inference for scheme 3	65
3.4.2.4	Inference for scheme 4	70
3.5	Concluding remarks	75
4	Causal inference from incomplete contingency table	77
4.1	Introduction	77
4.2	The data and the model	78
4.3	Bayesian inference	83
4.3.1	Using the IBF sampling	83
4.3.2	Using a stochastic version of the IBF sampling	86
4.4	A simulation study	88

4.5	Application	89
4.6	Concluding remarks	93
5	Prediction methods for catastrophic events	95
5.1	Introduction	95
5.2	Prediction with complete data: classical methods	96
5.2.1	Pivotal method	96
5.2.2	Bayesian method	97
5.2.3	Illustrative example	99
5.3	Prediction with mapped data	99
5.3.1	Mapping down rules	100
5.3.2	Mapping up rules	101
5.3.3	Illustrative example	104
5.4	Prediction with uncertain data	106
5.4.1	Uncertainty on observed data	106
5.4.2	Uncertainty on the exposure time	109
5.4.3	Uncertainty on both observed data and exposure time	111
5.5	Concluding remarks	115
6	General conclusion and perspectives	117
	Appendices	121
A	The EM algorithm	123
B	The Inverse Bayesian formula	125
C	The modified – Beta distribution	129
D	Metropolis-Hastings algorithm	133
E	Mean and variance of the posterior distribution of Λ	135
E.1	Uncertainties on both overved count and observation time	135
E.2	Uncertainties on observation time	136

System Reliability: Inference for Common Cause Failure Models in Contexts of Missing Data

Abstract

The effective operation of an entire industrial system is sometimes strongly dependent on the reliability of its components. A failure of one of these components can lead to the failure of the system with consequences that can be catastrophic, especially in the nuclear industry or in the aeronautics industry. To reduce this risk of catastrophic failures, a redundancy policy, consisting in duplicating the sensitive components in the system, is often applied. When one of these components fails, another will take over and the normal operation of the system can be maintained. However, some situations that lead to simultaneous failures of components in the system could be observed. They are called **common cause failure** (CCF). Analyzing, modeling, and predicting this type of failure event are therefore an important issue and are the subject of the work presented in this thesis.

We investigate several methods to deal with the statistical analysis of CCF events. Different algorithms to estimate the parameters of the models and to make predictive inference based on various type of missing data are proposed.

We treat *confounded data* using a BFR (Binomial Failure Rate) model. An EM algorithm is developed to obtain the maximum likelihood estimates (MLE) for the parameters of the model. We introduce the *modified-Beta distribution* to develop a Bayesian approach.

The α -factors model is considered to analyze *uncertainties* in CCF. We suggest a new formalism to describe uncertainty and consider Dirichlet distributions (nested, grouped) to make a Bayesian analysis.

Recording of CCF cause data leads to incomplete contingency table. For a Bayesian analysis of this type of tables, we propose an algorithm relying on inverse Bayes formula (IBF) and Metropolis-Hasting algorithm. We compare our results with those obtained with the α -decomposition method, a recent method proposed in the literature.

Prediction of catastrophic event is addressed and mapping strategies are described to suggest upper bounds of prediction intervals with pivotal method and Bayesian techniques.

Recent events have highlighted the importance of reliability redundant systems and we hope that our work will contribute to a better understanding and prediction of the risks of major CCF events.

Keywords: System reliability, common cause failure, missing data, contingency table, Poisson process, maximum of likelihood, Bayes estimator, EM algorithm, Metropolis-Hasting algorithm, inverse Bayes formula

Fiabilité des Systèmes : Inference pour des modèles de défaillance de cause commune dans des contextes de données manquantes

Résumé

Le bon fonctionnement de l'ensemble d'un système industriel est parfois fortement dépendant de la fiabilité de certains éléments qui le composent. Une défaillance de l'un de ces éléments peut conduire à une défaillance totale du système avec des conséquences qui peuvent être catastrophiques en particulier dans le secteur de l'industrie nucléaire ou dans le secteur de l'industrie aéronautique. Pour réduire ce risque de panne catastrophique, une stratégie consiste à dupliquer les éléments sensibles dans le dispositif. Ainsi, si l'un de ces éléments tombe en panne, un autre pourra prendre le relais et le bon fonctionnement du système pourra être maintenu. Cependant, on observe couramment des situations qui conduisent à des défaillances simultanées d'éléments du système : on parle de **défaillance de cause commune**. Analyser, modéliser, prédire ce type d'événement revêt donc une importance capitale et sont l'objet des travaux présentés dans cette thèse. Il existe de nombreux modèles pour les défaillances de cause commune. Des méthodes d'inférence pour étudier les paramètres de ces modèles ont été proposées. Dans cette thèse, nous considérons la situation où l'inférence est menée sur la base de données manquantes. Nous étudions en particulier le modèle BFR (Binomial Failure Rate) et la méthode des α -facteurs. En particulier, une approche bayésienne est développée en s'appuyant sur des techniques algorithmiques (algorithme de Metropolis, théorème Inverse de Bayes (IBF)). Dans le domaine du nucléaire, les données de défaillances sont peu abondantes et des techniques particulières d'extrapolations de données doivent être mis en oeuvre pour augmenter l'information. Nous proposons dans le cadre de ces stratégies, des techniques de prédiction des défaillances de cause commune. L'actualité récente a mis en évidence l'importance de la fiabilité des systèmes redondants et nous espérons que nos travaux contribueront à une meilleure compréhension et prédiction des risques de catastrophes majeures.

Keywords: Fiabilité des systèmes, défaillance de cause commune, données manquantes, tableau de contingence, processus de Poisson, maximum de vraisemblance, estimateur de Bayes, algorithme EM, algorithme de Metropolis-Hasting, Inverse Bayes formula.

Acronyms

BFR	Binomial Failure rate
CCF	Common Cause Failure
EDG	Emergency Diesel Generator
EM	Expectation Maximization
GDD	Grouped Dirichlet Distribution
HPP	Homogeneous Poisson Process
IBF	Inverse Bayesian Formula
MCMC	Markov Chain Monte Carlo
MSE	Mean Square Error
MLE	Maximum Likelihood Estimation
NDD	Nested Dirichlet Distribution
PRA	Probabilistic Risk Assessment

Notation

T	length of observation window
$N_I(T)$	process of independent failures
$N_i(T)$	process of failures of order $i, i = 0, 1, 2 \dots m$
$N_1^*(T)$	process of failures that CCF of order 1 and independent failure are not distinguishable
$N_L(T)$	process of lethal shock
$N_m^*(T)$	process of failures that CCF of order m and lethal shock are not distinguishable
\mathbf{N}_u	vector of data with uncertainties
\mathbf{N}	vector of data from margins of a contingency table
$\mathcal{B}(n, p)$	Binomial distribution with parameters (n, p)
$\mathcal{G}(\alpha, \beta)$	Gamma distribution with parameters (α, β)
$\mathcal{P}(\lambda)$	Poisson distribution with parameter λ
$\mathcal{M}(n, p_1, \dots, p_k)$	Multinomial distribution with parameters (n, p_1, \dots, p_k)
$\mathcal{D}(p_1, \dots, p_k)$	Dirichlet distribution with parameters (p_1, \dots, p_k)
$NB(r, p)$	Negative Binomial distribution with parameters (r, p)
$ND_{n,n-1}(\mathbf{a}, \mathbf{b})$	Nested Dirichlet distribution with parameters (\mathbf{a}, \mathbf{b})
$GD_{n,n-1}(\mathbf{a}, \mathbf{b})$	Grouped Dirichlet distribution with parameters (\mathbf{a}, \mathbf{b})
$B(x, y)$	value of Beta function at (x, y)
$\tilde{\theta}$	Bayesian estimator of the parameter θ
$\hat{\theta}$	Maximum likelihood estimator of the parameter θ

CONTENTS

Résumé

Introduction

Les défaillances de cause commune (CCF) sont des phénomènes que l'on observe lors du suivi longitudinal d'un système à plusieurs composants. Il s'agit de défaillances qui se produisent sur plusieurs composants du système de manière simultanée et qui ont été provoquées par une même cause. On les qualifie donc de **cause commune**.

L'évaluation de la propension d'un système à être sujet à ce type de défaillance est très importante lorsque le système est une ensemble d'éléments redondants supposés assurer le maintien d'un certain niveau de sécurité pour un dispositif industriel. C'est pour cette raison que l'étude des DCC s'est initialement développé dans l'industrie nucléaire et l'industrie aéronautique. Dans ces domaines, une DCC où tous les composants tombent en panne conduit à une catastrophe.

Les travaux présentés dans ce mémoire de thèse, portent sur l'analyse statistique de ces phénomènes de défaillances de cause commune dans différentes situations d'observation engendrant une perte d'information.

On peut considérer que l'objectif fondamental de l'étude des DCC est calculer la probabilité d'occurrence d'une DCC d'un ordre donné. Autrement dit, il s'agit de caractériser la distribution du nombre de DCC d'un ordre donné.

De nombreux modèles ont été proposés. Nous nous intéresserons au modèle BFR (Binomial Failure Rate) [2] et au modèle des α -facteurs [35]. Le modèle BFR repose sur une modélisation par des processus stochastiques et on mène l'inférence sur les intensités de ces procesus. Le modèle des α -facteurs se propose de caractériser les DCC par des probabilités α déduites de la probabilité élémentaire de défaillance de chaque composant et sous certaines hypothèses.

Les modèles que nous allons étudier, dépendent de la nature de l'observation. Idéalement, on peut disposer des dates exactes des défaillances successives se produisant sur le système. Ces dates peuvent être accompagnées d'information sur la nature de la défaillance : type de DCC, cause (erreur humaine, événement climatique, problème technique, par exemple).

En pratique, on dispose rarement de ce schéma d'observation dit *complet* et le praticien est souvent confronté à toutes sortes de situations d'observation

dites *dégradée* : données groupées, données manquantes, informations incomplètes sur les causes et le type de DCC, etc. Une autre situation particulière qui retiendra notre attention, est celle où l’observation consiste en un tableau de contingence dont seules les marges sont connues. Pour un nombre total de défaillance donné, seules les effectifs des différents causes possibles et les effectifs des différents types de DCC sont enregistrés [71]. On a en quelque sorte un tableau de contingence partiellement observé.

Nous résumons dans les sections suivantes les travaux présentés dans chacun des chapitres de ce mémoire.

Inférence sur le modèle BFR pour des données confondues

Dans le premier chapitre, nous considérons la situation où les données sont *confondues* et nous proposons une solution d’inférence par maximum de vraisemblance et par une approche bayésienne pour les paramètre du modèle BFR. Dans le cas du modèle BFR, le comptage des défaillances de cause commune est modélisé par un processus de Poisson. On note $\{N_i(t), t \geq 0\}$, le processus de comptage des DCC d’ordre i , $i = 1, \dots, m$, pour un système à m composants et λ_i son intensité. Cependant, une panne impliquant un seul composant n’est pas nécessairement une DCC. Ce peut être une défaillance intrinsèque du composant, défaillance qui n’a pas de “cause”. En pratique, il n’est pas possible de distinguer une défaillance intrinsèques d’une DCC d’ordre 1. De façon équivalente, une panne impliquant tous les composants n’est pas toujours une DCC d’ordre m . Ce peut être une défaillance létale (propre au système) et on ne peut distinguer une défaillance létale d’une DCC d’ordre m . On a des données **confondues**. On dira que N_1 est confondu, tout comme l’est N_m .

Cette situation fait apparaître la dualité “*données complètes / données incomplètes*”. Les données complètes qui sont non observées, sont représentées par le vecteur $(N_I, N_1, \dots, N_m, N_L)$ où N_I désigne le nombre de défaillances intrinsèques et N_L , le nombre de défaillances létales.

Les données incomplètes, qui sont effectivement observées, sont le vecteurs $(N_1^*, N_2, \dots, N_{m-1}, N_m^*)$ où $N_1^* = N_I + N_1$ et $N_m^* = N_m + N_L$. Nous appliquons un algorithme EM pour obtenir les estimateurs du maximum de vraisemblance des paramètres caractérisant chaque type de défaillance. La recherche d’une loi conjuguée pour une approche bayésienne de l’estimation, nous conduit à définir la *loi bêta-modifiée*. Nous proposons alors des stratégies d’éllicitation des hyperparamètres. Ce travail a été publié dans une revue internationale [44].

Inférence sur le modèle α -facteurs pour des données incertaines

Le second chapitre s’intéresse à la situation où il y a *incertitude* sur les données. Nous considérons quelques schémas d’incertitude pour lesquels nous

développons une méthode bayésienne qui s'appuie sur des lois de Dirichlet particulière.

Pour un système à m composants, une observation incertaine est classiquement représentée par un vecteur $\bar{I} = (p_0, p_1, \dots, p_m)$ dont la composante p_i est la probabilité que la DCC observée soit d'ordre i . On a $\sum_{i=0}^m p_i = 1$. p_0 s'interprète comme étant la probabilité qu'il n'y ait pas eu défaillance de cause commune. S'il n'y a pas d'incertitude, ce vecteur devient un vecteur binaire c'est-à-dire un vecteur dont la seule composante non nulle est égale à 1. A toute observation incertaine est associée une observation certaine de sorte qu'à l'observation d'un échantillon de données incertaines peut être associé un ensemble d'échantillons de données certaines possibles. La somme des vecteurs binaires associées à un échantillon de données certaines correspond à un vecteur $\mathbf{N} = (N_0, N_1, \dots, N_m)$ qui représente maintenant des données complètes possibles. On note \mathcal{S} , l'ensemble de tous les vecteurs \mathbf{N} possibles associés à l'observation \bar{I} , un échantillon de vecteur \bar{I} . L'incertitude ayant été quantifiée par les probabilités p_i , on peut calculer la probabilité des vecteurs (N_0, N_1, \dots, N_m) possibles. Cette distribution s'interprète donc comme la loi du complet sachant l'incomplet.

Ces données peuvent être analysées en considérant une modèle α -facteurs. Pour un système de taille m , les paramètres du modèle α -facteurs sont les probabilités α_i qu'une défaillance lorsqu'elle survient, implique i composants parmi les m . On suppose que la probabilité qu'un i -uplet tombe en panne, est la même quelque soit les i composants impliqués et on notera Q_i cette probabilité. Ces Q_i sont appelés les *paramètres de base* du modèle α -facteurs.

Les paramètres α_i peuvent être exprimés en fonction des paramètres de base et la loi du vecteur (N_1, \dots, N_m) exprimée en fonction des paramètres α_i prend alors la forme d'une loi multinomiale. Les estimateurs du maximum de vraisemblance pour les α_i s'obtiennent sans difficultés. Il en est de même pour des estimateurs de Bayes en considérant une loi de Dirichlet comme loi a priori.

Nous proposons une approche alternative au problème de l'incertitude. L'incertitude sur les données sera caractérisée par l'observation de variables de la forme N_{i_1, i_2, \dots, i_k} signifiant que la DCC observée peut être d'ordre i_1 ou i_2 ... ou i_k . L'approche que nous proposons dépend de la forme des données disponibles. Elle consiste en un algorithme EM partiel ou elle s'appuie sur la formule de Bayes inverse (IBF). Pour illustrer notre approche, nous considérons 4 situations de données avec incertitudes. L'approche bayésienne utilise la loi nested-Dirichlet ou encore la loi de Dirichlet dite groupée (grouped-Dirichlet) particulièrement adaptées à la formalisation que nous suggérons de l'incertitude. Les propriétés de ces lois permettent d'obtenir les estimateurs de Bayes du modèles α -facteurs. Nous en étudions le comportement en fonction de différents types d'a priori à partir de données simulées.

Tableaux de contingence et causes de défaillance

Ce chapitre est consacré à l'analyse de données de DCC sous forme de tableau de contingence contenant des informations concernant la cause associée à la défaillance. Comme nous l'avons mentionné précédemment la cause de la défaillance peut être de différente nature. Ce peut être une erreur humaine, un événement climatique, un dysfonctionnement logiciel par exemple. Lorsque cette information est renseignée. Les données peuvent mises sous forme d'un tableau de contingence croisant ordre de la DCC et cause. Les éléments du tableau sont des nombres N_{ij} représentant le nombre de DCC impliquant i composants ayant été provoquées par la cause j . Si tous ces nombres sont enregistrés, l'inférence peut être menée en utilisant des méthodes classiques et bien connues. Cependant, il est courant que ne soit observé que les marges du tableau. Zheng et al. [71] propose une extension du modèle α -facteurs qu'il appelle modèle α -décomposition, pour traiter cette situation. Nous considérons une approche s'appuyant sur une modélisation par les processus de Poisson et développons un algorithme IBF conduisant à une estimation bayésienne des intensités de ces processus. Nous comparons les résultats obtenues sur les paramètres du modèle α -décomposition par analogie.

Méthodes de prédiction des défaillances catastrophiques

Dans le dernier chapitre, nous envisageons le problème de la prédiction des défaillances catastrophiques en utilisant des stratégies de mapping. Les méthodologies suggérées s'appuient sur la méthode pivotale et sur une approche bayésienne.

On appelle *défaillances catastrophiques* les défaillances qui impliquent tous les composants du système. Elles correspondent donc aux DDC d'ordre m pour un système à m composants ou bien aux défaillances létales. Ce type de défaillance est rare. Pour compenser ce manque d'information, la stratégie consiste à considérer les événements provenant de système de différentes tailles. Pour que l'analyse soit cohérente, les données sont transformées en leur équivalent pour un système d'une même taille fixée. Cette opération appelée "*mapping*". Il s'agit d'un projection des données. Lorsqu'on projette les données vers des données correspond à un système de plus petite taille, on parle d'une opération de *mapping-down*. La projection vers un système de plus grande taille est appelée *mapping-up*. Nous nous sommes intéressé au problème de la prédiction dans ce contexte de données projetées. Dans un second temps, nous avons développé une méthodologie de prédiction pour différentes situations d'incertitudes : incertitudes sur les ordres des DCC observées, incertitudes sur longueur de la fenêtre d'observations et incertitudes sur ces deux informations.

Nous avons appliqué la méthode pivotale et une méthode bayésienne et effectué une comparaison des résultats obtenus en fonction des lois a priori

considérées.

Conclusion et perspectives

Nous présentons dans cette thèse, un ensemble de méthodes pour l'analyse statistique des défaillances de cause commune dans des situations où l'observation est incomplète. Les travaux montrent que l'approche bayésienne est particulièrement adaptée à cette problématique industrielle. De nombreuses pistes de travail se dégagent des différents stratégies que nous avons proposées. Nous envisageons, par exemple, de développer un cadre rigoureux s'appuyant sur la théorie de l'information pour optimiser l'opération de mapping. Nous envisageons également d'étudier plus finement la loi bêta modifiée introduite pour l'analyse bayésienne du modèle BFR. Nous préparons actuellement un article à partir des résultats du chapitre 4, concernant l'inférence pour des tableaux de contingence incomplets.

Nous espérons que nos travaux contribueront à une meilleure connaissances et compréhension du phénomène de défaillance de cause commune et permettront d'améliorer la sécurité d'installation sensible et la prévision des risques de catastrophe.

CONTENTS

Chapter 1

Introduction

Probability Risk Assessment

Some real catastrophic events such as the Air France Flight 447 accident (2009) and the Fukushima Daiichi nuclear disaster (2011) have posed higher and stricter requirements in protecting system safety against extraneous shocks which can defeat multiple levels of redundancy.

Probability Risk Assessment (PRA) refers to the comprehensive, structured, and disciplined approaches to identify and evaluate risks associated with complex engineering and technological systems. The aim of PRA is to quantify rare event probabilities of failures like the two events mentioned above. All possible events or influences that could reasonably affect the system is investigated. PRA has been applied in many contexts. It is one of the main tools of the U.S. Nuclear Regulatory Commission which controls the safety in the nuclear power industry. Several advantages and disadvantages of using PRA to evaluate risks from beyond-design-basis external events have been discussed by Council et al. [8].

In general, PRA seeks answers to the following basic questions [6]:

- What kinds of events or scenarios can occur?
- What are the likelihoods and associated uncertainties of the events or scenarios?
- What consequences could result from these events or scenarios?

The PRA is often characterized by event tree models, fault tree models, and other important analysis tools like common cause failure analysis. The common-cause failure analysis is a set of methods for evaluating the effect of inter-system and intra-system dependencies which tend to cause simultaneous failures and thus significant increase in overall risk. In the following, we discuss in more detail about the common cause failure concept, which is a major concern of the works presented in this study.

What is a Common Cause Failure?

To answer the question of "what is a common cause failure", let us start by considering a reliability problem in nuclear industry, which is discussed by Kvam [27]. Components in redundancy systems are used to enhance additional reliability as an individual component is considered incompetent for the operation. At a nuclear power plant, emergency diesel generators (EDGs) are employed to prepare for contingency in case of a temporary loss of power. In the case of power failure, two or more EDGs should be ready to start up and provide emergency power until the electric problems are fixed, even though only one available EDG is adequate to resupply power. When components in these redundant systems fail, they do not necessarily fail independently of each other. Both external shocks, like thunderstorms, earthquakes or floods, and internal sources, like poor maintenance procedures or design flaws, can induce the simultaneous failure of several components. In this case the failure has a common cause; it is said to be a **common cause failure**.

Common Cause Failures (CCFs) are multiple failures of a number of components due to a common event. These types of failures are able to disable multiple layers of redundancy and cause unexpected coincidental consequences for the safety critical systems. They may contribute between 20% and 80% to the unavailability of safety systems of nuclear power reactors, as mentioned in some PRA studies [69]. It has been shown that the occurrence of CCF tends to increase system failure probabilities. If CCF modeling is not included within the PRA or without careful insight of CCF, the system reliability model can result in a gross overestimation of system safety, which makes reliability analysis less effective in the system design, for example. Therefore, treatment of CCFs has been a key topic in reliability assessment of systems. It is also the major concern in a large number of reports of U.S Nuclear Regulatory Commission ([36], [45], [46], [47]).

In analyzing the effect of CCF on a system reliability, it is important to understand the concept of CCF. However, the definition of CCF is not unique and there are still some inconsistencies over its definition. Let us outline a number of different definitions of CCF found in the literature.

The first formal definition of CCF appeared in nuclear industry, under the term Common Mode Failures (CMF). In 1975, Rasmussen [53] define CCF as multiple failures which are not independent. People in different industry sectors may have different opinions of what a CCF event is ([54]).

By comparing nine different definitions which used 12 attributes to describe CCF, Smith and Watson [57] select the following attributes:

- (1) *the components affected are unable to perform as required;*
- (2) *multiple failures exist within (but not limited to) redundant configurations;*
- (3) *the failures are “first in line” type or failures and not the result of cascading failures;*
- (4) *the failures occur within a defined critical time period;*
- (5) *The failures are due to a single underlying defect or a physical phenomenon; and*
- (6) *The effect of failures must lead to some major disabling of the system’s ability to perform as required.*

The definitions of CCF is described differently in some references related to specific sectors:

- In nuclear industry, the CCF is defined as *dependent failures in which two or more components fault states exist at the same time, or in a short time interval, and are a direct result of a shared cause ([47]).*
- In oil and gas industry, CCFs are *the components/items within the same component group that fail due to the same root cause within a specified time ([14])*
- In electronic safety-related systems, a CCF *is the failure that is the result of one or more events, causing concurrent failures of two or more separate channels in a multiple channel system, leading to system failure ([16]).*

A technical note on the definition of common-cause failures was provided by Paula [52]. The author acknowledged that CCF is defined by the analyst and both general and narrow definitions may be true. Moreover, the author explained further: *CCF events are dependent failures resulting from causes that are not explicitly modeled.*

A more detail definition of CCF was provided by [45], that is:

A CCF event consists of component failures that meet four criteria:

- (1) *two or more individual components fail or are degraded, including failures during demand, in-service testing, or deficiencies that would have resulted in a failure if a demand signal had been received;*

- (2) components fail within a selected period of time such that success of the PRA mission would be uncertain;*
- (3) component failures result from a single shared cause and coupling mechanism; and*
- (4) a component failure occurs within the established component boundary.*

Despite some main points in common from these above definitions, O'Connor [51] shows that there are still uses and misuses of the term in current literature, adding to the confusion. He proposed a general definition of CCF as follows.

A CCF event consists of component failures that meet five criteria:

- (1) two or more individual components fail or are degraded, including failures during demand, in-service testing, or deficiencies that would have resulted in a failure if a demand signal had been received;*
- (2) components fail within a selected period of time such that multiple components are unable to perform their intended function or success of the PRA mission would be uncertain;*
- (3) component failures result from a single shared cause and coupling mechanism;*
- (4) a component failure occurs within the established component boundary; and*
- (5) the dependency between components has not already been explicitly modeled.*

Since we are going to drive the statistical analysis of CCF, we will consider a CCF as a random variable and model it with counting processes. This approach will allow us to handle different situations of missing information that are current in CCF study.

Causes of CCF

In general, CCF results from the availability of two factors, involving failure cause and coupling factor.

The failure cause of an event is a condition in which a change in the state of a component can be attributed. The definition of cause can be applied to different levels such as *proximity cause* or *root cause*. A proximity cause is associated with a component failure that is readily identifiable condition leading to the failure. A root cause is an initiating cause of a causal chain resulting in eventual failure of a component. Compared to root causes,

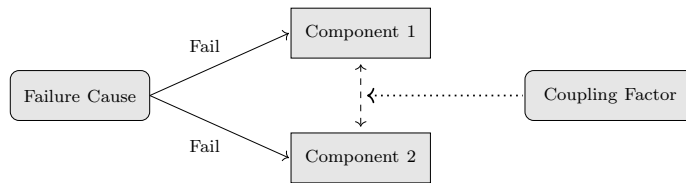


Figure 1.1: Failure cause and coupling factor of CCF [51]

proximity causes are more identifiable but do not necessarily reflect the complete understanding of the failure mechanism. Failure cause is related to design (design error, manufacturing error, installation error), operations or human (accidental action, incorrect procedure, failure to follow procedure, inadequate maintenance), external environment (fire or smoke, humidity or moisture, high or low temperature, acts of nature like wind, flood, lightning, snow and ice), internal to component (normal wear, internal environment, early failure) and miscellaneous [47].

The coupling factor is the propagation mechanism that allows failure of multiple components. Without coupling factor, a single root cause cannot cause multiple component failures. Coupling factors usually include the similarity in design, location, environment, mission and operational, maintenance, and test procedures. Coupling factor is related to environment, hardware design, hardware quality and operations maintenance [47].

Figure 1.1 illustrate the relation between root cause and coupling factor. More detail and discussion about causes of CCF can be found in [51].

CCF data

Many situations of CCF data could be observed in practice. They can be classified into two major categories, complete data and missing data. However, the definition of each category is not always the same: it depends on the context of study.

The *complete data* means that the needed information for analyzing the CCF data in a specific context is available. For example, in several situations, complete data means that the number of failed components in each CCF is known exactly. In other situations, it means that not only the number of failed components but also the cause of each CCF event need to be reported precisely.

In contrast, the *incomplete data* refers to data where some required information is not available, i.e., incomplete or missing. It could be the case when the observed failure of several components is not a CCF; when one does not know accurately the number of failed component in each CCF event; or when the cause of CCF event is not defined. Data with uncertainties and mapped

data, that we are going to define latter, are some examples of incomplete data.

There are two characteristics of the database used in common-cause failure events analysis. They are (a) the data consist mainly of narrative descriptions of observed failure events and (b) the number of events in the data is small. Both of them usually lead to uncertainties in the analysis of CCF events as the first characteristic requires some interpretation of the narratives and the second one forces us to pool data from other systems in creating a database for the system being analyzed. In these situations, it is difficult to determine exactly the number of components failed in CCF events. The data are called *data with uncertainties*.

The CCF data are often sparse. Collecting generic CCF data from others systems is widely used as an efficient strategy to deal with the problem of lacking information. But since the collected data come from systems of different size, they need to be transformed in order to be considered as coming from systems of the same size. Many procedures can be found in the literature to do that [21]. The obtained data are called *mapped data*.

CCF parametric models

Parametric models have been the primary approach for modeling and quantifying the CCF events. There are different CCF models corresponding to various definitions and assumptions of CCF events. In the literature, the parameteric CCF models are classified into five groups, including direct estimate model, ratio models, shock models, inference model and others. The advantages and the limitations of each kind of models have been discussed broadly by O'Connor [51]. In this dissertation, the Binomial Failure Rate (BFR) model and the α -factor model are applied. We also introduce a model based on Poisson distribution to deal with CCF data that contain not only occurence frequencies but also the triggering causes.

Statistical tools for CCF inference

A number of statistical methods are applied in this study. In dealing with a problem of missing data, we suggest using an EM algorithm (Appendix A) to obtain the MLE of parameters. Since the CCF data are usually sparse, the Bayesian scheme is widely applied. We address the problem of prior hyper-parameters elicitation.

To deal with the duality complete data versus incomplete data, we develop strategies relying on the Inverse Bayes Formula (IBF - Appendix B) and the Metropolis-Hastings algorithm (Appendix D).

Contributions and the structure of the thesis

Since treating CCF event plays an important role in PRA, the objective of this thesis is to develop statistical methods to analyze CCF data based on

different kind of data.

In chapter 2, we present the CCF inference for the BFR model in the case of incomplete data. Our work leads us to introduce a new distribution to create a conjugate prior for Bayesian analysis of the BFR model. We provide an efficient strategy to elicit prior hyper parameters based on expert's experience. This work has been published in an international journal [44].

In chapter 3, we propose a method to deal with data with uncertainties and to reduce the subjectivity in the traditional impact vector method based on the α -factor model. A new representation of data with uncertainties has been suggested. Considering some particular schemes, we provide several approaches to treat this new form of data with uncertainties.

In chapter 4, we present a novel method to handle the causality-based CCF data. We combine an IBF algorithm with a Metropolis-Hastings algorithm to drive the inference in the context of contingency table where only the margins are observed. In some certain conditions, we obtain an explicit formula for the posterior distribution of parameters interested.

Finally, a framework for predicting the number of catastrophic events is given in chapter 5. Still in the context of uncertainties, we have found that the approximate distributions can work efficiently to avoid the numerical integration.

Chapter 2

Inference on BFR model from confounded data

This chapter considers the statistical analysis of the Binomial Failure Rate (BFR) model in details. Computational aspects of maximum likelihood and Bayesian methods are investigated. An EM algorithm to obtain maximum likelihood estimates is suggested to deal with missing data inherent for common-cause failures. A Bayesian approach is developed and the modified-Beta distribution is introduced to characterize the posterior distribution of one of the parameters of the model. All the suggested methods are applied and compared on both simulated and real data.

2.1 Introduction

The BFR model was firstly proposed by Vesely [68] to describe the underlying failure processes generated by CCF events. This model can be considered as a specialization of the multivariate exponential model introduced by Marshall & Olkin [31]. The main idea is to model component failures in terms of individual item failures and outside shocks that affect the survival status of all the components in a system. Individual components failures are described using independent and identically distributed exponential lifetimes with common failure rate λ . Other parameters in the first version of BFR model are μ , the rate of CCF, and p , the probability that each component fails because of outside common cause shocks. When an outside shock occurs, it is assumed that each component has the same probability to fail, independently from each others. The number of components failing due to a shock follows a binomial distribution, leading to the name of Binomial Failure Rate (BFR) model.

According to a discussion in [26], some specific situations are not covered

by the first BFR model. In order to handle this practical problem, Atwood [2] suggested adding an independent lethal shock modeled by a Poisson distribution with a rate ω . A lethal shock provokes simultaneous failures of all the components of the system. The model became a four-parameter BFR model and has been readily accepted in the nuclear industry [26].

An important assumption in the BFR model is that the probability to fail for a component due to external shock is constant from shock-to-shock. This is sometimes controversial as shocks are typically produced with varying strength. As a consequence, p can conceivably change from one shock to others. This suggests a mixture model investigated in Kvam [22, 23] where a Beta distribution is associated with the parameter of the binomial. As another way to deal with this issue, Hauptmanns [15] considers a Multi-Class BFR model.

In practice, the number of failures observed is small and in the same time prior information on the system is available (from related databases or experts). This naturally leads to the application a Bayesian approach. Kvam and Martz [24] develop a simple Bayesian model for estimating component reliability and the corresponding probability of common cause failures in small operating systems. Atwood and Kelly [5] propose to use WinBUGS, a popular Markov Chain Monte Carlo (MCMC) program, for Bayesian estimation in the context of common cause failures. Table 2.1 presents a short overview of the studies related to the BFR model found in the literature.

Another important point in the BFR model is that it does not require any analysis about observed data but counting the number of failed components. In this chapter, we consider a practical situation described as follows:

1. when one component fails, it is not possible to know whether the failure is independent or due to a shock,
2. when all the components fail, it is not possible to know whether the failure corresponds to a CCF or whether it is due to a lethal shock.

We explore some methodologies to deal with this situation using both frequentist and Bayesian approaches. The main results in this chapter have been presented at the 4th Conference on the Interface between Statistics and Engineering (ICISE 2016 [42]), the 10th International Conference on Mathematical Methods in Reliability (MMR 2017 [43]), and published in Reliability Engineering and System Safety [44].

2.2 The data and the model

The model

Consider a system of m identical components. When a failure occurs, the system is supposed to be repaired immediately and the failed components

Reference	The model description	Estimation method
Vesely [68]	BFR model with three parameters: λ, μ and p	MLE
Atwood [2]	Adding the lethal shocks with the rate ω	Discussion several perspectives
Kvam [26]	BFR model with four parameters: λ, μ, ω and p	MLE, method of moment.
Kvam [27]	A variant of BFR with only two parameters: p and q , where q the probability of lethal shocks	MLE
Kvam and Martz [24]	A variant of BFR with only two parameters: p and q , where q the probability of lethal shocks	Bayesian method
Hauptmanns [15]	Multil-class BFR	MLE
Kvam [22]	Nonparametric BFR mixture model	MLE
Kvam [23]	Parametric BFR mixture model, p is allowed to change with shocks	MLE
Atwood and Kelly [5]	Three different versions of BFR model	Bayesian method with WinBUGS

Table 2.1: The studies related to BFR model in the literature

are replaced with no delay. As mentioned before, failures occurring on a system can be of different natures. They can be due to a shock or they can be independent. In general, failures of the system can be classified into three following categories:

- independent failures,
- common cause failures,
- lethal failures.

Independent failures

Each component of the system can fail independently at random times with the same constant failure rate, λ . Such failures will be termed independent failures to avoid the confusion with failures occurring from outside shocks. Assuming that each component is replaced immediately whenever it fails, the independent failure counting process of a component is a homogeneous Poisson process with intensity λ . We denote this process by $\{N_I(t), t \geq 0\}$.

Common cause failures

Sometimes the system may be exposed to extraneous and unexpected events. For example, a human error or an extreme environment such as abnormal temperature or vibration can provoke simultaneous failure of many components. We call these synchronized failures common cause failures.

Assume that the rate of occurrence of shocks that cause failures is constant equal to μ and when a shock occurs, each component may fail, independently of each others, with probability p . Denote Y the number of failed components due to a shock. Since the shocks with no failures are not observed, we suppose that Y only takes its values in $\{1, \dots, m\}$. The event $\{Y = i\}$ corresponds to the occurrence of a CCF of order i , that means a CCF implying i components. Y is a random variable that follows a truncated binomial distribution with parameters (m, p) :

$$P(Y = i) = \frac{\binom{m}{i} p^i (1-p)^{m-i}}{1 - (1-p)^m}, \quad i = 1, \dots, m. \quad (2.2.1)$$

This is the conditional distribution of the CCF order given that the order is greater than or equal to 1.

Lethal failures

To end with, we consider another type of shock that causes simultaneously failures of all components. The failure induced by a lethal shock should be distinguished from CCF order m . Let ω be the lethal-shock rate which is assumed to be constant. The process of this type of shock will be a homogeneous Poisson process with intensity ω , denoted by $\{N_L(t), t \geq 0\}$.

From this classification of failures, we propose a general definition for BFR model as random processes.

Definition 1. – *A BFR model for a m -component system includes $m + 2$ independent homogeneous Poisson processes*

- $\{N_I(t), t \geq 0\}$, the process of independent failures with intensity λ ,
- $\{N_i(t), t \geq 0\}$, the process of CCF of order i with intensity $\mu U_i(p)$, where $U_i(p)$ is the probability (2.2.1), $i = 1, \dots, m$,
- $\{N_L(t), t \geq 0\}$, the process of lethal shocks with intensity ω .

The purpose is to estimate:

- λ , the rate of occurrence of independent failures,
- μ , the rate of occurrence of the nonlethal shocks that produce at least one failure,
- p , the probability that a component fails due to the occurrence non-lethal shock,
- ω , the rate of occurrence of nonlethal shocks.

The estimation strategy will depend on the nature of observed data.

Data

Suppose a time window $[0, T]$ where T is fixed, if the numbers $N_I(T), N_1(T), \dots, N_m(T)$ and $N_L(T)$ are observed, the data are said to be complete. In the sequel, we omit T to lighten the notation. We denote $\mathbf{N} = (N_I, N_1, \dots, N_m, N_L)$ and $\mathbf{n} = (n_I, n_1, \dots, n_m, n_L)$ as a realization of \mathbf{N} .

In practice, given that it is not possible to directly observe the shocks arrival, it is not possible to distinguish the lethal shocks from CCF of order m and the CCF of order 1 from independent failures. These observed data are of the form $\mathbf{N}^* = (N_1^*, N_2, \dots, N_{m-1}, N_m^*)$, where $N_1^* = N_I + N_1$ and $N_m^* = N_m + N_L$. The data are said to be *incomplete*. Let $\mathbf{n}^* = (n_1^*, \dots, n_m^*)$ denote a realization of \mathbf{N}^* . An example of data of this form is given in [23]. It is a set of real failure data from a 4-component safety system of a nuclear power plant (NPP), where we observe $N_1^* = 11, N_2 = 10, N_3 = 7$ and $N_4^* = 6$.

In the following, we investigate different methods to make estimation. Firstly, we consider the maximum likelihood method with complete data and with incomplete data. Then we develop a Bayesian approach in the case of incomplete data introducing conjugate priors.

2.3 Inference

2.3.1 Maximum likelihood method

Complete data

With the observed data $\mathbf{n} = (n_I, n_1, \dots, n_m, n_L)$, the likelihood function is:

$$\begin{aligned} L(\theta) &= P(N_I(T) = n_I) \prod_{i=1}^m P(N_i(T) = n_i) P(N_L(T) = n_L) \\ &\propto \lambda^{n_I} e^{-\lambda T} \left\{ \prod_{i=1}^m [\mu U_i(p)]^{n_i} e^{-\mu U_i(p)T} \right\} \omega^{n_L} e^{-\omega T}, \end{aligned} \quad (2.3.1)$$

where $U_i(p) = \frac{\binom{m}{i} p^i (1-p)^{m-i}}{1-p^m}$, $i = 1, \dots, m$ and θ stands for $(\lambda, \mu, \omega, p)$. The log-likelihood is

$$\begin{aligned} \log L(\theta) &= n_I \log \lambda + s_C \log \mu + \sum_{i=1}^m n_i \log U_i(p) + n_L \log \omega \\ &\quad - (\lambda + \mu + \omega)T. \end{aligned}$$

where $s_C = \sum_{i=1}^m n_i$. It is easy to solve the likelihood equations for λ, μ and ω , which lead to the estimates as:

$$\hat{\lambda} = \frac{n_I}{T}, \hat{\mu} = \frac{s_C}{T}, \hat{\omega} = \frac{n_L}{T}.$$

The solution for p is not straightforward. Indeed, the term involving p in the likelihood function can be expressed as:

$$\begin{aligned} \sum_{i=1}^m n_i \log U_i(p) &= \sum_{i=1}^m n_i \log[(1-q)^i q^{m-i} / (1-q^m)] \\ &= s_D \log(1-q) + (m s_C - s_D) \log q - s_C \log(1-q^m) \end{aligned}$$

where $s_D = \sum_{i=1}^m i n_i$ and $q = 1 - p$.

The corresponding likelihood equation is

$$-\frac{s_D}{1-q} + \frac{m s_C - s_D}{q} + \frac{m s_C q^{m-1}}{1-q^m} = 0.$$

This equation is equivalent to $(1-q) m s_C - (1-q^m) s_D = 0$.

Since $0 < q < 1$, we obtain

$$s_D (q^{m-1} + q^{m-2} + \dots + q + 1) - m s_C = 0.$$

The last equation has a unique solution in $[0, 1]$ that is obtained using a numerical method.

From the calculation for the mean of a Poisson distribution, we have:

$$\begin{aligned} E(\hat{\lambda}) &= \frac{E(N_I(T))}{T} = \frac{\lambda T}{T} = \lambda \\ E(\hat{\omega}) &= \frac{E(N_L(T))}{T} = \frac{\omega T}{T} = \omega \\ E(\hat{\mu}) &= \frac{E(s_C(T))}{T} = \frac{\sum_{i=1}^m \mu U_i(p) T}{T} = \mu \sum_{i=1}^m \frac{C_m^i p^i (1-p)^{m-i}}{1-(1-p)^m} = \mu \end{aligned}$$

That means the estimators of μ , λ and ω are unbiased.

