



Generalised Bayesian Inference under Prior-Data Conflict

Gero Walter

Department of Statistics
Ludwig-Maximilians-Universität München (LMU)

October 25th, 2013



Institut
für
Statistik





Overview

1. Bayesian inference & prior-data conflict
2. Generalised Bayesian inference with sets of priors
(joint work with Thomas Augustin and Frank Coolen)
3. Common-cause failure modeling
(joint work with Matthias Troffaes and Dana Kelly)



Overview

Contents

Executive

1. Overview

1.1. Introduction

1.2. Bayesian Inference

1.3. Common-Cause Failure Modelling

2. Generalised Bayesian Inference

2.1. Introduction

2.2. Bayesian Inference

2.3. Common-Cause Failure Modelling

3. Common-Cause Failure Modelling

3.1. Introduction

3.2. Bayesian Inference

3.3. Common-Cause Failure Modelling

1. Overview

1.1. Introduction

1.2. Bayesian Inference

1.3. Common-Cause Failure Modelling

2. Generalised Bayesian Inference

2.1. Introduction

2.2. Bayesian Inference

2.3. Common-Cause Failure Modelling

3. Common-Cause Failure Modelling

3.1. Introduction

3.2. Bayesian Inference

3.3. Common-Cause Failure Modelling

2. Generalised Bayesian Inference

2.1. Introduction

2.2. Bayesian Inference

2.3. Common-Cause Failure Modelling

3. Common-Cause Failure Modelling

3.1. Introduction

3.2. Bayesian Inference

3.3. Common-Cause Failure Modelling

3. Common-Cause Failure Modelling

3.1. Introduction

3.2. Bayesian Inference

3.3. Common-Cause Failure Modelling

1. Bayesian inference & prior-data conflict
2. Generalised Bayesian inference with sets of priors (joint work with Thomas Augustin and Frank Coolen)
3. Common-cause failure modeling (joint work with Matthias Troffaes and Dana Kelly)



Overview

Contents

Contents	viii
List of Figures	ix
List of Tables	xi
Abstract	xiii
Zusammenfassung	xv
Eidesstattliche Versicherung	xvii
1. Introduction	1
1.1. Preliminaries	1
1.1.1. Overview	1
1.1.2. Sources	2
1.1.3. Notation	3
1.2. Some Fundamentals	3
1.2.1. Statistical Inference	3
1.2.2. Parametric Models	5
1.2.3. Statistical Inference with the Bayesian Paradigm	7
1.2.3.1. Regular Conjugate Families of Distributions	8
1.2.3.2. Inference Tasks	9
1.2.3.3. The Beta-Binomial Model	11
1.2.3.4. The Normal/Normal Model	14
1.2.3.5. The Dirichlet/Multinomial Model	16
1.3. Dirichlet-Multinomial Model for Common-Cause Failure	21
1.3.1. An Example for the Need to Model Common-Cause Failure	21
1.3.2. The Basic Parameter Model	23
1.3.3. The Alpha-Factor Model	25
1.3.4. Dirichlet Prior for Alpha-Factors	25
1.3.5. Usual Handling of Epistemic Information for Alpha-Factors	24
1.3.6. Cautious Epistemic Information for Alpha-Factors	26
1.3.6.1. Fixed Learning Parameter	27
1.3.6.2. Interval for Learning Parameter	28
1.3.6.3. Conclusion	31

iii	
ii	
i	
0	
-1	
-2	
-3	
-4	
-5	
-6	
-7	
-8	
-9	
-10	
-11	
-12	
-13	
-14	
-15	
-16	
-17	
-18	
-19	
-20	
-21	
-22	
-23	
-24	
-25	
-26	
-27	
-28	
-29	
-30	
-31	
-32	
-33	
-34	
-35	
-36	
-37	
-38	
-39	
-40	
-41	
-42	
-43	
-44	
-45	
-46	
-47	
-48	
-49	
-50	
-51	
-52	
-53	
-54	
-55	
-56	
-57	
-58	
-59	
-60	
-61	
-62	
-63	
-64	
-65	
-66	
-67	
-68	
-69	
-70	
-71	
-72	
-73	
-74	
-75	
-76	
-77	
-78	
-79	
-80	
-81	
-82	
-83	
-84	
-85	
-86	
-87	
-88	
-89	
-90	
-91	
-92	
-93	
-94	
-95	
-96	
-97	
-98	
-99	
-100	

Contents	ii
1. Introduction	1
1.1. Preliminaries	1
1.1.1. Overview	1
1.1.2. Sources	2
1.1.3. Notation	3
1.2. Some Fundamentals	3
1.2.1. Statistical Inference	3
1.2.2. Parametric Models	5
1.2.3. Statistical Inference with the Bayesian Paradigm	7
1.2.3.1. Regular Conjugate Families of Distributions	8
1.2.3.2. Inference Tasks	9
1.2.3.3. The Beta-Binomial Model	11
1.2.3.4. The Normal/Normal Model	14
1.2.3.5. The Dirichlet/Multinomial Model	16
1.3. Dirichlet-Multinomial Model for Common-Cause Failure	21
1.3.1. An Example for the Need to Model Common-Cause Failure	21
1.3.2. The Basic Parameter Model	23
1.3.3. The Alpha-Factor Model	25
1.3.4. Dirichlet Prior for Alpha-Factors	25
1.3.5. Usual Handling of Epistemic Information for Alpha-Factors	24
1.3.6. Cautious Epistemic Information for Alpha-Factors	26
1.3.6.1. Fixed Learning Parameter	27
1.3.6.2. Interval for Learning Parameter	28
1.3.6.3. Conclusion	31

iii	
ii	
i	
0	
-1	
-2	
-3	
-4	
-5	
-6	
-7	
-8	
-9	
-10	
-11	
-12	
-13	
-14	
-15	
-16	
-17	
-18	
-19	
-20	
-21	
-22	
-23	
-24	
-25	
-26	
-27	
-28	
-29	
-30	
-31	
-32	
-33	
-34	
-35	
-36	
-37	
-38	
-39	
-40	
-41	
-42	
-43	
-44	
-45	
-46	
-47	
-48	
-49	
-50	
-51	
-52	
-53	
-54	
-55	
-56	
-57	
-58	
-59	
-60	
-61	
-62	
-63	
-64	
-65	
-66	
-67	
-68	
-69	
-70	
-71	
-72	
-73	
-74	
-75	
-76	
-77	
-78	
-79	
-80	
-81	
-82	
-83	
-84	
-85	
-86	
-87	
-88	
-89	
-90	
-91	
-92	
-93	
-94	
-95	
-96	
-97	
-98	
-99	
-100	



Overview

1. Introduction

1.1. Preliminaries

Contents

Contents	iii
List of Figures	ix
List of Tables	xi
Abstract	xiii
Zusammenfassung	xv
Eidesstattliche Versicherung	xvii

1. Introduction	1
1.1. Preliminaries	1
1.1.1. Overview	1
1.1.2. Sources	2
1.1.3. Notation	3
1.2. Some Fundamentals	3
1.2.1. Statistical Inference	3
1.2.2. Parametric Models	5
1.2.3. Statistical Inference with the Bayesian Paradigm	7
1.2.3.1. Regular Conjugate Families of Distributions	8
1.2.3.2. Inference Tasks	9
1.2.3.3. The Beta-Binomial Model	11
1.2.3.4. The Normal/Normal Model	14
1.2.3.5. The Dirichlet/Multinomial Model	16
1.3. Dirichlet/Multinomial Model for Common-Cause Failure	21
1.3.1. An Example for the Need to Model Common-Cause Failure	21
1.3.2. The Basic Parametric Model	23
1.3.3. The Alpha-Factor Model	25
1.3.4. Dirichlet Prior for Alpha-Factors	25
1.3.5. Usual Handling of Epistemic Information for Alpha-Factors	24
1.3.6. Cautious Epistemic Information for Alpha-Factors	26
1.3.6.1. Fixed Learning Parameter	27
1.3.6.2. Interval for Learning Parameter	28
1.3.6.3. Conclusion	28



Overview

1. Introduction 1

1.1. Preliminaries 1

1.3. Dirichlet-Multinomial Model for Common-Cause Failure 20

1.3.1. An Example for the Need to Model Common-Cause Failure 20

1.3.2. The Basic Parameter Model 21

1.3.3. The Alpha-Factor Model 22

1.3.4. Dirichlet Prior for Alpha-Factors 23

1.3.5. Usual Handling of Epistemic Information for Alpha-Factors 24

1.3.6. Cautious Epistemic Information for Alpha-Factors 26

1.3.6.1. Fixed Learning Parameter 27

1.3.6.2. Interval for Learning Parameter 28

1.3.6.3. Conclusion 30

1.3.4. The Normal/Normal Model 34

1.3.4. The Dirichlet-Multinomial Model 36

11. Dirichlet-Multinomial Model for Common-Cause Failure 28

1.1.1. An Example for the Need to Model Common-Cause Failure 28

1.1.2. The Basic Parameter Model 21

1.1.3. The Alpha-Factor Model 22

1.1.4. Dirichlet Prior for Alpha-Factors 23

1.1.5. Usual Handling of Epistemic Information for Alpha-Factors 24

1.1.6. Cautious Epistemic Information for Alpha-Factors 26

1.1.6.1. Fixed Learning Parameter 27

1.1.6.2. Interval for Learning Parameter 28

1.1.6.3. Conclusion 30



Overview

vi	Contents
2.	Imprecise Probability as Foundation of Generalised Bayesian Inference 31
2.1.	Imprecise or Interval Probability 31
2.1.1.	General Concept and Basic Interpretation 32