Table 2.2 presents the results from a simulation study which are the average of the estimates obtained from 1000 simulations. The result allows us to conjecture that the estimator of p is asymptotically unbiased.

Parameters	λ	μ	ω	p
Input	0.12500	0.25000	0.08333	0.5000
Estimates	0.12558	0.25182	0.08342	0.49727

Table 2.2: Average of estimates for complete data based on 1000 simulations

Incomplete data

When the data are incomplete, we have $N_1^* = N_I + N_1$ and $N_m^* = N_m + N_L$. They are the summation of two Poisson random variables.

By the additive property of the Poisson distribution, N_1^* follows a Poisson with parameter $\lambda T + \mu U_1(p)T$ and N_m^* has a Poisson distribution with parameter $\omega T + \mu U_m(p)T$.

Thus, the likelihood function is:

$$\begin{aligned} L(\theta) &= P(N_1^*(T) = n_1^*) \prod_{i=2}^{m-1} P(N_i(T) = n_i) P(N_m^*(T) = n_m^*) \\ &\propto [\lambda + \mu U_1(p)]^{n_1^*} e^{-[\lambda + \mu U_1(p)]T} \prod_{i=2}^{m-1} \left([\mu U_i(p)]^{n_i} e^{-\mu U_i(p)T} \right) \\ &\quad [\omega + \mu U_m(p)]^{n_m^*} e^{-[\omega + \mu U_m(p)]T}. \end{aligned} \quad (2.3.2)$$

Therefore, the log-likelihood is:

$$\begin{aligned} \log L(\theta) &= n_1^* \log[\lambda + \mu U_1(p)] + \sum_{i=2}^{m-1} n_i \log \mu + \sum_{i=2}^{m-1} n_i \log U_i(p) \\ &\quad + n_m^* \log[\omega + \mu U_m(p)] - (\lambda + \mu + \omega)T. \end{aligned}$$

The likelihood equations are:

$$\begin{cases} \frac{\partial}{\partial \lambda} \log L(\theta) = \frac{n_1^*}{\lambda + \mu U_1(p)} - T = 0 \\ \frac{\partial}{\partial \mu} \log L(\theta) = \frac{n_1^* U_1(p)}{\lambda + \mu U_1(p)} + \frac{\sum_{i=2}^{m-1} n_i}{\mu} + \frac{n_m^* U_m(p)}{\omega + \mu U_m(p)} - T = 0 \\ \frac{\partial}{\partial \omega} \log L(\theta) = \frac{n_m^*}{\omega + \mu U_m(p)} - T = 0 \\ \frac{\partial}{\partial p} \log L(\theta) = \frac{n_1^* \mu U_1'(p)}{\lambda + \mu U_1(p)} + \sum_{i=2}^{m-1} \frac{n_i U_i'(p)}{U_i(p)} + \frac{n_m^* \mu U_m'(p)}{\omega + \mu U_m(p)} = 0 \end{cases}$$

The likelihood equations have no explicit solutions.

Since we have a situation of missing data, we suggest using an EM algorithm to solve the equations. The distribution of the observed data belongs to the exponential family. Therefore, applying EM consists in replacing in the likelihood of the complete data, the missing data with their conditional expectations given the observed data.

In our case, the conditional distributions are obtained from a well-know property of the Poisson distribution, leading to the following binomial distributions:

$$\begin{aligned} N_1(T) | N_1^*(T) = n_1^* &\sim \mathcal{B}(n_1^*, \lambda / [\lambda + \mu U_1(p)]), \\ N_1(T) | N_1^*(T) = n_1^* &\sim \mathcal{B}(n_1^*, \mu U_1(p) / [\lambda + \mu U_1(p)]), \\ N_L(T) | N_m^*(T) = n_m^* &\sim \mathcal{B}(n_m^*, \omega / [\omega + \mu U_m(p)]), \\ N_m(T) | N_m^*(T) = n_m^* &\sim \mathcal{B}(n_m^*, \mu U_m(p) / [\omega + \mu U_m(p)]). \end{aligned}$$

Therefore, the EM algorithm is performed as follows.

- Initialisation: $\theta^{(0)} = (\lambda^{(0)}, \mu^{(0)}, \omega^{(0)}, p^{(0)})$
- At step $(r+1)$, $\theta^{(r)} = (\lambda^{(r)}, \mu^{(r)}, \omega^{(r)}, p^{(r)})$ being available, compute:
 - $\lambda^{(r+1)} = \frac{1}{T} E[N_{ind}(T) | N_1^*(T) = n_1^*; \theta^{(r)}]$
 that is to say $\lambda^{(r+1)} = \frac{n_1^* \lambda^{(r)}}{[\lambda^{(r)} + \mu^{(r)} U_1(p^{(r)})] T}$
 - $\omega^{(r+1)} = \frac{1}{T} E[N_L(T) | N_m^*(T) = n_m^*; \theta^{(r)}]$
 that is to say $\omega^{(r+1)} = \frac{n_m^* \omega^{(r)}}{[\omega^{(r)} + \mu^{(r)} U_m(p^{(r)})] T}$
 - $\mu^{(r+1)} = \frac{1}{T} \left\{ E[N_1(T) | N_1^*(T) = n_1^*; \theta^{(r)}] + \sum_{i=2}^{m-1} n_i + E[N_m(T) | N_m^*(T) = n_m^*; \theta^{(r)}] \right\}$
 that is to say

$$\begin{aligned} \mu^{(r+1)} &= \frac{1}{T} \left\{ \frac{n_1^* \mu^{(r)} U_1(p^{(r)})}{[\lambda^{(r)} + \mu^{(r)} U_1(p^{(r)})] T} + \sum_{i=2}^{m-1} n_i + \frac{n_m^* \mu^{(r)} U_m(p^{(r)})}{[\omega^{(r)} + \mu^{(r)} U_m(p^{(r)})] T} \right\} \\ &= \frac{1}{T} \left(n_1^* + \sum_{i=2}^{m-1} n_i + n_m^* - \lambda^{(r+1)} - \omega^{(r+1)} \right) \end{aligned}$$
 - $p^{(r+1)} = 1 - q^{(r+1)}$ where $q^{(r+1)}$ is the unique solution of the equation

$$m s_C^{(r)} (1 - q) - s_D^{(r)} (1 - q^m) = 0,$$

$$s_C^{(r)} = \frac{n_1^* U_1(p^{(r)}) \mu^{(r)}}{\lambda^{(r)} + U_1(p^{(r)}) \mu^{(r)}} + \sum_{i=2}^{m-1} n_i + \frac{n_m^* U_m(p^{(r)}) \mu^{(r)}}{\omega^{(r)} + U_m(p^{(r)}) \mu^{(r)}}$$

and

$$s_D^{(r)} = \frac{n_1^* U_1(p^{(r)}) \mu^{(r)}}{\lambda^{(r)} + U_1(p^{(r)}) \mu^{(r)}} + \sum_{i=2}^{m-1} i n_i + m \frac{n_m^* U_m(p^{(r)}) \mu^{(r)}}{\omega^{(r)} + U_m(p^{(r)}) \mu^{(r)}}$$

- Stop when $\|\theta^{(r+1)} - \theta^{(r)}\|$ is sufficiently small.

A simulation study

We investigate the performance of the proposed EM algorithm through simulations. Considering a 5-component system and an observation time window $T = 120$ (10 years), the parameters are set to cover two situations.

- **Situation I:** $\lambda = 0.0833$, $\mu = 0.1$, $\omega = 0.0556$ and $p = 0.4$.

This combination corresponds to an expected numbers equal to one individual failure in a time interval of 12 months, one common-cause shock per 10 months and one lethal shock per 18 months. The probability that a component fails when a common-cause shock occurs is 0.4.

- **Situation II:** $\lambda = 0.025$, $\mu = 0.0333$, $\omega = 0.0167$ and $p = 0.25$.

In this situation, λ , μ and ω can be interpreted respectively as an average number of one individual failure every 40 months, one common-cause failure every 30 months and one lethal shock every 60 months, respectively. When a common-cause shock occurs, a component has a probability 0.25 to fail. In this case, we will obtain sparse observations.

The tables 2.4 and 2.5 provide the estimation results and the corresponding mean square error (MSE). Computations are performed with 1000 simulations using the EM algorithm with initial values $\lambda^{(0)} = .01$, $\mu^{(0)} = .015$, $\omega^{(0)} = .01$, $p^{(0)} = .1$. The initial values do not influence the convergence, which is usually obtained after less than 30 iterations. The estimates of the parameters in the first set of input are relatively close to the theoretical values. The results are a little less effective in the second set of input as might be anticipated because of the small expected number of failures. That outcome explains the poor results obtained.

Application on real data

We consider the CCF events involving emergency diesel generators at US nuclear power-plants presented mentioned previously. These data are extracted from an unpublished database from the US Nuclear Regulatory Commission. The length T of the time window observation is unknown. In the case of 4-component systems, we observe $N_1^* = 11$, $N_2 = 10$, $N_3 = 7$ and $N_4^* = 6$. Let us denote c_λ , c_μ and c_ω the MLE of λT , μT and ωT . Applying the EM algorithm we obtained $c_\lambda = 4.6483$, $c_\mu = 25.1883$, and $c_\omega = 4.1633$. Then the MLEs are $\hat{\lambda} = c_\lambda/T$, $\hat{\mu} = c_\mu/T$ and $\hat{\omega} = c_\omega/T$. For example, for $T = 15$ (years), we have: $\hat{\lambda} = 0.3098 \text{ year}^{-1}$, $\hat{\mu} = 1.6792 \text{ year}^{-1}$, $\hat{\omega} = 0.2775 \text{ year}^{-1}$ and $\hat{p} = 0.5121$. The mean time between CCFs can be obtained as T/c_λ , T/c_μ and T/c_ω .

2.3.2 Bayesian approach

In the context of CCF analysis, the information is sparse. Therefore considering a Bayesian approach could be a solution to make inference on the model. The distribution of the incomplete data \mathbf{N}^* can be rewritten in the form:

$$P(\mathbf{N}^* = \mathbf{n}^* | \theta) \propto \left(\sum_{k=0}^{n_1^*} C_{n_1^*}^k (\mu U_1(p))^k \lambda^{n_1^* - k} \right) \prod_{i=2}^{m-1} [\mu U_i(p)]^{n_i} \\ \times \left(\sum_{\ell=0}^{n_m^*} C_{n_m^*}^\ell (\mu U_m(p))^\ell \omega^{n_m^* - \ell} \right) e^{-(\lambda + \mu + \omega)T}$$

$$\begin{aligned} & \propto \sum_{k=0}^{n_1^*} \sum_{\ell=0}^{n_m^*} m^k C_{n_1^*}^k C_{n_m^*}^\ell \lambda^{n_1^*-k} e^{-\lambda T} \mu^{k+\ell+n_C} e^{-\mu T} \\ & \times \omega^{n_m^*-\ell} e^{-\omega t} \frac{p^{k+n_D+m\ell} q^{k(m-1)+mn_C-n_D}}{(1-q^m)^{k+n_C+\ell}} \end{aligned} \quad (2.3.3)$$

where $n_D = \sum_{i=2}^{m-1} iN_i$, $n_C = \sum_{i=2}^{m-1} N_i$ and $q = 1 - p$.

Assuming that the parameters are independent, we consider a gamma prior for the parameters λ , μ , ω with respective parameters $(\alpha_\lambda, \beta_\lambda)$, (μ_μ, β_μ) and $(\alpha_\omega, \beta_\omega)$ and a Beta distribution for p with parameters (α_p, β_p) .

Applying the Bayes formula, the posterior distribution for λ , μ and ω , is obtained as a mixture of Gamma distributions:

$$\lambda \mid \mathbf{N}^* = \mathbf{n}^* \sim \sum_{k=0}^{n_1^*} \eta_\lambda^{(k)} \mathcal{G}(n_1^* + \alpha_\lambda - k, T + \beta_\lambda), \quad (2.3.4)$$

$$\mu \mid \mathbf{N}^* = \mathbf{n}^* \sim \sum_{k=0}^{n_1^*} \sum_{\ell=0}^{n_m^*} \eta_\mu^{(k,\ell)} \mathcal{G}(n_C + k + \ell + \alpha_\mu, T + \beta_\mu), \quad (2.3.5)$$

$$\omega \mid \mathbf{N}^* = \mathbf{n}^* \sim \sum_{\ell=0}^{n_m^*} \eta_\omega^{(\ell)} \mathcal{G}(n_m^* + \alpha_\omega - \ell, T + \beta_\omega), \quad (2.3.6)$$

where $\eta_\lambda^{(k)}$, $\eta_\omega^{(\ell)}$ and $\eta_\mu^{(k,\ell)}$ are the weights calculated by integrating adequately the joint distribution $P(\mathbf{N}^* = \mathbf{n}^* \mid \theta)\pi(\theta)$ and computing $P(\mathbf{N}^* = \mathbf{n}^*)$.

To characterize the posterior distribution of p , we introduce the **modified-Beta distribution**.

Definition 1. – A continuous random variable X with support $]0,1[$ follows a **modified-Beta** distribution with parameters $(\alpha, \beta, \gamma, \delta)$ if and only if, its probability density function is

$$f(x) = \frac{1}{K(\alpha, \beta, \gamma, \delta)} \frac{x^{\alpha-1}(1-x)^{\beta-1}}{(1-(1-x)^\delta)^\gamma}$$

where $x \in]0, 1[$, $\delta \in \mathbb{N}$, $\alpha, \beta, \gamma > 0$, with $\gamma < \alpha$ and

$$K(\alpha, \beta, \gamma, \delta) = \int_0^1 \frac{x^{\alpha-1}(1-x)^{\beta-1}}{(1-(1-x)^\delta)^\gamma} dx.$$

We denote $X \sim \tilde{\mathcal{B}}(\alpha, \beta, \gamma, \delta)$.

Relying on this distribution, we can claim now that the posterior distribution of p is mixture of modified-Beta of the following form:

$$p \mid \mathbf{N}^* = \mathbf{n}^* \sim \sum_{k=0}^{n_1^*} \sum_{\ell=0}^{n_m^*} \eta_p^{(k,\ell)} \tilde{\mathcal{B}}(\alpha_{(k,\ell)}, \beta_{(k,\ell)}, \gamma_{(k,\ell)}, m) \quad (2.3.7)$$

where $\alpha_{(k,\ell)} = k + n_D + m\ell + \alpha_p$, $\beta_{(k,\ell)} = k(m-1) + mn_C - n_D + \beta_p$, $\gamma_{(k,\ell)} = k + n_C + \ell$ and $\eta_p^{(k,\ell)}$ is the weight calculated by integrating adequately the joint distribution $P(\mathbf{N}^* = \mathbf{n}^* | \theta)\pi(\theta)$ and computing $P(\mathbf{N}^* = \mathbf{n}^*)$.

More details on the modified-Beta distribution are given in Appendix (C). The expectation of the modified-Beta distribution is expressed as:

$$E(X) = \frac{K(\alpha + 1, \beta, \gamma, \delta)}{K(\alpha, \beta, \gamma, \delta)}. \quad (2.3.8)$$

The expression (2.3.8) will be used in the sequel to obtain the Bayes estimator of the parameter p .

Bayes estimators

Considering a quadratic loss function, the Bayes estimators of the parameters are the expected-values of the posterior distributions.

$$\tilde{\lambda} = \sum_{k=0}^{n_1^*} \eta_\lambda^{(k)} \frac{\alpha_\lambda + n_1^* - k}{T + \beta_\lambda}, \quad (2.3.9)$$

$$\tilde{\mu} = \sum_{k=0}^{n_1^*} \sum_{\ell=0}^{n_m^*} \eta_\mu^{(k,\ell)} \frac{n_C + k + \ell + \alpha_\mu}{T + \beta_\mu}, \quad (2.3.10)$$

$$\tilde{\omega} = \sum_{\ell=0}^{n_m^*} \eta_\omega^{(\ell)} \frac{\alpha_\omega + n_m^* - \ell}{T + \beta_\omega}, \quad (2.3.11)$$

$$\tilde{p} = \sum_{k=0}^{n_1^*} \sum_{\ell=0}^{n_m^*} \eta_p^{(k,\ell)} \frac{K(\alpha_{(k,\ell)} + 1, \beta_{(k,\ell)}, \gamma_{(k,\ell)}, m)}{K(\alpha_{(k,\ell)}, \beta_{(k,\ell)}, \gamma_{(k,\ell)}, m)}. \quad (2.3.12)$$

Note that one can also consider the mode of the posterior distributions as a possible estimate.

A $(1 - \alpha)$ credibility interval for each parameter can be numerically computed relying on the Newton-Raphson (NR) method. Let $F_{\theta|\mathbf{N}^*=\mathbf{n}^*}$ denote the posterior cumulative distribution function of the parameters. The bound of the interval $[a, b]$, a $(1 - \alpha)$ -credibility region, are obtained solving the two following equations

$$\begin{cases} F_{\theta|\mathbf{N}^*=\mathbf{n}^*}(a) = 1 - \alpha/2 \\ F_{\theta|\mathbf{N}^*=\mathbf{n}^*}(b) = \alpha/2 \end{cases}. \quad (2.3.13)$$

These equations can be solved using a NR algorithm. For example for the lower bound we use the following routine.

- Choose an initial value $a^{(0)}$. It could be the posterior mean.
- At setp $(n+1)$, $a^{(n)}$ being available, compute

$$a^{(n+1)} = a^{(n)} - \frac{F_{\theta|\mathbf{N}^*=\mathbf{n}^*}(a^{(n)})}{\pi(a^{(n)} | \mathbf{N}^* = \mathbf{n}^*)}$$

- Stop when $|a^{(n+1)} - a^{(n)}| \leq \varepsilon$, where ε is close to 0.

A similar routine can be applied for the upper bound. The table 2.8 displays 95% credibility region for λ, μ, ω and p from the data extracted from a USNRC report [23].

A simulation study

Bayesian inference needs to specify values for priors hyperparameters. One can consider many strategies to obtain these values. For example, these values can be deduced from guesses at the expected-times between failures of a given type or at the number of such failures in a given time-window. Considering a confidence level in these guesses through the standard-deviation, values for the hyperparameters are solutions of a simple system of two equations. These equations are obtained matching guesses and confidence with mean and variance of gamma and beta distributions. Suppose that the prior distribution for the rate is a gamma distribution with parameters (α, β) . and that an expert guess at this rate is a value η with a confidence ρ (that is $\eta \pm \rho\eta$).

We solve

$$\begin{cases} \alpha/\beta = \eta \\ \alpha/\beta^2 = \rho^2\eta^2 \end{cases} \text{ to obtain } \begin{cases} \alpha = 1/\rho^2 \\ \beta = 1/\rho^2\eta \end{cases}.$$

For p , considering a beta distribution as a prior, the same reasoning leads to the following solutions: $\alpha = \left[\frac{1-\eta}{\rho_p^2\eta} - 1 \right] \eta$ and $\beta = \left[\frac{1-\eta}{\rho_p^2\eta} - 1 \right] (1-\eta)$ where ρ_p is the degree of confidence for the guess at p .

Let us denote $\rho_\lambda, \rho_\mu, \rho_\omega$ the degrees of confidence respectively for λ, μ, ω . It should be considered that the naive estimate of p would be n_D/mn_C where $n_D = \sum_{i=2}^{m-1} in_i$, $n_C = \sum_{i=2}^{m-1} n_i$, the average proportion of failed components in multiple failure occurrences due to non-lethal shocks. This naive estimator can be used as an initial guess for solving the numerical equation.

The table 2.7 shows some results from this strategy for both situations of data. It gives the different values of the hyperparameters considering different natures of guesses (close or not to the input values) and confidence level in these guesses (low or high). The hyperparameters of type A correspond to noninformative priors [1]. The hyperparameters of type B and C correspond to values associated to guesses close to the input values with respectively high and low levels of confidence. In the case D, the guesses at λ and p are far from the input with respectively high and low confidence, the prior for ω is noninformative and the guess at μ is close with high degree of confidence. To end with, the type E corresponds to far guesses at λ and ω with respectively high and low confidence and noninformative priors for μ and p .

The Bayesian estimation results are displayed in table 2.4 for situation I and in table 2.5 for situation II. The method gives satisfactory overall results

Table 2.3: Possible values for the hyperparameters

Sit.	Type	ρ_λ	α_λ	β_λ	ρ_μ	α_μ	β_μ	ρ_ω	α_ω	β_ω	ρ_p	α_p	β_p
I	A	*	0.5	0	*	0.5	0	*	0.5	0	*	0.5	0.5
	B	0.7	2.04	22.67	0.7	2.04	18.55	0.7	2.04	40.81	0.7	0.87	1.38
	C	0.2	25	277.7	0.2	25	227.27	0.2	25	500	0.2	14.99	23.94
	D	0.2	25	416.66	0.2	25	227.27	*	0.5	0	0.7	1.28	3.84
	E	0.7	2.04	40.81	*	0.5	0	0.2	25	833	*	0.5	0.5
II	A	*	0.5	0	*	0.5	0	*	0.5	0	*	0.5	0.5
	B	0.7	2.04	88.73	0.7	2.04	68.02	0.7	2.04	127.55	0.7	1.34	4.49
	C	0.2	25	1086	0.2	25	833	0.2	25	1562	0.2	19.02	63.67
	D	0.2	25	555.55	0.2	25	833	*	0.5	0	0.7	1.58	8.97
	E	0.7	2.04	51	*	0.5	0	0.2	25	2777	*	0.5	0.5

with small MSE. The noninformative case, the type B and C provide estimates that are relatively close to the MLE with smaller MSE. If the guesses are fairly accurate, the bayes estimates are closer to the input than the MLE. We can observe that the degree of confidence do not have a significant effect.

Table 2.4: Means of MLE and Bayesian estimates for types of prior, with mean square error (MSE) computed on 1000 simulations for situation I

	λ	μ	ω	p
Input	0.083	0.1	0.0556	0.4
MLE	0.07321 (1.7e-3)	0.11015 (1.9e-3)	0.05513 (4.5e-4)	0.39070 (1.5e-2)
A	0.07167 (1.1e-3)	0.12285 (1.9e-3)	0.05695 (4.9e-4)	0.38897 (1.2e-2)
B	0.07798 (5.6e-4)	0.10897 (7.8e-4)	0.05341 (2.8e-4)	0.39583 (8.6e-3)
C	0.08756 (6.1e-5)	0.10729 (1.9e-3)	0.05089 (3.9e-5)	0.39141 (1.5e-3)
D	0.06157 (4.8e-4)	0.11238 (2.5e-4)	0.05851 (4.8e-4)	0.35154 (9.3e-3)
E	0.06124 (9e-4)	0.12630 (1.9e-3)	0.03283 (5.2e-4)	0.38371 (1.4e-2)

Application on real data

Once again we consider the data from a US NRC report concerning CCF events involving emergency diesel generators at US nuclear plants. Let us consider the 4-component systems data. The total number of failures observed is 34 over the time period T . Recall that T is not given in the report. We assume $T = 15$ years. To obtain values for the hyperparameters we suppose that experts are able to guess at the expected numbers of events in a given period that is to say guesses at $N_{ind}(T)$, $N_i(T)$, $i = 1, \dots, m$ and $N_L(T)$. We consider three scenarii and the noninformative case. The table 2.6 gives examples of experts guesses. In the first row the total number of events is under evaluated. The guess is that the system is more reliable. In the second row, experts guesses are close to what has been observed. The third row considers that we have observed less failures than we should.

From the guesses at the number of events, we compute the MLE of λ , μ ,

Table 2.5: Means of MLE and Bayesian estimates for types of priors, with mean square error (MSE) computed on 1000 simulations for situation II

	λ	μ	ω	p
Input	0.025	0.033	0.0167	0.25
MLE	0.01791 (4e-4)	0.04982 (5e-3)	0.01568 (9.5e-4)	0.27413 (3.6e-2)
A	0.02831 (2.1e-4)	0.03892 (3.7e-4)	0.01967 (1.3e-4)	0.33441 (2.2e-2)
B	0.02331 (4.3e-5)	0.03321 (9.8e-5)	0.01655 (3.3e-5)	0.24144 (5.6e-3)
C	0.02314 (4.5e-6)	0.03050 (1.1e-5)	0.01610 (1.0e-6)	0.23261 (5.0e-3)
D	0.04235 (3.0e-4)	0.02984 (1.4e-5)	0.02113 (1.6e-4)	0.20814 (6.4e-3)
E	0.03204 (1.3e-4)	0.03548 (2.9e-4)	0.00919 (5.5e-5)	0.37332 (3.7e-2)

Table 2.6: Examples of experts guesses in the case of real data (T=15 years).

Situation	$N_{ind}(T)$	$N_1(T)$	$N_2(T)$	$N_3(T)$	$N_4(T)$	$N_L(T)$	Total
1	3	5	6	4	1	1	20
2	5	8	10	7	3	2	35
3	4	15	13	10	5	3	50

ω and p . Then we proceed as described in section 2.3.2 to propose values for the hyperparameters.

The results of this strategy are given in table 2.7. The situation 4 corresponds to noninformative priors.

Table 2.7: Choices of values for the hyperparameters for data from US NRC report.

Situation	ρ_λ	α_λ	β_λ	ρ_μ	α_μ	β_μ	ρ_ω	α_ω	β_ω	ρ_p	α_p	β_p
1	0.4	6.25	375	0.2	25	281	0.4	6.25	1125	0.3	5.33	5.84
2	0.2	25	900	0.4	6.25	40.17	0.4	6.25	562.5	0.4	2.52	2.38
3	0.4	6.25	281.25	0.2	25	104.65	0.4	6.25	375	0.2	12.14	12.41
4	*	0.5	0	*	0.5	0	*	0.5	0	*	0.5	0.5

Applying formulas (2.3.9), (2.3.10), (2.3.11) and (2.3.12), we compute the Bayesian estimates. The table 2.8 displays the results.

One can see that the different estimate for p are close and are not really sensitive to the nature of the prior. For λ and μ , the noninformative case gives estimates close to the situation 2 corresponding to guesses close to what have been observed and the results are close to the MLE. The schemes 1 and 3 (respectively under and over guess at observation) lead to similar estimate for λ . It is not the case for ω . For all parameters, the noninformative case results and the MLE are close. The figure 2.1 displays the posterior distribution of the four parameters for the data extracted from US NRC Report with prior of situation 1 (table 2.8).

Table 2.8: Bayes estimates for data extracted from US NRC report

	λ	μ	ω	p
MLE	0.01936	0.10495	0.01734	0.51212
1	0.02035 (0.011;0.087)	0.11091 (0.086;0.14)	0.00652 (0.003;0.011)	0.54651 (0.436;0.654)
2	0.02811 (0.021;0.093)	0.14575 (0.1;0.193)	0.01218 (0.005;0.02)	0.54839 (0.431;0.661)
3	0.02095 (0.011;0.107)	0.18354 (0.141;0.229)	0.01725 (0.008;0.028)	0.50738 (0.41;0.6)
4	0.02729 (0.0015;0.266)	0.14789 (0.096;0.206)	0.02203 (0.0008;0.05)	0.52323 (0.379;0.677)

2.4 Concluding remarks

Obtaining accurate estimation of the occurrence rates of common cause failure is an important issue in the assessment of system reliability. The sparseness of observation which are usually also confounded, increases the difficulty. In this chapter, we have suggested different methods to make inference on the parameters of the BFR model in this context of incomplete data. An EM algorithm is described to obtain the MLE. A methodology to elicit hyperparameters for the prior relying on expert opinion, is outlined for a Bayesian approach. This methodology will be very helpful for practitioners. We have compared the results applying the methods on both simulated data and real data. The Bayesian approach leads us to introduce a new distribution: the modified-Beta distribution.

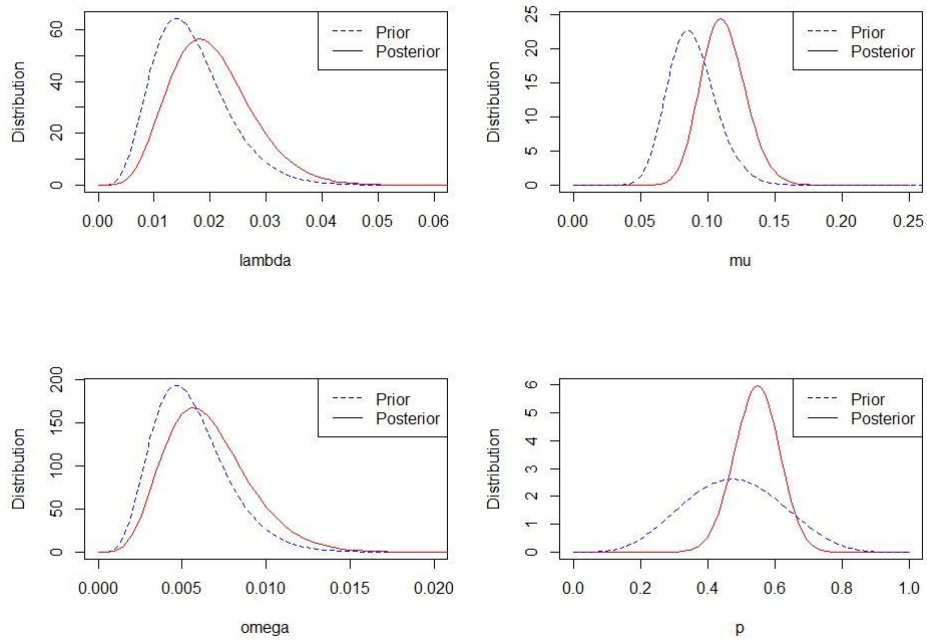


Figure 2.1: Posterior distributions of λ , μ , ω and p

Chapter 3

Inference for the α -factor model from uncertain data

In this chapter, we provide a novel method to treat data with uncertainties based on the α -factor model. The data with uncertainties refer to the cases where the order of each CCF event can not be defined exactly. We suggest a particular representation of the failure event. Both MLE and Bayesian estimation methods are considered. For the former, depending on different scenarios of data, the MLE can be obtained directly or via an EM algorithm. For the latter, the posterior distribution of the α -factor parameters are obtained by using the inverse Bayesian formula and a family of the Dirichlet distribution.

3.1 Introduction

The α -factor model is one of the most widely used model in the nuclear PRA. It is considered to be more realistic due to the ability of modeling real scenarios to a greater extent as well as analyzing various CCF events of different intensity. The model is a main concern for CCF analyzing in a number of reports from the U.S Nuclear Regulatory Commission ([45], [47],[48]).

The α -factor model is introduced by Mosleh et al. [36]. Zubair and Amjad [72] apply this model to calculate and update an unavailability due to CCF. Troffaes et al. [65] propose a Bayesian approach to the α -factor model by using a set of conjugate priors instead of a single prior. Extended versions of the α -factor model are also introduced such as the α -decomposition model [71], the partial α -factor model [49]. Duy et al. [29] suggest a practical methodology simulation-based for estimating the parameters of α -factor model for the case of incomplete or absent data. Varun et al. [12] present a pragmatic approach to estimate α -factors for CCF analysis based on mapping up technique. A consistent mapping of common cause failure rates and α -factors are

given by Vaurio [67]. Atwood [4] provides a detailed analysis of the consequences of mapping data in Bayesian common-cause inference using α -factor model.

An advantage of the α -factor model is the ability to handle data with uncertainties. The concept of data with uncertainties has been introduced in Chapter 1. In order to illustrate this concept, we reproduce the table provided by Mosleh [35]. In this example, Table 3.1, component states are represented by squares. All components were classified as either available or unavailable according to a particular success criterion. An unavailable component is either failed (\boxtimes) or functionally unavailable (\boxminus) to cover cases in which the nonfunctioning was due to the lack of required input. From the information about the failure event described in table 3.1, it seems that the third component was unavailable, but with no certainty. Thus, it is difficult for analysts to determine if it is a 2-failed component or a 3-failed component event. We call such data the data with uncertainties or uncertain data.

Plant (Date)	Event Description	Cause -Effect Diagram
Maine Yankee (August 1977)	Two diesel generators failed to run due to a plugged radiator. The third unit was also plugged	

Table 3.1: A real example of data with uncertainties [35]

A classical method to deal with data with uncertainties is to establish hypotheses about the order of each CCF event with corresponding probabilities, leading to the introduction of the **impact vector** concept. For uncertain data, estimating the parameters of α -factor model using impact vectors is applied in most of CCF references, see, for example, Siu and Mosleh [56] and Siu [55]. However, the impact vector method is based purely on the hypotheses about CCF events from the subjective experience of practitioners or experts.

The goal of this chapter is to introduce a novel method to avoid this subjectivity. We propose mixing all possible orders in each CCF event and introduce new representations for these events. In Section 2, we present the definition of the α -factor model and its parameters. Inference on α -factor model for complete data are briefly given in Section 3. In Section 4, we firstly explain the impact vector method and then describe the mixing method to provide an alternative way to handle data with uncertainties.

3.2 The α -factor model

In the α -factor model for a m -component system, the CCFs are characterized by the parameters α_k , $k = 1, \dots, m$ where α_k is the probability of the failure event that involves exactly k components, given that at least one component fails. Usually, these parameters are expressed with what is called the *basic parameters*. To understand more about notations in α -factor model, let us review briefly the basic parameters through a simple example which relies on Mosleh et al. [36].

Consider a system of three identical components, A, B and C , with two-out-of-three success logic, i.e., the system is operational if and only if at least two components out of the system's 3 component are operational. Figure 3.1 shows a classical fault tree representation of the system reliability. The min-

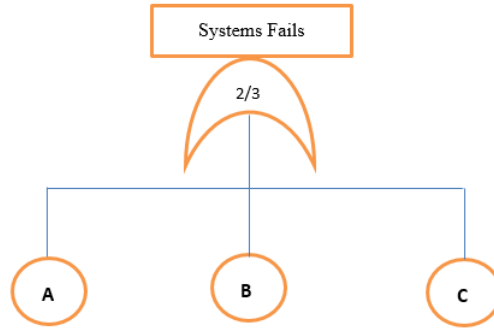


Figure 3.1: The component-level fault tree system

imal cutsets of this system failure, which are all the minimal sets of components whose failures provoke the failure of the system, are $\{A, B\}$, $\{A, C\}$, and $\{B, C\}$.

The fault tree is expanded to contain the *common-cause basic events*, which are the failure events that involved a given component. For instance, the common basic events for component A are displayed in Figure 3.2. They are $\{A_I\}$; $\{C_{AB}\}$; $\{C_{AC}\}$; $\{C_{ABC}\}$, where A_I represents the independent cause failures of component A , C_{AB} represents the CCFs implying components A and B , C_{AC} represents the CCFs implying components A and C , and C_{ABC} represents the CCFs implying all components A, B and C . The common-cause basic events for components B and C are obtained in a similar way.

According to the discussion in Mosleh et al. [36], it is reasonable to define the events, for example, C_{AB} and C_{AC} to be mutually exclusive. Using the rare event approximation, the probability of the system failure can be calculated approximately as

$$P(S) \approx P(A_I)P(B_I) + P(A_I)P(C_I) + P(B_I)P(C_I) + P(C_{AB}) + P(C_{AC}) + P(C_{BC}) + P(C_{ABC}).$$

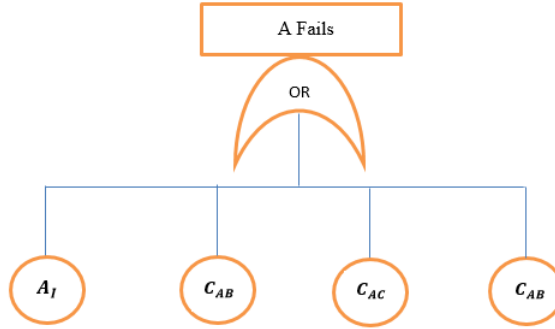


Figure 3.2: The expanded basic events fault tree for component A

Since the components of the system are supposed to be identical, the symmetry assumption discussed by Atwood in [4] is commonly used. In particular, it is assumed that

$$\begin{aligned} P(A_I) &= P(B_I) = P(C_I) = Q_1 \\ P(C_{AB}) &= P(C_{AC}) = P(C_{BC}) = Q_2 \\ P(C_{ABC}) &= Q_3 \end{aligned}$$

The parameters Q_1, Q_2, Q_3 are called the *basic parameters* for this 3-component system. The system failure probability can be expressed with these parameters as

$$P(S) = 3Q_1^2 + 3Q_2 + Q_3.$$

This definition can be extended to a system of any size. Let us consider a system with m identical components.

Let Q_k denote the probability that a particular set of k components will fail and no other components fails, $k = 1, \dots, m$. Then, Q_1, \dots, Q_k are the basic parameters. As mentioned before, the value of Q_k depends on k but not on the particular k components considered. The probability that exactly k out of m components fail is then

$$P_k = \binom{m}{k} Q_k, \quad k = 1, \dots, m,$$

and the probability that a CCF event occurs is

$$P_{tot} = \sum_{k=1}^m \binom{m}{k} Q_k.$$

As discussion in [71], the different testing schemes applied to the system lead to the different mathematical forms of the α factor model. The parameters of the α -factor model under the non-staggered testing scheme, i.e. all

components of the system are tested in the same period, are defined as

$$\alpha_k = \frac{P_k}{P_{tot}} = \frac{\binom{m}{k} Q_k}{\sum_{\ell=1}^m \binom{m}{\ell} Q_\ell}, \quad k = 1, \dots, m. \quad (3.2.1)$$

The parameters of the α -factor model under the staggered testing scheme are presented in the next chapter.

From this definition, it is easy to see that $\sum_{i=1}^m \alpha_i = 1$. Similar to Q_k , α_k also depends on k but not on the particular k components that are considered. For the simplicity in writing, let $\boldsymbol{\alpha}$ stands for $(\alpha_1, \dots, \alpha_m)$.

3.3 Inference on complete data

Before considering the data with uncertainties, we present briefly the traditional techniques to deal with complete data. Here, complete data means that we can be sure about the exact number of components failed in each CCF event. In this situation we do not isolate the independent cause failure and the lethal failure as in Chapter 2.

For a m -component system, the complete data are represented by a vector $\mathbf{N} = (N_1, \dots, N_m)$, where N_i is the number of CCF events of order i , $i = 1, \dots, m$.

3.3.1 Maximum likelihood method

We have defined previously α_i as the probability for a CCF to be of order i . Thus, for complete data, the likelihood function is:

$$L(\boldsymbol{\alpha} | \mathbf{N}) \propto \prod_{i=1}^m \alpha_i^{N_i},$$

where $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_m)$, $\alpha_i \geq 0$, $\sum_{i=1}^m \alpha_i = 1$.

The MLE of parameters are obtained directly based on the Lagrange multipliers method, that is:

$$\hat{\alpha}_i = \frac{N_i}{\sum_{i=1}^m N_i}, \quad i = 1, \dots, m. \quad (3.3.1)$$

3.3.2 Bayesian approach

The likelihood (3.3.1) has the form of a multinomial distribution. Therefore, the Dirichlet distribution with parameters $\boldsymbol{\beta} = (\beta_1, \dots, \beta_m)$ is chosen as a prior distribution for $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_m)$,

$$\pi(\boldsymbol{\alpha}) = \frac{\Gamma(\beta_1 + \dots + \beta_m)}{\Gamma(\beta_1) \dots \Gamma(\beta_m)} \prod_{i=1}^m \alpha_i^{\beta_i - 1},$$

where $\beta_i > 0$ and $\boldsymbol{\alpha} \in \Delta_m$ in which Δ_m is the close simplex,

$$\Delta_m = \{(\alpha_1, \dots, \alpha_m) : \alpha_1 \geq 0, \dots, \alpha_m \geq 0, \sum_{i=1}^m \alpha_i = 1\}.$$

It is a conjugate prior since the posterior distribution is also a Dirichlet distribution with parameters $\boldsymbol{\beta} + \mathbf{N} = (N_1 + \beta_1, \dots, N_m + \beta_m)$. The Bayes estimators for the parameters α_i under the quadratic loss are

$$\tilde{\alpha}_i = \frac{N_i + \beta_i}{\sum_{i=1}^m (N_i + \beta_i)}, \quad i = 1, \dots, m. \quad (3.3.2)$$

Remark that these estimators are convex combinations of the MLE and the expectation of α_i with respect to the prior distribution:

$$\tilde{\alpha}_i = q_i \hat{\alpha}_i + (1 - q_i) E(\alpha_i),$$

where $q_i = \frac{\sum_{i=1}^m N_i}{\sum_{i=1}^m (N_i + \beta_i)}$ and $E(\alpha_i) = \frac{\beta_i}{\sum_{i=1}^m \beta_i}$.

If the total number of observed CCFs is large, then q is close to 1 and the Bayes estimator is close to the MLE. Otherwise, when the number of observed CCFs is small, the prior information would have more influence on the Bayes estimator.

Troffaes et al. [65] suggest a specific form of hyperparameters that has the advantage of facilitating the interpretation of the parameters.

They consider a prior distribution of the form

$$\pi(\boldsymbol{\alpha} | \mathbf{s}, \mathbf{t}) \propto \prod_{i=1}^m \alpha_i^{st_i - 1},$$

where $s > 0$ and $\mathbf{t} \in \Delta_m$.

The posterior density for $\boldsymbol{\alpha}$ corresponding to this prior distribution is:

$$\pi(\boldsymbol{\alpha} | \mathbf{N} = \mathbf{n}, \mathbf{s}, \mathbf{t}) \propto \prod_{i=1}^m \alpha_i^{N_i + st_i - 1}.$$

From the properties of Dirichlet distribution, the marginal distribution of α_i is

$$\alpha_i | \mathbf{N} = \mathbf{n}, \mathbf{s}, \mathbf{t} \sim \text{Beta} \left(N_i + st_i, \sum_{i \neq j} (N_j + st_j) \right), \quad i = 1, \dots, m.$$

Therefore, under the quadratic loss function, the Bayesian estimate of each α_i is

$$\tilde{\alpha}_i = \frac{N_i + st_i}{\sum_{i=1}^m N_i + s} = (1 - q(s)) \hat{\alpha}_i + q(s) t_i, \quad i = 1, \dots, m,$$

where $q(s) = \frac{s}{\sum_{i=1}^m N_i + s}$. The balance between observation and prior information is now governed by one parameter: s .

We can make again the following interpretation of the Bayes estimators above:

- $\tilde{\alpha}_i$ is a weighted average of t_i and $\hat{\alpha}_i$ with weights $q(s)$ and $1 - q(s)$, respectively. $q(s)$ represents the portion of prior contribution to the posterior. The smaller the value of $\sum_{i=1}^m N_i$ compared to s , the larger the contribution of the prior to the posterior.
- If $\sum_{i=1}^m N_i$ is close to 0, i.e there is very few failure observations, then $\tilde{\alpha}_i = t_i$. That is to say, the value of t_i represents the prior expected number of CCF of order i .
- If $\sum_{i=1}^m N_i$ goes to infinity, then $\tilde{\alpha}_i$ is close to the MLE.

3.4 Inference on data with uncertainties

In many situations, the order of CCFs are reported with *uncertainties*. That is to say we do not know exactly how many components are really involved in the CCF events. A traditional method to treat data with uncertainties is to give a guess on the number of failed components and a corresponding associated probability, leading to what is called an *impact vector*. We are going to explain the impact vector method in the following before presenting a new method to deal with this kind of data.

3.4.1 Impact vector method

3.4.1.1 The definition of impact vector

With the narrative description of the failure events provided in Table 3.1, it is not clear whether or not the third diesel generator has actually failed. Thus, in this situation, there are two possibilities for the number of components involved in the CCF: 2 or 3.

Suppose a m -component system. To represent a CCF of order k , it is classical to consider a binary vector I with $m + 1$ components. If the CCF is of order k , the component $k + 1$ of the binary vector is 1 while the other components are zero. The vector I is denoted I_k . For the m -component system of the Table 3.1, a CCF of order 2 is represented as $I_2 = (0, 0, 1, 0)$, and a CCF of order 3 is represented as $I_3 = (0, 0, 0, 1)$.

Because of uncertainties, a CCF is now a random variable which outcomes are binary vectors I_0, I_1, \dots, I_m . With each binary vector I_k is associated a probability p_k , such that $\sum_{k=0}^m p_k = 1$.

In practice, these probabilities are attributed by engineers relying on the expert's opinion. For example, Mosleh et al. [36] associate a probability to each of the possibility in the example in Table 3.1 as in Table 3.2. Only two diesels generators failed (corresponding to the binary vector I_2) is more likely than all the three generators to fail (corresponding to the binary vector I_3).

We now give the definition of an **impact vector**.

Number of components involved	Binary vector	Probability
2	$I_2 = (0, 0, 1, 0)$	0.9
3	$I_3 = (0, 0, 0, 1)$	0.1

Table 3.2: Example of establishing an impact vector

Definition 2. Let us consider a m -component system. An **impact vector** for an uncertain CCF event is a $m + 1$ dimensional vector of probability $\bar{I} = (p_0, \dots, p_m)$ where p_i is the probability that i components are involved in the CCF, $p_i \geq 0, \forall i$ and $\sum_{i=0}^m p_i = 1$.

Remark that the impact vector can be obtained combining the probabilities and the binary vectors as an average vectors.

For example, with Table 3.2 we have:

$$\begin{aligned}\bar{I} &= 0.9I_2 + 0.1I_3 \\ &= (0, 0, 0.9, 0.1).\end{aligned}$$

This impact vector describes the uncertainty on the order of the observed CCF.

Table 3.2 presents an example of impact vectors for a set of real data, which is provided by Siu and Mosleh [56].

3.4.1.2 Inference for CCF with impact vector

A number of methods to use impact vectors in CCF analysis based on α -factor model are suggested by Siu and Mosleh [56]. One of these methods can be described as follows.

Suppose that n CCF events are observed. We have a sample $\bar{\mathbf{I}} = \{\bar{I}_1, \dots, \bar{I}_n\}$ of uncertain data which can be written as $n \times (m + 1)$ matrix.

To this sample of uncertain data corresponds a set of data with no uncertainties. By combining all the possible binary vectors for each uncertain event, we obtain a matrix with zero or one element such that the sum of the column terms gives the vectors $\mathbf{N} = \{N_0, \dots, N_m\}$ which represents the complete data. We can define the probability of each occurrence of these vectors using the impact vectors.

Consider a simple example (table 3.3) given by Siu & Mosleh [56] in which a total of 9 CCF events with uncertainties has been reported from a 3-component system. The table 3.4 resumes the situation.

The complete data deduced from the sample $\bar{\mathbf{I}} = \{\bar{I}_1, \dots, \bar{I}_5\}$ lies in a space which has less than $2^3 \times 3^2 = 72$ elements. For example, one of these elements

Event number	Plant (Date)	Status	Event description	p_0	p_1	p_2	p_3
1	Ginna (12/73)	Critical	Two motor-driven AFW pumps were inoperable due to air in common suction line	0.1	0	0	0.9
2	Zion 2 (2/74)	Power escalation test	Two motor-driven AFW pumps were inoperable due to air in common suction line	0.1	0	0	0.9
3	Turkey Point 3 (5/74)	98% power	AFW pumps A and B failed to start due to tight packing	0.9	0	0.05	0.05
4	Zion 2 (9/81)	Shutdown	AFW pumps 2B and 2C failed due to a faulty control switch modification	0	0	1	0
4I	Zion 2 (9/81)	Shutdown	AFW pumps 2A failed due to a pressure switch drift	0	1	0	0
5	Zion 2 (11/79)	Power	AFW pumps 2B and 2C failed due to miscalibrated pressure gauges	0.3	0	0.35	0.35
6	Zion 2 (12/79)	Poweer	AFW pumps 2B and 2C failed due to start circuitry design problems	0	0	1	0
7	Turkey Point 3 (6/73)	Prior to initial power testing	All 3 AFW failed to start due to missing fuses in pump autostart circuit	0	0	0	1
8	Arkansas 2(4/80)	0% power	Two emergency feedwater pumps lost suction due to steam flushing; system design problem	0.1	0	0	0.9

Table 3.3: An example of impact assessment of AFW pump failure to start [56].

Event	Narative description	Impact vector	Frequency
1-2-8	<i>no or 3 components fail</i>	0.1 0 0 0.9	3
3	<i>no or 2 or 3 components fail</i>	0.9 0 0.05 0.05	1
4-6	<i>2 components fail</i>	0 0 1 0	2
4I	<i>1 components fails</i>	0 1 0 0	1
5	<i>no or 2 or 3 components fail</i>	0.3 0 0.35 0.35	1
7	<i>3 components fail</i>	0 0 0 1	1

Table 3.4: Summary of data with uncertainties described in Table 3.3

N_0	N_1	N_2	N_3	Probability
5	1	2	1	0.00027
4	1	3	1	0.00033
4	1	2	2	0.00762
3	1	4	1	0.0000175
3	1	3	2	0.008945
3	1	2	3	0.0745375
2	1	4	2	0.0004725
2	1	3	3	0.081135
2	1	2	4	0.2774925
1	1	4	3	0.0042525
1	1	3	4	0.249075
1	1	2	5	0.2448225
0	1	4	4	0.0127575
0	1	3	5	0.025515
0	1	2	6	0.0127575

Table 3.5: The possible complete data for the data with uncertainties in Table 3.4 with their corresponding probabilities.

is:

$$\begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

for which $N_0 = 2$, $N_1 = 1$, $N_2 = 3$ and $N_3 = 3$. After computations, a total of 15 vectors of complete data are possible. Table 3.5 shows these vectors and the corresponding probabilities.

For the statistical analysis, we only consider $\{N_1, \dots, N_m\}$.

Suppose that w vectors of complete data are possible: $\{\mathbf{N}_1, \dots, \mathbf{N}_w\}$ with $\mathbf{N}_\ell = (N_{\ell,1}, \dots, N_{\ell,m})$, $\ell = 1, \dots, w$.

The log-likelihood with uncertain data can be expressed as:

$$\log L(\boldsymbol{\alpha} | \bar{\mathbf{I}}) = \sum_{\ell=1}^w \log L(\boldsymbol{\alpha} | \mathbf{N}_\ell) P(\mathbf{N}_\ell | \bar{\mathbf{I}}).$$

The likelihood equations are then:

$$\frac{\partial}{\partial \alpha_i} \log L(\boldsymbol{\alpha} | \bar{\mathbf{I}}) = \sum_{\ell=1}^w \frac{\partial}{\partial \alpha_i} \log L(\boldsymbol{\alpha} | \mathbf{N}_\ell) P(\mathbf{N}_\ell | \bar{\mathbf{I}}) = 0, \quad i = 1, \dots, m-1.$$

Because $L(\boldsymbol{\alpha} | \mathbf{N}_\ell)$ belongs to the exponential family and has a canonical form: $L(\boldsymbol{\eta} | \mathbf{N}_\ell)$ with $\eta_i = \log [\alpha_i / (1 - \sum_{j=1}^{m-1} \alpha_j)]$, $i = 1, \dots, m-1$, it can be shown that

$$\frac{\partial}{\partial \eta_i} \log L(\boldsymbol{\eta} | \mathbf{N}_\ell) = N_{\ell,i} - E(N_{\ell,i} | \boldsymbol{\eta}), \quad i = 1, \dots, m-1.$$

Thus

$$\frac{\partial}{\partial \eta_i} \log L(\boldsymbol{\eta} | \bar{\mathbf{I}}) = 0 \iff \sum_{\ell=1}^w N_{\ell,i} P(\mathbf{N}_\ell | \bar{\mathbf{I}}) - E(N_{\ell,i} | \boldsymbol{\eta}) = 0, \quad i = 1, \dots, m-1.$$

But $E(N_{\ell,i} | \boldsymbol{\eta})$ can be expressed as: $\alpha_i \sum_{i=1}^m N_{\ell,i}$.

Therefore

$$\sum_{\ell=1}^w N_{\ell,i} P(\mathbf{N}_\ell | \bar{\mathbf{I}}) = \alpha_i \sum_{i=1}^m N_{\ell,i}$$

And $\hat{\alpha}_{u,i}$, the MLE for α_i , is:

$$\hat{\alpha}_{u,i} = \sum_{\ell=1}^w \frac{N_{\ell,i}}{\sum_{j=1}^m N_{\ell,j}} P(\mathbf{N}_\ell | \bar{\mathbf{I}}), \quad i = 1, \dots, m. \quad (3.4.1)$$

$\hat{\alpha}_{u,i}$ is simply the weighted sum of the MLE for each possible complete data.

For Bayesian inference, the posterior distribution of $\boldsymbol{\alpha}$ is

$$\pi(\boldsymbol{\alpha} | \bar{\mathbf{I}}) = \sum_{\ell=1}^w \pi(\boldsymbol{\alpha} | \mathbf{N}_\ell) P(\mathbf{N}_\ell | \bar{\mathbf{I}}). \quad (3.4.2)$$

Under a quadratic loss function, the Bayesian estimators are:

$$\tilde{\alpha}_i = \sum_{\ell=1}^w \tilde{\alpha}_{\ell,i} P(\mathbf{N}_\ell | \bar{\mathbf{I}}) \quad (3.4.3)$$

where $\tilde{\alpha}_{\ell,i}$, $i = 1, \dots, m$ are the Bayesian estimators obtained with the complete data \mathbf{N}_ℓ .

A disadvantage of these approaches is that it requires the computation of the possible set of complete data which can lead to a combinatory explosion. To avoid this drawback, Siu and Mosleh [56] propose a data-averaging approximation approach for developing the posterior of $\boldsymbol{\alpha}$.

	α_1	α_2	α_3	α_4
MLE of the Situation 1	0.1666	0.1111	0.2666	0.4555
MLE of the Situation 2	0.3000	0.0555	0.3222	0.3222
Bayesian of the Situation 1	0.1818	0.1363	0.2636	0.4181
Bayesian of the Situation 2	0.2909	0.0909	0.3090	0.3090

Table 3.6: *MLE and Bayesian estimate of α corresponding to two situations of data and $\mathcal{D}(0.5, 0.5, 0.5, 0.5)$ prior.*

This method is widely used in CCF analysis, for example, Le & Vasseur [29] and O'Connor & Mosleh [49]. Suppose that n CCF events are observed and let $\bar{I}_j = (p_{0,j}, \dots, p_{m,j})$ be the impact vector corresponding to the j^{th} event, $j = 1, \dots, n$.

Then the expected number of CCF events of order i can be expressed as

$$E(N_i) = \sum_{j=1}^n p_{i,j}, \quad i = 1, \dots, m.$$

The vector $E(\mathbf{N}) = (E(N_1), \dots, E(N_m))$ is then used in the formula (3.3.1) for MLE and the formula (3.3.2) for Bayes estimators. Corresponding to the data in the table 3.4, say Situation 1, we have $E(\mathbf{N}) = (1.5, 1.0, 2.4, 4.1)$ and the estimates of α -factor parameters using this data appear in Table 3.6.

It should be considered that in the above analysis, we consider only the case when each I_j is precisely known, thus ignoring any uncertainties in these subjectively assessed probabilities. It could be true for the case if these probabilities are evaluated by a single expert. Nevertheless, in practice they could be elicited from a panel of experts, or even in the case of single expert, he or she may still be unsure about these values. This, obviously, affects the estimate of α .

In order to see more clearly this effect, let us consider another situation of data where the impact vectors for the events 3 to 7 in Table 3.4 are still the same, but not for the events 1-2, 4I and 8. Suppose that, the impact vector corresponding to the events 1 – 2 – 8 is $(0.5, 0, 0, 0.5)$ (instead of $(0.1, 0, 0, 0.9)$ as in Table 3.4) and the one corresponding to the event 4I is $(0, 0.5, 0.5, 0)$ (instead of $(0, 1, 0, 0)$). With this new data, say Situation 2, we obtain $E(\mathbf{N}) = (2.7, 0.5, 2.9, 2.9)$. Table 3.6 shows the estimate of α for these two situations of data with the MLE and Bayesian estimate using a Dirichlet prior distribution with parameters $(0.5, 0.5, 0.5, 0.5)$. It can be seen that the hypothesis set up in each CCF event has a significant influence on the estimate of α .

3.4.2 Confounded data method

The method presented in Section 3.4.1 relies on subjective hypotheses through impact vectors for estimating of α . However, in some situations, it is difficult

for experts to establish a weight for each hypothesis to formulate the value of the impact vectors.

In this section, we provide a new alternative to handle data with uncertainties. The idea is to introduce new random variables to describe the events corresponding to uncertain observations.

The data with uncertainties are now expressed as a set \mathbf{N}_u of number $N_{i_1 \dots i_k}$ where i_ℓ is a possible order of the CCF, $\ell = 1, \dots, k, i_\ell \in \{0, \dots, m\}$. With this new notation, engineers only need to treat data with uncertainties by assessing the number of components that could be involved in each CCF event based on its observed symptom, without establishing any related probabilities.

We are going to make inference on α based on this new proposed formalism of data with uncertainties.

For purpose of illustration, we consider a system of 4 components (which is more popular in CCF analysis). However, the reasoning will be the same regardless of the size.

The data analyzed includes two parts: the certain data, which are (N_1, N_2, N_3, N_4) , and the data with uncertainties. The latter is classified into two groups. The first one contains data of the form $\{N_{i_1 i_2}\}$, two possible CCF orders, and the second one contains data of the form $\{N_{i_1 i_2 i_3}\}$, three possible CCF orders).

The following specific schemes of data with uncertainties are considered:

- **Scheme 1:** $\mathbf{N}_u = \{N_1, N_2, N_3, N_4, N_{12}, N_{123}\}$.
- **Scheme 2:** \mathbf{N}_u includes $\{N_1, N_2, N_3, N_4, N_{12}, N_{123}\}$ and other terms combining different orders of CCF.
- **Scheme 3:** $\mathbf{N}_u = \{N_1, N_2, N_3, N_4, N_{12}, N_{34}\}$.
- **Scheme 4:** \mathbf{N}_u includes $\{N_1, N_2, N_3, N_4, N_{12}, N_{34}\}$ and other terms combining different orders of CCF.

The first two schemes correspond to picking up one uncertain element in the first group and the other in the second group. The last two schemes correspond to picking up two uncertain elements only in the first group.

These choices allow to illustrate the methods and it covers many situations of data with uncertainties. Various other schemes can be treated in the same way.

The introduction of terms of the form $N_{i_1 \dots i_k}$ means that the information is not complete. We are in a situation of incomplete data. A natural way to deal with this kind of data is to consider an EM algorithm. But the application of this algorithm requires the availability of the distribution of complete data given the incomplete data. In our case of uncertain data, the expression of this distribution will be difficult to manipulate. Therefore, we propose several methods to deal directly with uncertainties.

We are now going to present the methods with data from schemes 1-4. For each scheme, we consider the maximum likelihood estimation, then the Bayesian approach and we make comparison for the simulation study.

3.4.2.1 Inference for scheme 1

In this scheme, the data with uncertainties, denoted by \mathbf{N}_u , are of the form $\mathbf{N}_u = (N_1, N_2, N_3, N_4, N_{12}, N_{123})$. The corresponding likelihood function is:

$$L(\boldsymbol{\alpha}|\mathbf{N}_u) \propto \prod_{i=1}^4 \alpha_i^{N_i} (\alpha_1 + \alpha_2)^{N_{12}} (\alpha_1 + \alpha_2 + \alpha_3)^{N_{123}}. \quad (3.4.4)$$

Maximum likelihood method

The log-likelihood function is equal to

$$\log L(\boldsymbol{\alpha}|\mathbf{N}_u) = \sum_{i=1}^4 N_i \log \alpha_i + N_{12} \log(\alpha_1 + \alpha_2) + N_{123} \log(\alpha_1 + \alpha_2 + \alpha_3).$$

Computing the partial derivatives of $\log L(\boldsymbol{\alpha}|\mathbf{N}_u)$ with respect to α_i , the likelihood equations are

$$\begin{cases} \frac{N_1}{\alpha_1} - \frac{N_4}{\alpha_4} + \frac{N_{12}}{\alpha_1 + \alpha_2} + \frac{N_{123}}{\alpha_1 + \alpha_2 + \alpha_3} = 0, \\ \frac{N_2}{\alpha_2} - \frac{N_4}{\alpha_4} + \frac{N_{12}}{\alpha_1 + \alpha_2} + \frac{N_{123}}{\alpha_1 + \alpha_2 + \alpha_3} = 0, \\ \frac{N_3}{\alpha_3} - \frac{N_4}{\alpha_4} + \frac{N_{123}}{\alpha_1 + \alpha_2 + \alpha_3} = 0. \end{cases} \quad (3.4.5)$$

The first two equations lead to $\alpha_2 = \frac{N_2}{N_1} \alpha_1$. Similarly, the first equation and the third one give rise to

$$\frac{N_3}{\alpha_3} = \frac{N_1}{\alpha_1} + \frac{N_{12}}{\alpha_1 + \alpha_2}.$$

Using these two results we obtain

$$\begin{cases} \alpha_1 = \frac{N_1(N_1 + N_2 + N_{12})}{N_3(N_1 + N_2)} \alpha_3, \\ \alpha_2 = \frac{N_2(N_1 + N_2 + N_{12})}{N_3(N_1 + N_2)} \alpha_3. \end{cases} \quad (3.4.6)$$

Substituting equation (3.4.6) into the third equation of (3.4.5) and because $\sum_{i=1}^4 \alpha_i = 1$, we obtain the MLE:

$$\left\{ \begin{array}{l} \hat{\alpha}_4 = \frac{N_4}{\sum_{i=1}^4 N_i + N_{12} + N_{123}}, \\ \hat{\alpha}_3 = \frac{N_3(\sum_{i=1}^3 N_i + N_{12} + N_{123})}{((\sum_{i=1}^3 N_i + N_{12})(\sum_{i=1}^4 N_i + N_{12} + N_{123}))}, \\ \hat{\alpha}_2 = \frac{N_2(N_1 + N_2 + N_{12})(\sum_{i=1}^3 N_i + N_{12} + N_{123})}{(N_1 + N_2)(\sum_{i=1}^3 N_i + N_{12})(\sum_{i=1}^4 N_i + N_{12} + N_{123})}, \\ \hat{\alpha}_1 = \frac{N_1(N_1 + N_2 + N_{12})(\sum_{i=1}^3 N_i + N_{12} + N_{123})}{(N_1 + N_2)(\sum_{i=1}^3 N_i + N_{12})(\sum_{i=1}^4 N_i + N_{12} + N_{123})}. \end{array} \right. \quad (3.4.7)$$

Bayesian approach

Considering that the distribution of the observation $P(\mathbf{N}_u | \mathbf{a})$ as the form (3.4.4), we use a nested Dirichlet distribution as a prior distribution. It will be a natural conjugate prior. Let us recall the definition and some properties of the nested Dirichlet distribution (NDD).

Nested Dirichlet distribution

The nested Dirichlet distribution (NDD) was firstly introduced by Tian et al. [61] and then investigated more deeply by Ng et al. [40] and Tian et al. [64].

Definition 3. Let $\Delta_n = \{\mathbf{x} = (x_1, \dots, x_n) : x_i \geq 0, i = 1, \dots, n, \sum_{i=1}^n x_i = 1\}$, a random vector $\mathbf{x} \in \Delta_n$, is said to follow a NDD with parameters (\mathbf{a}, \mathbf{b}) if its density is of the form

$$f(\mathbf{x} | \mathbf{a}, \mathbf{b}) = c^{-1} \left(\prod_{i=1}^n x_i^{a_i - 1} \right) \prod_{j=1}^{n-1} \left(\sum_{k=1}^j x_k \right)^{b_j}, \quad (3.4.8)$$

where $\mathbf{a} = (a_1, \dots, a_n)$ is a positive parameter vector, $\mathbf{b} = (b_1, \dots, b_{n-1})$ is a non-negative parameter vector, c is the normalizing constant defined by:

$$c = \prod_{j=1}^{n-1} B(d_j, a_{j+1}),$$

with $B(x, y)$ stands for the value of the Beta function at (x, y) and $d_j = \sum_{k=1}^j (a_k + b_k)$.

It is written shortly by $\mathbf{x} \sim ND_{n,n-1}(\mathbf{a}, \mathbf{b})$ on Δ_n .

When $\mathbf{b} = \mathbf{0} = (0, \dots, 0)$, the $ND_{n,n-1}(\mathbf{a}, \mathbf{0})$ is the Dirichlet distribution $\mathcal{D}(\mathbf{a})$.

The NDD can be stochastically represented by a sequence of mutually independent Beta distribution by using the following result [40].

Proposition 1. *A random vector \mathbf{x} follows a nested Dirichlet distribution if and only if*

$$\begin{cases} x_i \stackrel{d}{=} (1 - y_{i-1}) \prod_{j=i}^{n-1} y_j, & i = 1, \dots, n-1, \\ x_n \stackrel{d}{=} 1 - y_{n-1}, \end{cases} \quad (3.4.9)$$

where $y_0 = 0, y_j \sim \text{Beta}(d_j, a_{j+1}), i = 1, \dots, n-1$ with $d_j = \sum_{k=1}^j (a_k + b_k)$ and y_1, \dots, y_{n-1} are mutually independent.

Proposition 1 suggests a simple procedure for generating independently and identically distributed samples from the NDD.

The mixed moment of a random vector that follows a NDD is given in proposition 2.

Proposition 2. *Let $\mathbf{x} \sim ND_{n,n-1}(\mathbf{a}, \mathbf{b})$ on Δ_n , then the **mixed moment** of \mathbf{x} is*

$$E\left(\prod_{i=1}^n x_i^{r_i}\right) = \prod_{i=1}^n \left[\frac{B(d_{i-1}, a_i + r_i)}{B(d_{i-1}, a_i)} \cdot \prod_{j=i}^{n-1} \frac{B(d_j + r_i, a_{j+1})}{B(d_j, a_{j+1})} \right], \quad (3.4.10)$$

where $d_j, j = 1, \dots, n$ are defined as in Proposition 1.

Moreover, a closed-form expression for the mode of an NDD is given by the following proposition.

Proposition 3. *Let $\mathbf{x} \sim ND_{n,n-1}(\mathbf{a}, \mathbf{b})$ on Δ_n , then the **mode** of \mathbf{x} is*

$$\begin{cases} m_n = \frac{a_{n-1}}{d_{n-1} + a_n - n}, \\ m_i = \frac{(a_i - 1)(1 - m_{i+1} - m_{i+2} - \dots - m_n)}{d_{i-1} + a_i - i}, & i = 2, \dots, n-1, \\ m_1 = 1 - m_2 - \dots - m_n. \end{cases} \quad (3.4.11)$$

□

We are going to use the NDD to conduct a Bayesian estimation for the scheme 1.

Let $\alpha \sim ND_{4,3}(\mathbf{a}, \mathbf{b})$ with hyperparameters $(\mathbf{a}, \mathbf{b}) = ((a_1, a_2, a_3, a_4), (b_1, b_2, b_3))$.

That is

$$\pi(\alpha) \propto \prod_{i=1}^4 a_i^{\alpha_i - 1} a_1^{b_1} (\alpha_1 + \alpha_2)^{b_2} (\alpha_1 + \alpha_2 + \alpha_3)^{b_3}. \quad (3.4.12)$$

The Bayesian formula leads to

$$\begin{aligned}\pi(\boldsymbol{\alpha} | \mathbf{N}_u) &\propto P(\mathbf{N}_u | \boldsymbol{\alpha}) \pi(\boldsymbol{\alpha}) \\ &\propto \prod_{i=1}^4 \alpha_i^{a_i + N_i - 1} \alpha_1^{b_1} (\alpha_1 + \alpha_2)^{b_2 + N_{12}} (\alpha_1 + \alpha_2 + \alpha_3)^{b_3 + N_{123}}\end{aligned}$$

Thus, the posterior distribution of $\boldsymbol{\alpha}$ is also a NDD with parameters $(\mathbf{a}^*, \mathbf{b}^*)$, where $\mathbf{a}^* = (a_1^*, a_2^*, a_3^*, a_4^*) = (N_1 + a_1, N_2 + a_2, N_3 + a_3, N_4 + a_4)^T$ and $\mathbf{b}^* = (b_1^*, b_2^*, b_3^*) = (b_1, b_2 + N_{12}, b_3 + N_{123})$.

Under a quadratic loss function, the Bayes estimators of $\boldsymbol{\alpha}$ are the expected-values of the posterior distributions, which are defined from the proposition 2:

$$\left\{ \begin{aligned}\tilde{\alpha}_1 &= \frac{d_1 d_2 d_3}{(d_1 + a_2^*)(d_2 + a_3^*)(d_3 + a_4^*)}, \\ \tilde{\alpha}_2 &= \frac{a_2^* d_2 d_3}{(d_1 + a_2^*)(d_2 + a_3^*)(d_3 + a_4^*)}, \\ \tilde{\alpha}_3 &= \frac{a_3^* d_3}{(d_2 + a_3^*)(d_3 + a_4^*)}, \\ \tilde{\alpha}_4 &= \frac{a_4^*}{d_3 + a_4^*},\end{aligned} \right. \quad (3.4.13)$$

where $d_i = \sum_{j=1}^i (a_j^* + b_j^*)$.

Determining hyperparameters for prior distributions is an important problem in Bayesian inference. When no information about the parameters is available, a noninformative prior can be considered. Here, we apply a NDD with parameters $\mathbf{a} = (0.5, 0.5, 0.5, 0.5)$ and $\mathbf{b} = (0, 0, 0)$, which is a Dirichlet distribution $\mathcal{D}(0.5, 0.5, 0.5, 0.5)$. Example of computations is shown in the sequel.

However, there are also some situations where the analysts can have experiences that allow to determine the prior. For example, they can be able to suggest values for each α_i , say $(\bar{\alpha}_1, \dots, \bar{\alpha}_4)$.

Assuming a NDD for $\boldsymbol{\alpha}$ with the parameters $\mathbf{a} = (a_1, \dots, a_4)$, $\mathbf{b} = (b_1, \dots, b_3)$ such that $\mathbf{a} = \mathbf{st} = (st_1, \dots, st_4)$, i.e. $a_i = st_i, i = 1, \dots, 4$, where $s > 0, t_i \geq 0$ and $\sum_{i=1}^4 t_i = 1$. It can be considered that the suggested value $\bar{\alpha}_i$ is equal to $E(\alpha_i)$, the expectation of α_i , for $i = 1, \dots, 4$, leading to the system of equations:

$$\left\{ \begin{aligned}st_1 + b_1 &= \frac{\bar{\alpha}_1}{\bar{\alpha}_2} st_2, \\ \frac{\bar{\alpha}_1 + \bar{\alpha}_2}{\bar{\alpha}_2} st_2 + b_2 &= \frac{\bar{\alpha}_1 + \bar{\alpha}_2}{\bar{\alpha}_3} st_3, \\ \frac{\bar{\alpha}_1 + \bar{\alpha}_2 + \bar{\alpha}_3}{\bar{\alpha}_3} st_3 + b_3 &= \frac{\bar{\alpha}_1 + \bar{\alpha}_2 + \bar{\alpha}_3}{\bar{\alpha}_4} st_4.\end{aligned} \right. \quad (3.4.14)$$

Obviously, the system of equations (3.4.14) does not have a unique solution. Given the value of (b_1, b_2, b_3) , we can find out the relation among st_1, st_2, st_3 and st_4 . For example, with $(b_1, b_2, b_3) = (0, 0, 0)$ (corresponding to Dirichlet prior), we obtain

$$\frac{\bar{\alpha}_1}{t_1} = \frac{\bar{\alpha}_2}{t_2} = \frac{\bar{\alpha}_3}{t_3} = \frac{\bar{\alpha}_4}{t_4}. \quad (3.4.15)$$

By combining this result with the constraint $\sum_{i=1}^4 t_i = \sum_{i=1}^4 \alpha_i = 1$, we have $t_i = \bar{\alpha}_i$. Using constrained non-informative prior distribution proposed in [3] and applied in [19] which maximizes uncertainty, Troffaes et al. [65] suggested choosing $s = \frac{1}{2(1 - \bar{\alpha}_1)}$.

Illustrative example

We illustrate the proposed methods by supposing a fictitious dataset with uncertainties: $\mathbf{N}_u = (5, 3, 1, 0, 3, 2)$. The MLE of α based on (3.4.7) is $\hat{\alpha} = (0.573, 0.344, 0.083, 0.0)$. Since there is no assumption about the failure events that all the components of the system failed simultaneously, the MLE of $\hat{\alpha}_4$ is equal to 0.

For the Bayesian inference, we use different prior distributions. Relying on the technique suggested by Nguyen & Gouno [44] and Gutiérrez-Pulido et al. [11], the following strategy is applied for eliciting the prior. Firstly, we suppose that a vector $\mathbf{N}_g = (9, 5, 1, 0.5)$ has been proposed by the analysts for the number of CCF events in a given time. Then, the MLE of α is calculated from (3.3.1) as $\hat{\alpha} = (0.58, 0.32, 0.07, 0.03)$. These values are assigned to (t_1, \dots, t_4) . Then, we pick up some values of s . Table 3.7 shows the obtained results of estimation. In general, using the $\mathcal{D}(0.5, 0.5, 0.5, 0.5)$ prior distribution gives the estimates close to those the MLE. Moreover, the results also confirm a conclusion in [65] that the smaller value of s makes the model learn faster from the data for the zero count.

		α_1	α_2	α_3	α_4
MLE		0.573	0.344	0.083	0.00
Prior	$\mathcal{D}(0.5, 0.5, 0.5, 0.5)$	0.527	0.338	0.103	0.032
	$s = 1.19$	0.574	0.341	0.082	0.0023
	$s = 2$	0.575	0.339	0.083	0.003
	$s = 10$	0.578	0.332	0.077	0.013

Table 3.7: MLE and Bayesian estimate of α for scheme 1 with different types of priors.

3.4.2.2 Inference for scheme 2

In scheme 2, we consider a situation where the data with uncertainties contains N_{12} , N_{123} and one or more other terms such as N_{23} , N_{14} , N_{234} , etc. For the sake of simplicity, we consider only one more term N_{23} , for instance. That means, the observed data with uncertainties are now

$$\mathbf{N}_u = (N_1, N_2, N_3, N_4, N_{12}, N_{23}, N_{123}).$$

Then, the likelihood function is

$$L(\boldsymbol{\alpha} | \mathbf{N}_u) = \prod_{i=1}^4 \alpha_i^{N_i} (\alpha_1 + \alpha_2)^{N_{12}} (\alpha_1 + \alpha_2 + \alpha_3)^{N_{123}} (\alpha_2 + \alpha_3)^{N_{23}}. \quad (3.4.16)$$

The likelihood function (3.4.16) is equivalent to the likelihood in (3.4.4) times the new term $(\alpha_2 + \alpha_3)^{N_{23}}$. A direct calculation shows that the corresponding log-likelihood equations have no close-form solutions.

The value N_{23} can be seen as the sum of a number of CCF event of order 2 and a number of CCF event of order 3. These two numbers are unknown. If these numbers were known, the likelihood (3.4.16) would have the form of (3.4.4), and inference can be conducted by following the procedure described in the previous section.

Let us introduce a variable Z which represents the number of CCF of order 2 amongst N_{23} CCF events, Z goes from 0 to N_{23} . As a result, $N_{23} - Z$ will be the number of CCF of order 3. The term $(\alpha_2 + \alpha_3)^{N_{23}}$ can be split in two parts and the likelihood (3.4.16) can be rewritten as:

$$L(\boldsymbol{\alpha} | \mathbf{N}_u, Z) = \alpha_1^{N_1} \alpha_2^{N_2+Z} \alpha_3^{N_3+N_{23}-Z} \alpha_4^{N_4} (\alpha_1 + \alpha_2)^{N_{12}} (\alpha_1 + \alpha_2 + \alpha_3)^{N_{123}}. \quad (3.4.17)$$

And we have

$$L(\boldsymbol{\alpha} | \mathbf{N}_u) = \sum_{z=0}^{n_{23}} L(\boldsymbol{\alpha} | \mathbf{N}_u, Z = z) P(Z = z | N_{23} = n_{23}). \quad (3.4.18)$$

Maximum likelihood method

Since the likelihood in (3.4.17) is similar to the one in (3.4.4), the MLE can be obtained as:

$$\left\{ \begin{array}{l} \hat{\alpha}_4 = \frac{N_4}{\sum_{i=1}^4 N_i + N_{12} + N_{23} + N_{123}}, \\ \hat{\alpha}_3 = \frac{(N_3 + N_{23} - Z)(\sum_{i=1}^3 N_i + N_{12} + N_{23} + N_{123})}{((\sum_{i=1}^3 N_i + N_{12} + N_{23})(\sum_{i=1}^4 N_i + N_{12} + N_{23} + N_{123}))}, \\ \hat{\alpha}_2 = \frac{(N_2 + Z)(N_1 + N_2 + N_{12} + Z)(\sum_{i=1}^3 N_i + N_{12} + N_{23} + N_{123})}{(N_1 + N_2 + Z)(\sum_{i=1}^3 N_i + N_{12} + N_{23})(\sum_{i=1}^4 N_i + N_{12} + N_{23} + N_{123})}, \\ \hat{\alpha}_1 = \frac{N_1(N_1 + N_2 + N_{12} + Z)(\sum_{i=1}^3 N_i + N_{12} + N_{23} + N_{123})}{(N_1 + N_2 + Z)(\sum_{i=1}^3 N_i + N_{12} + N_{23})(\sum_{i=1}^4 N_i + N_{12} + N_{23} + N_{123})}. \end{array} \right. \quad (3.4.19)$$

The expression (3.4.18) suggests the use of an EM algorithm where Z is considered as the missing data and applying EM will be replacing Z in the system (3.4.19) by the conditional expectation of Z . Since

$$Z \mid \boldsymbol{\alpha}, \mathbf{N}_u \sim \mathcal{B}\left(N_{23}, \frac{\alpha_2}{\alpha_2 + \alpha_3}\right), \quad (3.4.20)$$

we have

$$E(Z \mid \boldsymbol{\alpha}, \mathbf{N}_u) = N_{23} \frac{\alpha_2}{\alpha_2 + \alpha_3},$$

and the EM algorithm is performed as follows:

- Initialisation: $\boldsymbol{\alpha}^{(0)} = (\alpha_1^{(0)}, \alpha_2^{(0)}, \alpha_3^{(0)}, \alpha_4^{(0)})$
- At step $(r+1)$, $\hat{\boldsymbol{\alpha}}^{(r)} = (\hat{\alpha}_1^{(r)}, \hat{\alpha}_2^{(r)}, \hat{\alpha}_3^{(r)}, \hat{\alpha}_4^{(r)})$ being available, compute

$$\bar{z}^{(r)} = N_{23} \alpha_2^{(r)} / (\alpha_2^{(r)} + \alpha_3^{(r)})$$

and

$$\left\{ \begin{array}{l} \hat{\alpha}_4^{(r+1)} = \frac{N_4}{\sum_{i=1}^4 N_i + N_{12} + N_{23} + N_{123}} \\ \hat{\alpha}_3^{(r+1)} = \frac{(N_3 + N_{23} - \bar{z}^{(r)})(\sum_{i=1}^3 N_i + N_{12} + N_{23} + N_{123})}{((\sum_{i=1}^3 N_i + N_{12} + N_{23})(\sum_{i=1}^4 N_i + N_{12} + N_{23} + N_{123}))} \\ \hat{\alpha}_2^{(r+1)} = \frac{(N_2 + \bar{z}^{(r)})(N_1 + N_2 + N_{12} + \bar{z}^{(r)})(\sum_{i=1}^3 N_i + N_{12} + N_{23} + N_{123})}{(N_1 + N_2 + \bar{z}^{(r)})(\sum_{i=1}^3 N_i + N_{12} + N_{23})(\sum_{i=1}^4 N_i + N_{12} + N_{23} + N_{123})} \\ \hat{\alpha}_1^{(r+1)} = \frac{N_1(N_1 + N_2 + N_{12} + \bar{z}^{(r)})(\sum_{i=1}^3 N_i + N_{12} + N_{23} + N_{123})}{(N_1 + N_2 + \bar{z}^{(r)})(\sum_{i=1}^3 N_i + N_{12} + N_{23})(\sum_{i=1}^4 N_i + N_{12} + N_{23} + N_{123})} \end{array} \right.$$

- Stop when $\|\hat{\boldsymbol{\alpha}}^{(r+1)} - \hat{\boldsymbol{\alpha}}^{(r)}\|$ is sufficiently small.

An illustration for the method is given latter.

Bayesian approach

Considering that $P(\mathbf{N}_u, Z | \boldsymbol{\alpha})$, the distribution of the observation, has the form (3.4.17), a natural conjugate prior for $\boldsymbol{\alpha}$ is a NDD. We assume a NDD prior distribution for $\boldsymbol{\alpha}$ with parameters $(\mathbf{a}, \mathbf{b}) = ((a_1, a_2, a_3, a_4), (b_1, b_2, b_3))$ as in (3.4.12). Then, according to the Bayesian theorem, the distribution of $\boldsymbol{\alpha}$ given the data with uncertainties \mathbf{N}_u and the latent variable Z is

$$\begin{aligned} \pi(\boldsymbol{\alpha} | \mathbf{N}_u, Z) &\propto P(\mathbf{N}_u, Z | \boldsymbol{\alpha})\pi(\boldsymbol{\alpha}) \\ &\propto \alpha_1^{N_1+a_1-1} \alpha_2^{N_2+Z+a_2-1} \alpha_3^{N_3+N_{23}+a_3-Z-1} \alpha_4^{N_4+a_4-1} \times \\ &\quad \times \alpha_1^{b_1} (\alpha_1 + \alpha_2)^{N_{12}+b_2} (\alpha_1 + \alpha_2 + \alpha_3)^{N_{123}+b_3} \end{aligned}$$

Thus, $\boldsymbol{\alpha} | (\mathbf{N}_u, Z) \sim ND_{4,3}(\mathbf{a}^*, \mathbf{b}^*)$ where $\mathbf{a}^* = (a_1^*, \dots, a_4^*) = (N_1 + a_1, N_2 + a_2 + z, N_3 + N_{23} + a_3 - z, N_4 + a_4)$ and $\mathbf{b}^* = (b_1^*, b_2^*, b_3^*) = (b_1, N_{12} + b_2, N_{123} + b_3)$.