2.2. Motives for the Use of Imprecise Probability Methods 46

2.2.3.3. Prior-Data Conflict 50

2.2.	Motives for the Use of Imprecise Probability Methods 46
2.2.1.	The Fundamental Motives 46
2.2.1.1.	Risk and Ambiguity 46
2.2.1.2.	Motives from a Bayesian Perspective 46
2.2.1.3.	Foundational Motives 46
2.2.1.3.1.	Weak Information Bases 46
2.2.1.3.2.	Prior-Data Conflict 48
2.2.1.3.3.	Critique and Dimensions of Some Alternatives 48
2.2.1.3.4.	Objections to Imprecise Probability Models 48
2.2.1.4.	Hierarchical Models 52
3.	Generalised Bayesian Inference with Sets of Conjugate Priors in Exponential Families 55
3.1.	Model Overview and Discussion 55
3.1.1.	The General Framework 55
3.1.2.	Properties and Criteria 58
3.1.3.	The IDM and other Prior Non-Ignorance Models 63
3.1.4.	Substantial Prior Information and Sensitivity to Prior-Data Conflict 64
3.2.	Alternative Models Using Sets of Priors 65
3.2.1.	Some Alternative Model Frameworks 65
3.2.1.1.	Neighbourhood Models 65
3.2.1.1.1.	The Doubly Ratio Class 65
3.2.1.2.	Some Approaches Based on Conjugate Priors 70
3.2.1.2.1.	The Model by Coates [193b], [194] 70
3.2.1.2.1.1.	The Model by Birks [183] 72
3.2.1.2.1.2.	Some of the Models Studied by Pericchi and Walley [193 c] 72
3.2.1.3.	Some Other Approaches Using Sets of Priors 71
3.2.1.3.1.	The Model by Rinderknecht [181] 71

3.2.1.3.2.	Conditioning and the Generalised Bayes' Rule 33
3.2.1.3.3.	Prior-Data Conflict 50
3.2.1.3.4.	Critique and Dimensions of Some Alternatives 48
3.2.1.3.5.	Objections to Imprecise Probability Models 48
3.2.1.4.	Hierarchical Models 52
3.2.1.4.1.	Bayesian Inference with Sets of Conjugate Priors in Exponential Families 55
3.2.1.4.1.1.	The General Framework 55
3.2.1.4.1.2.	Properties and Criteria 58
3.2.1.4.1.3.	The IDM and other Prior Non-Ignorance Models 63
3.2.1.4.1.4.	Substantial Prior Information and Sensitivity to Prior-Data Conflict 64
3.2.1.4.2.	Alternative Models Using Sets of Priors 65
3.2.1.4.2.1.	Some Alternative Model Frameworks 65
3.2.1.4.2.1.1.	Neighbourhood Models 65
3.2.1.4.2.1.1.1.	The Doubly Ratio Class 65
3.2.1.4.2.1.2.	Some Approaches Based on Conjugate Priors 70
3.2.1.4.2.1.2.1.	The Model by Coates [193b], [194] 70
3.2.1.4.2.1.2.1.1.	The Model by Birks [183] 72
3.2.1.4.2.1.2.1.2.	Some of the Models Studied by Pericchi and Walley [193 c] 72
3.2.1.4.2.1.3.	Some Other Approaches Using Sets of Priors 71
3.2.1.4.2.1.3.1.	The Model by Rinderknecht [181] 71



Overview

2. Imprecise Probability as Foundation of Generalised Bayesian Inference	31
2.1. Imprecise or Interval Probability	32
2.1.1. General Concept and Basic Interpretation	32
2.2. Motives for the Use of Imprecise Probability Methods	46
2.2.3.3. Prior-Data Conflict	50
3. Generalised Bayesian Inference with Sets of Conjugate Priors in Exponential Families	55
3.1. Model Overview and Discussion	55
3.1.1. The General Framework	55
3.1.2. Properties and Criteria	58
3.1.3. The IDM and other Prior Non-Ignorance Models	63
3.1.4. Substantial Prior Information and Sensitivity to Prior-Data Conflict	64
3.2. Alternative Models Using Sets of Priors	65
3.2.1. Some Alternative Model Frameworks	65
3.2.1.1. Neighbourhood Models	65
3.2.1.2. The Density Ratio Class	65
3.2.2. Some Approaches Based on Conjugate Priors	70
3.2.2.1. The Model by Cozzari [193]; [394]	70
3.2.2.2. The Model by Birkis [183]	72
3.2.2.3. Some of the Models Studied by Pericchi and Walley [199]	72
3.2.3. Some Other Approaches Using Sets of Priors	73
3.2.3.1. The Model by Rinderknecht [181]	73



Overview

2. Imprecise Probability as Foundation of Generalised Bayesian Inference	31
2.1. Imprecise or Interval Probability	32
2.1.1. General Concept and Basic Interpretation	32
2.2. Motives for the Use of Imprecise Probability Methods	46
2.2.3.3. Prior-Data Conflict	50
3. Generalised Bayesian Inference with Sets of Conjugate Priors in Exponential Families	55
3.1. Model Overview and Discussion	56
3.1.1. The General Framework	56
3.1.2. Properties and Criteria	59
3.1.3. The IDM and other Prior Near-Ignorance Models	61
3.1.4. Substantial Prior Information and Sensitivity to Prior-Data Conflict	64



Overview

Contents	vii
1.3.1.3. The Model by Whitcomb (2011)	75
1.3. Imprecision and Prior-Data Conflict in Generalised Bayesian Inference	73
1.3.1. Introduction	73
1.3.1.1. Traditional Bayesian Inference and LUCC-models	81
1.3.1.2. Imperfect Priors for Inference in LUCC-models	84
1.3.1.3. LUCC-models and Prior-Data Conflict	89
1.3.1.4. Improving Imperfect Priors for Inference in LUCC-models	91
1.3.1.5. Illustration of the Generalised LUCC-model	92
1.3.1.6. Concluding Remarks	96
1.4. The Luck Package	98
1.5. On Prior-Data Conflict in Predictive Bernoulli Inference	103
1.5.1. Introduction	103
1.5.1.1. Imperfect Beta-Binomial Models	103
1.5.1.1.1. The Framework	103
1.5.1.1.2. Walley's psi-BBM	105
1.5.1.1.3. Anterior Shape Prior Sets	108
1.5.1.1.4. Intermediate Results	112
1.5.1.1.5. Weighted Inference	114
1.5.1.1.6. The Basic Model	114
1.5.1.1.7. The Extended Model	117
1.5.1.1.8. Weighted Inference Model Properties	118
1.5.1.1.9. Insights and Challenges	119
1.5.1.2. Concluding Remarks	123
1.5.1.3. Summary	123
1.5.1.4. Discussion	124
1.5.1.5. Outlook	125
A. Appendix	135
A.1. Bayesian Linear Regression: Divergence of Conjugate Models and Their (In)Sensitivity to Prior-Data Conflict	135
A.1.1. Introduction	135
A.1.1.1. Prior-Data Conflict in the Joint Case	135
A.1.1.1.1. Samples from a scaled Normal distribution	135
A.1.1.1.2. Samples from a Multinomial distribution	139
A.1.1.1.3. The Stumbled Approach for Bayesian Linear Regression (SCP)	140
A.1.1.1.4. Update of $\beta \sigma^2$	142
A.1.1.1.5. Update of σ^2	142
A.1.1.1.6. Update of β	143
A.1.1.2. An Alternative Approach for Conjugate Priors in Bayesian Linear Regression (CCCP)	144
A.1.1.2.1. Update of $\beta \sigma^2$	148

Contents	8
1.3.1.3. The Model by Whitcomb (2011)	75
1.3. Imprecision and Prior-Data Conflict in Generalised Bayesian Inference	73
1.3.1. Introduction	73
1.3.1.1. Traditional Bayesian Inference and LUCC-models	81
1.3.1.2. Imperfect Priors for Inference in LUCC-models	84
1.3.1.3. LUCC-models and Prior-Data Conflict	89
1.3.1.4. Improving Imperfect Priors for Inference in LUCC-models	91
1.3.1.5. Illustration of the Generalised LUCC-model	92
1.3.1.6. Concluding Remarks	96
1.4. The Luck Package	98
1.5. On Prior-Data Conflict in Predictive Bernoulli Inference	103
1.5.1. Introduction	103
1.5.1.1. Imperfect Beta-Binomial Models	103
1.5.1.1.1. The Framework	103
1.5.1.1.2. Walley's psi-BBM	105
1.5.1.1.3. Anterior Shape Prior Sets	108
1.5.1.1.4. Intermediate Results	112
1.5.1.1.5. Weighted Inference	114
1.5.1.1.6. The Basic Model	114
1.5.1.1.7. The Extended Model	117
1.5.1.1.8. Weighted Inference Model Properties	118
1.5.1.1.9. Insights and Challenges	119
1.5.1.2. Concluding Remarks	123
1.5.1.3. Summary	123
1.5.1.4. Discussion	124
1.5.1.5. Outlook	125
B. Bibliography	136
B.1. Bayesian Linear Regression: Divergence of Conjugate Models and Their (In)Sensitivity to Prior-Data Conflict	135
B.1.1. Introduction	135
B.1.1.1. Prior-Data Conflict in the Joint Case	135
B.1.1.1.1. Samples from a scaled Normal distribution	135
B.1.1.1.2. Samples from a Multinomial distribution	139
B.1.1.1.3. The Stumbled Approach for Bayesian Linear Regression (SCP)	140
B.1.1.1.4. Update of $\beta \sigma^2$	142
B.1.1.1.5. Update of σ^2	142
B.1.1.1.6. Update of β	143
B.1.1.2. An Alternative Approach for Conjugate Priors in Bayesian Linear Regression (CCCP)	144
B.1.1.2.1. Update of $\beta \sigma^2$	148

Contents	8
1.3.1.3. The Model by Whitcomb (2011)	75
1.3. Imprecision and Prior-Data Conflict in Generalised Bayesian Inference	73
1.3.1. Introduction	73
1.3.1.1. Traditional Bayesian Inference and LUCC-models	81
1.3.1.2. Imperfect Priors for Inference in LUCC-models	84
1.3.1.3. LUCC-models and Prior-Data Conflict	89
1.3.1.4. Improving Imperfect Priors for Inference in LUCC-models	91
1.3.1.5. Illustration of the Generalised LUCC-model	92
1.3.1.6. Concluding Remarks	96
1.4. The Luck Package	98
1.5. On Prior-Data Conflict in Predictive Bernoulli Inference	103
1.5.1. Introduction	103
1.5.1.1. Imperfect Beta-Binomial Models	103
1.5.1.1.1. The Framework	103
1.5.1.1.2. Walley's psi-BBM	105
1.5.1.1.3. Anterior Shape Prior Sets	108
1.5.1.1.4. Intermediate Results	112
1.5.1.1.5. Weighted Inference	114
1.5.1.1.6. The Basic Model	114
1.5.1.1.7. The Extended Model	117
1.5.1.1.8. Weighted Inference Model Properties	118
1.5.1.1.9. Insights and Challenges	119
1.5.1.2. Concluding Remarks	123
1.5.1.3. Summary	123
1.5.1.4. Discussion	124
1.5.1.5. Outlook	125
C. Index	143



Overview

Contents	vii
1.2.3.3. The Model by Whitcomb (2011)	71
1.3. Imprecision and Prior-Data Conflict in Generalised Bayesian Inference	73
1.3.1. Introduction	73
1.3.2. Traditional Bayesian Inference and LUCK-models	82
1.3.3. Imprecise Priors for Inference in LUCK-models	84
1.3.3.1. iLUCK-models	84
1.3.3.2. iLUCK-models and Prior-Data Conflict	88
1.3.4. Improved Imprecise Priors for Inference in LUCK-models	90
1.3.5. Illustration of the Generalised iLUCK-model	92
1.3.5.1. Concluding Remarks	94
1.4. The iuck Package	94
1.5. On Prior-Data Conflict in Predictive Bernoulli Inference	103
1.5.1. Introduction	103
1.5.1.1. Success Rate-Binomial Models	103

Contents	ix
1.2.3.3. The Model by Whitcomb (2011)	71
1.3. Imprecision and Prior-Data Conflict in Generalised Bayesian Inference	73
1.3.1. Introduction	73
1.3.2. Traditional Bayesian Inference and LUCK-models	82
1.3.3. Imprecise Priors for Inference in LUCK-models	84
1.3.3.1. iLUCK-models	84
1.3.3.2. iLUCK-models and Prior-Data Conflict	88
1.3.4. Improved Imprecise Priors for Inference in LUCK-models	90
1.3.5. Illustration of the Generalised iLUCK-model	92
1.3.5.1. Concluding Remarks	94
1.4. The iuck Package	94
1.5. On Prior-Data Conflict in Predictive Bernoulli Inference	103
1.5.1. Introduction	103
1.5.1.1. Success Rate-Binomial Models	103

Contents	xix
1.2.3.3. The Model by Whitcomb (2011)	71
1.3. Imprecision and Prior-Data Conflict in Generalised Bayesian Inference	73
1.3.1. Introduction	73
1.3.2. Traditional Bayesian Inference and LUCK-models	82
1.3.3. Imprecise Priors for Inference in LUCK-models	84
1.3.3.1. iLUCK-models	84
1.3.3.2. iLUCK-models and Prior-Data Conflict	88
1.3.4. Improved Imprecise Priors for Inference in LUCK-models	90
1.3.5. Illustration of the Generalised iLUCK-model	92
1.3.5.1. Concluding Remarks	94
1.4. The iuck Package	94
1.5. On Prior-Data Conflict in Predictive Bernoulli Inference	103
1.5.1. Introduction	103
1.5.1.1. Success Rate-Binomial Models	103

3.3. Imprecision and Prior-Data Conflict in Generalised Bayesian Inference	79
3.3.1. Introduction	79
3.3.2. Traditional Bayesian Inference and LUCK-models	82
3.3.3. Imprecise Priors for Inference in LUCK-models	84
3.3.3.1. iLUCK-models	84
3.3.3.2. iLUCK-models and Prior-Data Conflict	88
3.3.4. Improved Imprecise Priors for Inference in LUCK-models	90
3.3.5. Illustration of the Generalised iLUCK-model	92

A.1.1.4. An Alternative Approach for Conjugate Priors in Bayesian Linear Regression (CCCP)	144
A.1.1.4.1. Update of $\beta \sigma^2$	148



Overview

viii	Contents
	A.1.4.2. Update of α^2 140
	A.1.4.3. Update of β 133
	A.1.5. Discussion and Outlook 154
	A.2. A Parameter Set Shape for Strong Prior-Data Agreement Modelling 157
	A.2.1. A Novel Parametrisation of Canonical Conjugate Priors 157
	A.2.1.1. Informal Remarks for Non-Shaped Parameter Sets 159
	A.2.1.2. The Boundshape 163
	A.2.1.3. Basic Definition 164
	A.2.1.4. Finding the Toolpoints for the Basic Set 163
	A.2.1.5. Strong Prior-Data Agreement Property 164
	A.2.1.6. General Update with $\kappa > \frac{1}{2}$ 165
	A.2.4. Discussion and Outlook 168
Bibliography	171

viii	Contents
	A.1.4.2. Update of α^2 140
	A.1.4.3. Update of β 133
	A.1.5. Discussion and Outlook 154
	A.2. A Parameter Set Shape for Strong Prior-Data Agreement Modelling 157
	A.2.1. A Novel Parametrisation of Canonical Conjugate Priors 157
	A.2.1.1. Informal Remarks for Non-Shaped Parameter Sets 159
	A.2.1.2. The Boundshape 163
	A.2.1.3. Basic Definition 164
	A.2.1.4. Finding the Toolpoints for the Basic Set 163
	A.2.1.5. Strong Prior-Data Agreement Property 164
	A.2.1.6. General Update with $\kappa > \frac{1}{2}$ 165
	A.2.4. Discussion and Outlook 168
Bibliography	171

viii	Contents
	A.1.4.2. Update of α^2 140
	A.1.4.3. Update of β 133
	A.1.5. Discussion and Outlook 154
	A.2. A Parameter Set Shape for Strong Prior-Data Agreement Modelling 157
	A.2.1. A Novel Parametrisation of Canonical Conjugate Priors 157
	A.2.1.1. Informal Remarks for Non-Shaped Parameter Sets 159
	A.2.1.2. The Boundshape 163
	A.2.1.3. Basic Definition 164
	A.2.1.4. Finding the Toolpoints for the Basic Set 163
	A.2.1.5. Strong Prior-Data Agreement Property 164
	A.2.1.6. General Update with $\kappa > \frac{1}{2}$ 165
	A.2.4. Discussion and Outlook 168
Bibliography	171

viii	Contents
	A.1.4.2. Update of α^2 140
	A.1.4.3. Update of β 133
	A.1.5. Discussion and Outlook 154
	A.2. A Parameter Set Shape for Strong Prior-Data Agreement Modelling 157
	A.2.1. A Novel Parametrisation of Canonical Conjugate Priors 157
	A.2.1.1. Informal Remarks for Non-Shaped Parameter Sets 159
	A.2.1.2. The Boundshape 163
	A.2.1.3. Basic Definition 164
	A.2.1.4. Finding the Toolpoints for the Basic Set 163
	A.2.1.5. Strong Prior-Data Agreement Property 164
	A.2.1.6. General Update with $\kappa > \frac{1}{2}$ 165
	A.2.4. Discussion and Outlook 168
Bibliography	171

A.2. A Parameter Set Shape for Strong Prior-Data Agreement Modelling 157
A.2.1. A Novel Parametrisation of Canonical Conjugate Priors 157
A.2.1.1. Informal Remarks for Non-Shaped Parameter Sets 159
A.2.1.2. The Boundshape 163
A.2.1.3. Basic Definition 164
A.2.1.4. Finding the Toolpoints for the Basic Set 163
A.2.1.5. Strong Prior-Data Agreement Property 164
A.2.1.6. General Update with $\kappa > \frac{1}{2}$ 165
A.2.4. Discussion and Outlook 168



Bayesian Inference & Prior-Data Conflict

The Bayesian approach to statistical inference

prior $p(\vartheta)$ + likelihood $f(\mathbf{x} | \vartheta)$ \rightarrow posterior $p(\vartheta | \mathbf{x})$

All inferences are based on the posterior (e.g., point estimate, ...)

Bayesian Inference & Prior-Data Conflict

The Bayesian approach to statistical inference

prior $p(\vartheta)$ + likelihood $f(\mathbf{x} | \vartheta)$ \rightarrow posterior $p(\vartheta | \mathbf{x})$

All inferences are based on the posterior (e.g., point estimate, ...)

Assigning a certain prior distribution on ϑ

= defining a conglomerate of probability statements (on ϑ)



Bayesian Inference & Prior-Data Conflict

The Bayesian approach to statistical inference

prior $p(\vartheta)$ + likelihood $f(\mathbf{x} | \vartheta)$ \rightarrow posterior $p(\vartheta | \mathbf{x})$

All inferences are based on the posterior (e.g., point estimate, ...)