Using the law of total probability, the posterior distribution of $\boldsymbol{\alpha}$ given \mathbf{N}_u can be expressed as

$$\pi(\boldsymbol{\alpha} | \mathbf{N}_u) = \sum_{z=0}^{N_{23}} \pi(\boldsymbol{\alpha} | \mathbf{N}_u, z) P(Z = z | \mathbf{N}_u) \quad (3.4.21)$$

Therefore, in order to obtain $\pi(\boldsymbol{\alpha} | \mathbf{N}_u)$, the conditional distribution $\pi(Z | \mathbf{N}_u)$ is needed. We apply the inverse Bayesian formula (IBF), which is presented in detail in Appendix B, to get this distribution. The method is performed as follows.

Firstly, the conditional distribution of Z given \mathbf{N}_u and $\boldsymbol{\alpha}$ is known as a binomial distribution as in (3.4.20). According to the sampling IBF, for each $z \in \{0, \dots, N_{23}\}$ and with any $\boldsymbol{\alpha}^{(0)}$ we have

$$\begin{aligned} Pr(Z = z | \mathbf{N}_u) &\propto \frac{Pr(Z = z | \mathbf{N}_u, \boldsymbol{\alpha}^{(0)})}{\pi(\boldsymbol{\alpha}^{(0)} | \mathbf{N}_u, Z = z)} \\ &\propto \frac{C_{N_{23}}^z \prod_{j=1}^3 B(d_j, a_{j+1}^*) \prod_{i=1}^4 \alpha_i^{1-N_i-a_i}}{(\alpha_1 + \alpha_2)^{b_2^*} (\alpha_1 + \alpha_2 + \alpha_3)^{b_3^*} (\alpha_2 + \alpha_3)^{N_{23}}} \\ &\propto F(\boldsymbol{\alpha}^{(0)}) C_{N_{23}}^z B(d_1, a_2^*) B(d_2, a_3^*) \\ &\propto q_z(\boldsymbol{\alpha}^{(0)}) \end{aligned} \quad (3.4.22)$$

where $B(x, y)$ stands for the value of Beta function at (x, y) ; $F(\boldsymbol{\alpha}^{(0)})$ does not depend on z and

$$F(\boldsymbol{\alpha}^{(0)}) = \frac{B(d_3, a_4^*) \prod_{i=1}^4 \alpha_i^{1-a_i-N_i}}{(\alpha_1 + \alpha_2)^{b_2^*} (\alpha_1 + \alpha_2 + \alpha_3)^{b_3^*} (\alpha_2 + \alpha_3)^{N_{23}}}$$

Then, the conditional distribution of Z given \mathbf{N}_u can be achieved as

$$p_z = Pr(Z = z | \mathbf{N}_u) = \frac{q_z(\boldsymbol{\alpha}^{(0)})}{\sum_{k=0}^{N_{23}} q_k(\boldsymbol{\alpha}^{(0)})}, \quad z = 0, \dots, N_{23}, \quad (3.4.23)$$

where $q_z(\boldsymbol{\alpha}^{(0)})$ is defined in (3.4.22).

The probabilities p_z , $z = 0, \dots, N_{23}$ are independent from the choice of $\boldsymbol{\alpha}^{(0)}$.

Therefore, the conditional distribution of $\boldsymbol{\alpha}$ given \mathbf{N}_u has the form

$$\boldsymbol{\alpha} | \mathbf{N}_u \sim \sum_{i=0}^{N_{23}} p_i ND_{4,3}(\mathbf{a}^{(i)*}, \mathbf{b}^*), \quad (3.4.24)$$

where $\mathbf{a}^{(i)*} = (N_1 + a_1, N_2 + a_2 + i, N_3 + N_{23} + a_3 - i, N_4 + a_4)$ and $\mathbf{b}^* = (b_1, N_{12} + b_2, N_{123} + b_3)$, and $p_i = P(Z = i | N_{23} = N_{23})$, $i = 0, \dots, N_{23}$. That is to say, the posterior distribution of $\boldsymbol{\alpha}$ given the data with uncertainties $\mathbf{N}_u = (N_1, N_2, N_3, N_4, N_{12}, N_{23}, N_{123})$ is a mixture of NDD with weights p_i .

Illustrative example

To illustrate the proposed methods for dealing with the data belonging to scheme 2 we consider a hypothetical set of data: $\mathbf{N}_u = (5, 3, 1, 0, 3, 4, 2)$. Applying the EM algorithm we obtain the MLE of $\boldsymbol{\alpha}$, which is $\hat{\boldsymbol{\alpha}} = (0.395, 0.497, 0.106, 0.000)$. The choice of initial value $\boldsymbol{\alpha}^{(0)}$ has an insignificant effect on the converge of the algorithm, which is usually less than 30 iterations. For Bayesian inference, we also follow the strategy presented in Section 3.4.2.1 to get the hyperparameters of the prior distribution of $\boldsymbol{\alpha}$. In particular, we suppose that the number of CCF events according to the experience of analysts in a given time is $\mathbf{N}_g = (10, 6, 3, 1)$. The MLE of $\boldsymbol{\alpha}$ according to (3.3.1) is $\hat{\boldsymbol{\alpha}} = (0.50, 0.30, 0.15, 0.05)$. Then we assign $\mathbf{t} = (0.50, 0.30, 0.15, 0.05)$ and investigate some values of s . By calculating the weights in (3.4.23), we obtain an explicit prior distribution of $\boldsymbol{\alpha}$ as in (3.4.24). Table 3.8 presents the MLE and the mean of the posterior distribution of $\boldsymbol{\alpha}$ based on different values of s .

		α_1	α_2	α_3	α_4
MLE		0.395	0.498	0.107	0.00
Prior	$\mathcal{D}(0.5, 0.5, 0.5, 0.5)$	0.383	0.456	0.134	0.024
	$s = 1$	0.405	0.479	0.113	0.003
	$s = 2$	0.413	0.463	0.118	0.004
	$s = 10$	0.445	0.401	0.135	0.017

Table 3.8: MLE and Bayesian estimate of $\boldsymbol{\alpha}$ for scheme 2 with different priors.

A simulation study

In this section, we evaluate the performance of the proposed methods through simulations. Considering a system of $m = 4$ components with the input parameters $\boldsymbol{\alpha} = (0.56, 0.31, 0.11, 0.02)$, we simulate a total of 30 CCF events. We firstly draw (N_1^*, \dots, N_m^*) from a multinomial distribution with parameters $(30; 0.56, 0.31, 0.11, 0.02)$, where N_i^* represents the number of CCF events of

order i . The value N_1^* is then distributed to N_1, N_{12} and N_{123} . To be more precise, we simulate a multinomial distribution with parameters (N_1^*, p_1, p_2, p_3) to obtain N_1, u_1, u_2 where $N_1 + u_1 + u_2 = N_1^*$; u_1 is the portion of N_1^* which contributes to N_{12} ; u_2 is the portion of N_1^* which contributes to N_{123} ; and (p_1, p_2, p_3) is predetermined probabilities. Similarly, the value N_2^* is subdivided into N_2, N_{12}, N_{23} and N_{123} ; and the value N_3^* is subdivided into N_3, N_{23} and N_{123} .

The obtained confounded data, which have the form $(N_1, N_2, N_3, N_4, N_{12}, N_{23}, N_{123})$, are then used to estimate the α -factor parameters.

The process is implemented repeatedly as follows.

(1) Simulate $(N_1^*, \dots, N_4^*) \sim \mathcal{M}(30; 0.56, 0.31, 0.11, 0.02)$.

(2) Simulate:

$$\begin{aligned} (N_1, u_1, u_2) &\sim \mathcal{M}(N_1^*; p_1, p_2, p_3), \\ (N_2, v_1, v_2, v_3) &\sim \mathcal{M}(N_2^*; q_1, q_2, q_3, q_4) \\ (N_3, w_1, w_2) &\sim \mathcal{M}(N_3^*; r_1, r_2, r_3). \end{aligned}$$

(3) Set $N_{12} = u_1 + v_1$, $N_{23} = v_2 + w_1$, $N_{123} = u_2 + v_3 + w_2$.

(4) Estimate $\hat{\alpha} = (\hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3, \hat{\alpha}_4)$ using the EM algorithm or the Bayesian approach described in Section 3.4.2.2.

(5) Repeat Steps 1 - 4 and then calculate the mean of estimates and the mean square error from the estimated $\hat{\alpha}$ in each step.

In this procedure, the probabilities (p_1, p_2, p_3) , (r_1, r_2, r_3) and (q_1, q_2, q_3, q_4) represent the mechanism to degrade the data, which correspond to the uncertainties. Tables 3.9 - 3.10 show the results of simulation from $M = 10000$ iterations, where the Table 3.9 corresponds to a low level of data degradation and Table 3.10 corresponds to a high level of data degradation. Several possibilities of prior are considered. The obtained results show that the accuracy of MLE depends on the level of degradation of data: the closer are these probabilities to the inputs, the more accurate is the MLE.

The accuracy of Bayesian estimates depend on both the prior distribution and the level of data degradation, but the effect of prior distribution overwhelms the effect of the level of data degradation. From this point of view, Bayesian approach is more robust to the level of data degradation which represents the uncertainty data compared to the frequentist approach. We also calculate the average squared difference between MLE and Bayesian estimates (denoted by Δ_D) from each level of data degradation by the formula

$$\Delta_D = \frac{1}{M} \left((MLE^{(i)} - Bayesian^{(i)})^2 \right), \quad (3.4.25)$$

where $MLE^{(i)}$ and $Bayesian^{(i)}$ are the MLE and Bayesian estimates for the i^{th} simulation. In general, the lower level of data degradation gives smaller Δ_D , confirming again the fact that lower level of data degradation leads to better estimation. For example, using the noninformative prior of Dirichlet distribution, the value of Δ_D for the (p_1, p_2, p_3) , (r_1, r_2, r_3) and (q_1, q_2, q_3, q_4) in Tables 3.9 is $\Delta_D = 0.00042$, which is smaller to $\Delta_D = 0.00088$ corresponding to the probabilities in Table 3.10. In addition, for Bayesian approach, Type C (corresponding to a small value of s and better guess of α) and Type B (corresponding to a large value of s and worse guess of α) of prior in general lead to better estimates. That is to say, in the case the guess at α is reliable, the small value of s should be used. Otherwise, the larger values of s should be used.

	α_1	α_2	α_3	α_4
Input	0.56	0.31	0.11	0.02
MLE	0.5224 (0.0102)	0.3802 (0.0140)	0.0770 (0.0030)	0.0201 (0.0004)
$\mathcal{D}(0.5, 0.5, 0.5, 0.5)$	0.5091 (0.0106)	0.3769 (0.0126)	0.0852 (0.0023)	0.0286 (0.0004)
Guess \mathbf{t}	0.55	0.3	0.12	0.03
$s = 2$	0.5224 (0.0098)	0.3786 (0.0133)	0.0787 (0.0028)	0.0202 (0.0004)
$s = 30$	0.5455 (0.0019)	0.3250 (0.0017)	0.1032 (0.0005)	0.0259 (0.0002)
Guess \mathbf{t}	0.7	0.15	0.09	0.06
$s = 2$	0.5343 (0.0074)	0.3558 (0.0094)	0.0787 (0.0026)	0.0310 (0.0004)
$s = 30$	0.6290 (0.0116)	0.2600 (0.0118)	0.0867 (0.0010)	0.0241 (0.0007)

Table 3.9: MLE and Bayesian estimate of α for scheme 2 when $(p_1, p_2, p_3) = (r_1, r_2, r_3) = (0.5, 0.3, 0.2)$, $(q_1, q_2, q_3, q_4) = (0.5, 0.38, 0.09, 0.03)$. The values in parentheses are the MSE.

	α_1	α_2	α_3	α_4
Input	0.56	0.31	0.11	0.02
MLE	0.4664 (0.0207)	0.4359 (0.02931)	0.0777 (0.0040)	0.0199 (0.00037)
$\mathcal{D}(0.5, 0.5, 0.5, 0.5)$	0.4585 (0.0349)	0.4199 (0.0319)	0.0924 (0.0057)	0.0290 (0.0012)
Guess \mathbf{t}	0.55	0.3	0.12	0.03
$s=2$	0.4889 (0.0251)	0.4018 (0.0303)	0.0878 (0.0059)	0.0214 (0.0008)
$s=30$	0.5369 (0.0020)	0.3287 (0.0015)	0.1082 (0.0004)	0.0261 (0.0001)
Guess \mathbf{t}	0.7	0.15	0.09	0.06
$s=2$	0.5253 (0.0218)	0.3637 (0.0264)	0.0871 (0.0067)	0.0237 (0.0008)
$s=30$	0.6586 (0.0113)	0.2018 (0.0129)	0.0953 (0.0007)	0.0441 (0.0007)

Table 3.10: MLE and Bayesian estimate of α for scheme 2 when $(p_1, p_2, p_3) = (r_1, r_2, r_3) = (1/3, 1/3, 1/3)$, $(q_1, q_2, q_3, q_4) = (1/4, 1/4, 1/4, 1/4)$. The values in parentheses are the MSE.

3.4.2.3 Inference for scheme 3

In this scheme, the data with uncertainties are represented by $\mathbf{N}_u = (N_1, N_2, N_3, N_4, N_{12}, N_{34})$ and the likelihood function is

$$L(\boldsymbol{\alpha}|\mathbf{N}_u) = \prod_{i=1}^4 \alpha_i^{N_i} (\alpha_1 + \alpha_2)^{N_{12}} (\alpha_3 + \alpha_4)^{N_{34}}. \quad (3.4.26)$$

This likelihood has the form of a *grouped Dirichlet distribution*. We are going to use this argument to make inference. Let us recall the definition of the grouped Dirichlet distribution.

Grouped Dirichlet distribution

The grouped Dirichlet distribution is also firstly defined in [61] and then further studied in [39]. A random vector $\mathbf{x} \in \Delta_n$ is said to be a grouped Dirichlet distribution (*GDD*) with two partitions, denoted by $\mathbf{x} \sim GD_{n,2,s}(\mathbf{a}, \mathbf{b})$ on Δ_n , if its density is of the form [61]

$$f(\mathbf{x}|\mathbf{a}, \mathbf{b}) = c^{-1} \left(\prod_{i=1}^n x_i^{a_i-1} \right) \left(\sum_{i=1}^s x_i \right)^{b_1} \left(\sum_{i=s+1}^n x_i \right)^{b_2}, \quad (3.4.27)$$

where $\mathbf{a} = (a_1, \dots, a_n)^T$ and $\mathbf{b} = (b_1, b_2)^T$ are two non-negative parameter vectors, s is a known positive integer, $0 < s < n$ and c is the normalizing constant. The constant c is defined as

$$c = B(a_1, \dots, a_s) B(a_{s+1}, \dots, a_n) B\left(\sum_{i=1}^s a_i + b_1, \sum_{i=s+1}^n a_i + b_2\right),$$

where

$$B(a_1, \dots, a_s) = \prod_{i=1}^s \frac{\Gamma(a_i)}{\Gamma(\sum_{i=1}^s a_i)}.$$

Particularly, a $GD_{n,2,s}(\mathbf{a}, \mathbf{0})$ is a Dirichlet distribution with parameter \mathbf{a} . Let us partition vectors $\mathbf{x}_{n \times 1}$ and $\mathbf{a}_{n \times 1}$ into $(\mathbf{x}^{(1)T}, \mathbf{x}^{(2)T})$ and $(\mathbf{a}^{(1)T}, \mathbf{a}^{(2)T})$, each with s and $n - s$ elements, respectively. The following proposition [39] provides a procedure for generating *i.i.d.* samples from a *GDD*.

Proposition 4. *A random vector \mathbf{x} follows a grouped Dirichlet distribution if and only if*

$$\mathbf{x} = \begin{pmatrix} \mathbf{x}^{(1)T} \\ \mathbf{x}^{(2)T} \end{pmatrix} = \begin{pmatrix} R \cdot \mathbf{y}^{(1)T} \\ (1 - R) \cdot \mathbf{y}^{(2)T} \end{pmatrix},$$

where $\mathbf{y}^{(1)} \sim \mathcal{D}_s(\mathbf{a}^{(1)})$ on Δ_s , $\mathbf{y}^{(2)} \sim \mathcal{D}_{n-s}(\mathbf{a}^{(2)})$ on Δ_{n-s} , $R \sim \text{Beta}(\sum_{i=1}^s a_i^{(1)} + b_1, \sum_{j=1}^{n-s} a_j^{(2)} + b_2)$ and $y^{(1)}, y^{(2)}$ and R are mutually independent.

Proposition 5. *Let $\mathbf{x} \sim GD_{n,2,s}(\mathbf{a}, \mathbf{b})$ on Δ_n . for any $r_1, \dots, r_n \geq 0$, the mixed moments of \mathbf{x} are given by*

$$E\left(\prod_{i=1}^n x_i^{r_i}\right) = \frac{B(\mathbf{a}^{(1)} + \mathbf{r}^{(1)})}{B(\mathbf{a}^{(1)})} \frac{B(\mathbf{a}^{(2)} + \mathbf{r}^{(2)})}{B(\mathbf{a}^{(2)})} \frac{B(\sum_{i=1}^s (a_i + r_i) + b_1, \sum_{i=s+1}^n (a_i + r_i) + b_2)}{B(\mathbf{a}^{(1)})}$$

where $\mathbf{r} = (\mathbf{r}^{(1)T}, \mathbf{r}^{(2)T})$.

Proposition 6. Let $\mathbf{x} \sim GD_{n,2,s}(\mathbf{a}, \mathbf{b})$ on Δ_n , then the **mode** of \mathbf{x} is

$$m_i = \frac{a_i - 1}{\sum_{i=1}^n (a_i) + b_1 + b_2 - n} \left(\left(1 + \frac{b_1}{\sum_{i=1}^s a_i - s} \right) \cdot \mathbb{1}_{(1 \leq i \leq s)} \right. \\ \left. + \left(1 + \frac{b_2}{\sum_{i=s+1}^n a_i - n + s} \right) \cdot \mathbb{1}_{(s+1 \leq i \leq n)} \right), i = 1, \dots, n$$

where $\mathbb{1}_{(\cdot)}$ denotes the indicator function.

The proof for propositions 1- 5 has been given in Ng et al. [39]. We present here a proof for the proposition 6.

Proof. Let L denote the log-likelihood function corresponding to the GDD in (3.4.27). We have

$$L = \sum_{i=1}^{n-1} (a_i - 1) \log(x_i) + (a_n - 1) \log(x_n) \\ + b_1 \log\left(\sum_{i=1}^s x_i\right) + b_2 \log\left(1 - \sum_{i=1}^s x_i\right).$$

The partial derivatives of L with respect to x_i set to zero leads to

$$\begin{cases} \frac{a_i - 1}{x_i} - \frac{a_n - 1}{x_n} + \frac{b_1}{\sum_{i=1}^s x_i} - \frac{b_2}{1 - \sum_{i=1}^s x_i} = 0, & 1 \leq i \leq s, \\ \frac{a_i - 1}{x_i} - \frac{a_n - 1}{\sum_{i=s+1}^n x_i} = 0, & s + 1 \leq i \leq n. \end{cases} \quad (3.4.28)$$

The first equation of (3.4.28) yields $x_i = \frac{a_i - 1}{a_1 - 1} x_1 \forall 1 < i \leq s$ while the second one yields $x_i = \frac{a_i - 1}{a_n - 1} x_n \forall s + 1 \leq i < n$. Since $\sum_{i=1}^n x_i = 1$, we have

$$x_1 \left(\sum_{i=1}^s \frac{a_i - 1}{a_1 - 1} \right) + x_n \left(\sum_{i=s+1}^n \frac{a_i - 1}{a_n - 1} \right) = 1,$$

and then

$$x_n = \frac{(a_n - 1) \left(1 - x_1 \left(\sum_{i=1}^s \frac{a_i - 1}{a_1 - 1} \right) \right)}{\sum_{i=s+1}^n (a_i - 1)}.$$

Substituting these values of $x_i, 1 \leq i \leq s$ and x_n into the first equation of (3.4.28) for the case $i = 1$ we obtain

$$\frac{a_1 - 1}{x_1} - \frac{\sum_{i=s+1}^n (a_i - 1)}{1 - x_1 \left(\sum_{i=1}^s \frac{a_i - 1}{a_1 - 1} \right)} + \frac{b_1}{x_1 \left(\sum_{i=1}^s \frac{a_i - 1}{a_1 - 1} \right)} - \frac{b_2}{1 - x_1 \left(\sum_{i=1}^s \frac{a_i - 1}{a_1 - 1} \right)} = 0.$$

Simple computation from this equation leads to

$$x_1 = \frac{a_1 - 1}{\sum_{i=1}^n (a_i) + b_1 + b_2 - n} \cdot \frac{\sum_{i=1}^s a_i + b_1 - s}{\sum_{i=1}^s a_i - s}.$$

The formula for x_i is obtained from the relation between x_i , $i = 2, \dots, n$ with x_1 as expressed above. \square

Maximum likelihood method

The MLE of α from the likelihood function in (3.4.26) can be found by applying directly the proposition 6, which leads to

$$\begin{cases} \hat{\alpha}_i = \frac{N_i}{\sum_{i=1}^4 N_i + N_{12} + N_{34}} \left(1 + \frac{N_{12}}{N_1 + N_2}\right), & i = 1, 2, \\ \hat{\alpha}_i = \frac{N_i}{\sum_{i=1}^4 N_i + N_{12} + N_{34}} \left(1 + \frac{N_{34}}{N_3 + N_4}\right), & i = 3, 4. \end{cases} \quad (3.4.29)$$

Bayesian approach

Because the distribution of the observation $P(\mathbf{N}_u | \alpha)$ has the form (3.4.26), we use a GDD with hyperparameters $(\mathbf{a}, \mathbf{b}) = ((a_1, a_2, a_3, a_4), (b_1, b_2))$ as a conjugate prior distribution for α . That is

$$\pi(\alpha) \propto \prod_{i=1}^4 \alpha_i^{a_i-1} (\alpha_1 + \alpha_2)^{b_1} (\alpha_3 + \alpha_4)^{b_2}. \quad (3.4.30)$$

According to the Bayesian theorem, we have

$$\begin{aligned} \pi(\alpha | \mathbf{N}_u) &\propto P(\mathbf{N}_u | \alpha) \pi(\alpha) \\ &\propto \prod_{i=1}^4 \alpha_i^{a_i + N_i - 1} (\alpha_1 + \alpha_2)^{b_1 + N_{12}} (\alpha_3 + \alpha_4)^{b_2 + N_{34}} \end{aligned}$$

That means, the posterior distribution of α is also a GDD with parameters $(\mathbf{a}^*, \mathbf{b}^*)$, where $\mathbf{a}^* = (N_1 + a_1, N_2 + a_2, N_3 + a_3, N_4 + a_4)$ and $\mathbf{b}^* = (b_1 + N_{12}, b_2 + N_{34})$.

Considering a quadratic loss function, the Bayes estimators of α are the posterior means, which are defined from the proposition 5:

$$\left\{ \begin{array}{l} \tilde{\alpha}_1 = \frac{(a_1 + N_1)(\sum_{i=1}^2(a_i + N_i) + b_1 + N_{12})}{(\sum_{i=1}^4(a_i + N_i) + b_1 + b_2 + N_{12} + N_{34})(\sum_{i=1}^2(a_i + N_i))}, \\ \tilde{\alpha}_2 = \frac{(a_2 + N_2)(\sum_{i=1}^2(a_i + N_i) + b_1 + N_{12})}{(\sum_{i=1}^4(a_i + N_i) + b_1 + b_2 + N_{12} + N_{34})(\sum_{i=1}^2(a_i + N_i))}, \\ \tilde{\alpha}_3 = \frac{(a_3 + N_3)(\sum_{i=3}^4(a_i + N_i) + b_2 + N_{34})}{(\sum_{i=1}^4(a_i + N_i) + b_1 + b_2 + N_{12} + N_{34})(\sum_{i=3}^4(a_i + N_i))}, \\ \tilde{\alpha}_4 = \frac{(a_4 + N_4)(\sum_{i=3}^4(a_i + N_i) + b_2 + N_{34})}{(\sum_{i=1}^4(a_i + N_i) + b_1 + b_2 + N_{12} + N_{34})(\sum_{i=3}^4(a_i + N_i))}. \end{array} \right.$$

Similar to the expression in (3.4.2.1), we write $\mathbf{a} = \mathbf{st} = (st_1, \dots, st_4)$ in the conjugate GDD prior, i.e. $a_i = st_i, i = 1, \dots, 4$, where $s > 0, t_i \geq 0$ and $\sum_{i=1}^4 t_i = 1$. If a guess at the expectation value of α_i can be achieved by the analysts, we assign these values to the mean of the GDD prior, leading to:

$$\left\{ \begin{array}{l} \frac{\bar{\alpha}_1}{t_1} = \frac{\bar{\alpha}_2}{t_2}; \quad \frac{\bar{\alpha}_3}{t_3} = \frac{\bar{\alpha}_4}{t_4}, \\ st_1(1 + \frac{b_1}{st_1 + st_2}) = \bar{\alpha}_1 \sum_{i=1}^4 (st_i + b_i), \\ st_3(1 + \frac{b_3}{st_3 + st_4}) = \bar{\alpha}_3 \sum_{i=1}^4 (st_i + b_i). \end{array} \right. \quad (3.4.31)$$

There are many solutions for the equations (3.4.31). The simplest way is to choose a Dirichlet distribution for the prior, i.e. $\mathbf{b} = (0, 0, 0)$.

Substituting $\mathbf{b} = (0, 0, 0)$ into (3.4.31) we obtain:

$$\frac{\bar{\alpha}_1}{t_1} = \frac{\bar{\alpha}_2}{t_2} = \frac{\bar{\alpha}_3}{t_3} = \frac{\bar{\alpha}_4}{t_4}, \quad (3.4.32)$$

Since $\sum_{i=1}^4 \alpha_i = \sum_{i=1}^4 t_i = 1$, (3.4.32) leads to $t_i = \bar{\alpha}_i$. The variation of the prior is now represented by the choice of s .

Illustrative example

Let us consider another hypothetical situation of the data with uncertainties, $\mathbf{N}_u = (5, 3, 1, 0, 4, 2)$. Applying the formulas in (3.4.2.9) we obtain the MLE of α as $\hat{\alpha} = (0.5, 0.3, 0.2, 0.0)$. For the Bayesian scheme, similar to the strategy used in section 3.4.2.1, a set of $\mathbf{N}_g = (9, 5, 1, 0.5)$ has been assumed for the number of CCF events in a given time by the analysts. According to (3.3.1), the MLE of α corresponding to this set of complete data is $\hat{\alpha} = (0.58, 0.32, 0.07, 0.03)$.

These values are assigned to (t_1, t_2, t_3, t_4) . The posterior means are presented in Table 3.7 for different values of s . We also consider the case of the prior Dirichlet distribution (Type A) with parameters $(0.5, 0.5, 0.5, 0.5)$.

		α_1	α_2	α_3	α_4
MLE		0.5	0.3	0.2	0.0
Prior	$\mathcal{D}(0.5, 0.5, 0.5, 0.5)$	0.4673	0.2973	0.1764	0.0588
	$s = 1$	0.505	0.301	0.189	0.005
	$s = 2$	0.5102	0.3015	0.1788	0.0094
	$s = 10$	0.5336	0.3063	0.1360	0.0240

Table 3.11: MLE and Bayesian estimate of α for scheme 3 with different priors.

An alternative approach: binomial expansion

The binomial expansion can be applied to the expression (3.4.26). Indeed, the distribution of the observation in (3.4.26) can be rewritten as:

$$P(\mathbf{N}_u | \alpha) = \sum_{k=0}^{N_{12}} \sum_{\ell=0}^{N_{34}} C_{N_{12}}^k C_{N_{34}}^\ell \alpha_1^{N_1+k} \alpha_2^{N_2+N_{12}-k} \alpha_3^{N_3+\ell} \alpha_4^{N_4+N_{34}-\ell}$$

Suppose that $\alpha \sim \mathcal{D}(a_1, a_2, a_3, a_4)$, the Bayes' theorem leads to

$$\pi(\alpha | \mathbf{N}_u) \propto \sum_{k=0}^{N_{12}} \sum_{\ell=0}^{N_{34}} C_{N_{12}}^k C_{N_{34}}^\ell \alpha_1^{a_1^*(k,\ell)-1} \alpha_2^{a_2^*(k,\ell)-1} \alpha_3^{a_3^*(k,\ell)-1} \alpha_4^{a_4^*(k,\ell)-1}$$

Thus, the posterior distribution of α is a mixture of Dirichlet distribution:

$$\alpha | \mathbf{N}_u \sim \sum_{k=0}^{N_{12}} \sum_{\ell=0}^{N_{34}} w(k, \ell) \mathcal{D}(\mathbf{a}^*(k, \ell)),$$

where

$$w(k, \ell) = \frac{C_{N_{12}}^k C_{N_{34}}^\ell B(\mathbf{a}^*(k, \ell))}{\sum_{k=0}^{N_{12}} \sum_{\ell=0}^{N_{34}} C_{N_{12}}^k C_{N_{34}}^\ell B(\mathbf{a}^*(k, \ell))}$$

and $\mathcal{D}(\mathbf{a}^*(k, \ell))$ is a Dirichlet distribution with parameters $\mathbf{a}^*(k, \ell)$, in which

$$\begin{aligned} \mathbf{a}^*(k, \ell) &= (a_1^*(k, \ell), a_2^*(k, \ell), a_3^*(k, \ell), a_4^*(k, \ell)) \\ a_1^*(k, \ell) &= a_1 + N_1 + k, \\ a_2^*(k, \ell) &= a_2 + N_2 + N_{12} - k, \\ a_3^*(k, \ell) &= a_3 + N_3 + \ell, \\ a_4^*(k, \ell) &= a_4 + N_4 + N_{34} - \ell. \end{aligned}$$

Under a quadratic loss, the Bayesian estimators are:

$$\tilde{\alpha}_i = \frac{\sum_{k=0}^{N_{12}} \sum_{\ell=0}^{N_{34}} w(k, \ell) a_i^*(k, \ell)}{\sum_{i=1}^4 a_i^*(k, \ell)}, \quad i = 1, \dots, 4.$$

Consider the previous example of confounded data with $\mathbf{N}_u = (5, 3, 1, 0, 4, 2)$. Table 3.12 shows the estimate of α using both the GDD in the previous section and the binomial expansion method with the same prior, $\alpha \sim \mathcal{D}(1.16, 0.64, 0.14, 0.06)$ (corresponding to $s = 2$ from previous example).

The running time for each algorithm is recorded in the right most column. It can be seen that both two methods give the same estimates but the running time using the binomial expansion is twelve time higher than one using the GDD.

Method	α_1	α_2	α_3	α_4	Time(seconds)
GDD	0.5102	0.3015	0.1788	0.0094	0.0176
Binomial expansion	0.5102	0.3015	0.1788	0.0094	0.2122

Table 3.12: Compare the estimate of α and the runing time using the GDD and the binomial expansion method.

3.4.2.4 Inference for scheme 4

As in section 3.4.2.2, we suppose that one more term, that is N_{23} , is added to the data with uncertainties for scheme 4. Therefore, the data with uncertainties are now represented by:

$$\mathbf{N}_u = (N_1, N_2, N_3, N_4, N_{12}, N_{34}, N_{23}).$$

The likelihood function is

$$L(\alpha|\mathbf{N}_u) = \prod_{i=1}^4 \alpha_i^{N_i} (\alpha_1 + \alpha_2)^{N_{12}} (\alpha_3 + \alpha_4)^{N_{34}} (\alpha_2 + \alpha_3)^{N_{23}}. \quad (3.4.33)$$

In this case, the binomial expansion could be applied. However, it will be time-consuming since the likelihood function contains more terms. Moreover, we have checked that the solutions of the corresponding log-likelihood equations based on (3.4.33) have no close-form expressions.

Maximum likelihood approach

As for scheme 2, we introduce a latent variable Z for the term N_{23} and use a partial EM algorithm.

With the latent variable Z , the likelihood function is

$$L(\alpha|\mathbf{N}_u, Z) = \alpha_1^{N_1} \alpha_2^{N_2+Z} \alpha_3^{N_3+N_{23}-Z} \alpha_4^{N_4} (\alpha_1 + \alpha_2)^{N_{12}} (\alpha_3 + \alpha_4)^{N_{34}}, \quad (3.4.34)$$

and the MLE are:

$$\left\{ \begin{array}{l} \hat{\alpha}_1 = \frac{N_1}{\sum_{i=1}^4 N_i + N_{12} + N_{34} + N_{23}} \left(1 + \frac{N_{12}}{N_1 + N_2}\right), \\ \hat{\alpha}_2 = \frac{N_2 + Z}{\sum_{i=1}^4 N_i + N_{12} + N_{34} + N_{23}} \left(1 + \frac{N_{12}}{N_1 + N_2}\right), \\ \hat{\alpha}_3 = \frac{N_3 + N_{23} - Z}{\sum_{i=1}^4 N_i + N_{12} + N_{34} + N_{23}} \left(1 + \frac{N_{34}}{N_3 + N_4}\right), \\ \hat{\alpha}_4 = \frac{N_4}{\sum_{i=1}^4 N_i + N_{12} + N_{34} + N_{23}} \left(1 + \frac{N_{34}}{N_3 + N_4}\right). \end{array} \right. \quad (3.4.35)$$

Replacing Z with the conditional expectation

$$E(Z | \boldsymbol{\alpha}, \mathbf{N}_u) = N_{23} \frac{\alpha_2}{\alpha_2 + \alpha_3},$$

the partial EM algorithm (i.e. the EM algorithm is only applied to the latent variable Z) is performed as follows:

- Initialisation: $\boldsymbol{\alpha}^{(0)} = (\alpha_1^{(0)}, \alpha_2^{(0)}, \alpha_3^{(0)}, \alpha_4^{(0)})$
- At step $(r+1)$, $\boldsymbol{\alpha}^{(r)} = (\alpha_1^{(r)}, \alpha_2^{(r)}, \alpha_3^{(r)}, \alpha_4^{(r)})$ being available, compute

$$\bar{Z}^{(r)} = N_{23} \alpha_2^{(r)} / (\alpha_2^{(r)} + \alpha_3^{(r)})$$

and

$$\left\{ \begin{array}{l} \alpha_1^{(r+1)} = \frac{N_1}{\sum_{i=1}^4 N_i + N_{12} + N_{34} + N_{23}} \left(1 + \frac{N_{12}}{N_1 + N_2}\right), \\ \alpha_2^{(r+1)} = \frac{N_2 + \bar{Z}^{(r)}}{\sum_{i=1}^4 N_i + N_{12} + N_{34} + N_{23}} \left(1 + \frac{N_{12}}{N_1 + N_2}\right), \\ \alpha_3^{(r+1)} = \frac{N_3 + N_{23} - \bar{Z}^{(r)}}{\sum_{i=1}^4 N_i + N_{12} + N_{34} + N_{23}} \left(1 + \frac{N_{34}}{N_3 + N_4}\right), \\ \alpha_4^{(r+1)} = \frac{N_4}{\sum_{i=1}^4 N_i + N_{12} + N_{34} + N_{23}} \left(1 + \frac{N_{34}}{N_3 + N_4}\right). \end{array} \right.$$

- Iterate Step 2 until convergence.

In this procedure, the values of α_1 and α_4 do not change over the iterations. In the following, we present a Bayesian method for this scheme before providing a simulation study and an illustrative example for both methods.

Bayesian approach

Due to the distribution of the observation $P(\mathbf{N}_u, Z | \boldsymbol{\alpha})$ as in (3.4.34), the GDD will be a conjugate prior. We suppose that the prior distribution of $\boldsymbol{\alpha}$ is a GDD with parameters (\mathbf{a}, \mathbf{b}) where $\mathbf{a} = (a_1, a_2, a_3, a_4)$, $\mathbf{b} = (b_1, b_2)$ as described in (3.4.30).

The Bayes' theorem leads to

$$\begin{aligned} \pi(\boldsymbol{\alpha} | \mathbf{N}_u, Z) &\propto \pi(\mathbf{N}_u, Z | \boldsymbol{\alpha})\pi(\boldsymbol{\alpha}) \\ &\propto \alpha_1^{N_1+a_1-1} \alpha_2^{N_2+Z+a_2-1} \alpha_3^{N_3+N_{23}+a_3-Z-1} \alpha_4^{N_4+a_4-1} \times \\ &\quad \times (\alpha_1 + \alpha_2)^{N_{12}+b_1} (\alpha_3 + \alpha_4)^{N_{34}+b_2} \end{aligned}$$

That is to say, given the data with uncertainties \mathbf{N}_u and the latent variable Z , the conditional distribution of $\boldsymbol{\alpha}$ is a GDD with parameters $(\mathbf{a}^*, \mathbf{b}^*)$:

$$\boldsymbol{\alpha} | (\mathbf{N}_u, Z) \sim GD_{4,2,2}(\mathbf{a}^*, \mathbf{b}^*) \quad (3.4.36)$$

where $\mathbf{a}^* = (a_1^*, \dots, a_4^*) = (N_1 + a_1, N_2 + a_2 + Z, N_3 + N_{23} + a_3 - Z, N_4 + a_4)$ and $\mathbf{b}^* = (b_1^*, b_2^*) = (b_1 + N_{12}, b_2 + N_{34})$.

As for scheme 2, we write

$$\pi(\boldsymbol{\alpha} | \mathbf{N}_u) = \sum_{z=0}^{N_{23}} \pi(\boldsymbol{\alpha} | \mathbf{N}_u, z) P(Z = z | \mathbf{N}_u) \quad (3.4.37)$$

where $P(Z = z | \mathbf{N}_u)$ is a binomial distribution with parameters $(N_{23}, \alpha_2 / (\alpha_2 + \alpha_3))$ and we use the IBF sampling technique as follows.

For each $z \in \{0, \dots, N_{23}\}$, we have

$$\begin{aligned} P(Z = z | \mathbf{N}_u) &\propto \frac{\pi(Z = z | \mathbf{N}_u, \boldsymbol{\alpha}^{(0)})}{\pi(\boldsymbol{\alpha}^{(0)} | \mathbf{N}_u, Z = z)} \\ &= \frac{C_{N_{23}}^k B(a_1^*, a_2^*) B(a_3^*, a_4^*) B(\sum_{i=1}^2 a_i^* + b_1^*, \sum_{i=3}^4 a_i^* + b_2^*)}{(\alpha_1 + \alpha_2)^{b_1^*} (\alpha_3 + \alpha_4)^{b_2^*} (\alpha_2 + \alpha_3)^{N_{23}}} \times \\ &\quad \times \prod_{i=1}^4 \alpha_i^{1-N_i-a_i} \\ &\propto F(\boldsymbol{\alpha}^{(0)}) C_{N_{23}}^k B(a_1^*, a_2^*) B(a_3^*, a_4^*) B(\sum_{i=1}^2 a_i^* + b_1^*, \sum_{i=3}^4 a_i^* + b_2^*) \\ &\propto q_z(\boldsymbol{\alpha}^{(0)}) \end{aligned}$$

where $\boldsymbol{\alpha}^{(0)}$ is an arbitrary value of $\boldsymbol{\alpha}$; $F(\boldsymbol{\alpha}^{(0)})$ is independent from Z and

$$F(\boldsymbol{\alpha}^{(0)}) = \frac{\prod_{i=1}^4 \alpha_i^{1-N_i-a_i}}{(\alpha_1 + \alpha_2)^{b_1^*} (\alpha_3 + \alpha_4)^{b_2^*} (\alpha_2 + \alpha_3)^{N_{23}}}.$$

Thus, the conditional distribution of Z given \mathbf{N}_u can be defined as

$$p_z = P(Z = z | \mathbf{N}_u) = \frac{q_z(\boldsymbol{\alpha}^{(0)})}{\sum_{k=0}^{N_{23}} q_k(\boldsymbol{\alpha}^{(0)})}, \quad z = 0, \dots, N_{23}. \quad (3.4.38)$$

. Remark that the conditional distribution of $Z \mid \mathbf{N}_u$ in does not depend on $\boldsymbol{\alpha}^{(0)}$.