Assigning a certain prior distribution on ϑ

= defining a conglomerate of probability statements (on ϑ)

Prior-Data Conflict

- ▶ *informative prior beliefs* and *trusted data* (sampling model correct, no outliers, etc.) are in conflict
- ▶ “[...] the prior [places] its mass primarily on distributions in the sampling model for which the observed data is surprising” (Evans & Moshonov, 2006)
- ▶ there are not enough data to overrule the prior



Prior-Data Conflict: Basic Example

- ▶ Bernoulli observations: 0/1 observations (team wins no/yes)



Prior-Data Conflict: Basic Example

- ▶ Bernoulli observations: 0/1 observations (team wins no/yes)
- ▶ given: a set of observations and strong prior information



Prior-Data Conflict: Basic Example

- ▶ Bernoulli observations: 0/1 observations (team wins no/yes)
- ▶ given: a set of observations and strong prior information
- ▶ we are, e.g., interested in (predictive) probability P that team wins in the next match



Prior-Data Conflict: Basic Example

- ▶ Bernoulli observations: 0/1 observations (team wins no/yes)
- ▶ given: a set of observations and strong prior information
- ▶ we are, e.g., interested in (predictive) probability P that team wins in the next match

Beta-Binomial Model

data :	$s \mid \theta$	\sim	$\text{Binom}(n, \theta)$
conjugate prior:	$\theta \mid n^{(0)}, y^{(0)}$	\sim	$\text{Beta}(n^{(0)}, y^{(0)})$
posterior:	$\theta \mid n^{(n)}, y^{(n)}$	\sim	$\text{Beta}(n^{(n)}, y^{(n)})$

where s = number of wins in the n matches observed



Beta-Binomial Model (BBM)

Beta-Binomial Model

data :	$s \mid \theta$	\sim	$\text{Binom}(n, \theta)$
conjugate prior:	$\theta \mid n^{(0)}, y^{(0)}$	\sim	$\text{Beta}(n^{(0)}, y^{(0)})$
posterior:	$\theta \mid n^{(n)}, y^{(n)}$	\sim	$\text{Beta}(n^{(n)}, y^{(n)})$



Beta-Binomial Model (BBM)

Beta-Binomial Model

data :	$s \mid \theta$	\sim	$\text{Binom}(n, \theta)$
conjugate prior:	$\theta \mid n^{(0)}, y^{(0)}$	\sim	$\text{Beta}(n^{(0)}, y^{(0)})$
posterior:	$\theta \mid n^{(n)}, y^{(n)}$	\sim	$\text{Beta}(n^{(n)}, y^{(n)})$

$$P = E[\theta \mid n^{(n)}, y^{(n)}]$$



Beta-Binomial Model (BBM)

Beta-Binomial Model

data :	$s \mid \theta$	\sim	$\text{Binom}(n, \theta)$
conjugate prior:	$\theta \mid n^{(0)}, y^{(0)}$	\sim	$\text{Beta}(n^{(0)}, y^{(0)})$
posterior:	$\theta \mid n^{(n)}, y^{(n)}$	\sim	$\text{Beta}(n^{(n)}, y^{(n)})$

$$P = E[\theta \mid n^{(n)}, y^{(n)}] = y^{(n)} = \frac{n^{(0)}}{n^{(0)} + n} \cdot y^{(0)} + \frac{n}{n^{(0)} + n} \cdot \frac{s}{n}$$



Beta-Binomial Model (BBM)

Beta-Binomial Model

data :	$s \mid \theta$	\sim	$\text{Binom}(n, \theta)$
conjugate prior:	$\theta \mid n^{(0)}, y^{(0)}$	\sim	$\text{Beta}(n^{(0)}, y^{(0)})$
posterior:	$\theta \mid n^{(n)}, y^{(n)}$	\sim	$\text{Beta}(n^{(n)}, y^{(n)})$

$$P = E[\theta \mid n^{(n)}, y^{(n)}] = y^{(n)} = \frac{n^{(0)}}{n^{(0)} + n} \cdot y^{(0)} + \frac{n}{n^{(0)} + n} \cdot \frac{s}{n}$$

$$n^{(n)} = n^{(0)} + n$$



Beta-Binomial Model (BBM)

Beta-Binomial Model

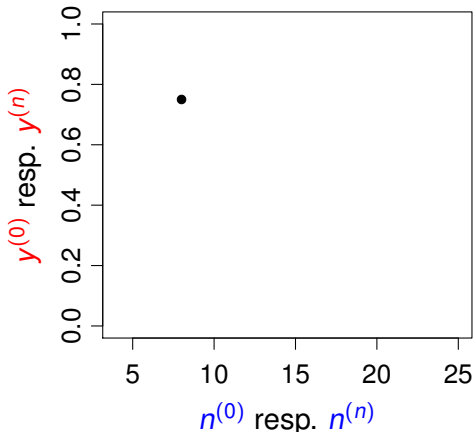
data :	$s \mid \theta$	\sim	$\text{Binom}(n, \theta)$
conjugate prior:	$\theta \mid n^{(0)}, y^{(0)}$	\sim	$\text{Beta}(n^{(0)}, y^{(0)})$
posterior:	$\theta \mid n^{(n)}, y^{(n)}$	\sim	$\text{Beta}(n^{(n)}, y^{(n)})$

$$P = E[\theta \mid n^{(n)}, y^{(n)}] = y^{(n)} = \frac{n^{(0)}}{n^{(0)} + n} \cdot y^{(0)} + \frac{n}{n^{(0)} + n} \cdot \frac{s}{n}$$

$$n^{(n)} = n^{(0)} + n \quad \text{Var}(\theta \mid n^{(n)}, y^{(n)}) = \frac{y^{(n)}(1 - y^{(n)})}{n^{(n)} + 1}$$



Beta-Binomial Model (BBM)



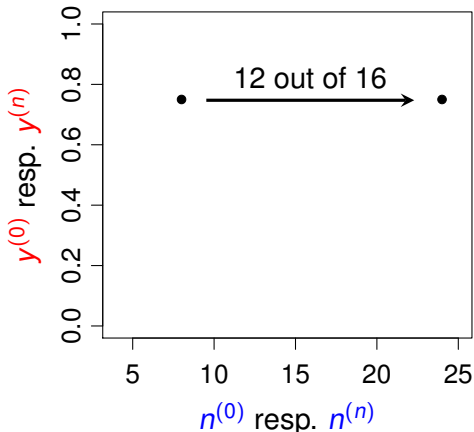
no conflict:

prior $n^{(0)} = 8$, $y^{(0)} = 0.75$

data $s/n = 12/16 = 0.75$



Beta-Binomial Model (BBM)



no conflict:

prior $n^{(0)} = 8$, $y^{(0)} = 0.75$

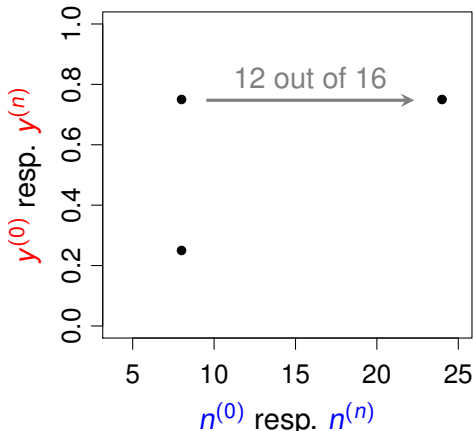
data $s/n = 12/16 = 0.75$



$n^{(n)} = 24$, $y^{(n)} = 0.75$



Beta-Binomial Model (BBM)



no conflict:

prior $n^{(0)} = 8$, $y^{(0)} = 0.75$
data $s/n = 12/16 = 0.75$

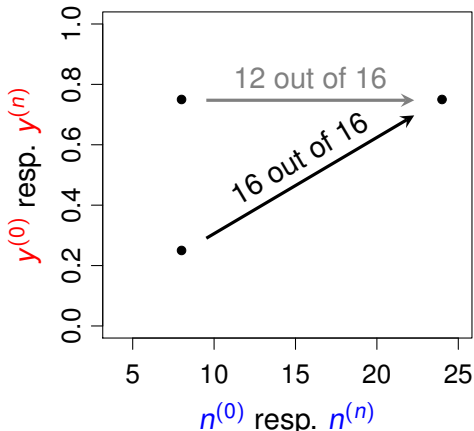
$n^{(n)} = 24$, $y^{(n)} = 0.75$

prior-data conflict:

prior $n^{(0)} = 8$, $y^{(0)} = 0.25$
data $s/n = 16/16 = 1$



Beta-Binomial Model (BBM)



no conflict:

prior $n^{(0)} = 8, y^{(0)} = 0.75$
 data $s/n = 12/16 = 0.75$

$n^{(n)} = 24, y^{(n)} = 0.75$

prior-data conflict:

prior $n^{(0)} = 8, y^{(0)} = 0.25$
 data $s/n = 16/16 = 1$

Canonical Conjugate Priors

Weighted average structure is underneath **all common conjugate priors** for exponential family sampling distributions!

$(x_1, \dots, x_n) = \mathbf{x} \stackrel{iid}{\sim}$ canonical exponential family

$$p(\mathbf{x} \mid \vartheta) \propto \exp \left\{ \langle \psi, \tau(\mathbf{x}) \rangle - nb(\psi) \right\} \quad \left[\psi \text{ transformation of } \vartheta \right]$$

(includes Binomial, Multinomial, Normal, Poisson, Exponential, ...)

Canonical Conjugate Priors

Weighted average structure is underneath **all common conjugate priors** for exponential family sampling distributions!

$(x_1, \dots, x_n) = \mathbf{x} \stackrel{iid}{\sim}$ canonical exponential family

$$p(\mathbf{x} \mid \vartheta) \propto \exp \left\{ \langle \boldsymbol{\psi}, \boldsymbol{\tau}(\mathbf{x}) \rangle - nb(\boldsymbol{\psi}) \right\} \quad \left[\boldsymbol{\psi} \text{ transformation of } \vartheta \right]$$

(includes Binomial, Multinomial, Normal, Poisson, Exponential, ...)

► conjugate prior: $p(\boldsymbol{\psi} \mid n^{(0)}, \mathbf{y}^{(0)}) \propto \exp \left\{ n^{(0)} \left[\langle \boldsymbol{\psi}, \mathbf{y}^{(0)} \rangle - b(\boldsymbol{\psi}) \right] \right\}$



Canonical Conjugate Priors

Weighted average structure is underneath **all common conjugate priors** for exponential family sampling distributions!

$(x_1, \dots, x_n) = \mathbf{x} \stackrel{iid}{\sim}$ canonical exponential family

$$p(\mathbf{x} | \vartheta) \propto \exp \left\{ \langle \psi, \tau(\mathbf{x}) \rangle - nb(\psi) \right\} \quad \left[\psi \text{ transformation of } \vartheta \right]$$

(includes Binomial, Multinomial, Normal, Poisson, Exponential, ...)

- ▶ conjugate prior: $p(\psi | n^{(0)}, y^{(0)}) \propto \exp \left\{ n^{(0)} \left[\langle \psi, y^{(0)} \rangle - b(\psi) \right] \right\}$
- ▶ (conjugate) posterior: $p(\psi | n^{(0)}, y^{(0)}, \mathbf{x}) \propto \exp \left\{ n^{(n)} \left[\langle \psi, y^{(n)} \rangle - b(\psi) \right] \right\}$

where $y^{(n)} = \frac{n^{(0)}}{n^{(0)} + n} \cdot y^{(0)} + \frac{n}{n^{(0)} + n} \cdot \frac{\tau(\mathbf{x})}{n}$ and $n^{(n)} = n^{(0)} + n$



Canonical Conjugate Priors

► (conjugate) posterior: $p(\psi \mid n^{(n)}, \mathbf{y}^{(n)}) \propto \exp \left\{ n^{(n)} \left[\langle \psi, \mathbf{y}^{(n)} \rangle - b(\psi) \right] \right\}$

where $\mathbf{y}^{(n)} = \frac{n^{(0)}}{n^{(0)} + n} \cdot \mathbf{y}^{(0)} + \frac{n}{n^{(0)} + n} \cdot \frac{\tau(\mathbf{x})}{n}$ and $n^{(n)} = n^{(0)} + n$

Canonical Conjugate Priors

► (conjugate) posterior: $p(\psi \mid n^{(n)}, \mathbf{y}^{(n)}) \propto \exp \left\{ n^{(n)} \left[\langle \psi, \mathbf{y}^{(n)} \rangle - b(\psi) \right] \right\}$

where $\mathbf{y}^{(n)} = \frac{n^{(0)}}{n^{(0)} + n} \cdot \mathbf{y}^{(0)} + \frac{n}{n^{(0)} + n} \cdot \frac{\tau(\mathbf{x})}{n}$ and $n^{(n)} = n^{(0)} + n$

Interpretation

- $n^{(0)}$ = determines **spread** and **learning speed**
- $\mathbf{y}^{(0)}$ = **prior expectation** of $\tau(\mathbf{x})/n$



Canonical Conjugate Priors

► (conjugate) posterior: $p(\psi \mid n^{(n)}, \mathbf{y}^{(n)}) \propto \exp \left\{ n^{(n)} \left[\langle \psi, \mathbf{y}^{(n)} \rangle - b(\psi) \right] \right\}$

where $\mathbf{y}^{(n)} = \frac{n^{(0)}}{n^{(0)} + n} \cdot \mathbf{y}^{(0)} + \frac{n}{n^{(0)} + n} \cdot \frac{\tau(\mathbf{x})}{n}$ and $n^{(n)} = n^{(0)} + n$

Interpretation

- $n^{(0)}$ = determines spread and learning speed
- $\mathbf{y}^{(0)}$ = prior expectation of $\tau(\mathbf{x})/n$

Example: Scaled Normal Data

Data :	$\mathbf{x} \mid \mu$	\sim	$N(\mu, 1)$
conjugate prior:	$\mu \mid n^{(0)}, \mathbf{y}^{(0)}$	\sim	$N(\mathbf{y}^{(0)}, 1/n^{(0)})$
posterior:	$\mu \mid n^{(n)}, \mathbf{y}^{(n)}$	\sim	$N(\mathbf{y}^{(n)}, 1/n^{(n)}) \quad \left(\frac{\tau(\mathbf{x})}{n} = \bar{\mathbf{x}} \right)$



Why Generalise Bayesian Inference?

Bayesian theory lacks the ability to specify the degree of uncertainty in probability statements encoded in a (prior, posterior) distribution.

Foundational arguments regarding over-precision of the classical Bayesian framework

Comments on hierarchical modelling



Why Generalise Bayesian Inference?

Bayesian theory lacks the ability to specify the degree of uncertainty in probability statements encoded in a (prior, posterior) distribution.

Variance or stretch of a distribution for describing uncertainty?

Foundational arguments regarding over-precision of the classical Bayesian framework

Comments on hierarchical modelling



Why Generalise Bayesian Inference?

Bayesian theory lacks the ability to specify the degree of uncertainty in probability statements encoded in a (prior, posterior) distribution.

Variance or stretch of a distribution for describing uncertainty?

- ➔ Does not work in the case of prior-data conflict:
In conjugate updating, the posterior variance does not depend on the degree of prior-data conflict in most cases.
- ➔ How to express the precision of a probability statement?

Foundational arguments regarding over-precision of the classical Bayesian framework

Comments on hierarchical modelling



Imprecision

Add **imprecision** as new model dimension:
Sets of priors model uncertainty in probability statements
and allow to better model partial information on ϑ

Imprecision

Add **imprecision** as new model dimension:
Sets of priors model uncertainty in probability statements
and allow to better model partial information on ϑ

Interpretation

smaller sets = more precise probability statements

Lottery A

Number of winning tickets:
exactly known as 5 out of 100

$$\rightarrow P(\text{win}) = 5/100$$

Lottery B

Number of winning tickets:
not exactly known, supposedly
between 1 and 7 out of 100

$$\rightarrow P(\text{win}) = [1/100, 7/100]$$

Bayesian Inference with Sets of Priors

Standard Bayesian inference procedure

prior + likelihood \rightarrow posterior

using Bayes' Rule

All inferences are based on the posterior

Bayesian Inference with Sets of Priors

Standard Bayesian inference procedure

prior + likelihood \rightarrow posterior

using Bayes' Rule

All inferences are based on the posterior

Generalised Bayesian inference procedure

set of priors + likelihood \rightarrow set of posteriors

Coherence (consistency of inferences) ensured by using

Generalised Bayes' Rule (GBR, Walley 1991)

= element-wise application of Bayes' Rule

All inferences are based on the set of posteriors



Bayesian Inference with Sets of Priors

Standard Bayesian inference procedure

prior + likelihood \rightarrow posterior

using Bayes' Rule

All inferences are based on the posterior

Generalised Bayesian inference procedure

set of priors + likelihood \rightarrow set of posteriors

Coherence (consistency of inferences) ensured by using

Generalised Bayes' Rule (GBR, Walley 1991)

= element-wise application of Bayes' Rule

All inferences are based on the set of posteriors

Let hyperparameters $(n^{(0)}, y^{(0)})$ vary in a set \rightarrow set of priors $\mathcal{M}^{(0)}$

Bayesian Inference with Sets of Priors

Set of posteriors $\mathcal{M}^{(n)}$ via $= \left\{ (n^{(n)}, \mathbf{y}^{(n)}) : (n^{(0)}, \mathbf{y}^{(0)}) \in \quad \right\}$

single posterior $p(n^{(n)}, \mathbf{y}^{(n)}) \rightarrow$ set of posteriors $\mathcal{M}^{(n)}$ (via \quad)

$E[\psi \mid n^{(n)}, \mathbf{y}^{(n)}] \rightarrow [\underline{E}[\psi \mid \quad], \bar{E}[\psi \mid \quad]]$

$P(\psi \in A \mid n^{(n)}, \mathbf{y}^{(n)}) \rightarrow [\underline{P}[\psi \in A \mid \quad], \bar{P}[\psi \in A \mid \quad]]$

HPD interval \rightarrow union of HPD intervals

Lower/upper bounds by min/max over set of posteriors



Taking the Convex Hull as the Set of Priors

Convex Set of Priors

$$\mathcal{M}^{(0)} = \text{conv}\left(\left\{p(\psi \mid n^{(0)}, \mathbf{y}^{(0)}) : (n^{(0)}, \mathbf{y}^{(0)}) \in \quad \right\}\right)$$

$\mathcal{M}^{(0)}$ = finite convex mixtures of canonical conjugate priors defined by



Taking the Convex Hull as the Set of Priors

Convex Set of Priors

$$\mathcal{M}^{(0)} = \text{conv}\left(\left\{p(\psi \mid n^{(0)}, \mathbf{y}^{(0)}) : (n^{(0)}, \mathbf{y}^{(0)}) \in \quad \right\}\right)$$

$\mathcal{M}^{(0)}$ = finite convex mixtures of canonical conjugate priors defined by

Updating & mixture commute → set of posteriors can be written as...