After finding this distribution, combining with (3.4.36), we obtain an explicit formula for the posterior distribution of $\boldsymbol{\alpha}$ given \mathbf{N}_u as

$$\boldsymbol{\alpha} \mid \mathbf{N}_u \sim \sum_{i=0}^{N_{23}} p_i GD_{4,2,2}(\mathbf{a}^{(i)*}, \mathbf{b}^*), \quad (3.4.39)$$

where $\mathbf{a}^{(i)*} = (N_1 + \alpha_1, N_2 + \alpha_2 + i, N_3 + N_{23} + \alpha_3 - i, N_4 + \alpha_4)$ and $\mathbf{b}^* = (b_1 + N_{12}, b_2 + N_{34})$. This is a mixture of the GDD.

Illustrative example

Consider another fictitious set of the data with uncertainties, that $\mathbf{N}_u = (5, 3, 1, 0, 3, 2, 4)$. The EM algorithm leads to the MLE of $\boldsymbol{\alpha}$ as $\hat{\boldsymbol{\alpha}} = (0.395, 0.497, 0.106, 0.00)$. The algorithm converged after less than 25 iterations regardless of the choice of initial value $\boldsymbol{\alpha}^{(0)}$. For Bayesian inference, we firstly elicit some values of hyperparameters following the strategy presented in section 3.4.2.1. We suppose that analysts agree a guess at the number of CCF events of $\mathbf{N}_g = (10, 6, 3, 1)$. The MLE of $\boldsymbol{\alpha}$ according to (3.3.1) is $\hat{\boldsymbol{\alpha}} = (0.50, 0.30, 0.15, 0.05)$. Then we assign $\mathbf{t} = (0.50, 0.30, 0.15, 0.05)$ and investigate some values of s . We also consider the noninformative prior of Dirichlet distribution which is a GDD with paramters $\mathbf{a} = (0.5, 0.5, 0.5, 0.5)$ and $\mathbf{b} = (0, 0)$. Table 3.13 displays the MLE and the mean of the posterior distribution of $\boldsymbol{\alpha} \mid \mathbf{N}_u$ with several choices of prior.

		α_1	α_2	α_3	α_4
Prior	MLE	0.3578	0.3873	0.2548	0.0000
	$\mathcal{D}(0.5, 0.5, 0.5, 0.5)$	0.3478	0.3743	0.2370	0.0407
	$s = 1$	0.368	0.380	0.247	0.005
	$s = 2$	0.3754	0.3762	0.2393	0.0089
	$s = 10$	0.4151	0.3538	0.2046	0.0263

Table 3.13: MLE and Bayesian estimate of $\boldsymbol{\alpha}$ for scheme 4 with different priors.

A simulation study

Similar to the section 3.4.2.2, the performance of proposed methods for scheme 4 of data with uncertainties is evaluated through simulations. We also suppose a system of $m = 4$ identical components with a set of $N = 30$ CCF events observed. The input parameters are $\boldsymbol{\alpha} = (0.56, 0.31, 0.11, 0.02)$.

The following steps are performed in the simulation procedure:

- (1) Generate $(N_1^*, \dots, N_4^*) \sim \mathcal{M}(30; 0.56, 0.31, 0.11, 0.02)$

(2) Generate:

$$\begin{aligned} (N_1, u_1) &\sim \mathcal{M}(N_1^*; p_1, p_2) \\ (N_2, v_1, v_2) &\sim \mathcal{M}(N_2^*; q_1, q_2, q_3) \\ (N_3, w_1, w_2) &\sim \mathcal{M}(N_3^*; r_1, r_2, r_3) \\ (N_4, t_1) &\sim \mathcal{M}(N_4^*; s_1, s_2) \end{aligned}$$

(3) Set $N_{12} = u_1 + v_1, N_{23} = v_2 + w_1, N_{34} = w_2 + t_1$.

(4) Estimate $\hat{\boldsymbol{\alpha}} = (\hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3, \hat{\alpha}_4)$ using the EM algorithm or the Bayesian inference in this section.

(5) Repeat Steps 1 - 4 and then calculate the mean and the mean square error from the estimate of $\hat{\boldsymbol{\alpha}}$ in each step.

The probabilities (p_1, p_2) , (s_1, s_2) , (q_1, q_2, q_3) and (r_1, r_2, r_3) represent the mechanism to degrade the data that leads to the uncertainties. Tables 3.14 show the average of MLE and Bayesian estimate of $\boldsymbol{\alpha}$ from $M = 10000$ iterations for a low level of data degradation while Table 3.15 shows the same results for a higher level of data degradation. The obtained results confirm the same conclusion as in previous section: the MLE is sensitive to the choice of the level of data degradation while the the Baysian estimates is more robust in the sense that the prior can reduce the dependence of this data degradation on the estimates. In addition, with the low level of data degradation, the average squared difference between MLE and Bayesian estimates defined in 3.4.25 is smaller. For example, using the same GDD prior with parameters $\mathbf{a} = (0.5, 0.5, 0.5, 0.5)$, $\mathbf{b} = (0, 0)$, we have $\Delta_D = 0.00053$ for the level of data degradation in Table 3.15 which is larger than $\Delta_D = 0.00034$ for the level of data degradation in Table 3.14.

	α_1	α_2	α_3	α_4
Input	0.56	0.31	0.11	0.02
MLE	0.5578 (0.0068)	0.3221 (0.0071)	0.1017 (0.0026)	0.0182 (0.0005)
$\mathcal{D}(0.5, 0.5, 0.5, 0.5)$	0.5445 (0.0063)	0.3201 (0.0064)	0.1060 (0.0022)	0.0293 (0.0005)
Guess \mathbf{t}	0.55	0.3	0.12	0.03
$s = 2$	0.5565 (0.0061)	0.3209 (0.0063)	0.1033 (0.0022)	0.0191 (0.0004)
$s = 30$	0.5561 (0.0020)	0.3115 (0.0018)	0.1093 (0.0007)	0.0229 (0.0001)
Guess \mathbf{t}	0.7	0.15	0.09	0.06
$s = 2$	0.5660 (0.0061)	0.3128 (0.0062)	0.1005 (0.0023)	0.0206 (0.0004)
$s = 30$	0.6256 (0.0063)	0.2402 (0.0067)	0.0978 (0.0009)	0.0362 (0.0004)

Table 3.14: MLE and Bayesian estimate of $\boldsymbol{\alpha}$ when $(p_1, p_2) = (s_1, s_2) = (0.7, 0.3)$, $(q_1, q_2, q_3) = (r_1, r_2, r_3) = (0.5, 0.3, 0.2)$. The values in parentheses are the MSE.

	α_1	α_2	α_3	α_4
Input	0.56	0.31	0.11	0.02
MLE	0.4976 (0.0124)	0.3785 (0.0145)	0.1074 (0.0031)	0.0163 (0.0006)
$\mathcal{D}(0.5, 0.5, 0.5, 0.5)$	0.4890 (0.0130)	0.3699 (0.0125)	0.1116 (0.0026)	0.0292 (0.0005)
Guess \mathbf{t}	0.55	0.3	0.12	0.03
$s = 2$	0.5025 (0.0107)	0.3703 (0.0121)	0.1094 (0.0027)	0.0176 (0.0005)
$s = 30$	0.5289 (0.0030)	0.3339 (0.0024)	0.1147 (0.0008)	0.0224 (0.0001)
Guess \mathbf{t}	0.7	0.15	0.09	0.06
$s = 2$	0.5148 (0.0094)	0.3578 (0.01098)	0.1075 (0.0029)	0.0197 (0.0004)
$s = 30$	0.6072 (0.0044)	0.2527 (0.0053)	0.1034 (0.0009)	0.0365 (0.0004)

Table 3.15: MLE and Bayesian estimate of $\boldsymbol{\alpha}$ when $(p_1, p_2) = (s_1, s_2) = (0.5, 0.5)$, $(q_1, q_2, q_3) = (r_1, r_2, r_3) = (1/3, 1/3, 1/3)$. The values in parentheses are the MSE.

3.5 Concluding remarks

In this chapter, we have proposed a new method to handle the data with uncertainties. We have used a confounded representation of the data with uncertainties. Several techniques to treat this kind of data have been provided.

One can see that the strategy will strongly depends on the nature of the available data. It could be the partial EM algorithm, it could be the IBF sampling technique with the Nested Dirichlet distribution or it could be the IBF sampling technique with the Grouped Dirichlet distribution.

Clearly, it is not possible to give a general method but by combining the proposed techniques, many situations of data with uncertainties could be covered. In addition, several perspectives can be released from the framework presented in this chapter. For example, we can investigate the properties of the MLE of the α -factor model. Investigating noninformative prior distributions for the NDD and the GDD could be also a challenging problem.

Chapter 4

Causal inference from incomplete contingency table

In this chapter, we present a model to deal with the CCF data that contain not only occurrence frequencies but also the triggering causes. These data form a contingency table. We consider a situation of incomplete data where only the margins of the contingency table are observed. We apply an IBF sampling technique to obtain the posterior distribution of the intensity of the HPP representing the frequency of each cell of the table. A relationship between the parameters of our model and the parameters of the α -decomposition model [71] is pointed out to make comparison.

4.1 Introduction

The previous models are based purely on the order of CCF events, i.e., the number of failed components, but not on the causes of each event.

In practice, potential causes of CCF can be recorded. Examples for systems in U.S. commercial NPPs such as emergency diesel generators (EDG), motor-operated valves, pumps, and circuit breakers are given in [48].

One of these examples considers 41 EDG CCF events. 15 events are classified in the cause group: *Design / Construction / Installation / Manufacture*, 9 events are classified in the cause group: *Internal to Component*, 9 events are classified in the cause group: *Operational / Human Error*, 5 events are classified in the cause group: *External Environment* and 3 events are classified in the cause group: *Other causes*.

In [47], the authors provide a way to achieve further understanding of the occurrence of CCF events by presenting the process of data collection and grouping the ranking of proximate failure causes (A proximate cause refers to a characterization of the condition that is considered as having led to the failure). Using Bayesian networks, Kelly et al. [18] propose the preliminary

framework of a causality-based inference, providing cause-specific quantitative insights into likely causes of failures. O'Connor and Mosleh [50] develop a general cause based methodology for analysis of common cause and dependent failures in system risk and reliability assessments. Zheng et al. [71] suggest the α -decomposition model to estimate CCF parameters corresponding to the α -factor model to deal with order of CCF frequencies and causes frequencies. The data are on the form of an incomplete contingency table.

Our purpose in this chapter is to make inference on contingency table crossing orders of CCF and causes of CCF where only the margins are observed.

Assuming that each cell of the contingency table follows a Poisson distribution with a constant rate. We use an IBF sampling technique to obtain the posterior distribution of the parameters of the model. A new stochastic version of the IBF sampling technique based on the Metropolis-Hastings (M-H) algorithm is suggested to avoid a cumbersome calculation when the space of complete data is large. We compare our results with the results obtained with the α -decomposition model [71].

The chapter is organized as follows. In section 2, we describe the incomplete data and propose a new model to analyze the CCF data in a contingency table form. For purpose of comparison, we briefly recall the α -decomposition factor model suggested in [71]. Section 3 gives a Bayesian inference based methodology for the proposed model. In section 4, we present a stochastic version of the IBF sampling technique to overcome the difficulty when the space of complete data is large which might lead to heavy computations. A simulation study and an application of the proposed method are displayed in section 4 and section 5, respectively.

4.2 The data and the model

The data

Suppose that from a m -component system, the potential causes of failures can be classified into k main groups, C_1, C_2, \dots, C_k . Then the contingency table crossing orders of CCF and causes has a size: $m \times k$. Table 4.1 displays a general and complete representation of a contingency table, where $N_{i,j}$ represents the number of CCF events of order i from the cause j .

In this table, the sum of the terms in column j , corresponding to the total number of components failed by cause C_j , is equal to:

$$\sum_{i=1}^m N_{ij} = N_{\bullet j}, \quad j = 1, \dots, k.$$

The sum of the terms in row i , corresponding to the total number of CCF

	Cause 1	Cause 2	...	Cause k	Total
Order 1	N_{11}	N_{12}	...	N_{1k}	$N_{1\bullet}$
Order 2	N_{21}	$N_{2,2}$...	N_{2k}	$N_{2\bullet}$
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
Order m	N_{m1}	N_{m2}	...	N_{mk}	$N_{m\bullet}$
Total	$N_{\bullet 1}$	$N_{\bullet 2}$...	$N_{\bullet k}$	

Table 4.1: A complete contingency table

events of order i from any causes, is equal to:

$$\sum_{j=1}^k N_{ij} = N_{i\bullet}, \quad i = 1, \dots, m.$$

The sum of all the terms in the table which is the total number of observations, is equal to:

$$\sum_{i=1}^m \sum_{j=1}^k N_{ij} = \sum_{j=1}^k N_{\bullet j} = \sum_{i=1}^m N_{i\bullet}.$$

The matrix $\{N_{ij}\}$, $i = 1, \dots, m$, $j = 1, \dots, k$ that we denote \mathbf{N} , represents the **complete data**. By incomplete data in a contingency table, we mean that the exact value in each cell of the table is not observed. Suppose that we only know the sum of each column and each row of the table, i.e. the margins, that is to say, we observe $\mathbf{N}_{\bullet} = \{N_{i\bullet}, N_{\bullet j}\}_{i=1, \dots, m; j=1, \dots, k}$. \mathbf{N}_{\bullet} represents the **incomplete data**.

Inference methods for contingency tables with missing data are suggested by Little & Rubin [30]. In the context of CCF, this problem is addressed by Zheng et al. [71]. The authors simulate CCF events and causes for 16 systems with 3 components (Table 4.2). For each system, only the total number of CCF events of a given order and the total number of specific cause are recorded. We do not know exactly the number of failures of a specific order due to a specific cause. Remark that the data for each system can be organized as a contingency table where only the margins are known. For example, the data from the first system in Table 4.2 can be expressed as in Table 4.3.

In the following, we present two models to deal with this kind of data. The first model is the α -decomposition model and the second one is introduced based on the Poisson distribution.

The α -decomposition model

Consider a system of m identical components. Recall that the parameters of the α -factor model have been defined in Chapter 3, which are $(\alpha_1, \dots, \alpha_m)$, where α_i refers to the probability that exact i components are involved in a CCF event given that at least one component failed. The parameters are

	Common causes's occurrence			Total	CCF event		
	Cause 1	Cause 2	Cause 3		Order 1	Order 2	Order 3
System 1	32	28	67	127	113	11	3
System 2	17	78	11	106	98	7	1
System 3	18	19	50	87	73	9	5
System 4	29	6	31	66	53	5	8
System 5	7	33	10	50	45	4	1
System 6	15	9	17	41	33	3	5
System 7	12	15	7	34	32	2	0
System 8	2	22	7	31	29	2	0
System 9	7	4	11	22	20	2	0
System 10	10	8	3	21	20	2	0
System 11	3	6	10	19	16	2	1
System 12	7	3	6	16	14	1	1
System 13	3	5	7	15	13	1	1
System 14	5	3	7	15	12	1	2
System 15	4	5	2	11	9	1	1
System 16	1	6	2	9	7	1	1

Table 4.2: A hypothetical database of CCF events provided in [71]

	Cause 1	Cause 2	Cause 3	Total
Order 1	x	x	x	113
Order 2	x	x	x	11
Order 3	x	x	x	3
Total	32	28	67	127

Table 4.3: An expression of observed incomplete data from the first system in Table 4.2

then decomposed into the parameters of the α -decomposition model as proposed in [71]. In order to understand the parameters of the model and figure out the link between the parameters of this model and the parameters in the proposed model (presented in the next section), we briefly explain the α -decomposition model in the following.

Let A_1, \dots, A_m denote m components of the system. Suppose that potential causes that may affect the system can be divided into k main causes, say C_1, C_2, \dots, C_k . The conditional probability of the failure event that involves a specific component of the system, say A_1 (or any other component of the system), given the occurrence of a specific cause, say C_j , is

$$\begin{aligned}
 P(A_1 | C_j) = & P(A_{1,I} | C_j) + \sum_{i=2}^m P(A_1 A_i | C_j) + \dots + \sum_{2 \leq i < s}^m P(A_1 A_i A_s | C_j) \\
 & + P(A_1 \dots A_m | C_j)
 \end{aligned} \tag{4.2.1}$$

where $A_{1,I}$ refers to the failures involved only A_1 , $A_1 A_i$ refers to the CCFs involved 2 components including A_1 (a given failed component) and so on. It should be considered that under this definition the considered effects are supposed to be mutually exclusive.

Let us define α factor related to a specific cause. We denote

$$\begin{aligned}
 \alpha_1^{C_j} &= \frac{P(A_{1,I} | C_j)}{P(A_1 | C_j)}, \\
 \alpha_2^{C_j} &= \sum_{i=2}^m \frac{P(A_1 A_i | C_j)}{P(A_1 | C_j)}, \\
 \alpha_3^{C_j} &= \sum_{2 \leq i < s}^m P \frac{P(A_1 A_i A_s | C_j)}{P(A_1 | C_j)}, \\
 &\vdots \\
 \alpha_m^{C_j} &= \frac{P(A_1 A_2 \dots A_m | C_j)}{P(A_1 | C_j)}.
 \end{aligned} \tag{4.2.2}$$

From (4.2.1) we have

$$\sum_{i=1}^m \alpha_i^{C_j} = 1.$$

According to the law of total probability, the probability of a failure event where only A_1 fails is

$$P(A_{1,I}) = \sum_{i=1}^k P(A_{1,I} | C_i) P(C_i). \tag{4.2.3}$$

The first equation in (4.2.2) leads to $P(A_{1,I} | C_i) = \alpha_1^{C_i} P(A_1 | C_i)$.

Substituting this equation into (4.2.3), we get:

$$P(A_{1,I}) = \sum_{i=1}^k \alpha_1^{C_i} P(A_1 | C_i) P(C_i). \tag{4.2.4}$$

Dividing both sides of (4.2.4) by $P(A_1)$, the probability of a CCF event that involves A_1 is obtained as

$$\frac{P(A_{1,I})}{P(A_1)} = \sum_{i=1}^k \alpha_1^{C_i} \frac{P(A_1 | C_i) P(C_i)}{P(A_1)} \tag{4.2.5}$$

In [71], the authors considered the staggered testing scheme, i.e., only one component is tested in a test episode. The rest of components will be tested only when a failure from the first test is detected. Under this scheme, the parameter α_1 of the α -factor model is defined as:

$$\alpha_1 = \frac{P(A_{1,I})}{P(A_1)}.$$

Then (4.2.5) can be rewritten by

$$\alpha_1 = \sum_{i=1}^k \alpha_1^{C_i} r_i, \tag{4.2.6}$$

where

$$r_i = \frac{P(A_1 | C_i)P(C_i)}{P(A_1)} = P(C_i | A_1),$$

r_i is the occurrence rate of Cause i that generated the failure of A_1 .

In the same way, under the assumption that the causes and the corresponding probabilities are the same for each identical component all the parameters α_i of the α -factor model can be decomposed into $\alpha_i^{C_j}$ by the formula:

$$\alpha_i = \sum_{j=1}^k \alpha_i^{C_j} r_j, \quad i = 1, \dots, m. \quad (4.2.7)$$

The parameters $\{\alpha_i^{C_j}\}_{i=1, \dots, m; j=1, \dots, k}$ are now the new parameters of interest. They are the parameters of the α -decomposition model.

The inference on $\alpha_i^{C_j}$ in [71] is drawn by using a hierarchical Bayesian modeling via OpenBUGS.

The Poisson distribution based model

We propose modelling the data in a contingency table form by using the Poisson distribution. In particular, we suppose that the cell counts are independent homogeneous Poisson processes. That is to say, we assume that $N_{ij}(T)$, the number of CCF events of order i due to Cause j over the time window $[0, T]$, follows a Poisson distribution with parameter $\lambda_{ij}T$.

In the case of complete data, the distribution of \mathbf{N} is obtained as a product of Poisson distributions:

$$P(\mathbf{N} = \mathbf{n} | \boldsymbol{\lambda}) = \prod_{i,j} \frac{(\lambda_{ij}T)^{n_{ij}}}{n_{ij}!} e^{-\lambda_{ij}T}, \quad (4.2.8)$$

where $\mathbf{n} = \{n_{ij}\}_{i=1, \dots, m; j=1, \dots, k}$ is a representation of \mathbf{N} .

The distribution of the incomplete data can then be expressed as:

$$P(\mathbf{N}_\bullet = \mathbf{n}_\bullet | \boldsymbol{\lambda}) = \sum_{\mathbf{u} \in \mathcal{D}_{\mathbf{n}_\bullet}} P(\mathbf{N} = \mathbf{u} | \boldsymbol{\lambda}). \quad (4.2.9)$$

where \mathbf{n}_\bullet denotes a representation of \mathbf{N}_\bullet and $\mathcal{D}_{\mathbf{n}_\bullet}$ denotes the set of all $m \times k$ matrices $\mathbf{u} = \{u_{ij}\}_{i=1, \dots, m; j=1, \dots, k}$ such that $\sum_{i=1}^m u_{ij} = n_{\bullet j}$ and $\sum_{j=1}^k u_{ij} = n_{i \bullet}$, i.e. the space of complete data that satisfy the condition of the observed margins.

When N_{ij} is observed, the estimate for the λ_{ij} based on (4.2.8) is easy to obtain. However, in the case of incomplete data, i.e. only margins are observable, the inference on these parameters is not straightforward.

Our goal is to provide a method to estimate the matrix $m \times k$ with terms λ_{ij} , denoted by $\boldsymbol{\lambda}$, based on the incomplete data of the contingency table.

We present in the next section a Bayesian method for making inference on $\boldsymbol{\lambda}$.

4.3 Bayesian inference

4.3.1 Using the IBF sampling

In order to obtain the posterior distribution of $\boldsymbol{\lambda}$ given the margins of the contingency table \mathbf{N}_\bullet , we firstly rewrite $\pi(\boldsymbol{\lambda} | \mathbf{N}_\bullet)$ in the following form based on the law of total probability:

$$\pi(\boldsymbol{\lambda} | \mathbf{N}_\bullet) = \sum_{\mathbf{n} \in \mathcal{D}_{\mathbf{n}_\bullet}} \pi(\boldsymbol{\lambda} | \mathbf{N}_\bullet, \mathbf{N} = \mathbf{n}) \pi(\mathbf{N} = \mathbf{n} | \mathbf{N}_\bullet) \quad (4.3.1)$$

Each element in summation (4.3.1) is a product of two terms: $\pi(\boldsymbol{\lambda} | \mathbf{N}_\bullet, \mathbf{N} = \mathbf{n})$ and $\pi(\mathbf{N} = \mathbf{n} | \mathbf{N}_\bullet)$; where the first term is the conditional distribution of $\boldsymbol{\lambda}$ given the complete data \mathbf{N} (but missing) and the observed data \mathbf{N}_\bullet , the second term is the conditional distribution of complete data \mathbf{N} given the incomplete data \mathbf{N}_\bullet (observable). Each of them is calculated as follows.

Since the distribution of complete data in (4.2.8) is the product of Poisson distributions, we consider a Gamma distribution $\mathcal{G}(\alpha_{ij}, \beta_{ij})$ as a conjugate prior for each λ_{ij} and suppose that the priors of parameters are independent. Thus, apriori independent the prior of $\boldsymbol{\lambda}$ is a product of Gamma distributions:

$$\pi(\boldsymbol{\lambda}) = \prod_{i,j} \frac{\beta_{ij}^{\alpha_{ij}}}{\Gamma(\alpha_{ij})} \lambda_{ij}^{\alpha_{ij}-1} e^{-\beta_{ij}\lambda_{ij}}. \quad (4.3.2)$$

The posterior distribution of $\boldsymbol{\lambda}$ given $\mathbf{N} = \mathbf{n} \in \mathcal{D}_{\mathbf{n}_\bullet}$ and \mathbf{N}_\bullet is now obtained according to the Bayes' theorem:

$$\begin{aligned} \pi(\boldsymbol{\lambda} | \mathbf{N}_\bullet, \mathbf{N} = \mathbf{n}) &\propto P(\mathbf{N} = \mathbf{n}, \mathbf{N}_\bullet | \boldsymbol{\lambda}) \pi(\boldsymbol{\lambda}) \\ &\propto \prod_{i,j} \lambda_{ij}^{n_{ij} + \alpha_{ij} - 1} e^{-\lambda_{ij}(T + \beta_{ij})}. \end{aligned} \quad (4.3.3)$$

That means, the first term $\pi(\boldsymbol{\lambda} | \mathbf{N}_\bullet, \mathbf{N} = \mathbf{n})$ in the summation (4.3.1) is defined as a product of Gamma distributions with parameters $(n_{ij} + \alpha_{ij}, T + \beta_{ij})$.

In order to calculate the second term, we apply the sampling-wise IBF (Appendix B), that is

$$P(\mathbf{N} = \mathbf{n} | \mathbf{N}_\bullet) \propto \frac{P(\mathbf{N} = \mathbf{n} | \mathbf{N}_\bullet, \boldsymbol{\lambda}^*)}{\pi(\boldsymbol{\lambda}^* | \mathbf{N} = \mathbf{n}, \mathbf{N}_\bullet)}, \quad (4.3.4)$$

where $\boldsymbol{\lambda}^*$ is an arbitrary value in the support of $\boldsymbol{\lambda}$.

The denominator in (4.3.4) is defined as in (4.3.3), which is also a product of Gamma distributions with the same parameters, where $\boldsymbol{\lambda}$ is repalced by $\boldsymbol{\lambda}^*$.

The numerator in (4.3.4) is defined as

$$\begin{aligned} P(\mathbf{N} = \mathbf{n} | \mathbf{N}_\bullet; \boldsymbol{\lambda}^*) &= \frac{P(\mathbf{N} = \mathbf{n} | \boldsymbol{\lambda}^*)}{\sum_{\mathbf{u} \in \mathcal{D}_{\mathbf{n}_\bullet}} P(\mathbf{N} = \mathbf{u} | \boldsymbol{\lambda}^*) \pi(\boldsymbol{\lambda}^*)} \\ &\propto P(\mathbf{N} = \mathbf{n} | \boldsymbol{\lambda}^*) \\ &\propto \prod_{i,j} \frac{(\lambda_{ij}^* T)^{n_{ij}}}{n_{ij}!} e^{-\lambda_{ij}^* T} \end{aligned} \quad (4.3.5)$$

With the numerator defined in (4.3.5) and the denominator defined in (4.3.3), (4.3.4) becomes:

$$P(\mathbf{N} = \mathbf{n} \mid \mathbf{N}_\bullet) \propto \prod_{i,j} \frac{\Gamma(\alpha_{ij} + n_{ij})}{n_{ij}!(T + \beta_{ij})^{n_{ij} + \alpha_{ij}}} \lambda_{ij}^{*1 - \alpha_{ij}} e^{\lambda_{ij}^* \beta_{ij}}. \quad (4.3.6)$$

Setting $\mathcal{D}_{\mathbf{n}_\bullet} = \{\mathbf{n}_1, \dots, \mathbf{n}_K\}$, for each $\mathbf{n}_\ell \in \mathcal{D}_{\mathbf{n}_\bullet}$, we denote the term on the right of (4.3.6) by q_ℓ . Then, the conditional probability of the event $\mathbf{N} = \mathbf{n}_\ell$ given the observed data \mathbf{N}_\bullet is

$$P(\mathbf{N} = \mathbf{n}_\ell \mid \mathbf{N}_\bullet) = \frac{q_\ell(\boldsymbol{\lambda}^*)}{\sum_{\ell=1}^K q_\ell(\boldsymbol{\lambda}^*)} = p_\ell \quad (4.3.7)$$

and this conditional probability is independent from $\boldsymbol{\lambda}^*$.

After finding both distributions in each term of the summation in (4.3.1), we can rewrite the posterior distribution of $\boldsymbol{\lambda}$ given the observed incomplete data \mathbf{N}_\bullet in (4.3.1) as

$$\pi(\boldsymbol{\lambda} \mid \mathbf{N}_\bullet) = \sum_{\ell=1}^K p_\ell \pi(\boldsymbol{\lambda} \mid \mathbf{N} = \mathbf{n}_\ell). \quad (4.3.8)$$

Moreover, (4.3.8) can be applied for each λ_{ij} , i.e, the posterior distribution of λ_{ij} given \mathbf{N}_\bullet can be expressed by

$$\pi(\lambda_{ij} \mid \mathbf{N}_\bullet) = \sum_{\ell=1}^K p_\ell \pi(\lambda_{ij} \mid \mathbf{N} = \mathbf{n}_\ell), \quad (4.3.9)$$

which is mixture of Gamma distributions. That is to say, we obtain an explicit formula for the posterior of each λ_{ij} .

In the case when the space $\mathcal{D}_{\mathbf{n}_\bullet}$ is large, the explicit formulas for the λ'_{ij} 's could be cumbersome, the IBF sampling approach could be conducted as an alternative way to get the posterior of $\boldsymbol{\lambda}$. The main idea of the approach is to simulate a set of samples $\{\boldsymbol{\lambda}^{(i)}\}$ from $\pi(\boldsymbol{\lambda} \mid \mathbf{N}_\bullet, \mathbf{N}^{(i)})$ where $\mathbf{N}^{(i)} \sim \pi(\mathbf{N} \mid \mathbf{N}_\bullet)$.

In particular, this sampling IBF method is performed by the following steps:

- Determine the space $\mathcal{D}_{\mathbf{n}_\bullet} = \{\mathbf{n}_1, \dots, \mathbf{n}_K\}$ of the complete data that satisfy the condition of observed incomplete data $\mathbf{N}_\bullet = \mathbf{n}_\bullet$, and calculate the corresponding probability p_1, \dots, p_K as in (4.3.7).
- Generate *i.i.d.* samples $\mathbf{n}^{(1)}, \dots, \mathbf{n}^{(M)}$ of \mathbf{N} from the space with corresponding probabilities $\{p_1, \dots, p_K\}$.
- Simulate $\boldsymbol{\lambda}^{(i)} \sim \pi(\boldsymbol{\lambda} \mid \mathbf{N}_\bullet, \mathbf{N} = \mathbf{n}^{(i)})$ for $i = 1, \dots, M$, then $\{\boldsymbol{\lambda}^{(1)}, \dots, \boldsymbol{\lambda}^{(M)}\}$ are considered as a *i.i.d.* samples from the posterior distribution $\pi(\boldsymbol{\lambda} \mid \mathbf{N}_\bullet)$.

The inference for the parameters of interest $\boldsymbol{\lambda}$ is now made directly from these samples.

Illustrative example

Let us consider a simple illustrative example with a hypothetical set of CCF data from a 2-component system over the time window $T = 10$ (unit of time). Suppose that the potential causes are divided into three main groups, say Cause 1, Cause 2 and Cause 3. The data is presented in Table 4.4. From this

	Cause 1	Cause 2	Cause 3	Total
Order 1	x	x	x	4
Order 2	x	x	x	6
Total	3	5	2	10

Table 4.4: An illustrative example of incomplete data

set of data, the margins, or observed data is $\mathbf{n}_\bullet = \{(3, 5, 2), (4, 6)\}$. The value in each cell of the table is unknown. The space $\mathcal{D}_{\mathbf{n}_\bullet}$ can be described as:

$$\mathcal{D}_{\mathbf{n}_\bullet} = \left\{ \begin{bmatrix} 0 & 2 & 2 \\ 3 & 3 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 3 & 1 \\ 3 & 2 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 4 & 0 \\ 3 & 1 & 2 \end{bmatrix}, \begin{bmatrix} 1 & 1 & 2 \\ 2 & 4 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 2 & 1 \\ 2 & 3 & 1 \end{bmatrix}, \right. \\ \left. \begin{bmatrix} 1 & 3 & 0 \\ 2 & 2 & 2 \end{bmatrix}, \begin{bmatrix} 2 & 0 & 2 \\ 1 & 5 & 0 \end{bmatrix}, \begin{bmatrix} 2 & 1 & 1 \\ 1 & 4 & 1 \end{bmatrix}, \begin{bmatrix} 2 & 2 & 0 \\ 1 & 3 & 2 \end{bmatrix}, \begin{bmatrix} 3 & 0 & 1 \\ 0 & 5 & 1 \end{bmatrix}, \right. \\ \left. \begin{bmatrix} 3 & 1 & 0 \\ 0 & 4 & 2 \end{bmatrix} \right\}.$$

The probabilities $p_\ell = P(\mathbf{N} = \mathbf{n}_\ell | \mathbf{N}_\bullet = \mathbf{n}_\bullet)$, $\mathbf{n}_\ell \in \mathcal{D}_{\mathbf{n}_\bullet}$ are calculated according to equation (4.3.7) and displayed in Table 4.5. For purpose of illustration,

ℓ	p_ℓ	ℓ	p_ℓ
1	0.10608204	7	0.13366337
2	0.07072136	8	0.04950495
3	0.12376238	9	0.06364922
4	0.07425743	10	0.14851485
5	0.04243281	11	0.12376238
6	0.06364922		

Table 4.5: The conditional distribution $P(\mathbf{N} | \mathbf{N}_\bullet = \mathbf{n}_\bullet)$

we consider a noninformative distribution for each λ_{ij} , i.e., $\lambda_{ij} \sim \mathcal{G}(0.5, 0)$.

For example, the posterior distribution of λ_{11} is

$$\pi(\lambda_{11} | \mathbf{N}_\bullet = \mathbf{n}_\bullet) = \sum_{N_{11}=0}^3 p_{n_{11}} \mathcal{G}(\lambda_{11}, 0.5 + N_{11}, 10),$$

where the weight $p_{N_{11}}$'s are given in Table 4.6. The Table 4.7 presents the posterior distribution, the mean and the standard deviation for each λ_{ij} . For illustration, we also simulate samples of $\boldsymbol{\lambda} | \mathbf{N}_\bullet = \mathbf{n}_\bullet$ according to the sampling IBF method as follows.

N_{11}	0	1	2	3
$p_{N_{11}}$	0.30	0.18	0.25	0.27

 Table 4.6: *The conditional distribution $\pi(N_{11} | \mathbf{N}_\bullet = \mathbf{n}_\bullet)$*

- (i) Draw $M = 30000$ independent samples $\mathbf{n}^{(i)}, i = 1, \dots, M$ of missing data \mathbf{N} from the space $\mathcal{D}_{\mathbf{n}_\bullet}$ using the discrete distribution given in Table 4.5.
- (ii) Generate $\boldsymbol{\lambda}^{(i)}$ from the posterior distribution $\pi(\boldsymbol{\lambda} | \mathbf{N}_\bullet, \mathbf{N} = \mathbf{n}^{(i)})$ as in (4.3.3).

Under a quadratic loss, the Bayes estimators are the mean of the posterior distribution that we approximate using the samples of $\boldsymbol{\lambda}$ generated by the sampling IBF method. The obtained results are presented in the rightmost column of Table 4.7.

	Posterior $\lambda_{ij} \mathbf{N}_\bullet = \mathbf{n}_\bullet$	Posterior Mean	Sampling mean
λ_{11}	$0.3\mathcal{G}_0 + 0.18\mathcal{G}_1 + 0.25\mathcal{G}_2 + 0.27\mathcal{G}_3$	0.199 (0.194)	0.199 (1.195)
λ_{12}	$0.28\mathcal{G}_0 + 0.25\mathcal{G}_1 + 0.21\mathcal{G}_2 + 0.14\mathcal{G}_3 + 0.12\mathcal{G}_4$	0.207 (0.209)	0.207 (0.209)
λ_{13}	$0.38\mathcal{G}_0 + 0.31\mathcal{G}_1 + 0.31\mathcal{G}_2$	0.143 (0.139)	0.142 (0.138)
λ_{21}	$0.27\mathcal{G}_0 + 0.25\mathcal{G}_1 + 0.18\mathcal{G}_2 + 0.3\mathcal{G}_3$	0.201 (0.195)	0.200 (0.194)
λ_{22}	$0.12\mathcal{G}_1 + 0.14\mathcal{G}_2 + 0.12\mathcal{G}_3 + 0.25\mathcal{G}_4 + 0.28\mathcal{G}_5$	0.366 (0.338)	0.391 (0.351)
λ_{23}	$0.31\mathcal{G}_0 + 0.31\mathcal{G}_1 + 0.38\mathcal{G}_2$	0.157 (0.147)	0.156 (0.145)

 Table 4.7: *Explicit posterior distributions, posterior mean of λ_{ij} , $\mathcal{G}_i = \mathcal{G}(0.5 + i, 10)$. The values in brackets are the corresponding variance.*

The above procedures allows us to get exactly the posterior distribution or to simulate the samples from the posterior distribution $\pi(\boldsymbol{\lambda} | \mathbf{N}_\bullet = \mathbf{n}_\bullet)$. They require the space $\mathcal{D}_{\mathbf{n}_\bullet}$ to be tractable. However, in practice it is not usually the case. The size of the space $\mathcal{D}_{\mathbf{n}_\bullet}$ can be very large. This fact leads to computational limits because of the need to explore the whole space of $\mathcal{D}_{\mathbf{n}_\bullet}$. To overcome this challenge, we suggest to use a stochastic version of IBF sampling method.

4.3.2 Using a stochastic version of the IBF sampling

To simulate from the distribution $\{p_k\}_{k=1, \dots, K}$ when K is large, we use a Metropolis-Hastings (M-H) algorithm. This strategy has an advantage of not needing to compute the denominators in (4.3.7). The algorithm is described as follows. Initially, we pick up an arbitrary element in $\mathcal{D}_{\mathbf{n}_\bullet}$, and denote this element by $\mathbf{n}^{(1)}$.

Let \mathbf{Q} be a specified transition matrix from one element of the space to another. After p cycles of the following scheme, $\mathbf{n}^{(p)}$ is available, do the following.

- Generate a candidate $\mathbf{n}^{(*)} \sim Q(\cdot | \mathbf{n}^{(p)})$.
- Compute

$$\rho = \min \left\{ 1, \frac{p_{\mathbf{n}^{(*)}}}{p_{\mathbf{n}^{(p)}}} = \frac{\prod_{i=1}^m \prod_{j=1}^k \frac{\Gamma(\alpha_{ij} + n_{ij}^{(*)})}{n_{ij}^{(*)}!(T + \beta_{ij})^{n_{ij}^{(*)} + \alpha_{ij}}}{\prod_{i=1}^m \prod_{j=1}^k \frac{\Gamma(\alpha_{ij} + n_{ij}^{(p)})}{n_{ij}^{(p)}!(T + \beta_{ij})^{n_{ij}^{(p)} + \alpha_{ij}}} \right\}.$$

- Set $\mathbf{n}^{(p+1)} = \mathbf{n}^{(*)}$ with probability ρ ; else, set $\mathbf{n}^{(p+1)} = \mathbf{n}^{(p)}$.

The matrix Q is called the generating-candidate matrix. Many choices are possible for Q . In our case, the following scheme is applied to generate the candidate.

- Choose randomly a 2×2 submatrix of $\underline{n}^{(p)}$ with probability $1/(C_k^2 \cdot C_m^2)$, say

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}.$$

- Draw a number x from a uniform distribution in $(\max\{0, a_{\bullet 1} - a_{2\bullet}\}, \dots, \min\{a_{1\bullet}, a_{\bullet 1}\})$, where $a_{1\bullet} = a_{11} + a_{12}$ and $a_{\bullet 1} = a_{11} + a_{21}$.
- Replace the corresponding matrix A in $\underline{n}^{(p)}$ by matrix B where

$$B = \begin{pmatrix} x & a_{1\bullet} - x \\ a_{\bullet 1} - x & a_{2\bullet} - a_{\bullet 1} + x \end{pmatrix}$$

and $a_{2\bullet} = a_{2,1} + a_{2,2}$.

After generating M *i.i.d.* samples $\mathbf{n}^{(1)}, \dots, \mathbf{n}^{(M)}$ according to the method described above, we generate $\boldsymbol{\lambda}^{(\ell)} \sim \pi(\boldsymbol{\lambda} | \mathbf{N}_{\bullet}, \mathbf{n}^{(\ell)})$ from the posterior distribution as in (4.3.3). Thus $\{\boldsymbol{\lambda}^{(\ell)}\}_{\ell=1, \dots, M}$ are *i.i.d.* samples from the posterior $\pi(\boldsymbol{\lambda} | \mathbf{N}_{\bullet})$.

Figure 4.1 compares the samples generated from conditional discrete distribution $P(\mathbf{N} | \mathbf{N}_{\bullet} = \mathbf{n}_{\bullet})$ corresponding to both the exact probabilities p_{ℓ} in Table 4.5 and the probabilities obtained with the M-H algorithm from the data $\mathbf{n}_{\bullet} = \{(3, 5, 2), (4, 6)\}$.