Convex Set of Posteriors

$$\mathcal{M}^{(n)} = \text{conv}\left(\left\{p(\psi \mid n^{(n)}, \mathbf{y}^{(n)}) : (n^{(n)}, \mathbf{y}^{(n)}) \in \quad \right\}\right)$$

$\mathcal{M}^{(n)}$ = finite convex mixtures of canonical conjugate posteriors defined by set of updated hyperparameters



Taking the Convex Hull as the Set of Priors

Convex Set of Priors

$$\mathcal{M}^{(0)} = \text{conv}\left(\left\{p(\psi \mid n^{(0)}, \mathbf{y}^{(0)}): (n^{(0)}, \mathbf{y}^{(0)}) \in \quad \right\}\right)$$

$\mathcal{M}^{(0)}$ = finite convex mixtures of canonical conjugate priors defined by

Updating & mixture commute \rightarrow set of posteriors can be written as...

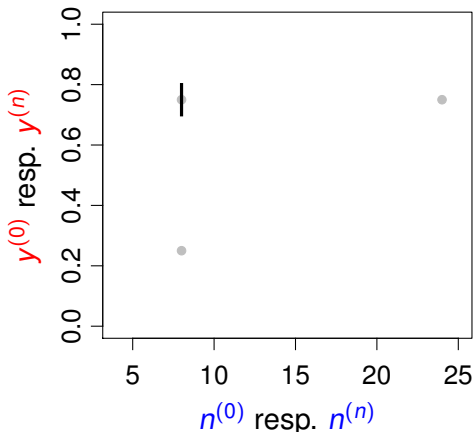
Convex Set of Posteriors

$$\mathcal{M}^{(n)} = \text{conv}\left(\left\{p(\psi \mid n^{(n)}, \mathbf{y}^{(n)}): (n^{(n)}, \mathbf{y}^{(n)}) \in \quad \right\}\right)$$

$\mathcal{M}^{(n)}$ = finite convex mixtures of canonical conjugate posteriors defined by set of updated hyperparameters

Convex sets make the procedure very general (mixture distributions), but are useful only for inferences that are *linear* in the posteriors (expectations: yes, variances: no)

Imprecise BBM with $n^{(0)}$ fixed: IDM (Walley 1996) Quaghebeur & de Cooman (2005)

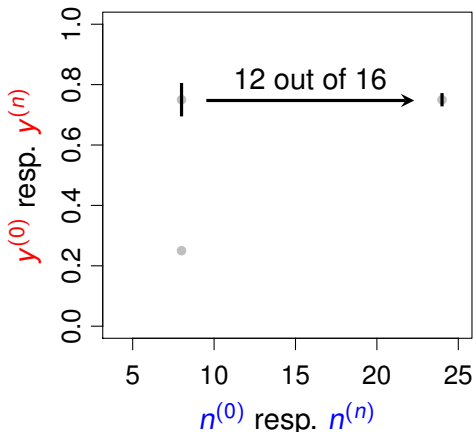


no conflict:

prior $n^{(0)} = 8$, $y^{(0)} \in [0.7, 0.8]$
data $s/n = 12/16 = 0.75$



Imprecise BBM with $n^{(0)}$ fixed: IDM (Walley 1996) Quaghebeur & de Cooman (2005)



no conflict:

prior $n^{(0)} = 8$, $y^{(0)} \in [0.7, 0.8]$
data $s/n = 12/16 = 0.75$

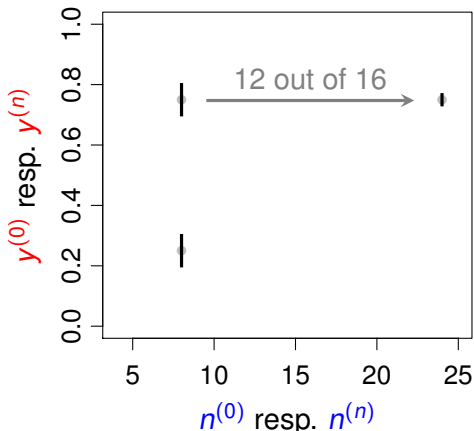
$n^{(n)} = 24$, $y^{(n)} \in [0.73, 0.77]$



Imprecise BBM with $n^{(0)}$ fixed:

IDM (Walley 1996)

Quaghebeur & de Cooman (2005)



no conflict:

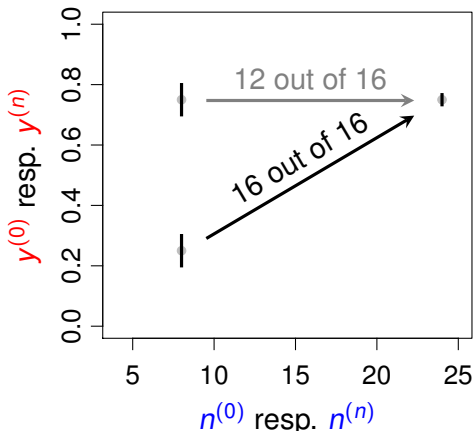
prior $n^{(0)} = 8$, $y^{(0)} \in [0.7, 0.8]$
data $s/n = 12/16 = 0.75$

$n^{(n)} = 24$, $y^{(n)} \in [0.73, 0.77]$

prior-data conflict:

prior $n^{(0)} = 8$, $y^{(0)} \in [0.2, 0.3]$
data $s/n = 16/16 = 1$

Imprecise BBM with $n^{(0)}$ fixed: IDM (Walley 1996) Quaghebeur & de Cooman (2005)



no conflict:

prior $n^{(0)} = 8$, $y^{(0)} \in [0.7, 0.8]$
data $s/n = 12/16 = 0.75$

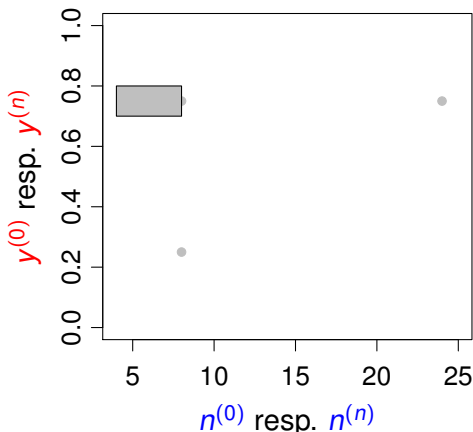
$n^{(n)} = 24$, $y^{(n)} \in [0.73, 0.77]$

prior-data conflict:

prior $n^{(0)} = 8$, $y^{(0)} \in [0.2, 0.3]$
data $s/n = 16/16 = 1$



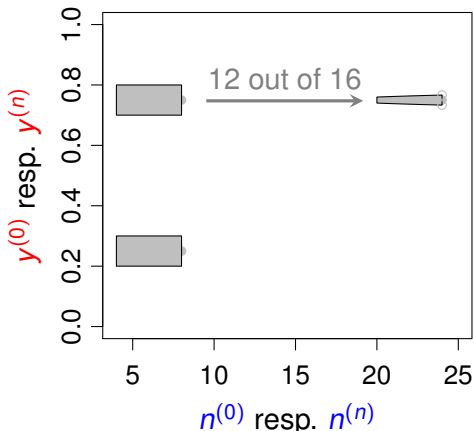
Imprecise BBM with $[\underline{n}^{(0)}, \bar{n}^{(0)}]$: Walley (1991, §5.4.3) Walter & Augustin (2009)



no conflict:

prior $n^{(0)} \in [4, 8]$, $y^{(0)} \in [0.7, 0.8]$
data $s/n = 12/16 = 0.75$

Imprecise BBM with $[\underline{n}^{(0)}, \bar{n}^{(0)}]$: Walley (1991, §5.4.3) : Walter & Augustin (2009)



no conflict:

prior $n^{(0)} \in [4, 8]$, $y^{(0)} \in [0.7, 0.8]$
 data $s/n = 12/16 = 0.75$

$y^{(n)} \in [0.73, 0.77]$

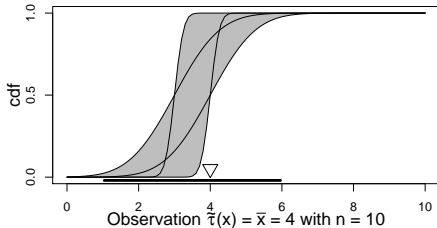
prior-data conflict:

prior $n^{(0)} \in [4, 8]$, $y^{(0)} \in [0.2, 0.3]$
 data $s/n = 16/16 = 1$

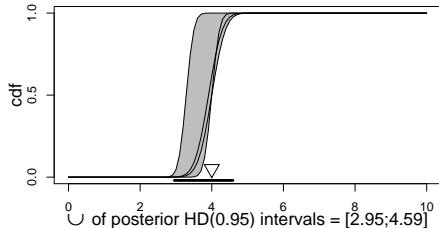


Scaled Normal Data Example ($x \stackrel{iid}{\sim} N(\mu, 1)$)

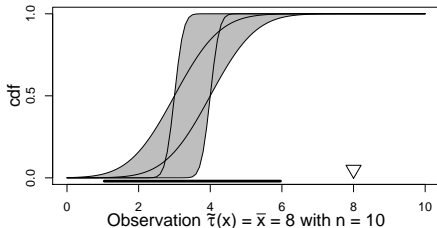
Set of priors: $y^{(0)} \in [3;4]$ and $n^{(0)} \in [1;25]$



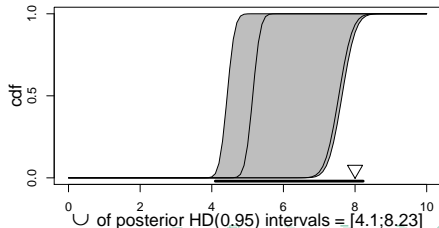
Set of posteriors: $y^{(1)} \in [3.29;4]$ and $n^{(1)} \in [11;35]$



Set of priors: $y^{(0)} \in [3;4]$ and $n^{(0)} \in [1;25]$



Set of posteriors: $y^{(1)} \in [4.43;7.64]$ and $n^{(1)} \in [11;35]$





General Model Properties

Favourable inference properties (cf. other models based on sets of priors)