Sampling from $\pi(\boldsymbol{\lambda} | \mathbf{N}_{\bullet} = \mathbf{n}_{\bullet})$ with the M-H algorithm and computing the empirical mean, we obtain the following estimates for $\boldsymbol{\lambda}$:

$$\hat{\boldsymbol{\lambda}} = \begin{bmatrix} 0.1978159 & 0.2069460 & 0.1431974 \\ 0.2018113 & 0.3933634 & 0.1561672 \end{bmatrix}.$$

which are close to the estimates obtained with the exact distribution p_{ℓ} , $\ell = 1, \dots, K$.

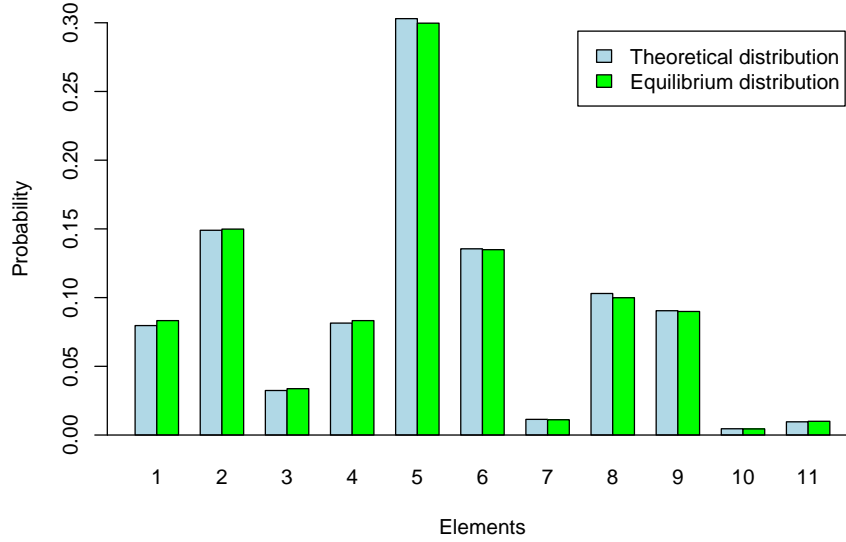


Figure 4.1: Comparison of the theoretical distribution $P(\mathbf{N}|\mathbf{N}_\bullet = \mathbf{n}_\bullet)$ and the equilibrium distribution of the $M-H$ samples for the simple case

4.4 A simulation study

We evaluate the performance of the proposed method through simulations. The simulating procedure is performed according to the following steps:

1. Assign an input value for $\boldsymbol{\lambda} = (\lambda_{ij})$ and a time window T .
2. Simulate the value in each cell of contingency table which follows a Poisson distribution with parameter $\lambda_{ij}T$ and calculate the corresponding margins of the contingency table generated.
3. Use the stochastic version of IBF sampling method to estimate $\boldsymbol{\lambda}$ based on the two margins calculated above, denote this estimate by $\hat{\boldsymbol{\lambda}}^{(1)}$.
4. Repeat steps 2-3 W times to obtain the estimates $\hat{\boldsymbol{\lambda}}^{(1)}, \dots, \hat{\boldsymbol{\lambda}}^{(W)}$.
5. Consider mean of these values, $\hat{\boldsymbol{\lambda}} = \frac{1}{W} \sum_{i=1}^W \hat{\boldsymbol{\lambda}}^{(i)}$, as the estimates of $\boldsymbol{\lambda}$ and calculate the corresponding mean square error.

For purpose of illustration, we consider a 3×3 contingency table. This size of table is also consistent with the data provided in [71]. In our experiment, the matrix input $\boldsymbol{\lambda}$ is defined by assigning a time window $T = 120$ and the average number of CCF events in each cell of contingency table ($\lambda_{ij}T$) as

		n_{11}	n_{12}	n_{13}	n_{21}	n_{22}	n_{23}	n_{31}	n_{32}	n_{33}	Total
	Input	6	5	1	4	3	1	4	1	2	27
Guess	Situation 1	7	4	1	3	4	2	4	3	1	29
	Situation 2	4	3	1	3	2	1	2	1	1	17
	Situation 3	9	7	3	6	4	2	5	3	1	40

Table 4.8: The average number of CCF events in each cell of contingency table

in the first row in Table 4.8. Bayesian inference needs to specify values for prior hyperparameters. We are dealing here with subjective probability and it is well-known that their elicitation leads to difficulties. Using the strategy suggested in [44], we firstly suppose three situations of guess for the average numbers of CCF events in the time window T as in Table 4.8. The first situation corresponds to a guess at the total number of failures close to the input number. The second and the third one correspond to situations where the guess at the total number underestimate (resp. overestimate) the input number. Then, the prior hyperparameters for each λ_{ij} is defined by solving a simple system of two equations matching guesses and uncertainty with mean and variance of the Gamma distribution:

$$\begin{cases} \alpha_{ij}/\beta_{ij} = \eta_{ij} \\ \alpha_{ij}/\beta_{ij}^2 = \rho_{ij}^2 \eta_{ij}^2 \end{cases}, \quad (4.4.1)$$

where η_{ij} represents the guess at λ_{ij} and ρ_{ij} is the corresponding uncertainty of the guess. The hyperparameters for each situation of the guess and the corresponding uncertainty are given in Table 4.9. We present two level of uncertainty: low and high. We choose ρ_{ij} from an uniform distribution in [0.05,0.3] for the low level and [0.7,0.95] for the high level of uncertainty. Type A and B correspond to the guess in the first situation with low and high level of uncertainty; type C and D correspond to the guess in the second situation with low and high level of uncertainty; type E and F correspond to the guess in the third situation with low and high level uncertainty, respectively. The corresponding estimates are presented in Table 4.10. In general, when the guess is close to the input, the small value of ρ leads to better estimates. Otherwise, when the guess is far from input, the large value of ρ give better results. Moreover, the noninformative prior give acceptable results with small RMSE.

4.5 Application

In this section, we apply the proposed algorithm to the data in Table 4.2. This database contains the occurrence of causes and CCF events from 16 systems of the same size. Because of the additive property of the Poisson distribution (sums of independent Poisson-distributed random variables are also Poisson

ρ	Situation 1				Situation 2				Situation 3			
	Low		High		Low		High		Low		High	
	A		B		C		D		E		F	
$(\alpha_{11}, \beta_{11})$	34.59	593.07	1.30	22.41	23.09	692.70	1.99	59.98	15.18	202.52	1.36	18.16
$(\alpha_{12}, \beta_{12})$	51.12	2045.17	1.46	58.47	33.27	1331.10	1.43	57.50	32.18	643.70	1.83	36.65
$(\alpha_{13}, \beta_{13})$	17.54	526.46	1.66	49.89	24.96	1497.81	1.77	106.33	21.74	521.89	1.58	37.99
$(\alpha_{21}, \beta_{21})$	12.15	364.56	1.18	35.32	33.79	1351.95	1.35	54.33	11.06	189.62	1.17	20.18
$(\alpha_{22}, \beta_{22})$	47.81	1434.53	1.12	33.62	34.70	2082.51	1.24	74.56	12.12	363.80	1.22	36.89
$(\alpha_{23}, \beta_{23})$	15.16	606.50	1.52	60.89	18.18	2182.50	1.14	136.99	20.16	806.68	1.53	61.43
$(\alpha_{31}, \beta_{31})$	89.46	10736.35	1.88	226.11	60.09	7211.36	1.34	160.85	10.64	425.81	1.15	46.23
$(\alpha_{32}, \beta_{32})$	12.45	747.06	1.41	84.72	32.23	3868.41	1.13	136.52	13.71	823.03	1.30	78.10
$(\alpha_{33}, \beta_{33})$	18.97	2276.96	1.75	210.21	21.65	2598.92	1.20	144.76	14.12	1694.50	1.32	158.26

Table 4.9: Hyperparameters of prior based on the guess in Table 4.8.

random variable), we aggregate the failure data from these 16 systems to obtain a contingency table with observed margins displayed in Table 4.2.

It is clear that the size of the space $\mathcal{D}_{\mathbf{n}}$ in this case is so large that it is difficult to explore all the elements of the space. The stochastic version is then suggested. Firstly, we choose an element of the space, namely, we find a contingency table with specific values in each cell so that they satisfy the margin conditions. For example, a possible contingency table could be

$$\mathbf{n}^{(1)} = \begin{bmatrix} 150 & 212 & 225 \\ 18 & 15 & 20 \\ 4 & 23 & 3 \end{bmatrix}.$$

Next, we use the $M-H$ algorithm described in previous section to simulate $M = 100.000$ samples $\mathbf{n}^{(1)}, \dots, \mathbf{n}^{(M)}$. The samples $\lambda^{(i)}$ is then generated from $\pi(\lambda | \mathbf{N} = \mathbf{n}^{(i)}, i = 1, \dots, M)$. The summary of posterior distribution of each λ is displayed in Table 4.12. Considering mean of each parameters, it can be seen that

$$\begin{aligned} \lambda_{11} &< \lambda_{13} < \lambda_{12}, \\ \lambda_{22} &< \lambda_{21} < \lambda_{23}, \\ \lambda_{32} &< \lambda_{33} < \lambda_{31}. \end{aligned} \tag{4.5.1}$$

Thus, the order among λ_{ij} 's is the same the order among $\alpha_i^{C_j}$'s in α -decomposition model. Similar to the discussion in [71], Cause 2 is of least CCF risk, Cause 3 is best at provoking partial CCF and Cause 1 is best at provoking complete CCF. However, the impact of Cause 2 and Cause 3 are similar as the difference between them are not significant. Based on this result, the most hazardous causes could be recognized, which has a practical engineering meanings. Figure 4.2 shows the posterior densites of each λ_{ij} based on $M = 100.000$ *i.i.d.* samples generated by using the sampling IBF method.

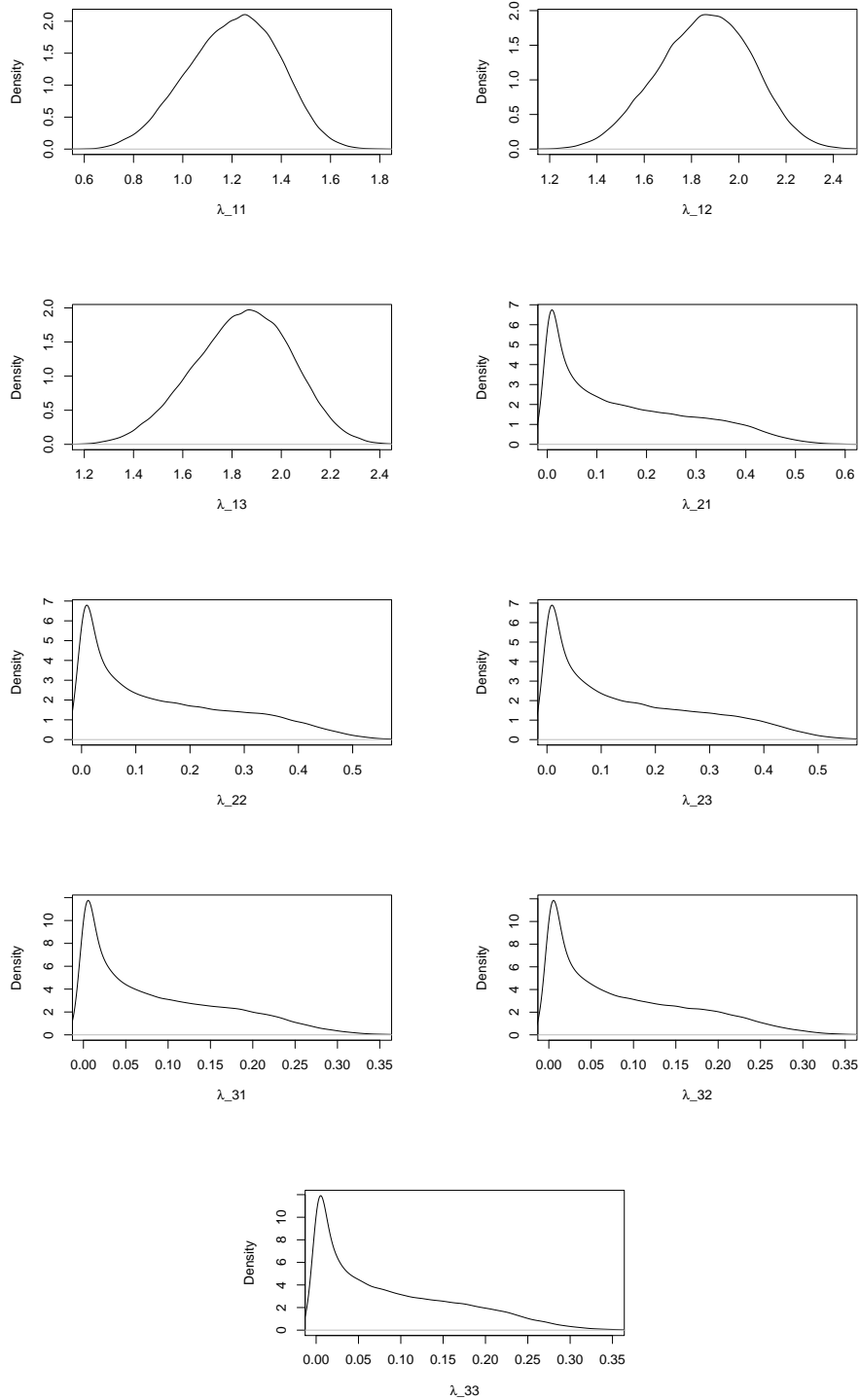


Figure 4.2: The posterior ditribution of each λ_{ij} with noninformative prior distribu-
tion and data in Table 4.2.

	Input	Type A	Type B	Type C	Type D	Type E	Type F	Type G
λ_{11}	0.05	0.05887 (8.93e-5)	0.06104 (4.74e-4)	0.03604 (2.02e-4)	0.04546 (1.80e-4)	0.06528 (2.78e-4)	0.05336 (3.31e-4)	0.0612 (6.91e-4)
λ_{12}	0.0416	0.03236 (9.86e-5)	0.03038 (2.56e-4)	0.02603 (2.45e-4)	0.03424 (1.94e-4)	0.04973 (9.34e-5)	0.03807 (2.07e-4)	0.0353 (2.89e-4)
λ_{13}	0.0083	0.00834 (3.44e-8)	0.00858 (2.45e-6)	0.00836 (1.03e-8)	0.00912 (5.40e-6)	0.02301 (2.18e-4)	0.01802 (1.35e-4)	0.0160 (1.01e-4)
λ_{21}	0.0333	0.02495 (7.06e-5)	0.02525 (1.31e-4)	0.02577 (5.84e-5)	0.03113 (1.08e-4)	0.04761 (20.8e-4)	0.03835 (1.50e-4)	0.03622 (2.62e-4)
λ_{22}	0.025	0.03282 (6.20e-5)	0.02728 (1.01e-4)	0.01699 (6.43e-5)	0.020275 (5.98e-5)	0.03209 (3.43e-5)	0.02394 (5.39e-5)	0.02781 (1.14e-4)
λ_{23}	0.0083	0.01653 (6.87e-5)	0.01607 (8.85e-5)	0.00834 (3.09e-8)	0.00903 (6.45e-6)	0.01592 (5.82e-5)	0.01330 (3.67e-5)	0.01482 (7.31e-5)
λ_{31}	0.0333	0.03291 (6.76e-6)	0.03085 (1.04e-4)	0.01768 (2.45e-4)	0.02418 (1.30e-6)	0.04010 (5.24e-5)	0.03421 (1.21e-4)	0.03155 (2.00e-4)
λ_{32}	0.0083	0.02399 (2.47e-4)	0.02172 (2.08e-4)	0.00866 (2.32e-8)	0.01104 (1.47e-5)	0.02418 (2.52e-4)	0.02134 (1.93e-4)	0.02432 (3.25e-4)
λ_{33}	0.0167	0.00833 (6.94e-5)	0.00832 (7.11e-5)	0.00854 (6.61e-5)	0.01058 (5.01e-5)	0.00822 (7.14e-5)	0.00775 (8.15e-5)	0.01481 (3.52e-5)

Table 4.10: Means of Bayesian estimates for various types of prior, with MSE computed for 100,000 simulations. Type G corresponds to noninformative prior distribution.

	Cause 1	Cause 2	Cause 3	Total
CCF of order 1	x	x	x	587
CCF of order 2	x	x	x	53
CCF of order 3	x	x	x	30
Total	172	250	248	670

Table 4.11: Data extracted from the database in Table 4.2

Comments

The $\alpha_i^{C_j}$ represents the conditional probability that exactly i components of the system fail because of the Cause C_j given that a CCF event occurs because of C_j . Assuming Poisson distributions for the contingency table cells, $\lambda_{ij}T$ represents the mean of the number of CCF events of order i due to cause C_j , and $\lambda_{\cdot j}T$, where $\lambda_{\cdot j} = \sum_{i=1}^3 \lambda_{ij}$, represents the mean of the number of CCF events from Cause C_j over the time window T .

Therefore, the α -decomposition parameters can be approximated by:

$$\alpha_i^{C_j} \approx \frac{\lambda_{ij}T}{\lambda_{\cdot j}T} = \frac{\lambda_{ij}}{\lambda_{\cdot j}}, \quad i = 1, \dots, m, \quad j = 1, \dots, k, \quad (4.5.2)$$

With the estimates of λ_{ij} 's in Table (4.12), the parameters $\alpha_i^{C_j}$ in α -factor model can be approximated by (4.5.2). Table 4.13 compares the estimates

Posterior λ_{ij}	Mean	SD	0.025-quantile	Median	0.975-quantile
CCF order 1					
λ_{11}	1.2028	0.1849	0.8321	1.2108	1.5398
λ_{12}	1.8598	0.1973	1.4639	1.8652	2.2268
λ_{13}	1.8413	0.1977	1.4444	1.8517	2.2076
CCF order 2					
λ_{21}	0.1545	0.1394	0.00029	0.1177	0.4545
λ_{22}	0.1492	0.1373	0.00027	0.1097	0.4498
λ_{23}	0.1594	0.1379	0.00028	0.1093	0.4518
CCF order 2					
λ_{31}	0.0886	0.0815	0.00016	0.0642	0.2705
λ_{32}	0.0866	0.0808	0.00015	0.0645	0.2682
λ_{33}	0.0873	0.0811	0.00016	0.0628	0.2692

Table 4.12: Summary of posterior distributions of each λ_{ij} , ($i, j = 1, 2, 3$).

of $\alpha_i^{C_j}$ obtained from the method suggested by Zheng et al. [71] and the approximation in (4.5.2). In general, the estimates using two methods are close. The order among $\alpha_i^{C_j}$'s when j is fixed (similar to (4.5.1)) is still held. That means, $\alpha_i^{C_j}$ can be well approximated through λ_{ij} .

Method	$\alpha_1^{C_1}$	$\alpha_1^{C_2}$	$\alpha_1^{C_3}$	$\alpha_2^{C_1}$	$\alpha_2^{C_2}$	$\alpha_2^{C_3}$	$\alpha_3^{C_1}$	$\alpha_3^{C_2}$	$\alpha_3^{C_3}$
Zheng et al. [71]	0.813	0.910	0.820	0.0988	0.0729	0.0112	0.0879	0.0174	0.0679
Proposed method	0.8318	0.8874	0.8818	0.1068	0.0711	0.0763	0.0612	0.0413	0.0418

Table 4.13: Compare the estimates of $\alpha_i^{C_j}$ using the α -decomposition method and our proposed method.

4.6 Concluding remarks

We have presented in this chapter a new method for making causality-based inference on incomplete CCF data. When considering the causes of CCF event, the data is recorded in the form of a contingency table. We have considered a situation when only margins of the contingency table are observed. A Poisson distribution based model has been suggested. The parameters of the model are estimated based on Bayesian method using the sampling IBF technique. We have also found that the parameters in α -factor model proposed in [71] can be well approximated through the parameters of our proposed model.

Chapter 5

Prediction methods for catastrophic events

Making predictive inference on the number of CCF events in a given future time is a major concern in PRA. In this chapter, we provide several methods for dealing with this problem. In particular, we focus on predicting the number of catastrophic events, that is to say CCF events where all the components of the system fail at the same time. The failure count is supposed to be a homogeneous Poisson process. We suggest different approaches depending on the nature of the available data.

5.1 Introduction

In the previous chapters, we have presented several models to analyze the CCF events and to estimate the parameters of CCF models. In this chapter, we will investigate another problem in PRA, which is to answer the question of how to predict the number of CCF events in a given future time. In particular, we present a framework to make predictive inference on the upper prediction bound (UPB) of the number of catastrophic events that a system could be suffered in the future. The term "*catastrophic event*" refers to the CCF event where all the components of the system fail simultaneously. By this definition, we do not distinguish the lethal failures from the CCF events of order m for a m -component system (as in Chapter 2): we simply count the number of events where all the components of the system fail simultaneously.

Let $N_m(T)$ denote the number of catastrophic events of the system over the observation time $[0, T]$. We suppose that $N_m(T)$ follows a Poisson distribution with an unknown constant rate Λ . Our goal is to make prediction on the UPB of the future number of catastrophic events over the future time window $[T, T + T_{new}]$, say $N_m(T_{new})$, given the value of $N_m(T)$.

Finding the $100\%(1 - \gamma)$ UPB of $N_m(T_{new})$, $0 < \gamma < 1$, is equivalent to find

the smallest integer b_γ such that

$$Pr(N_m(T_{new}) \geq b_\gamma + 1 | N_m(T)) < \gamma. \quad (5.1.1)$$

As discussed in Chapter 1, there are several types of CCF data. In this Chapter, we consider three types of data: complete data, mapped data, and data with uncertainties.

For complete data, we apply two available methods for making prediction on Poisson count: the pivotal method and the Bayesian method. In the literature, the pivotal method is suggested in [37] and applied to predict the UPB of the number of Poisson counts in ([20],[28], [34]). The Bayesian method for predicting the number of catastrophic events for a nuclear plant in US is presented in [25]. We will see that the Bayesian method is a generalization of the pivotal method.

The mapped data are applied for predicting the number of catastrophic failures in [28]. However, in this study, the authors used only the mapping down technique. We present both mapping up and mapping down rules and then use these rules for making predictive inference.

For data with uncertainties, to our best knowledge, the problem of prediction has not been considered in the literature yet. We provide several strategies to deal with this problem.

The chapter is organized as follows. Section 2 outlines the existing methods of predictive inference for a Poisson failure data and then applying for finding UPB of the number of catastrophic events in the case of complete data. The mapped data are treated in Section 3 using both mapping down and mapping up. In Section 4, we propose the procedure for making inference on UPB of catastrophic events based on data with uncertainties.

5.2 Prediction with complete data: classical methods

Similar to Chapter 3, by complete data, we mean that the order of each CCF event is known exactly. That is to say, a set of complete data for a m -component system is of the form $\mathbf{N} = (N_1, \dots, N_m)$ with the same meaning of N_i as discussed previously. Under the assumption that the catastrophic event is independent from other CCF events, only N_m is used in our predictive inference.

In the following, we present two classical methods to predict the value $N_m(T_{new})$.

5.2.1 Pivotal method

With the assumptions discussed in the introduction, $\{N_m(t), t \geq 0\}$ is a HPP with an unknown constant intensity Λ . A classical method for making prediction on $(N_m(T_{new}))$ is the pivotal method. In this method, a conditional

distribution of $N_m(T_{new})$ which does not depend on the parameter Λ is used.

According to a well-known property of the Poisson distribution, the conditional distribution of $N_m(T)$ given $N_m(T) + N_m(T_{new}) = n_{tot}$, say $N_m(T) + N_m(T_{new}) = n_{tot}$, is a binomial distribution with parameters $(n_{tot}, T/(T + T_{new}))$.

Thus $N_m(T) | N_m(T) + N_m(T_{new}) = n_{tot}$ forms a pivotal quantity; its distribution does not depend on Λ .

This distribution is used to determine the UPB of $N_m(T_{new})$ as defined in (5.1.1). Let n denote the value of N_m over the time window $[0, T]$, i.e., $N_m(T) = n$. The 100%(1 - γ) UPB of $N_m(T_{new})$ is smallest integer b_γ , such that

$$Pr(N_m(T_{new}) \geq b_\gamma + 1 | N_m(T) + N_m(T_{new}) = b_\gamma + n) < \gamma, \quad (5.2.1)$$

where

$$N_m(T_{new}) | N_m(T) + N_m(T_{new}) = b_\gamma + n \sim \mathcal{B}(b_\gamma + n, \frac{T_{new}}{T + T_{new}}). \quad (5.2.2)$$

The relation between this pivotal method and the Bayesian method is presented in the next section.

5.2.2 Bayesian method

In this section, we will make predictive inference based on the posterior distribution of the parameter Λ . Since $N_m(T)$ is the Poisson distribution with parameter ΛT , the distribution of the observation is:

$$P(N_m(T) = n | \Lambda) = \frac{(\Lambda T)^n}{n!} e^{-\Lambda T}. \quad (5.2.3)$$

Assuming a Gamma prior distribution with parameters (α_0, β_0) for Λ , the posterior distribution $\pi(\Lambda | N_m(T) = n)$ is a Gamma distribution with parameters $(\alpha_0 + n, \beta_0 + T)$,

$$\Lambda | N_m(T) = n \sim \mathcal{G}(n + \alpha_0, T + \beta_0).$$

The posterior predictive distribution of $N_m(T_{new})$ is defined as:

$$\begin{aligned} & P(N_m(T_{new}) = z | N_m(T) = n) \\ &= \int P(N_m(T_{new}) = z | N_m(T) = n, \Lambda) \pi(\Lambda | N_m(T) = n) d\Lambda \\ &= \int_0^{+\infty} \frac{(\Lambda T_{new})^z e^{-\Lambda T_{new}}}{z!} \frac{(\beta_0 + T)^{n+\alpha_0}}{\Gamma(n + \alpha_0)} \Lambda^{n+\alpha_0-1} e^{-(\beta_0+T)\Lambda} d\Lambda \\ &= \frac{\Gamma(z + n + \alpha_0)}{\Gamma(n + \alpha_0)z!} \left(\frac{\beta_0 + T}{\beta_0 + T + T_{new}} \right)^{n+\alpha_0} \left(\frac{T_{new}}{\beta_0 + T + T_{new}} \right)^z. \end{aligned} \quad (5.2.4)$$

The last expression in (5.2.4) indicates that the posterior predictive distribution is a negative binomial (NB) distribution with parameters $\left(\alpha_0 + n, \frac{\beta_0 + T}{\beta_0 + T + T_{new}}\right)$. Based on this posterior predictive distribution, the expected-value and the variance of the number of catastrophic events are:

$$E(N_m(T_{new}) | N_m(T) = n) = \frac{(\alpha_0 + n) T_{new}}{\beta_0 + T},$$

$$Var(N_m(T_{new}) | N_m(T) = n) = \frac{(\alpha_0 + n)(\beta_0 + T + T_{new}) T_{new}}{(\beta_0 + T)^2}.$$

Therefore, the $100\%(1 - \gamma)$ UPB of $N_m(T_{new})$ is the smallest integer b_γ satisfying (5.1.1) in which

$$N_m(T_{new}) | N_m(T) = n \sim NB\left(n + \alpha_0, \frac{\beta_0 + T}{\beta_0 + T + T_{new}}\right). \quad (5.2.5)$$

Remark: We can show that the pivotal method can be obtained from the Bayesian method by selecting the hyperparameters appropriately. Indeed, the following propositions show the relation between the binomial distribution, the negative binomial distribution and the beta distribution [17]:

Proposition 7. Let $X \sim \mathcal{B}(n, p)$, then $P(X \geq x) = P(W \leq p)$ where $W \sim Beta(x, n - x + 1)$.

Proposition 8. Let $X \sim NB(n, p)$, then $P(X \geq x) = P(Y \leq 1 - p)$ where $W \sim Beta(x, n)$.

According to proposition 7, the smallest integer b_γ satisfying (5.1.1) with respect to the distribution in (5.2.2) is equivalent to the smallest integer b_γ satisfying

$$Pr\left(W_1 \leq \frac{T_{new}}{T + T_{new}}\right) < \gamma, \quad (5.2.6)$$

where $W_1 \sim Beta(b_\gamma + 1, n)$.

According to proposition 8, the smallest integer b_γ satisfying (5.1.1) with respect to the posterior predictive distribution in (5.2.5) is equivalent to the smallest integer b_γ satisfying

$$Pr\left(W_2 \leq \frac{T_{new}}{\beta_0 + T + T_{new}}\right) < \gamma, \quad (5.2.7)$$

where $W_2 \sim Beta(b_\gamma + 1, n + \alpha_0)$.

By comparing the equations (5.2.6) and (5.2.7), and the corresponding distributions W_1 and W_2 , we deduce that the Bayesian method is a generalization of the pivotal method. The pivotal method can be obtained from the Bayesian method by choosing $\alpha_0 = 0, \beta_0 = 0$.

In the next section, we provide an application of these two methods.

5.2.3 Illustrative example

In this Section, we apply the methods presented in previous sections to draw an UPB for the number of catastrophic events based on a set of CCF data involving EDGs at US nuclear power-plants [22]. The data were observed from a system of five components. They consist of 2 failure events involving one component, 2 failure events involving two components, 1 event involving three components, 2 failures involving four out of five components, and 1 catastrophic event. That is, these data can be summarized by a vector $\mathbf{N} = (2, 2, 1, 2, 1)$. There was no information about the observation time, we suppose a window $T = 1$ (year).

The 95% UPB of the number of catastrophic event over the future time $T_{new} \in \{0.5, 1, 2, 3, 4, 5\}$ is found by finding the smallest integer $b_{0.05}$ satisfying (5.1.1) with respect to the distribution defined in (5.1.1), where $N_m(T) = n = 1$ and $T = 1$. Table 5.9 presents these values of UPB considering several choices of prior: a $\mathcal{G}(0, 0)$ prior (corresponding to pivotal method), a $\mathcal{G}(0.5, 0)$ prior (noninformative) and a $\mathcal{G}(4.67, 4.48)$ prior (which is suggested by Engelhardt in [10]). We also calculate the average number of catastrophic events based on its predictive distribution. The obtained results show that in a short time period of prediction, the UPB makes no numerical difference. However, these value make a larger difference as the future observation time increases. The pivotal method tends to highest bound.

$T_{new}(\text{years})$	Pivotal method $\mathcal{G}(0, 0)$ prior		$\mathcal{G}(0.5, 0)$ prior		$\mathcal{G}(4.67, 4.78)$ prior	
	UPB	Mean	UPB	Mean	UPB	Mean
0.5	2	0.5	3	0.75	2	0.49
1	4	1.0	5	1.5	3	0.98
2	7	2.0	9	3.0	5	1.96
3	10	3.0	13	4.5	7	2.94
4	13	4.0	17	6.0	9	3.92
5	16	5.0	21	7.5	11	4.90

Table 5.1: The 95% UPB and the posterior meand of the number of catastrophic events involving EDG at US NPP.

5.3 Prediction with mapped data

The CCFs are rare events. In practice, for a given size of a system, few number of CCFs are observed. Therefore the accuracy of the inference will be poor. In order to overcome this problem, a strategy is to use failure information coming from systems of different size. This strategy has been introduced by Mosleh et al. [36], it is called **mapping**. The principle is to project the available data from different size systems to a given target size system. If

the projection is from a bigger size system compared to the target size system, the procedure is called mapping up. Conversely, if the projection is from a smaller size system compared to the target size system, the procedure is called mapping down. Then the inference can be driven on systems considered as being of the same size. Many rules has been proposed for mapping CCF data. They are expressed through different approaches: probability, rate or count. Some of them are presented in the sequel considering the count approach.

5.3.1 Mapping down rules

Suppose that we would like to make inference on a system of size k and that the available data are from a system of a bigger size m , $m > k$. The rules for mapping down are firstly introduced in [36] for systems with specific sizes. In particular, the authors gave the rules for mapping down data from the systems of the size $m = 2, 3, 4$ to the system of the size $k = 1, 2, 3$. These rules are reproduced in Table 5.2. The idea is to distribute $N_{i|m}$, the number of CCF events of order i from the system of size m , to $N_{j|k}$, the number of CCF events of order j from the system of target size k , where $j \leq i$. The portion of $N_{i|m}$ contributing to $N_{j|k}$ is defined based on the possibility that a CCF event of order i from the system of size m could bring to the CCF event of order j from the system of size k .

m/k	3	2	1
4	$N_{1 3} = \frac{3}{4}N_{1 4} + \frac{1}{2}N_{2 4}$ $N_{2 3} = \frac{1}{2}N_{2 4} + \frac{3}{4}N_{3 4}$ $N_{3 3} = \frac{1}{4}N_{3 4} + N_{4 4}$	$N_{1 2} = \frac{1}{2}N_{1 4} + \frac{2}{3}N_{2 4} + \frac{1}{2}N_{3 4}$ $N_{2 2} = \frac{1}{6}N_{1 4} + \frac{1}{2}N_{2 4} + N_{3 4}$	$N_{1 1} = \frac{1}{4}N_{1 4} + \frac{1}{2}N_{2 4}$ $+ \frac{3}{4}N_{3 4} + N_{4 4}$
3		$N_{1 2} = \frac{2}{3}N_{1 3} + \frac{2}{3}N_{2 3}$ $N_{2 2} = \frac{1}{3}N_{2 3} + N_{3 3}$	$N_{1 1} = \frac{1}{3}N_{1 3} + \frac{2}{3}N_{2 3}$ $+ N_{3 3}$
2			$N_{1 1} = \frac{1}{2}N_{1 2} + N_{2 2}$

Table 5.2: Mapping down rules for systems of the sizes up to 4

In [67], Vaurio propose to map down the data from a system of size $m + 1$ to the system of size m by a general formula resting on rate, which is equivalent to

$$N_{i|m} = \frac{m + 1 - i}{m + 1}N_{i|m+1} + \frac{i + 1}{m + 1}N_{i+1|m+1}, \quad i = 1, \dots, m. \quad (5.3.1)$$

By repeating (5.3.1), we can obtain the mapping down rules for the data from any system to the systems with smaller size. Applying this rule to $m = 2, 3, 4$ and $k = 1, 2, 3$ we get the same formulas as presented in Table 5.2.

Based on hypergeometric sampling, Kvam [21] provide a general expression for mapping down rules as

$$N_{j|k} = \sum_{i=1}^m C_{j,i} N_{i|m}, \quad j = 1, \dots, k, \quad (5.3.2)$$

where

$$C_{j,i} = \frac{C_i^j C_m^{k-j}}{C_m^k} I_{(j \leq i \leq m-k+j)}. \quad (5.3.3)$$

Assigning $m = 2, 3, 4$ and $k = 1, 2, 3$ again, we achieve the same formulas as in Table 5.2. That means these three approaches are consistent.

The mapping down transformation can be expressed on the form of matrix notations. For example, the mapping down rule from the system of 4 components to the system of 3 components can be rewritten as

$$\begin{bmatrix} N_{1|3} \\ N_{2|3} \\ N_{3|3} \end{bmatrix} = \begin{bmatrix} \frac{3}{4} & \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & \frac{3}{4} & 0 \\ 0 & 0 & \frac{1}{4} & 1 \end{bmatrix} \begin{bmatrix} N_{1|4} \\ N_{2|4} \\ N_{3|4} \\ N_{4|4} \end{bmatrix}.$$

The general formula will be $\mathbf{N}_{\bullet|k} = \mathbf{Q}_{k \times m} \mathbf{N}_{\bullet|m}$, where $\mathbf{N}_{\bullet|k} = (N_{1|k}, \dots, N_{k|k})'$, $\mathbf{N}_{\bullet|m} = (N_{1|m}, \dots, N_{m|m})'$, and $\mathbf{Q}_{k \times m}$ is a transformation matrix. Each element q_{ij} in $\mathbf{Q}_{k \times m}$ is calculated by $q_{ij} = C_{j,i}$ where $C_{j,i}$ is defined in (5.3.3). That is to say, the existing mapping down rules are based on hypergeometric sampling with the transformation matrix $\mathbf{Q}_{k \times m}$ defined based on (5.3.3).

Going from one size to another means losing information. One can imagine finding the best matrix in order to maintain the level of information in the mapping of operation. In that case, the transformation matrix $\mathbf{Q}_{k \times m}$ could be defined in a different way. This problem has not been studied yet in the literature.

5.3.2 Mapping up rules

Unlike the mapping down rule, mapping up is much more challenging because it is obviously an extrapolation; it requires additional assumptions or even parameters to transfer the data from systems of smaller size to the systems of larger size. In this context, we suppose that the existing data from a system of size k are mapped up to a system of size m which is the target system size where $m > k$. There are a number of applicable methods to make upward data transformation presented in the literature such as the BFR-indirect (BFR-INDIR) method (Mosleh et al. [36]), the BFR-mapped (BFR-MAP) method (Kvam [21]) and the method of using mapping up ratio (Vaurio [67]). These methods are described in the following.

BFR-INDIR

In the BFR-INDIR method (Mosleh et al. [36]), no specific mapping rule is given for an arbitrary choice of k and m , but plausible transformation are identified only for cases which $k = 1, 2, 3$ and $m = 2, 3, 4$. Table 5.3 reproduces these rules for systems of the size from 2 to 4.

Ideally, the BFR-INDIR mapping up rules are based on the BFR equations in each system. The parameters of the model are supposed to be independent of the system size. By expressing the relation between the frequency of CCF events that occur within each system and the parameters of the BFR model, we can derive the mapping rules from system to system. However, in many situations, it is difficult to write down the $N_{i|m}$ as a function of $N_{j|k}$ based on the intermediate relationship with BFR parameters. As a consequence, the BFR-INDIR method is only applied for a few systems with small size. Other limitations of the method has been also discussed by Kvam [21].

BFR-MAP

To avoid the excessive dependence on BFR model, Kvam [21] introduces the new BFR-MAP method. The idea is to calculate the conditional probability of the event that j components fail in the system of size m , given that i components in the system of size k failed, $j \geq i$. In particular, this probability is equal to $C_{m-k}^{j-i} \hat{p}^{j-i} (1 - \hat{p})^{(m-k)-(j-i)}$, $i \leq j \leq m - k - i$, where \hat{p} is an estimate of the probability that each component fails due to a shock (it is independent from the system size). In this method, the author also considers a practical situation that the shock that causes no failure in the system of size k could potentially cause from zero to $m - k$ failed components in the system of size m . Then, the following steps of the BFR-MAP method are suggested.

- Estimate p , a parameter of BFR model.
- Estimate $N_0^{(k)}$, the number of events that a non-lethal shock causes no failure in the original system, by the formula

$$\hat{N}_{0|k} = \sum_{i=1}^k \frac{i (1 - \hat{p})^k}{\hat{p}} N_{i|k}. \quad (5.3.4)$$

The reason for the formula of $N_0^{(k)}$ in (5.3.4) is to make sure that $E(\hat{N}_{0|k} | \hat{p}) = \mu T (1 - \hat{p})^k$, which is the mean of Poisson distribution with parameter $\mu T (1 - \hat{p})^k$, where T is the observation time.

- Map data from the system of size k up to the system of size m by the formula

$$\begin{aligned} N_{j|m} &= \hat{N}_{0|k} C_{m-k}^j \hat{p}^j (1 - \hat{p})^{(m-k)-j} \mathbb{1}_{(j \leq m-k)} \\ &+ \sum_{i=\max\{1, k+j-m\}}^{\min\{j, k\}} N_{i|k} C_{m-k}^{j-i} \hat{p}^{j-i} (1 - \hat{p})^{(m-k)-(j-i)}. \end{aligned} \quad (5.3.5)$$

The first term in (5.3.5) is the contribution of $N_{0|k}$ while the others is the contribution of $N_{i|k}$ to $N_{j|m}$ based on the conditional probability mentioned above.

Some representations of 5.3.5 for the system of the size from 2 to 4 are presented in Table 5.3.

Modification of BFR-INDIR

In the BFR-INDIR method, the following relation holds: $N_{m+1|m+1} = pn_{m|m}$ for all the system size $m \geq 1$. That means the parameter p is independent from the system size. This could be not always true in practice. Vaurio [67] makes this assumption more realistic by introducing a new parameter as

$$N_{m+1|m+1} = \eta_{m/m+1} N_{m|m} \quad \forall m \geq 1. \quad (5.3.6)$$

where the mapping up ratio $\eta_{m/m+1}$ represents the relation between $N_{m+1|m+1}$ and $N_{m|m}$. This new parameter is allowed to vary from system to system. By substituting successively the relation in (5.3.6) into (5.3.1) for $i = m, m-1, m-2, \dots$, we obtain all data for the next higher system. Table 5.3 shows the mapping up formulas based on the mapping up ratio for systems of the size form 2 to 4.

Method	3 → 4	2 → 4
BFR-INDIR	$N_{1 4} = \frac{4}{3}(1-p)N_{1 3}$ $N_{2 4} = pN_{1 3} + (1-p)N_{2 3}$ $N_{3 4} = pN_{2 3} + (1-p)N_{3 3}$ $N_{4 4} = pN_{3 3}$	$N_{1 4} = 2(1-p)^2 N_{1 2}$ $N_{2 4} = \frac{5}{2}p(1-p)N_{1 2} + (1-p)^2 N_{2 2}$ $N_{3 4} = p^2 N_{1 2} + 2p(1-p)N_{2 2}$ $N_{4 4} = p^2 N_{2 2}$
Modified BFR-INDIR	$N_{1 4} = \frac{4}{3}N_{1 3} - \frac{4}{3}N_{2 3} + 4(1-\eta_{3/4})N_{3 3}$ $N_{2 4} = 2N_{2 3} - 6(1-\eta_{3/4})N_{3 3}$ $N_{3 4} = 4(1-\eta_{3/4})N_{3 3}$ $N_{4 4} = \eta_{3/4}N_{3 3}$	$N_{1 4} = 2n_1^{(2)} - 4(2-3\eta_{2/3} + \eta_{2/3}\eta_{3/4})N_{2 2}$ $N_{2 4} = 6(1-2\eta_{2/3} + \eta_{2/3}\eta_{3/4})N_{2 2}$ $N_{3 4} = 4(1-\eta_{3/4})\eta_{2/3}N_{2 2}$ $N_{4 4} = \eta_{2/3}\eta_{3/4}N_{2 2}$
BFR-MAP	$N_{1 4} = (1-p)^3(\frac{1}{3}N_{1 3} + \frac{2}{3}N_{2 3} + N_{3 3})$ $\quad + (1-p)N_{2 3}$ $N_{2 4} = pN_{1 3} + (1-p)N_{2 3}$ $N_{3 4} = pN_{2 3} + (1-p)N_{3 3}$ $N_{4 4} = pN_{3 3}$	$N_{1 4} = 2(1-p)^3(\frac{1}{2}N_{1 2} + N_{2 2}) + (1-p)^2 N_{1 2}$ $N_{2 4} = p(1-p)^2(\frac{1}{2}N_{1 2} + N_{2 2}) + 2p(1-p)N_{1 2}$ $\quad + (1-p)^2 N_{2 2}$ $N_{3 4} = p^2 N_{1 2} + 2p(1-p)N_{2 2}$ $N_{4 4} = p^2 N_{2 2}$

Table 5.3: Mapping up rules for systems of the sizes up to 4

The modified BFR-INDIR method has a disadvantage that we have to evaluate subjectively the mapping ratio from system to system, which does not depend on the observed data. The implementation of BFR-INDIR is complicated when the size of systems involved is large. Moreover, in this method, the effect of the non-lethal shocks that cause no failure in the original system is not considered. Therefore, the BFR-MAP method is applied in our study.

Similar to the previous section,, we can express the formulas presented in Table 5.3 in the matrix form. For example, the mapping up rule for the systems of the size from 3 to 4 in the BFR-MAP method can be rewritten by

$$\begin{bmatrix} N_{1|4} \\ N_{2|4} \\ N_{3|4} \\ N_{4|4} \end{bmatrix} = \begin{bmatrix} \frac{1}{3}(1-p)^3 & \frac{2}{3}(1-p)^3 + 1-p & (1-p)^3 \\ p & 1-p & 0 \\ 0 & p & 1-p \\ 0 & 0 & p \end{bmatrix} \begin{bmatrix} N_{1|3} \\ N_{2|3} \\ N_{3|3} \end{bmatrix}.$$

Or, in a general case, $\mathbf{N}_{\cdot|m} = \mathbf{Q}_{m \times k} \mathbf{N}_{\cdot|k}$, where $\mathbf{N}_{\cdot|k} = (N_{1|k}, \dots, N_{k|k})'$, $\mathbf{N}_{\cdot|m} = (N_{1|m}, \dots, N_{m|m})'$, and $\mathbf{Q}_{m \times k}$ is a transformation matrix. Finding a matrix that preserves the information when projecting data from a system to others in this case means either using another way to find an optimal value of p or using a new form for $\mathbf{Q}_{m \times k}$. This could be an interesting problem for the further research.

5.3.3 Illustrative example

In this Section, we illustrate the use of mapped data to make predictive inference on the number of catastrophic events based on a data set involving simultaneous failures of EDGs introduced in [25]. The second column in Table 5.4 shows this data set, in which the system size ranges from 2 to 5. In [25], the authors use these data to predict the number of catastrophic events for a 2-component system. As a result, only mapping down rules are applied. We now suppose that the data are used for predicting the number of catastrophic events for a 4-component system. By this consideration, the data from the systems of size 2 and 3 need to be mapped up while the data from the system of size 5 need to be mapped down. Moreover, in [25], instead of treating directly the data obtained after mapping as the complete data, the authors suggested to manipulate the time on test and fix the number of failures to what was observed in the data.

In the following, we will treat the data obtained after mapping directly as complete data. The method suggested in [25] is also described and applied to compare the obtained results from two methods. For the data set in Table

System size	Data	Data mapped to a 4-component system
2	(17, 4)	(1.94, 9.18, 14.7, 7.66)
3	(9, 5, 6)	(3.58, 7.08, 5.48, 4.68)
4	(11, 10, 7, 6)	(11, 10, 7, 6)
5	(2, 2, 1, 2, 1)	(1.4, 2, 1.8, 2.4)

Table 5.4: Common cause failure data for EDGs at US NPP

5.4, the data from the systems of the size 2 and 3 are mapped up using the formulas in Table 5.3, where the probability p_i from i -component system is estimated using an EM algorithm as discussed in Chapter 2: $\hat{p}_2 = 0.74$ and $\hat{p}_3 = 0.52$.

The data from the system of size 4 do not need to be translated. The data from the system of size 5 are mapped down using the following formulas, which come from (5.3.2) with $k = 4$ and $m = 5$:

$$\begin{aligned}
 N_{1|4} &= \frac{4}{5}N_{1|5} + \frac{2}{5}N_{2|5}, & N_{3|4} &= \frac{2}{5}N_{3|5} + \frac{4}{5}N_{3|5}, \\
 N_{2|4} &= \frac{3}{5}N_{2|5} + \frac{3}{5}N_{3|5}, & N_{4|4} &= \frac{1}{5}N_{4|5} + N_{5|5}.
 \end{aligned}$$

Type of observed failure	Frequency	Time on test	Contribution	Transformed time on test
2 out of 2	14	T_2	0.74^2	$1.82T_2$
3 out of 3	6	T_3	0.52	$1.92T_3$
4 out of 4	6	T_4	1	T_4
4 out of 5	2	T_5	1/5	$5T_5$
5 out of 5	1	T_5	1	T_5

Table 5.5: *Transformed time on test for EDG failure data into the 4-component system*

The results of mapped data are presented in the third column in Table 5.4. There is no information about the observation time of each system in the data set in Table 5.4: but we know that these failure data concern CCF events from 1980 to 1995 [25]. Therefore, we suppose that the time window for each system is $T_j = 15$ (years), $j = 2, \dots, 5$, where T_j represents the observation time for the system of size j .

There are two ways to deal with the mapped data in Table 5.4. The first way is to use directly the mapped data as complete data. By this method, in a period of 60 years we count the CCF data of $\mathbf{N} = (17.92, 28.26, 28.98, 20.74)$, which is the sum of mapped data in the third column.

The second way is to use the transformed time method suggested by Kvam and Miller [25]. Detail of this method is as follows.

The time on test has been transformed so that the number of failures is fixed as observed from the original data. This is done based on an intuitive appealing property of Poisson distribution. That is, the data generated by a Poisson process with constant rate θ over the time observation T_1 can be regarded as the data counted by a process with rate $T_1\theta/T_2$ over the time T_2 . As a result, mapping $pN_{j|m}$ CCFs to the period T is equivalent to mapping $N_{j|m}$ failures to the period T/p , where p is a mapped weight, $N_{j|m}$ is the number of CCFs need to be mapped.

In our case, for 2-component system, 14 failures of CCF of order 2 is mapped to $N_{4|4} = \hat{p}_2^2 N_{2|2} = 7.66$ CCFs of order 4 in the 4-component system over the time T_2 . This is equivalent to transform this 14 failures to 14 failures in the 4-component system over the time $T_2/\hat{p}_2^2 = 1.82T_2$. It is similar for other systems.

Table 5.5 shows the transformed time on test for EDD failure data into the 4-component system. In Table 5.5, we only show the failures that contribute to the CCFs of order 4 in 4-component system, i.e, catastrophic events, because for the events that do not contribution to CCFs of order 4 in 4-component system, the time on test will be zero.

From Table 5.5, we count 29 failure events in the total time

$$T = 1.82T_2 + 1.92T_3 + T_4 + 6T_5 = 161 \text{ (years)}.$$

□

The UPBs of the number of catastrophic events based on the pivotal method and the Bayesian method using both the direct mapped data and the transformed time data are displayed in Table 5.6. It can be seen that when the future time short compared the total time considered, the obtained UPBs are close for different choices of hyperparameters. Moreover, the UPBs using the transformed time data are slightly smaller compared to those using directly data from mapping procedure.

T_{new} (years)	$\mathcal{G}(1, 0)$ prior				$\mathcal{G}(0.5, 0)$ prior				$\mathcal{G}(4.67, 4.78)$ prior			
	Direct		Transformed time		Direct		Transformed time		Direct		Transformed time	
	UPB	Mean	UPB	Mean	UPB	Mean	UPB	Mean	UPB	Mean	UPB	Mean
1	2	0.36	1	0.19	2	0.35	1	0.18	2	0.39	1	0.20
2	2	0.72	2	0.37	2	0.70	2	0.36	2	0.78	2	0.40
5	4	1.81	3	0.93	4	1.77	3	0.91	5	1.96	3	1.01
10	7	3.62	4	1.86	7	3.54	4	1.83	8	3.92	5	2.03
15	10	5.4	6	2.79	10	5.31	6	2.74	11	5.88	6	3.04

Table 5.6: 95% The UPB and posterior mean of the number of catastrophic events for a 4-component system based on data set in Table 5.4.

5.4 Prediction with uncertain data

In this Section, the data with uncertainties are treated to predict the number of catastrophic events using the Bayesian method. We consider three scenarios of data with uncertainties

1. uncertainty on observed count,
2. uncertainty on exposure time, and
3. uncertainty on both observed data and exposure time.

The data with uncertainties will be denoted by \mathbf{N}_u .

5.4.1 Uncertainty on observed data

As discussed in Chapter 1 and Chapter 3, the uncertainty on observed count means that we do not know exactly the number of failed components in each CCF event. As a consequence, the exact number of catastrophic events over the observation time T is also unknown.

We will apply the impact vector method to predict the UPB of $N_m(T_{new})$. Suppose that the available information is a set of impact vectors. We assume k impact vectors $\bar{I}_j = (p_{0,j}, \dots, p_{m,j}), j = 1, \dots, k$, where $p_{m,j}$ is the probability that the CCF event in the j^{th} -observation is a catastrophic events.

In fact, the exact number of catastrophic events is not observed. We only know that it is a number between 0 and k . The number of catastrophic events, say $N_{m,u}(T)$, is the sum of k independent Bernoulli random variables $N_{m,j}$ with parameter $p_{m,j}$. Therefore, the distribution of this numbers has support $\{0, 1, \dots, k\}$. Let us denote $\zeta_v = P(\sum_{j=1}^k N_{m,j} = v)$.

The event $\{N_{m,j} = 0\}$ means that no catastrophic event occurs. The probability of this event is $P(N_{m,j} = 0) = 1 - p_{m,j}$. Thus,

$$\zeta_0 = P\left(\sum_{j=1}^k N_{m,j} = 0\right) = \prod_{j=1}^k (1 - p_{m,j})$$

In the same way, $\{\sum_{j=1}^k N_{m,j} = 1\}$ means that amongst the k CCF events, one event i is a catastrophic event, probability $p_{m,i}$ and the other CCF events are not catastrophic, probability $\prod_{j \neq i} p_{m,j}$. Then $\zeta_1 = \sum_{i=1}^k p_{m,i} \prod_{j \neq i} (1 - p_{m,j})$.

Repeating the same reasoning leads to the complete expression of the distribution of $N_{m,u}(T) = \sum_{j=1}^k N_{m,j}$ as:

$$\begin{aligned} \zeta_0 &= \prod_{i=1}^k (1 - p_{m,i}), \\ \zeta_1 &= \sum_{i=1}^k p_{m,i} \prod_{j=1, j \neq i}^k (1 - p_{m,j}), \\ \zeta_2 &= \sum_{i=1}^k \sum_{j=i+1}^k p_{m,i} p_{m,j} \prod_{s=1, s \neq i, j}^k (1 - p_{m,s}), \\ &\vdots \\ \zeta_k &= \prod_{i=1}^k p_{m,i}. \end{aligned} \tag{5.4.1}$$

If a specific number v of catastrophic events is recorded, the posterior distribution of Λ is a Gamma distribution with parameters $(v + \alpha_0, T + \beta_0)$, denoted by $\pi_v(\Lambda)$, where (α_0, β_0) are the hyperparameters of the prior for Λ .

Then, the posterior distribution of Λ in a general case can be expressed as:

$$\begin{aligned} \pi(\Lambda | N_{m,u}(T)) &= \sum_{v=0}^k (\Lambda | N_{m,u}(T) = v) P(N_{m,u}(T) = v) \\ &= \sum_{v=0}^k \zeta_v \pi_v(\Lambda). \end{aligned} \tag{5.4.2}$$

This is a mixture of gamma distribution with weights ζ_v , $v = 0, \dots, k$.

The posterior predictive distribution of $N_m(T_{new})$ given $N_{m,u}(T)$ is now

defined as:

$$\begin{aligned}
 P(N_m(T_{new}) = z | N_{m,u}(T)) &= \int P(N_m(T_{new}) = z | N_{m,u}(T), \Lambda) \pi(\Lambda | N_{m,u}(T)) d\Lambda \\
 &= \sum_{v=0}^k \zeta_v \int P(N_m(T_{new}) = z | N_{m,u}(T), \Lambda) \pi_v(\Lambda) d\Lambda \\
 &= \sum_{v=0}^k \zeta_v P(NB_v(T_{new}) = z), \tag{5.4.3}
 \end{aligned}$$

where

$$NB_v(T_{new}) \sim \text{NB} \left(\alpha_0 + v, \frac{\beta_0 + T}{\beta_0 + T + T_{new}} \right).$$

That means the posterior predictive distribution of $N_m(T_{new})$ is a mixture of negative binomial distributions. A 100%(1 - γ) UPB of $N_m(T_{new})$ is the smallest integer b_γ satisfying (5.1.1) with respect to the posterior predictive distribution of $N_m(T_{new})$ defined in (5.4.3).

Example

To illustrate the method, we consider a set of data that includes $k = 5$ multiple-failure events from a system of 4 EDGs provided in [66]. The observation time is assumed to be one unit of time, say $T = 1$ (year). The data set is displayed in Table 5.7. From the impact vectors in Table 5.7 and according to

Events (i)	$p_{1,i}$	$p_{2,i}$	$p_{3,i}$	$p_{4,i}$
1	0.0	0.9	0.1	0.0
2	0.0	0.45	0.5	0.05
3	0.0	0.25	0.5	0.25
4	0.0	0.5	0.5	0.0
5	0.0	0.25	0.5	0.25

Table 5.7: The failure data from a system of 4 diesel generators

(5.4.1) we can calculate the distribution of $N_{m,u}(T)$ as in Table 5.8. Then, the posterior predictive distribution in (5.4.3) is computed. The 95% UPB of the number of catastrophic events for different values of future time is calculated as the smallest value $b_{0.05}$ satisfying (5.1.1) with respect to this distribution. Similar to previous the section, we consider three situations of the prior distribution. The obtained results are shown in Table 5.7. In general, when T_{new} is small compared to T , the three prediction limits are very close. However, when the new time period is large compared to the observation time, the choice of the prior hyperparameters has an important influence.

$N_{m,u}(T)$	0	1	2	3
ζ	0.534375	0.384375	0.078125	0.003125

Table 5.8: The distribution of $N_m(T)$ based on the data in Table 5.4.1.

$T_{new}(\text{years})$	$\mathcal{G}(1, 0)$ prior	$\mathcal{G}(0.5, 0)$ prior	$\mathcal{G}(4.67, 4.78)$ prior
0.5	3	2	2
1	5	4	3
2	10	8	5
3	14	11	7
4	18	15	9
5	23	18	10

Table 5.9: 95% UPB of the number of catastrophic events based on data in Table 5.7.

5.4.2 Uncertainty on the exposure time

The uncertainties on data are sometimes related to the observation time. In these situations, one can be certain about the observed count but not on the collecting time.

Let T_u denote the unknown time window and suppose that the uncertainty on T_u is characterized by a subjective probability distribution $h(t)$.

Denote $N_m(T_u)$ the number of catastrophic events in this time window. The distribution of this observation is

$$P(N_m(T_u) = n | \Lambda) = \int_0^{+\infty} \frac{(\Lambda t)^n}{n!} e^{-\Lambda t} h(t) dt \quad (5.4.4)$$

With a $\mathcal{G}(\alpha_0, \beta_0)$ prior distribution for Λ , the Bayes' theorem leads to

$$\begin{aligned} \pi(\Lambda | N_m(T_u) = n) &\propto \pi(N_m(T_u) = n | \Lambda) \pi(\Lambda) \\ &\propto \left(\int_0^{+\infty} \frac{(\Lambda t)^n}{n!} e^{-\Lambda t} h(t) dt \right) \frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)} \Lambda^{\alpha_0-1} e^{-\beta_0 \Lambda} \\ &\propto \int_0^{+\infty} \frac{t^n \beta_0^{\alpha_0}}{n! \Gamma(\alpha_0)} \Lambda^{n+\alpha_0-1} e^{-\Lambda(\beta_0+t)} h(t) dt \end{aligned} \quad (5.4.5)$$

Then the posterior distribution of Λ is now an average distribution with respect to $h(t)$:

$$\Lambda | N_m(T_u) = n \sim \int_0^{+\infty} \mathcal{G}(n + \alpha_0, t + \beta_0) h(t) dt. \quad (5.4.6)$$

Using this posterior, the posterior predictive distribution of $N_m(T_{new})$ is

$$\begin{aligned}
 & Pr(N_m(T_{new}) = z | N_m(T_u) = n) \\
 &= \int Pr(N_m(T_{new}) = z | N_m(T_u) = n, \Lambda) \pi(\Lambda | N_m(T_u) = n) d\Lambda \\
 &= \int \frac{(\Lambda T_{new})^z}{z!} e^{-\Lambda T_{new}} \left(\int_0^{+\infty} \frac{(\beta_0 + t)^{n+\alpha_0}}{\Gamma(n + \alpha_0)} \Lambda^{n+\alpha_0-1} e^{-(\beta_0+t)\Lambda} \right) d\Lambda \\
 &= \int_0^{+\infty} \left(\frac{(\beta_0 + t)^{n+\alpha_0} T_{new}^z}{z! \Gamma(n + \alpha_0)} \int \Lambda^{z+n+\alpha_0-1} e^{-\Lambda(T_{new}+t+\beta_0)} d\Lambda \right) h(t) dt
 \end{aligned} \tag{5.4.7}$$

The integration with respect to Λ in the last equation in (5.4.7) is calculated similarly in (5.2.4), relying on the negative binomial distribution with parameters $(n + \alpha_0, \frac{\beta_0 + t}{\beta_0 + t + T_{new}})$. That means

$$N_m(T_{new}) | N_m(T_u) = n \sim \int_0^{+\infty} \text{NB}\left(n + \alpha_0, \frac{t + \beta_0}{T_{new} + t + \beta_0}\right) h(t) dt. \tag{5.4.8}$$

The 100%(1 - γ) UPB of $N_m(T_{new})$ is the smallest integer b_γ satisfying (5.1.1) with respect to the distribution in (5.4.8) found by using numerical intergration.

There is a way to avoid the numerical intergration in (5.4.8) by using an approximation for the posterior distribution of Λ . That is, we replace the distribution in (5.4.6) by another distribution with the same mean and variance. Let $E_{\pi_1}(\Lambda)$ and $V_{\pi_1}(\Lambda)$ denote the posterior mean and the posterior variance of Λ with respect to π_1 , where π_1 stands for the posterior distribution in (5.4.6), respectively. Matz and Picard [32] advocate to approximate the posterior distribution $\Lambda | N_m(T_u) = n$ by a Gamma distribution with parameters (α_1, β_1) where

$$\alpha_1 = \frac{E_{\pi_1}^2(\Lambda)}{V_{\pi_1}(\Lambda)} \quad \text{and} \quad \beta_1 = \frac{E_{\pi_1}(\Lambda)}{V_{\pi_1}(\Lambda)}. \tag{5.4.9}$$

This choice of (α_1, β_1) is to make sure that the approximate distribution $\mathcal{G}(\alpha_1, \beta_1)$ has the same mean E_{π_1} and variance V_{π_1} .

The Appendix E shows that these value can be approximated by:

$$\begin{cases} E_{\pi_1}(\Lambda) \approx \frac{n + \alpha_0}{E_h(T) + \beta_0} \left(1 + \frac{V_h(t)}{(E_h(T) + \beta_0)^2} \right), \\ V_{\pi_1}(\Lambda) \approx \frac{n + \alpha_0 + (n + \alpha_0)^2}{(E_h(t) + \beta_0)^2} \left(1 + \frac{3V_h(t)}{(E_h(t) + \beta_0)^2} \right) - E_{\pi_1}^2(\Lambda), \end{cases} \tag{5.4.10}$$

where $E_h(t)$ and $V_h(t)$ denote the mean and the variance of T_u with respect to the distribution $h(t)$, respectively.

Replacing the posterior distribution in (5.4.6) by its approximate distribution above, the posterior predictive distribution of $N_m(T_{new})$ is defined as in the case of complete data, which leads to

$$N_m(T_{new}) | N_m T(u) = n \sim \text{NB} \left(\alpha_1, \frac{\beta_1}{\beta_1 + T_{new}} \right). \quad (5.4.11)$$

In the below example, we compare the UPB of $N_m(T_{new})$ using both the direct numerical integration method and the approximation method.

Example

We consider again the data set applied in section 5.2.3 with no hypothesis made on the observation time. That is, one catastrophic event has been reported but we do not know the observation time. According to the discussion in [32], we characterize the uncertainty on T_u by a lognormal distribution $h(t)$.

With a little information that the failure data of emergency diesel generators is collected from US nuclear power-plants over the past 20 years, we assume a finite support lognormal distribution for $h(t)$ with median 10 and error factor 1.5. With this distribution of T_u , we can calculate the approximate Gamma distribution for the posterior distribution of Λ with the parameters denfined in (5.4.9) and (5.4.10).

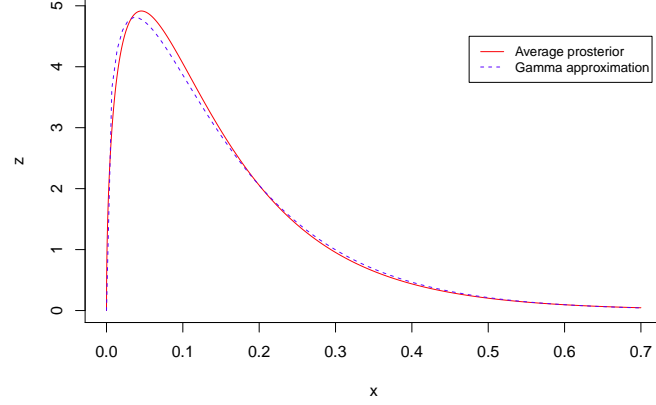
For example, when a Gamma distribution with parameters (0.5, 0) is chosen as a prior distribution, we obtain $\alpha_1 = 1.32$ and $\beta_1 = 8.54$. Figure 5.1 displays the density function of the posterior distribution $\Lambda | N_m(T_u) = n$ and the corresponding approximate Gamma distribution. The 95% UPBs of $N_m(T_{new})$ using both the exact posterior and the approximate posterior are given in Table 5.10. We also consider three situations of the piror distribution. It can be seen that the UPBs using both kinds of posterior distribution are almost the same.

T_{new} (years)	$\mathcal{G}(1, 0)$ prior		$\mathcal{G}(0.5, 0)$ prior		$\mathcal{G}(4.67, 4.78)$ prior	
	Exact prior	Approximation	Exact prior	Approximation	Exact prior	Approximation
1	1	1	1	1	2	2
2	2	2	2	2	3	3
5	4	4	3	3	5	5
10	6	6	5	5	9	9
15	9	9	7	7	14	13

Table 5.10: 95% UPB of the number of catastrophic events for a 4-component system based on data set section 5.2.3 .

5.4.3 Uncertainty on both observed data and exposure time

In this section, we consider the highest level of uncertainties of data: uncertainties on both observed count and exposure time. We suppose that from


 Figure 5.1: The posterior distribution of Λ

k CCF events, the observed count of catastrophic events is represented by the probabilities $p_{m,j}, j = 1, \dots, k$ while the time window T_u is not reported. Similar to the section 5.4, the number of observations of the catastrophic event, $N_{m,u}$, can take the value in $\{0, 1, \dots, k\}$ with corresponding probabilities $\zeta_v = P(N_{m,u} = v)$ defined in (5.4.1). The uncertainties on the time window is also characterized by a distribution $h(t)$.

In this case, the distribution of the observation can be written as:

$$P(N_{m,u}(T_u) | \Lambda) = \sum_{v=0}^k \zeta_v \int_0^{+\infty} \frac{(\Lambda t)^v}{v!} e^{-\Lambda t} h(t) dt \quad (5.4.12)$$

Suppose that $\Lambda \sim \mathcal{G}(\alpha_0, \beta_0)$. Applying the Bayes' theorem, we obtain

$$\begin{aligned} \pi(\Lambda | N_{m,u}(T_u)) &\propto P(N_{m,u}(T_u) | \Lambda) \pi(\Lambda) \\ &\propto \left(\sum_{v=0}^k \zeta_v \int_0^{+\infty} \frac{(\Lambda t)^v}{v!} e^{-\Lambda t} h(t) dt \right) \frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)} \Lambda^{\alpha_0-1} e^{-\beta_0 \Lambda} \\ &\propto \sum_{v=0}^k \zeta_v \int_0^{+\infty} \frac{t^v \beta_0^{\alpha_0}}{v! \Gamma(\alpha_0)} \Lambda^{v+\alpha_0-1} e^{-\Lambda(\beta_0+t)} h(t) dt \quad (5.4.13) \end{aligned}$$

Therefore, the posterior distribution of Λ is

$$\Lambda | N_{m,u}(T_u) \sim \sum_{v=0}^k \zeta_v \int_0^{+\infty} \mathcal{G}(v + \alpha_0, t + \beta_0) h(t) dt. \quad (5.4.14)$$

The posterior predictive distribution of $N_m(T_{new})$ is

$$\begin{aligned}
 & Pr(N_m(T_{new}) = z | N_m(T_u)) \\
 &= \int Pr(N_m(T_{new}) = z | N_{m,u}(T_u), \Lambda) \pi(\Lambda | N_{m,u}(T_u)) d\Lambda \\
 &= \sum_{v=0}^k \zeta_v \int \frac{(\Lambda T_{new})^z}{z!} e^{-\Lambda T_{new}} \left(\int_0^{+\infty} \frac{(\beta_0 + t)^{v+\alpha_0}}{\Gamma(v + \alpha_0)} \Lambda^{v+\alpha_0-1} e^{-(\beta_0+t)\Lambda} \right) d\Lambda \\
 &= \sum_{v=0}^k \zeta_v \int_0^{+\infty} \left(\frac{(\beta_0 + t)^{v+\alpha_0} T_{new}^z}{z! \Gamma(v + \alpha_0)} \int \Lambda^{z+v+\alpha_0-1} e^{-\Lambda(T_{new}+t+\beta_0)} d\Lambda \right) h(t) dt
 \end{aligned} \tag{5.4.15}$$

The last equation in (5.4.15) shows that

$$N_m(T_{new}) | N_{m,u}(T_u) \sim \sum_{v=0}^k \zeta_v \int_0^{+\infty} \text{NB} \left(v + \alpha_0, \frac{t + \beta_0}{T_{new} + t + \beta_0} \right) h(t) dt. \tag{5.4.16}$$

Numerical integration must be used to find the UPB for the distribution in (5.4.16). This may be avoided by approximating the posterior distribution (5.4.14) with a Gamma distribution with the same mean and variance. That is to say, we advocate using the $\mathcal{G}(\alpha_1, \beta_1)$ to approximate the posterior distribution of Λ . The parameters (α_1, β_1) are also calculated applying the formula (5.4.9), but the values E_{π_1} and V_{π_1} should be replaced by the mean E_{π_2} and the variance V_{π_2} of Λ with respect to π_2 , where π_2 stands for the posterior distribution in (5.4.14). The Appendix E shows that the values E_{π_2} and V_{π_2} can be approximated by:

$$\begin{cases} E_{\pi_2}(\Lambda) \approx \frac{E_{\zeta} + \gamma_0}{E_h(t) + \beta_0} \left(1 + \frac{V_h(t)}{(E_h(t) + \beta_0)^2} \right), \\ V_{\pi_2}(\Lambda) \approx \frac{E_{\zeta} + \alpha_0 + \sum_{v=0}^k \zeta_v (v + \alpha_0)^2}{(E_h(t) + \beta_0)^2} \left(1 + \frac{3V_h(t)}{(E_h(t) + \beta_0)^2} \right) - E_{\pi_2}^2(\Lambda), \end{cases} \tag{5.4.17}$$

where $E_{\zeta} = \sum_{v=0}^k v \zeta_v$. The proof for this approximation is given in the appendix E.

The posterior predictive distribution of $N_m(T_{new})$ with respect to the approximated posterior distribution $\mathcal{G}(\alpha_1, \beta_1)$ is also a negative binomial with parameters $(\alpha_1, \frac{\beta_1}{\beta_1 + T_{new}})$. The procedure to find UPB of $N_m(T_{new})$ is performed similarly to the previous sections.

An example to show the performance of using the approximate distribution for the posterior distribution of Λ is given in the next section.

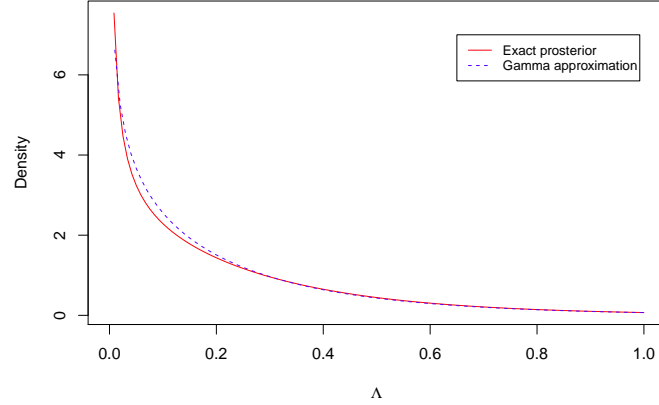


Figure 5.2: The posterior distribution of Λ

Example

Consider again the failure data of diesel generators presented in Table 5.7. In section 5.4.1, it is supposed that the time observation is one year. In fact, the true time window is not reported. We are in the situation that both the observed count of the catastrophic event and the observation time are uncertain. Similar to previous sections, we suppose a lognormal distribution with median 10 and error factor 1.5 for T_u .

Based on these data and using a prior distribution $\mathcal{G}(0.5, 0)$, we can compute the approximate distribution of the posterior distribution of Λ in (5.4.14), which is a Gamma distribution with paramters (0.72, 3.3).

Figure 5.2 displays the exact posterior distribution of Λ in (5.4.14) and its approximate distribution. The 95% UPBs of $N_m(T_{new})$ for three choices of prior distribution are presented in Table 5.11 using both exact posterior of Λ and its approximation. We can see that the approximate distribution provides an efficient alternative for the posterior distribution of Λ in making predictive inference.

T_{new} (years)	$\mathcal{G}(1, 0)$ prior		$\mathcal{G}(0.5, 0)$ prior		$\mathcal{G}(4.67, 4.78)$ prior	
	Exact prior	Approximation	Exact prior	Approximation	Exact posterior	Approximation
1	1	1	1	1	2	2
2	2	2	1	1	3	3
5	3	3	3	2	5	5
10	6	6	4	4	9	9
15	8	8	6	6	13	12

Table 5.11: 95% UPB of the number of catastrophic events for a 4-component system based on data set in Table 5.4.

5.5 Concluding remarks

We have considered in this chapter the problem of making predictive inference on the number of catastrophic failures based on several situation of observed data. We have shown that the pivotal method appears as a particular case of the Bayesian method. The flexibility of the Bayesian scheme allows us to handle data with uncertainties. The approximate distribution with the same mean and variance for the posterior is proposed to avoid the numerical integration. In the context of mapping, the problem of finding a transformation matrix to best maintain the information could be interesting for further research.

Chapter 6

General conclusion and perspectives

Throughout this dissertation, we have presented several statistical models to deal with the problem of analysing CCF events. Different algorithms to estimate the parameters of the models or to make predictive inference based on various kind of observed data have been proposed.

In Chapter 2, the full version of Binomial Failure Rate (BFR) model has been investigated in details. We have considered a practical situation of data where the independent failures can not be distinguished from the CCF of order 1 as well as the lethal shocks from the CCF involving all the components of the system. The observed data are then treated as missing data.

An EM algorithm is proposed to obtain the maximum likelihood estimates of the parameters of the BFR model. In the Bayesian scheme, we have introduced a new distribution, called the *modified-Beta distribution* to obtain a conjugate prior distribution for one of the parameters of the model. A simple strategy for eliciting the hyperparameters of prior distributions is also investigated.

In Chapter 3, we have studied the α -factor model for data with uncertainties. In the literature, the impact vector method is the most widely used to handle the data with uncertainties. Using this method, the uncertainties is represented by an impact vector, which is established based on the assessment of the analysts about the probability of all the possible orders of each CCF event.

This leads to another uncertainties: the subjectivity of the assessment. We have provided an example to show the sensitivity of the impact vector method to the probabilities established on the number of failed components in each CCF event. In order to avoid this subjective property of the impact vector method, we have proposed a novel method by introducing a new representation of data with uncertainties. By this representation, the analysts group all possible orders in a CCF event without assuming any probabilities.

A number of scenarios of this new form of data with uncertainties has been investigated. We have presented several methods to estimate the α parameters depending on the nature of available data. The proposed methods work efficiently on simulated data.

In Chapter 4, we have considered the causality-based inference problem for CCF data, that is to say, the situation where the CCF database includes not only the order but also the cause of each CCF event. This kind of data is given in the form of a contingency table. The value in each cell of the contingency table, corresponding to the number of CCF events of a specific order triggered off by a specific cause, is modeled with a homogenous Poisson process (HPP). In many practical situations, the exact value in each cell of the table is unobservable. Instead, only the margins of the contingency table are observed. In order to deal with this situation of missing information, we have suggested a Bayesian approach. The inverse Bayes formula is applied to obtain the conditional distribution of missing data given the observed data. Based on this conditional distribution, in some situations when the space of missing data is not too large, we have obtained an explicit formula for the posterior distribution of the rates in the HPPs.

In the case where the space of missing data given observed data is large, we have provided a stochastic version of IBF to obtain samples from the posterior distributions of the parameters. The link between parameters in the α -decomposition model suggested in [71] and parameters in our proposed model is pointed out. We have shown that the α -decomposition can be well approximated by using the estimated parameters in our proposed model.

In Chapter 5, we have studied the problem of making predictive inference for the number of catastrophic events that could occur in a system. We have shown that the Bayesian method is a generalization of the classical pivotal method. That is, by choosing a specific prior in Bayesian methodology, we obtain the pivotal method. In addition, a number of situations about the uncertainties on data has been investigated in the prediction. We have found that the approximative distribution could be applied to avoid numerical integration with very good performance.

Perspectives

In the present time, we are preparing a paper on the incomplete contingency table problem handled with the IBF/Metropolis-Hastings algorithm. Many other directions of research are revealed by the work presented in this dissertation.

We are planning to investigate the mapping technique in a more theoretical way using information theory.

We would like to study in more details the modified-Beta distribution. For example, finding an approximate distribution could be fruitful to avoid the numerical integration involved in its use.

In the Bayesian inference based on the inverse Bayes formula and the families of Dirichlet distribution like the nested Dirichlet or the grouped Dirichlet, most of the current studies are using a uniform Dirichlet prior. It would be interesting to investigate different choices for the parameters values of those Dirichlet. Finding an noninformative prior distribution for the mixed representation method is also a promising project.

In this dissertation, we have covered many situations of missing information for the sensitive problem of common cause failure connected with real life. We hope that our work will be useful for practitioners and will contribute to the safety of the environment.

Appendices

Appendix A

The EM algorithm

The EM (Expectation-Maximization) algorithm is perhaps the most widely used approach for finding maximum likelihood estimates (MLEs) in the context of missing data. Since the seminal paper of Dempster, Laird and Rubin [9], many papers and books has been published related to this algorithm. The book of McLachlan & Krishnan [33] gives a complete overview on the topic. In the following, we briefly present the algorithm.

Suppose that Y_{obs} is the observed data and that we have some missing data Z . Y_{obs} are then the incomplete data while (Y_{obs}, Z) represents the complete data.

Let $f(y_{obs} | \theta)$ be the distribution of Y_{obs} and suppose that $g(z | Y_{obs}; \theta)$, the distribution of X given Y_{obs} , is available.

Our goal is to find the maximum likelihood estimate of θ , that is to say the value of θ that maximize $\log L(\theta | Y_{obs})$, called the *incomplete likelihood*.

To this problem, considered as difficult even impossible in some cases, the idea of EM is to substitute the problem: maximizing the expectation of the complete likelihood given the incomplete data, that is to say the quantity

$$\int \log L(\theta | Y_{obs}, z) g(z | Y_{obs}; \theta) dz.$$

This is performed iteratively, considering the following two steps:

- **E-step:** Compute the Q function defined by

$$\begin{aligned} Q(\theta, \theta^{(t)}) &= \int \log L(\theta | Y_{obs}, z) g(z | Y_{obs}; \theta^{(t)}) dz, \\ &= E \left[\log L(\theta | Y_{obs}, Z) | Y_{obs}, \theta^{(t)} \right] \end{aligned}$$

where $\theta^{(t)}$ is the current estimate of the MLE $\hat{\theta}$.

- **M-step:** Maximize Q with respect to θ to obtain

$$\theta^{(t+1)} = \underset{\theta}{\operatorname{arg\,max}} Q(\theta, \theta^{(t)})$$

The two-step process is repeated until convergence occurs, i.e. until $\|\theta^{(t+1)} - \theta^{(t)}\| \leq \delta$, where δ is a sufficiently small quantity predetermined.

Wu [70] provides conditions for the convergence of the algorithm.

A very convenient feature of the EM algorithm is that when the distribution belongs to the exponential family, applying EM is simply replacing in the complete data likelihood, the missing data by their conditional expectation in the likelihood of the complete data.

Appendix B

The Inverse Bayesian formula

The inverse Bayes formulae (IBF) was introduced by Ng [38]. It has been widely used in a large number of studies, especially in problems of missing data, see for example [58, 62, 63]. The natural connection between the IBF and the data augmentation algorithm introduced by Tanner and Wong [60] is shown in [62]. The summary of this method and its numerous applications can be found in [41, 59]. For purpose of application, we describe in the following this formula.

Let Y_{obs} and Z denote the observed data and the missing data, respectively. The set (Y_{obs}, Z) represents the complete data. Let θ stands for the parameter of interest. Because of the missing data Z , the posterior distribution of θ given the observed data, $\pi(\theta|Y_{obs})$, is not straightforward to obtain. We assume that both the distribution of θ given the complete data, i.e. $\pi(\theta|Y_{obs}, Z)$, and the conditional distribution of the missing data, i.e. $\pi(Z|Y_{obs}, \theta)$, are known.

The goal now is to find $\pi(\theta | Y_{obs})$ based on these two distributions. In this context, Y_{obs} can be seen as a given constant throughout, acting like an indexing parameter in the family of joint distributions for (θ, Z) . Therefore, without loss of generality on the conceptual level we may drop the given Y_{obs} in all density functions. That is to say, in this abstraction we are going in the opposite direction of Bayes' theorem - finding $\pi(\theta)$ (the prior) from $\pi(\theta | Z)$ (the posterior) with $\pi(Z | \theta)$ (the likelihood).

This naturally leads to the name of *Inverse Bayesian formula*. Figure B.1 provides a graphical representation of this idea, which is an extension of the graph presented in [41].