▶ $n \rightarrow \infty$



General Model Properties

Favourable inference properties (cf. other models based on sets of priors)

▶ $n \rightarrow \infty$ \rightarrow $y^{(n)}$ stretch in $\rightarrow 0$



General Model Properties

Favourable inference properties (cf. other models based on sets of priors)

- ▶ $n \rightarrow \infty$ → $y^{(n)}$ stretch in
- 0 → precise inferences



General Model Properties

Favourable inference properties (cf. other models based on sets of priors)

- ▶ $n \rightarrow \infty \rightarrow y^{(n)}$ stretch in $\rightarrow 0 \rightarrow$ precise inferences
- ▶ larger $n^{(0)} \rightarrow$ larger \rightarrow more vague inferences



General Model Properties

Favourable inference properties (cf. other models based on sets of priors)

- ▶ $n \rightarrow \infty \rightarrow y^{(n)}$ stretch in $\rightarrow 0 \rightarrow$ precise inferences
- ▶ larger $n^{(0)} \rightarrow$ larger \rightarrow more vague inferences
- ▶ larger range of $y^{(0)}$ in \rightarrow larger range of $y^{(n)}$ in \rightarrow more vague inferences



General Model Properties

Favourable inference properties (cf. other models based on sets of priors)

- ▶ $n \rightarrow \infty \rightarrow y^{(n)}$ stretch in $\rightarrow 0 \rightarrow$ precise inferences
- ▶ larger $n^{(0)} \rightarrow$ larger \rightarrow more vague inferences
- ▶ larger range of $y^{(0)}$ in \rightarrow larger range of $y^{(n)}$ in \rightarrow more vague inferences

Model very easy to handle:

- ▶ Hyperparameter set defines set of priors $\mathcal{M}^{(0)}$



General Model Properties

Favourable inference properties (cf. other models based on sets of priors)

- ▶ $n \rightarrow \infty \rightarrow y^{(n)}$ stretch in $\rightarrow 0 \rightarrow$ precise inferences
- ▶ larger $n^{(0)} \rightarrow$ larger \rightarrow more vague inferences
- ▶ larger range of $y^{(0)}$ in \rightarrow larger range of $y^{(n)}$ in \rightarrow more vague inferences

Model very easy to handle:

- ▶ Hyperparameter set defines set of priors $\mathcal{M}^{(0)}$
- ▶ Hyperparameter set defines set of posteriors $\mathcal{M}^{(n)}$

General Model Properties

Favourable inference properties (cf. other models based on sets of priors)

- ▶ $n \rightarrow \infty \rightarrow y^{(n)}$ stretch in $\rightarrow 0 \rightarrow$ precise inferences
- ▶ larger $n^{(0)} \rightarrow$ larger \rightarrow more vague inferences
- ▶ larger range of $y^{(0)}$ in \rightarrow larger range of $y^{(n)}$ in \rightarrow more vague inferences

Model very easy to handle:

- ▶ Hyperparameter set defines set of priors $\mathcal{M}^{(0)}$
- ▶ Hyperparameter set defines set of posteriors $\mathcal{M}^{(n)}$
- ▶ \rightarrow is easy: $n^{(n)} = n^{(0)} + n$, $y^{(n)} = \frac{n^{(0)}}{n^{(0)}+n} y^{(0)} + \frac{n}{n^{(0)}+n} \frac{\tau(\mathbf{x})}{n}$



General Model Properties

Favourable inference properties (cf. other models based on sets of priors)

- ▶ $n \rightarrow \infty \rightarrow y^{(n)}$ stretch in $\rightarrow 0 \rightarrow$ precise inferences
- ▶ larger $n^{(0)} \rightarrow$ larger \rightarrow more vague inferences
- ▶ larger range of $y^{(0)}$ in \rightarrow larger range of $y^{(n)}$ in \rightarrow more vague inferences

Model very easy to handle:

- ▶ Hyperparameter set defines set of priors $\mathcal{M}^{(0)}$
- ▶ Hyperparameter set defines set of posteriors $\mathcal{M}^{(n)}$
- ▶ \rightarrow is easy: $n^{(n)} = n^{(0)} + n$, $y^{(n)} = \frac{n^{(0)}}{n^{(0)}+n} y^{(0)} + \frac{n}{n^{(0)}+n} \frac{\tau(\mathbf{x})}{n}$
- ▶ For quantities linear in posteriors, bounds are attained at “pure” posteriors $p(\psi | n^{(n)}, y^{(n)}) \rightarrow$ optimise over (not over $\mathcal{M}^{(n)}$)



General Model Properties

Favourable inference properties (cf. other models based on sets of priors)

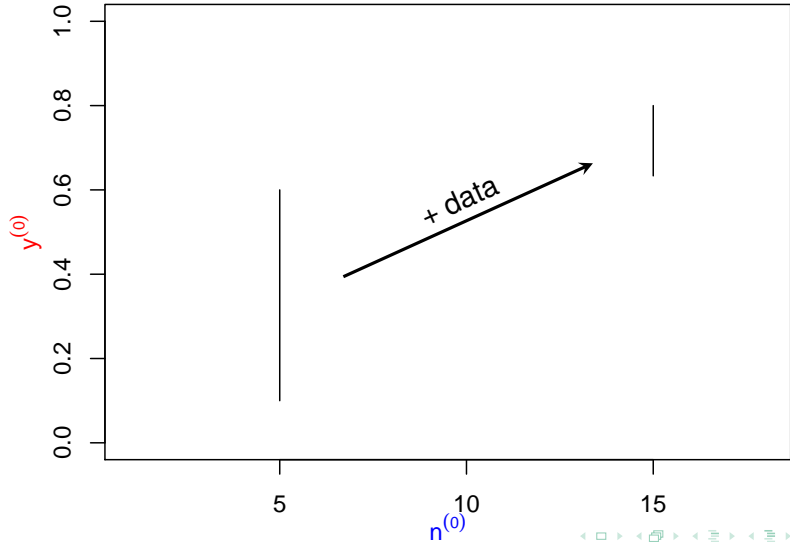
- ▶ $n \rightarrow \infty \rightarrow y^{(n)}$ stretch in $\rightarrow 0 \rightarrow$ precise inferences
- ▶ larger $n^{(0)} \rightarrow$ larger \rightarrow more vague inferences
- ▶ larger range of $y^{(0)}$ in \rightarrow larger range of $y^{(n)}$ in \rightarrow more vague inferences

Model very easy to handle:

- ▶ Hyperparameter set defines set of priors $\mathcal{M}^{(0)}$
- ▶ Hyperparameter set defines set of posteriors $\mathcal{M}^{(n)}$
- ▶ \rightarrow is easy: $n^{(n)} = n^{(0)} + n, y^{(n)} = \frac{n^{(0)}}{n^{(0)}+n} y^{(0)} + \frac{n}{n^{(0)}+n} \frac{\tau(\mathbf{x})}{n}$
- ▶ For quantities linear in posteriors, bounds are attained at “pure” posteriors $p(\psi | n^{(n)}, y^{(n)}) \rightarrow$ optimise over (not over $\mathcal{M}^{(n)}$)
- ▶ Often, optimising over $(n^{(n)}, y^{(n)}) \in$ is also easy:
 closed form solution for $y^{(n)}$ = posterior ‘guess’ for $\frac{\tau(\mathbf{x})}{n}$ (think: $\bar{\mathbf{x}}$)
 given has ‘nice’ shape

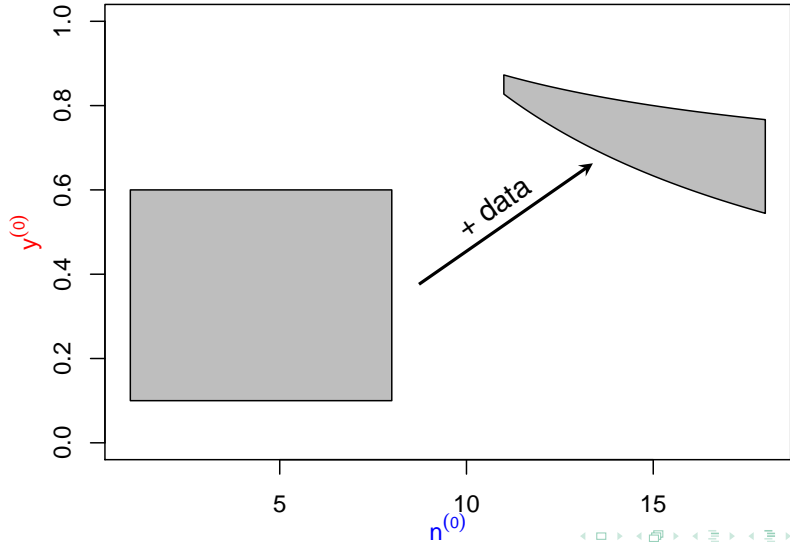


Parameter Set Shapes



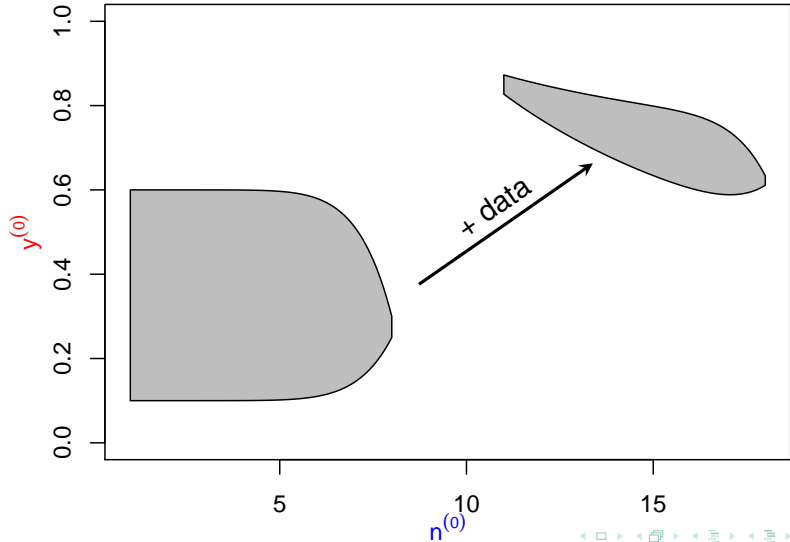


Parameter Set Shapes





Parameter Set Shapes





Parameter Set Shapes

- ▶ Set shape is crucial modeling choice:
trade-off between model complexity and model behaviour