The following equation is a solution to answer the question of how to find $\pi(\theta | Y_{obs})$ from $\pi(\theta|Y_{obs}, Z)$ and $\pi(Z|Y_{obs}, \theta)$.

$$\pi(\theta|Y_{obs}) = \int \pi(\theta|Y_{obs}, z)\pi(z|Y_{obs})dz. \quad (\text{B.0.1})$$

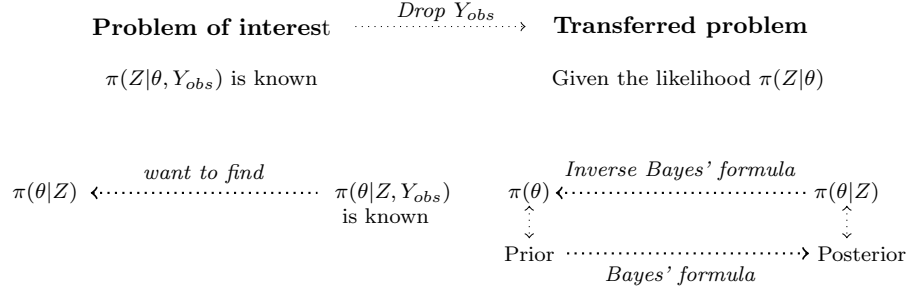


Figure B.1: A framework for the idea of the introduction of IBF.

In order to use (B.0.1), the conditional distribution $\pi(z|Y_{obs})$ is required. Tian and Tan [62] provide three versions for $\pi(z|Y_{obs})$ described as follows.

Let $\mathcal{S}(\theta|Y_{obs}, Z)$ and $\mathcal{S}(Z|Y_{obs}, \theta)$ denote the conditional supports of θ given Y_{obs} and Z given Y_{obs} , respectively.

According to the Bayes' theorem, we have

$$\pi(\theta | Y_{obs}) = \frac{\pi(\theta | Y_{obs}, Z)}{\pi(Z | Y_{obs}, \theta)} \pi(Z | Y_{obs}), \quad \theta \in \mathcal{S}(\theta|Y_{obs}), \quad Z \in \mathcal{S}(Z|Y_{obs}). \quad (\text{B.0.2})$$

Integrating both sides of the identity (B.0.2) with respect to θ on its support $\mathcal{S}(\theta|Y_{obs}, Z)$, we obtain the following formula:

$$\pi(Z | Y_{obs}) = \left\{ \int_{\mathcal{S}(\theta|Y_{obs}, Z)} \frac{\pi(\theta | Y_{obs}, Z)}{\pi(Z | Y_{obs}, \theta)} d\theta \right\}^{-1} \int_{\mathcal{S}(\theta|Y_{obs}, Z)} \pi(\theta | Y_{obs}) d\theta. \quad (\text{B.0.3})$$

For any value z_0 satisfying $\mathcal{S}(\theta|Y_{obs}, z_0) = \mathcal{S}(\theta|Y_{obs})$, we have

$$\int_{\mathcal{S}(\theta|Y_{obs}, z_0)} \pi(\theta|Y_{obs}) = 1.$$

As a result, the identity (B.0.3) reduces to

$$\pi(z_0 | Y_{obs}) = \left\{ \int_{\mathcal{S}(\theta|Y_{obs})} \frac{\pi(\theta | Y_{obs}, z_0)}{\pi(z_0 | Y_{obs}, \theta)} d\theta \right\}^{-1}, \quad (\text{B.0.4})$$

which is called the *point-wise IBF*.

Similarly, for any θ_0 satisfying $\mathcal{S}(Z|Y_{obs}, \theta_0) = \mathcal{S}(Z|Y_{obs})$, by the symmetry we obtain the point-wise IBF for θ as,

$$\pi(\theta_0 | Y_{obs}) = \left\{ \int_{\mathcal{S}(Z|Y_{obs})} \frac{\pi(Z | Y_{obs}, \theta_0)}{\pi(\theta_0 | Y_{obs}, Z)} dZ \right\}^{-1}. \quad (\text{B.0.5})$$

Substituting (B.0.5) into (B.0.2), we obtain the *function-wise IBF* for $\pi(Z | Y_{obs})$ as

$$\pi(Z | Y_{obs}) = \left\{ \int_{\mathcal{S}(Z|Y_{obs})} \frac{\pi(Z | Y_{obs}, \theta_0)}{\pi(\theta_0 | Y_{obs}, Z)} dZ \right\}^{-1} \frac{\pi(Z | Y_{obs}, \theta_0)}{\pi(\theta_0 | Y_{obs}, Z)}. \quad (\text{B.0.6})$$

By dropping the constant, (B.0.6) can be rewritten in the form

$$\pi(Z | Y_{obs}) \propto \frac{\pi(Z | Y_{obs}, \theta_0)}{\pi(\theta_0 | Y_{obs}, Z)}, \quad (\text{B.0.7})$$

which is called the *sampling IBF*.

After finding $\pi(Z | Y_{obs})$, we can achieve $\pi(\theta | Y_{obs})$ by calculating the integral in (B.0.1). When the direct calculation is not straightforward, the sampling method can be applied. That is, we obtain $\pi(Y_{obs})$ through simulating M samples $\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(M)}$ where $\theta^{(i)} \sim \pi(\theta | Y_{obs})$. In particular, we firstly simulate M samples $z^{(1)}, z^{(2)}, \dots, z^{(M)}$ from $\pi(Z | Y_{obs})$ in (B.0.7), and then generate $\theta^{(i)}$ from $\pi(\theta | Y_{obs}, z^{(i)})$, $i = 1, \dots, M$ which is supposed to be known.

Consider now the case when Z is a discrete random variable. Suppose that Z can take the value on the finite domain $\{z_1, \dots, z_\ell\}$. Thus, the integral in (B.0.1) becomes a finite summation where $\pi(Z | Y_{obs})$ is a probability mass function calculated from (B.0.7). Then, (B.0.1) can be rewritten as

$$\pi(\theta | Y_{obs}) = \sum_{i=1}^{\ell} \pi(\theta | Y_{obs}, z_i) \pi(Z = z_i | Y_{obs}). \quad (\text{B.0.8})$$

It should be consider that in general, the values in the set $\{z_i\}_{1, \dots, \ell}$ do not depend on θ , i.e, the condition $\mathcal{S}(Z | Y_{obs}) = \mathcal{S}(Z | Y_{obs}, \theta)$ is satisfied.

The probability $\pi(Z = z_i | Y_{obs})$ in (B.0.7) can be calculated as follows. Firstly, choose a value θ_0 in the support of $\theta | Y_{obs}$ and compute

$$q_i(\theta_0) = \frac{Pr(Z = z_i | Y_{obs}, \theta_0)}{\pi(\theta_0 | Y_{obs}, z_i)}, \quad i = 1, \dots, \ell. \quad (\text{B.0.9})$$

The value $q_i(\theta_0)$ obviously depends on θ_0 . Apply in the sampling IBF, we obtain

$$\pi(Z = z_i | Y_{obs}) = \frac{q_i(\theta_0)}{\sum_{i=1}^{\ell} q_i(\theta_0)}, \quad i = 1, \dots, \ell. \quad (\text{B.0.10})$$

which is independet of θ_0 .

The posterior distribution $\pi(\theta | Y_{obs})$ now is calculated as in (B.0.8) with the weights defined in (B.0.10).

In the case the support of Z is large, the expression of posterior distribution $\pi(\theta | Y_{obs})$ is cumbersome. The following procedure is suggested to simulate samples from $\pi(\theta | Y_{obs})$:

- Generate M samples $Z^{(1)}, \dots, Z^{(M)}$ where $Z^{(i)} \sim \pi(Z | Y_{obs})$ defined in (B.0.10).
- Generate M samples $\theta^{(1)}, \dots, \theta^{(M)}$ with $\theta^{(i)} \sim \pi(\theta | Y_{obs}, Z^{(i)})$.

Appendix C

The modified – Beta distribution

Definition 2. – A continuous random variable X with support $]0,1[$ follows a **modified–Beta** distribution with parameters $(\alpha, \beta, \gamma, \delta)$ if and only if, its probability density function is

$$f(x) = \frac{1}{K(\alpha, \beta, \gamma, \delta)} \frac{x^{\alpha-1}(1-x)^{\beta-1}}{(1-(1-x)^\delta)^\gamma}$$

where $x \in]0, 1[$, $\delta \in \mathbb{N}$, $\alpha, \beta, \gamma > 0$, with $\gamma < \alpha$ and

$$K(\alpha, \beta, \gamma, \delta) = \int_0^1 \frac{x^{\alpha-1}(1-x)^{\beta-1}}{(1-(1-x)^\delta)^\gamma} dx.$$

We denote $X \sim \tilde{\mathcal{B}}(\alpha, \beta, \gamma, \delta)$.

The figure C.1 represents the modified–Beta distribution for different values of the parameters. From the definition, when $\delta = 0$ or $\gamma = 0$, the modified–Beta distribution is the classical beta distribution.

When δ is large, the distribution can be approximated by the beta distribution.

The condition $\alpha > \gamma$ is needed for the constant K to be finite. Indeed, for $x \in]0, 1[$, $1 - x^m \geq 1 - x$ and $(1 - x^m)^\delta \geq (1 - x)^\delta$, therefore:

$$\frac{(1-x)^{\alpha-1}x^{\beta-1}}{(1-x^m)^\beta} \leq (1-x)^{\alpha-\gamma-1}x^{\beta-1}.$$

Integrating both parts, we have:

$$K(\alpha, \beta, \gamma, \delta) \leq B(\beta, \alpha - \gamma),$$

and if $\alpha > \gamma$ then $K(\alpha, \beta, \gamma, \delta)$ is finite.

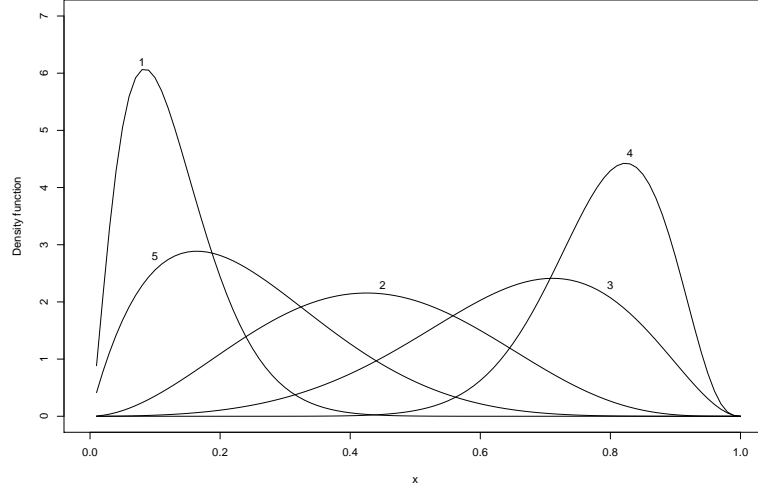


Figure C.1: Modified-beta distribution. **1.** $\alpha = 3, \beta = 18, \gamma = 0.5, \delta = 4$; **2.** $\alpha = 5, \beta = 5, \gamma = 2, \delta = 3$; **3.** $\alpha = 6, \beta = 3, \gamma = 3, \delta = 5$; **4.** $\alpha = 15, \beta = 4, \gamma = 5, \delta = 6$; **5.** $\alpha = 4, \beta = 7, \gamma = 2, \delta = 2$.

The n^{th} moment of a r.v. following a modified-Beta distribution with parameters $(\alpha, \beta, \gamma, \delta)$ is:

$$E(X^n) = \frac{K(\alpha + n, \beta, \gamma, \delta)}{K(\alpha, \beta, \gamma, \delta)} \quad (\text{C.0.1})$$

Proposition 9. *The normalization constant of modeified truncated distribution K can be expressed by the combination of Beta function as*

$$K(\alpha, \beta, \gamma, \delta) = B(\alpha, \beta) + \sum_{k=0}^{\infty} \frac{\gamma(\gamma+1)\dots(\gamma+k)}{(k+1)!} B(\alpha, \beta + (k+1)\delta).$$

Proof. The Maclaurin expansion of the function $\frac{1}{(1-x)^\gamma}$ with $0 < x < 1$ is:

$$\frac{1}{(1-x)^\gamma} = 1 + \sum_{k=0}^{\infty} \frac{\gamma(\gamma+1)\dots(\gamma+k)}{(k+1)!} x^{k+1}.$$

Since $x \in [0, 1]$, $1-x$ and then $(1-x)^\delta$ also belongs to $[0, 1]$. Substituting x by $(1-x)^\delta$, the expansion becomes:

$$\frac{1}{(1-(1-x)^\delta)^\gamma} = 1 + \sum_{k=0}^{\infty} \frac{\gamma(\gamma+1)\dots(\gamma+k)}{(k+1)!} (1-x)^{(k+1)\delta}.$$

The proposition is achieved by this expansion:

$$\begin{aligned}
 K &= \int_0^1 \frac{x^{\alpha-1}(1-x)^{\beta-1}}{(1-(1-x)^\delta)^\gamma} dx \\
 &= \int_0^1 x^{\alpha-1}(1-x)^{\beta-1} \left(1 + \sum_{k=0}^{\infty} \frac{\gamma(\gamma+1)\dots(\gamma+k)}{(k+1)!} (1-x)^{(k+1)\delta}\right) dx \\
 &= \int_0^1 x^{\alpha-1}(1-x)^{\beta-1} dx + \sum_{k=0}^{\infty} \frac{\gamma(\gamma+1)\dots(\gamma+k)}{(k+1)!} \int_0^1 x^\alpha (1-x)^{\beta+(k+1)\delta-1} dx \\
 &= B(\alpha, \beta) + \sum_{k=0}^{\infty} \frac{\gamma(\gamma+1)\dots(\gamma+k)}{(k+1)!} B(\alpha, \beta + (k+1)\delta).
 \end{aligned}$$

□

Proposition 10. *The modified–Beta distribution has at most one mode. Particularly, with $\alpha > \gamma + 1, \beta > 1$, the distribution is unimodal.*

Proof. Let us show that $\frac{\partial}{\partial x} \log f(x)$ is zero at at most one point and the sign change can only be from positive to negative. Indeed, the logarit of $f(x)$ is:

$$\log f(x) = (\alpha - 1)\log x + (\beta - 1)\log(1 - x) - \gamma \log(1 - (1 - x)^\delta) - \log K,$$

and then

$$\frac{\partial}{\partial x} \log f(x) = \frac{\alpha - 1}{x} - \frac{\beta - 1}{1 - x} - \frac{\delta \gamma (1 - x)^{\delta-1}}{1 - (1 - x)^\delta}.$$

To make the calculation succinct, y is used to refer $(1 - x)$ only in this proof. Since $\frac{\partial}{\partial x} \log f(x)$ has the same sign as $(1 - x) \frac{\partial}{\partial x} \log f(x)$, it is sufficient to show that

$$\begin{aligned}
 y \frac{\partial}{\partial x} \log f(x) &= y \left(\frac{\alpha - 1}{x} - \frac{\beta - 1}{1 - x} - \frac{\delta \gamma y^{\delta-1}}{1 - y^\delta} \right) \\
 &= \frac{\alpha - 1}{x} - \frac{\delta \gamma}{1 - y^\delta} - \alpha - \beta + \delta \gamma + 2 \\
 &= \frac{\alpha - \gamma - 1}{x} + \gamma \left(\frac{1}{x} - \frac{\delta}{1 - y^\delta} \right) - \alpha - \beta + \delta \gamma + 2.
 \end{aligned}$$

is strictly decreasing in x for $0 < x < 1$. By assumption $\alpha - 1 > \gamma$, the first term of $y \frac{\partial}{\partial x} \log f(x)$ is decreasing in x . Therefore it is only need to prove that the function in brackets is strictly decreasing in x for $0 < x < 1$. The derivative of the function $g(x) = \frac{1}{x} - \frac{\delta}{1 - y^\delta}$ is:

$$g'(x) = \frac{\delta^2 x^2 y^{\delta-1} - (1 - y^\delta)^2}{x^2 (1 - y^\delta)^2}.$$

Denote the numerrator by $A_\delta(x)$, we calculate:

$$\begin{aligned} A_1(x) &= x^2 - x^2 = 0, \\ A_{\delta+1}(x) - A_\delta(x) &= (\delta + 1)^2 x^2 y^\delta - (1 - y^{\delta+1})^2 - [\delta^2 x^2 y^{\delta-1} - (1 - y^\delta)^2] \\ &= x y^{\delta-1} [x((\delta + 1)^2 y - \delta^2) - y(2 - y^\delta - y^{\delta+1})] \\ &= x y^{\delta-1} B_\delta(x). \end{aligned}$$

where $B_\delta(x) = x((\delta + 1)^2 y - \delta^2) - y(2 - y^\delta - y^{\delta+1})$ which is satisfied:

$$\begin{aligned} B_1(x) &= x(2y - 1) - y(2 - y - y^2) \\ &= y^3 - y^2 + y - 1 < 0 \quad \forall y \in [0, 1], \\ B_{\delta+1}(x) - B_\delta(x) &= x[(2\delta + 3)x - 2\delta - 1 - y^{\delta+1} - y^{\delta+2}] \\ &= x C_\delta(x). \end{aligned}$$

in which $C_\delta(x) = [(2\delta + 3)y - 2\delta - 1 - y^{\delta+1} - y^{\delta+2}]$ satisfies:

$$\begin{aligned} C_1(x) &= 5y - 3 - y^2 - y^3 \\ &= x(y^2 + 2y - 3) < 0 \quad \forall y \in [0, 1], \\ C_{\delta+1}(x) - C_\delta(x) &= 2y - 2 + y^{\delta+1} - y^{\delta+3} \\ &= x(y^{\delta+1} - y^{\delta+2} - 2) < 0 \quad \forall y \in [0, 1]. \end{aligned}$$

Evidently, $C_\delta(x) < 0$ for $\delta \geq 1, 0 < x < 1$, then $B_\delta(x) < 0$ for $\delta \geq 1, 0 < x < 1$, and finally $A_\delta(x) < 0$ for $\delta \geq 1, 0 < x < 1$. That means $g(x)$ and then $y \frac{\partial}{\partial x} \log f(x)$ is also strictly decreasing. Hence it follows that $f'(x)$ is zero at at most one point and this function only changed sign from positive to negative. To locate x_0 , consider the limit of $y \frac{\partial}{\partial x} \log f(x)$ as x tends to 0 and 1. First, the limit

$$\lim_{x \rightarrow 1} y \frac{\partial}{\partial x} \log f(x) = \lim_{x \rightarrow 1} \left(\frac{\alpha - 1}{x} - \frac{\delta \gamma}{1 - y^\delta} - \alpha - \beta + \delta \gamma + 2 \right) = -\beta + 1$$

is negative as $\beta > 1$, then $x_0 < 1$.

Since $(1 - x)^\delta = 1 - \delta x + o(x)$, we obtain:

$$\begin{aligned} y \frac{\partial}{\partial x} \log f(x) &= \frac{\alpha - 1}{x} - \frac{\delta \gamma}{1 - y^\delta} - \alpha - \beta + \delta \gamma + 2 \\ &\stackrel{x \rightarrow 0}{=} \frac{\alpha - 1}{x} - \frac{\delta \gamma}{\delta x + o(x)} - \alpha - \beta + \delta \gamma + 2 \\ &\stackrel{x \rightarrow 0}{=} \frac{\alpha - \gamma - 1}{x} - \alpha - \beta + \delta \gamma + 2. \end{aligned}$$

By assumption $\alpha > \gamma + 1$ and $x \in (0, 1)$, the last term results in that $y \frac{\partial}{\partial x} \log f(x)$ is positive as $x \rightarrow 0$. So $x_0 > 0$. \square

Appendix D

Metropolis-Hastings algorithm

The Metropolis-Hastings (M-H) algorithm belongs to the family of the Markov Chain Monte Carlo (MCMC) methods. The purpose of those methods is to obtain a realizations sequence of a random variable having a probability distribution π for which the direct sampling is difficult, even impossible. This strategy is to build a Markov chain with a stationary distribution which is the distribution of interest π . In the sequel, we shortly describe the algorithm [7].

As for acceptance-rejection method, the M-H algorithm needs a density that can generate candidates. The candidate generating density denoted $q(x, y)$ is chosen as the transition kernel of a Markov chain.

Therefore $q(x, y)$ describes the movement from a state x to a state y and $q(x, y)$ is such that $\int q(x, y)dx = 1$.

If $q(x, y)$ satisfies the reversibility condition:

$$\pi(x)q(x, y) = \pi(y)q(y, x)$$

the Metropolis algorithm is described as follows:

Given that the chain is in the state $x^{(t)} = x$,

- Generate $y^* \sim q(x, y)$,
- If $\pi(y^*) > \pi(x)$, choose $x^{(t+1)} = y^*$; else choose $x^{(t+1)} = x$.

The choice depends on the ratio $\frac{\pi(y^*)}{\pi(x)}$.

If $\frac{\pi(y^*)}{\pi(x)} > 1$, $x^{(t+1)} = y^*$, else $x^{(t+1)} = x$.

Then the probability of move also called the *probability of acceptance*, can be described as:

$$\alpha(x, y) = \min \left\{ 1, \frac{\pi(y)}{\pi(x)} \right\}.$$

The transition probability $p_M(x, y)$ in the Metropolis algorithm are then expressed as:

$$p_M(x, y) = q(x, y) \alpha(x, y).$$

In the case where $q(x, y)$ is an arbitrary transition probability that means a transition probability which does not necessarily satisfy the reversibility condition, Hasting [13] suggests to take:

$$\alpha(x, y) = \begin{cases} \min \left\{ \frac{\pi(y)q(y, x)}{\pi(x)q(x, y)}, 1 \right\} & \text{if } \pi(x)q(x, y) > 0, \\ 1 & \text{if } \pi(x)q(x, y) = 0. \end{cases}$$

The M-H algorithm is then:

Given $x^{(t)}$,

1. generate $y^* \sim q(x^{(t)}, y)$,
2. choose

$$x^{(t+)} = \begin{cases} y^* & \text{with probability } \rho(x^{(t)}, y^*) \\ x^{(t)} & \text{with probability } 1 - \rho(x^{(t)}, y^*) \end{cases}$$

$$\text{where } \rho(x, y) = \min \left\{ \frac{\pi(y)q(y, x)}{\pi(x)q(x, y)}, 1 \right\}.$$

To summarize, the M-H algorithm in the algorithmic form is:
Starting from an arbitrary value $x^{(0)}$:

- repeat for $j = 1, \dots, M$
- generate y from $q(x^{(j)}, \cdot)$ and u from $\mathcal{U}[0, 1]$,
 - if $u \leq \alpha(x^{(j)}, y)$, set $x^{(j+1)} = y$,
 - else, set $x^{(j+1)} = x^{(j)}$
- return the values $x^{(1)}, \dots, x^{(M)}$.

Appendix E

Mean and variance of the posterior distribution of Λ

E.1 Uncertainties on both overved count and observation time

Recall that the posterior distribution of Λ (5.4.14) in the case of uncertainties on both observed count and observation time, which is denoted by π_2 , is

$$\Lambda | N_{m,u}(T_u) \sim \sum_{v=0}^k \zeta_v \int_0^{+\infty} \mathcal{G}(v + \alpha_0, t + \beta_0) h(t) dt.$$

The posterior mean is

$$\begin{aligned} \mathbf{E}_{\pi_2}(\Lambda) &= \int \Lambda \left(\sum_{v=0}^k \zeta_v \int_0^{+\infty} \mathcal{G}(v + \alpha_0, t + \beta_0) h(t) dt \right) d\Lambda \\ &= \int_0^{+\infty} \left(\sum_{v=0}^k \zeta_v \int \Lambda \mathcal{G}(v + \alpha_0, t + \beta_0) d\Lambda \right) h(t) dt \\ &= \int_0^{+\infty} \left(\sum_{v=0}^k \zeta_v \frac{v + \alpha_0}{t + \beta_0} \right) h(t) dt \\ &= \mathbf{E}_h \left(\frac{\mathbf{E}_\zeta + \alpha_0}{t + \beta_0} \right). \end{aligned} \tag{E.1.1}$$

Upon expanding the function $\frac{\mathbf{E}_\zeta + \alpha_0}{t + \beta_0}$ of t in a second-order Taylor series about the meand $\mathbf{E}_h(t)$ of t we obtain

$$\frac{1}{t + \beta_0} = \frac{1}{\mathbf{E}_h(t) + \beta_0} - (t - \mathbf{E}_h(t)) \frac{1}{(\mathbf{E}_h(t) + \beta_0)^2} + \frac{1}{2} (t - \mathbf{E}_h(t))^2 \frac{2}{(\mathbf{E}_h(t) + \beta_0)^3}.$$

Therefore, the posterior mean of Λ is

$$\begin{aligned} E_{\pi_2}(\Lambda) &= (E_{\zeta} + \alpha_0) E_h \left(\frac{1}{t + \beta_0} \right) \\ &= \frac{E_{\zeta} + \alpha_0}{E_h(t) + \beta_0} \left(1 + \frac{V_h(t)}{(E_h(t) + \beta_0)^2} \right). \end{aligned} \quad (\text{E.1.2})$$

The posterior variance of Λ is

$$\begin{aligned} V_{\pi_2}(\Lambda) &= E_{\pi_2}(\Lambda^2) - E_{\pi_2}^2(\Lambda) \\ &= \int \Lambda^2 \left(\sum_{v=0}^k \zeta_v \int_0^{+\infty} \mathcal{G}(v + \alpha_0, t + \beta_0) h(t) dt \right) d\Lambda - E_{\pi_2}^2(\Lambda) \\ &= \int_0^{+\infty} \left(\sum_{v=0}^k \zeta_v \int \Lambda^2 \mathcal{G}(v + \alpha_0, t + \beta_0) d\Lambda \right) h(t) dt - E_{\pi_2}^2(\Lambda) \\ &= \int_0^{+\infty} \left(\sum_{v=0}^k \zeta_v \left(\frac{v + \alpha_0}{(t + \beta_0)^2} + \left(\frac{v + \alpha_0}{t + \beta_0} \right)^2 \right) \right) h(t) dt - E_{\pi_2}^2(\Lambda) \\ &= E_h \left(\sum_{v=0}^k \zeta_v \left(\frac{v + \alpha_0}{(t + \beta_0)^2} + \left(\frac{v + \alpha_0}{t + \beta_0} \right)^2 \right) \right) - E_{\pi_2}^2(\Lambda) \\ &= E_h \left(\frac{E_{\zeta} + \alpha_0 + \sum_{v=0}^k \zeta_v (v + \alpha_0)^2}{(t + \beta_0)^2} \right) - E_{\pi_2}^2(\Lambda). \end{aligned} \quad (\text{E.1.3})$$

Using a second-order Taylor series expansion of the function $\frac{1}{(t + \beta_0)^2}$ about the mean $E_h(t)$ of t , we have

$$\frac{1}{(t + \beta_0)^2} = \frac{1}{(E_h(t) + \beta_0)^2} - (t - E_h(t)) \frac{2}{(E_h(t) + \beta_0)^3} + \frac{1}{2} (t - E_h(t))^2 \frac{6}{(E_h(t) + \beta_0)^4}.$$

Substituting this expansion into (E.1.3) we obtain

$$\begin{aligned} V_{\pi_2}(\Lambda) &= (E_{\zeta} + \alpha_0 + \sum_{v=0}^k \zeta_v (v + \alpha_0)^2) E_h \left(\frac{1}{(t + \beta_0)^2} \right) - E_{\pi_2}^2(\Lambda) \\ &= \frac{E_{\zeta} + \alpha_0 + \sum_{v=0}^k \zeta_v (v + \alpha_0)^2}{(E_h(t) + \beta_0)^2} \left(1 + \frac{3V_h(t)}{(E_h(t) + \beta_0)^2} \right) - E_{\pi_2}^2(\Lambda). \end{aligned} \quad (\text{E.1.4})$$

E.2 Uncertainties on observation time

Consider that π_1 is a special case of π_2 when the observed count is certain, i.e., $\zeta_n = Pr(N_{m,u}(T_u) = n) = 1, n \leq k$ and $\zeta_i = 0, i \neq n$, where $N_m(T_u) = n$. Thus, in the case of uncertainties only on the observation time, we have $E_{\zeta} = n$ and

$\sum_{v=0}^k \zeta_v (v + \alpha_0)^2 = (n + \alpha_0)^2$. Therefore, (E.1.4) and (E.2.1) becomes

$$\begin{aligned} E_{\pi_1}(\Lambda) &= \frac{n + \alpha_0}{E_h(t) + \beta_0} \left(1 + \frac{V_h(t)}{(E_h(t) + \beta_0)^2} \right), \\ V_{\pi_1}(\Lambda) &= \frac{n + \alpha_0 + (n + \alpha_0)^2}{(E_h(t) + \beta_0)^2} \left(1 + \frac{3V_h(t)}{(E_h(t) + \beta_0)^2} \right) - E_{\pi_1}^2(\Lambda). \end{aligned}$$

□

Bibliography

- [1] ATWOOD, C. (1980). Estimators for the binomial failure rate common cause model. *EG & G Idaho, NUREG/CR-1401* .
- [2] ATWOOD, C. (1986). The binomial failure rate common-cause model. *Technometrics* **28**, 139–148.
- [3] ATWOOD, C. (1996). Constrained noninformative priors in risk assessment. *Reliability Engineering and System Safety* **53**, 37–46.
- [4] ATWOOD, C. (2013). Consequences of mapping data or parameters in bayesian common-cause analysis. *Reliability Engineering and System Safety* **118**, 118–131.
- [5] ATWOOD, C. & KELLY, D. (2009). The binomial failure rate common-cause model with WinBUGS. *Reliability Engineering and System Safety* **46**, 990–999.
- [6] BOYER, R. (2017). Probabilistic risk assessment (PRA): The basis for recognizing emerging operational risks. Tech. rep., NASA Johnson Space Center Safety and Mission Assurance.
- [7] CHIB, S. & GREENBERG, E. (1995). Understanding the metropolis - hasting algorithm. *Amer. Statistician* **49**, 327–335.
- [8] COUNCIL, N. R. et al. (2014). *Lessons learned from the Fukushima nuclear accident for improving safety of US nuclear plants*. National Academies Press.
- [9] DEMPSTER, A., LAIRD, N. & RUBIN, D. (1977). Maximum likelihood from incomplete data with the EM algorithm. *J. Roy. Statist. Soc. Series B(Methodol.)* **39**, 1–38.
- [10] ENGELHARDT, M. (1994). Events in time: Basic analysis of Poisson data. Tech. rep., EG and G Idaho, Inc., Idaho Falls, ID (United States).
- [11] GUTIÉRREZ, P. H., AGUIRRE, T. V. & CHRISTEN, J. A. (2005). A practical method for obtaining prior distributions in reliability. *IEEE Transactions on Reliability* **54**, 262–269.

- [12] HASSIJA, V., KUMAR, C. S. & VELUSAMY, K. (2017). A pragmatic approach to estimate alpha factors for common cause failure analysis. *Annals of Nuclear Energy* **63**, 317–325.
- [13] HASTING, W. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**, 97–109.
- [14] HAUGE, S., HOEM, A. S., HOKSTAD, P., HABREKKE, S. & LUNDTEIGEN, M. (2015). Common cause failures in safety instrumented systems. Tech. rep., SINTEF, Trondheim.
- [15] HAUPTMANN, U. (1996). The multi-class binomial failure rate model. *Reliability Engineering and System Safety* **53**, 85–90.
- [16] IEC-61508 (2010). Functional safety of electrical/electronic/ programmable electronic safety related systems. Tech. rep., International Electrotechnical Commission.
- [17] JOHNSON, N. L., KOTZ, S. & KEMP, A. W. (1993). *Univariate Discrete Distributions*. 2nd Ed., Wiley.
- [18] KELLY, D. (2010). Common-cause failure treatment in event assessment: basis for a proposed new model. *Proceedings of Probabilistic Safety Assessment and Management (PSAM) 10*.
- [19] KELLY, D. & ATWOOD, C. (2011). Finding a minimally informative Dirichlet prior distribution using least squares. *Reliability Engineering & System Safety* **96**, 398–402.
- [20] KRISHNAMOORTHY, K., PENG J. (2011). Improved closed-form prediction intervals for Binomial and Poisson distributions. *Journal of Statistical Planning and Inference* **141**, 1709–1718.
- [21] KVAM, P. (1996). Estimation techniques for common cause failure data with different system sizes. *Technometrics* **38**, 382–388.
- [22] KVAM, P. (1998). The binomial failure rate mixture model for common cause failure data from the nuclear industry. *Appl. Statist.* **47**, 49–61.
- [23] KVAM, P. (1998). A parametric mixture model for common cause failure data. *IEEE Transaction on reliability* **47**, 30–34.
- [24] KVAM, P. & MARTZ, H. (1995). Bayesian inference in a discrete shock model using confounded common cause data. *Reliability Engineering and System Safety* **48**, 19–25.
- [25] KVAM, P. & MILLER, J. (2002). Common cause failure prediction using data mapping. *Reliability Engineering and System Safety* **76**, 273–278.

- [26] KVAM, P. (1993). Computational problems with the binomial failure rate model and incomplete common cause failure reliability data. Tech. rep., Los Alamos National Lab., NM (United States).
- [27] KVAM, P. (1996). Maximum likelihood estimation and the multivariate Bernoulli distribution: An application to reliability. In *Lifetime Data: Models in Reliability and Survival Analysis*. Springer, pp. 187–194.
- [28] KVAM, P. & MILLER, J. (2002). Discrete predictive analysis in probabilistic safety assessment. *Journal of quality technology* **34**, 106–117.
- [29] LE, D. T. & VASSEUR, D. (2018). A practical methodology for modeling and estimation of common cause failure parameters in multi-unit nuclear psa model. *Reliability Engineering and System Safety* **170**, 159–174.
- [30] LITTLE, R. J. & RUBIN, D. B. (2002). *Statistical analysis with missing data*, 2nd Edition. Wiley.
- [31] MARSHALL, A. W. & OLKIN, I. (1967). A multivariate exponential distribution. *Journal of the American statistical association* **62**, 30–44.
- [32] MARTZ, H. & PICARD, R. (1995). Uncertainty in Poisson event counts and exposure time in rate estimation. *Reliability Engineering and System Safety* **48**, 181–190.
- [33] MCLACHLAN, G. & KRISHNAN, T. (2007). *The EM algorithm and extensions*, vol. 382. John Wiley & Sons.
- [34] MEEKER, W., HAHN, G. & ESCOBAR, L. (2017). *Statistical Intervals: A Guide for practitioners and researchers*. Wiley Series in Probability and Statistics. Wiley, 2nd ed.
- [35] MOSLEH, A. (1991). Common cause failure: An analysis methodology and examples. *Reliability Engineering and System Safety* **34**, 249–292.
- [36] MOSLEH, A., FLEMING, K., PARRY, G., PAULA, H., WORLEDGE, D. & RASMUSON, D. (1988). Procedures for treating common cause failures in safety and reliability studies. Tech. rep., NUREG-4780, V1 & V2, US Nuclear Regulatory Commission.
- [37] NELSON, W. (1970). Confidence intervals for the ratio of two Poisson means and Poisson predictor intervals. *IEEE Transactions on Reliability* **19**, 42–49.
- [38] NG, K. (1997). Inversion of Bayes formula: explicit formulae for unconditional. In *Advance in the Theory and Practice in Statistics*, Wiley, New York. **19**, 571–584.

- [39] NG, K., TANG, M., TAN, M. & TIAN, G. (2008). Grouped Dirichlet distribution: A new tool for incomplete categorical data analysis. *Journal of multivariate analysis* **99**, 490–509.
- [40] NG, K., TANG, M., TIAN, G. & TAN, M. (2009). The nested Dirichlet distribution and incomplete categorical data analysis. *Statistical Sinica* **19**, 251–271.
- [41] NG, K. N., TIAN, G. & TANG, M. (2011). *Dirichlet and related distributions: Theory, Methods and Applications*. John Wiley & Sons.
- [42] NGUYEN, H. D. & GOUNO, E. (2016). Inference for common cause of failure models with incomplete data. In *The 4th Conference on the Interface between Statistics and Engineering, Parlemo, Italy*.
- [43] NGUYEN, H. D. & GOUNO, E. (2017). Bayesian estimation for a common-cause failure model. In *10th International Conference on Mathematical Methods in Reliability, Grenoble, France*.
- [44] NGUYEN, H. D. & GOUNO, E. (2019). Maximum likelihood and Bayesian inference for common-cause of failure model. *Reliability Engineering & System Safety* **182**, 56–62.
- [45] NUREG/CR-5485 (1998). Guidelines on modeling common-cause failures in probabilistic risk assessment. Tech. rep., U.S. Nuclear Regulatory Commission.
- [46] NUREG/CR-5500 (1987-1993). Reliability study: High-pressure coolant injection system. Tech. rep. 10, U.S. Nuclear Regulatory Commission.
- [47] NUREG/CR-6268 (2007). Common-cause failure database and analysis system: Event data collection, classification, and coding. Tech. rep., U.S. Nuclear Regulatory Commission.
- [48] NUREG/CR-6819 (2003). Common-cause failure event insights. *U.S. Nuclear Regulatory Commission* **1-4**.
- [49] O’CONNOR, A. & MOSLEH, A. (2014). Extending the alpha factor model for cause based treatment of common cause failure events in pra and event assesment. *Proceedings of probabilistic safety assessment and management conference, Honolulu, HI* .
- [50] O’CONNOR, A. & MOSLEH, A. (2016). A general cause based methodology for anaysis of common cause and dependent failures in system risk and reliability assessments. *Reliability Engineering and System Safety* **145**, 341–350.

- [51] O'CONNOR, A. N. (2013). *A general cause based methodology for analysis of dependent failures in system risk and reliability assessments*. Ph.D. thesis, University of Maryland, College Park.
- [52] PAULA, H. (1995). Technical note: On the definition of common-cause failures. *Nuclear Safety* **36**.
- [53] RASMUSSEN, N. (1975). Reactor safety study: An assessment of accident risks in u.s. commerical nuclear power plants. Tech. rep., Washington DC: U.S. Nuclear Regulatory Commission.
- [54] RAUSAND, M. (2013). *Risk assessment: theory, methods, and applications*. John Wiley & Sons.
- [55] SIU, N. (1990). A Monte Carlo method for multiple parameter estimation in the presence of uncertain data. *Reliability Engineering and System Safety* **28**, 59–98.
- [56] SIU, N. & MOSLEH, A. (1989). Treating data uncertainties in common-cause failure analysis. *Nuclear technology* **84**, 265–281.
- [57] SMITH, A. M. & WATSON, I. A. (1980). Common cause failure—a dilemma in perspective. *Reliability Engineering* **1**, 127–142.
- [58] TAN, M., TIAN, G. & NG, K. (2003). A noniterative sampling method for computing posteriors in the structure of em-type algorithms. *Statistical Sinica* **13**, 625–639.
- [59] TAN, M., TIAN, G. & NG, K. (2010). *Bayesian missing data problems: EM, data augmentation and noniterative computation*. Biostatistics Series. Takir abd Francis Group, LLC, 1st ed.
- [60] TANNER, M. & WONG, W. (1987). The calculation of posterior distributions by data augmentation (with discussion). *Journal of the American Statistical Association* **82**, 528–540.
- [61] TIAN, G., NG, K. & GENG, Z. (2003). Bayesian computation for contingency tables with incomplete cell-counts. *Statistical Sinica* **13**, 189–206.
- [62] TIAN, G. & TAN, M. (2003). Exact statistical solutions using the inverse bayes formulae. *Statistics & Probability Letters* **62**, 305–315.
- [63] TIAN, G., TAN, M. & NG, K. (2007). An exact non-iterative sampling procedure for discrete missing data problems. *Statistical Neerlandica* **61**, 232–242.
- [64] TIAN, G., TANG, M., YUEN, K. & NG, K. (2010). Further properties and new applications of the nested Dirichlet distribution. *Computational statistics and data analysis* **54**, 394–405.

- [65] TROFFAES, M., WALTER, G. & KELLY, D. (2014). A robust bayesian approach to modeling epistemic uncertainty in common-cause failure models. *Reliability Engineering and System Safety* **125**, 13–21.
- [66] VAURIO, J. (2002). Extentions of the uncertainty quantification of common cause failure rates. *Reliability Engineering and System Safety* **78**, 63–69.
- [67] VAURIO, J. (2007). Consistent mapping of common cause failure rates and alpha factors. *Reliability Engineering and System Safety* **92**, 628–945.
- [68] VESESLY, W. (1977). Estimating common cause failure probabilities in reliability and risk analysis: Marshall-olkin specializations. *Nuclear Systems Reliability Engineering and Risk Assessment* , 314–341.
- [69] WERNER, W. (1995). Results of recent risk studies in France, Germany, Japan, Sweden and the United States. Tech. rep., Bundesministerium fuer Umwelt.
- [70] WU, C. F. (1983). On the convergence properties of the EM algorithm. *The Annals of statistics* **11**, 95–103.
- [71] ZHENG, X., YAMAGUCHI, A. & TAKATA, T. (2013). α -Decomposition for estimating parameters in common cause failure modeling based on causal inference. *Reliability Engineering and System Safety* **116**, 20–27.
- [72] ZUBAIR, M. & AMJAD, Q. (2016). Calculation and updating of common cause failure unavailability by using alpha factor model. *Annals of Nuclear Energy* **90**, 106–114.