Parameter Set Shapes

- ▶ Set shape is crucial modeling choice:
trade-off between model complexity and model behaviour
- ▶ $= n^{(0)} \times [\underline{y}^{(0)}, \bar{y}^{(0)}]$ (Walley 1996; Quaghebeur & de Cooman 2005):
 $= n^{(n)} \times [\underline{y}^{(n)}, \bar{y}^{(n)}]$ \rightarrow optimise over $[\underline{y}^{(n)}, \bar{y}^{(n)}]$ only,
but no prior-data conflict sensitivity



Parameter Set Shapes

- ▶ Set shape is crucial modeling choice:
 trade-off between model complexity and model behaviour
- ▶ $= n^{(0)} \times [\underline{y}^{(0)}, \bar{y}^{(0)}]$ (Walley 1996; Quaghebeur & de Cooman 2005):
 $= n^{(n)} \times [\underline{y}^{(n)}, \bar{y}^{(n)}]$ \rightarrow optimise over $[\underline{y}^{(n)}, \bar{y}^{(n)}]$ only,
 but no prior-data conflict sensitivity
- ▶ $= [\underline{n}^{(0)}, \bar{n}^{(0)}] \times [\underline{y}^{(0)}, \bar{y}^{(0)}]$ (Walley 1991; Walter & Augustin 2009):
 have non-trivial forms (banana / spotlight), but prior-data
 conflict sensitivity and closed form for min / max $y^{(n)}$ over \dots .
 For other inferences, **R** package `luck` implements optimisation
 over \dots via box-constraint optimisation over

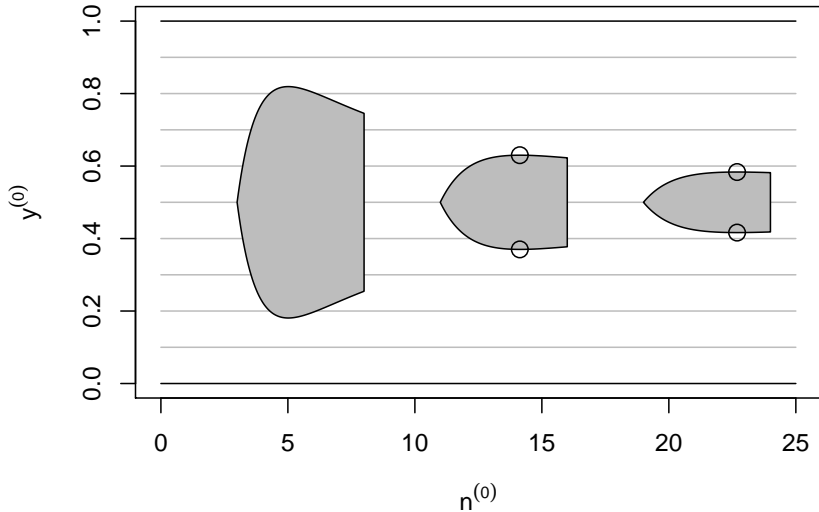


Parameter Set Shapes

- ▶ Set shape is crucial modeling choice:
 trade-off between model complexity and model behaviour
- ▶ $= n^{(0)} \times [\underline{y}^{(0)}, \bar{y}^{(0)}]$ (Walley 1996; Quaghebeur & de Cooman 2005):
 $= n^{(n)} \times [\underline{y}^{(n)}, \bar{y}^{(n)}]$ \rightarrow optimise over $[\underline{y}^{(n)}, \bar{y}^{(n)}]$ only,
 but no prior-data conflict sensitivity
- ▶ $= [\underline{n}^{(0)}, \bar{n}^{(0)}] \times [\underline{y}^{(0)}, \bar{y}^{(0)}]$ (Walley 1991; Walter & Augustin 2009):
 have non-trivial forms (banana / spotlight), but prior-data
 conflict sensitivity and closed form for min / max $y^{(n)}$ over \dots .
 For other inferences, **R** package `luck` implements optimisation
 over \dots via box-constraint optimisation over \dots
- ▶ Other set shapes are possible, but may be more difficult to handle

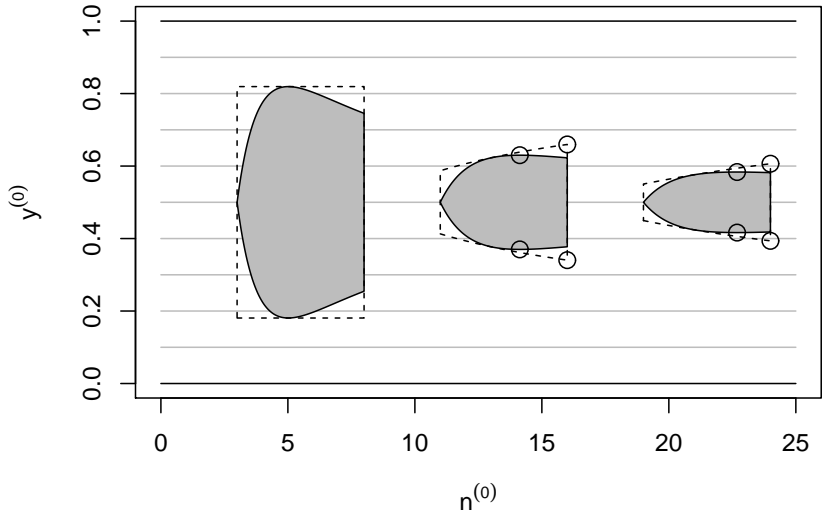


Parameter Set Shape for Strong Prior-Data Agreement





Parameter Set Shape for Strong Prior-Data Agreement



Common-Cause Failures



Source: Wikimedia Commons, http://commons.wikimedia.org/wiki/File:Fukushima_I_by_Digital_Globe.jpg

Common-Cause Failures

common-cause failure

*simultaneous failure of several redundant components
due to a common or shared root cause (Høyland & Rausand, 1994)*

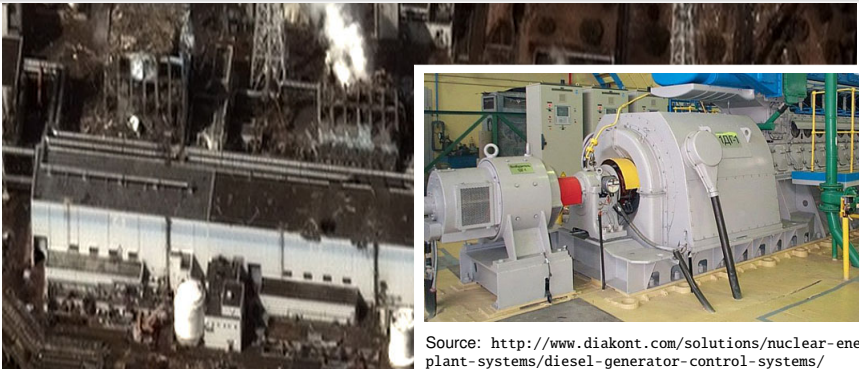


Source: Wikimedia Commons, http://commons.wikimedia.org/wiki/File:Fukushima_I_by_Digital_Globe.jpg

Common-Cause Failures

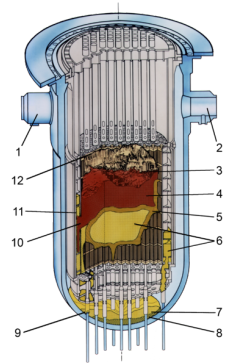
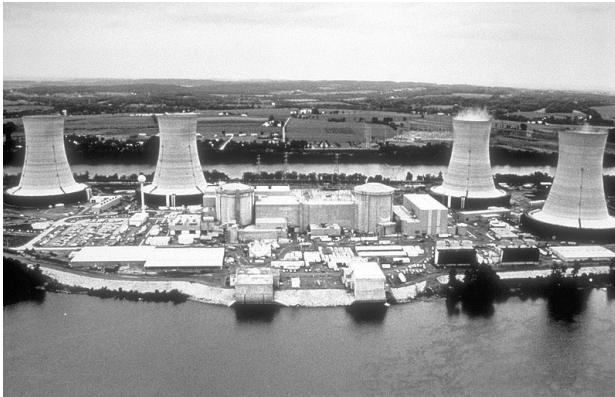
common-cause failure

simultaneous failure of several redundant components due to a common or shared root cause (Høyland & Rausand, 1994)



Source: Wikimedia Commons, http://commons.wikimedia.org/wiki/File:Fukushima_I_by_Digital_Globe.jpg

Common-Cause Failure Modelling



Above: CDC, <http://phil.cdc.gov/phil/ID1194>

Right: Wikimedia Commons,
http://commons.wikimedia.org/wiki/File:Graphic_TMI-2_Core_End-State_Configuration.png



Alpha-Factor Model: Definition

Alpha-Factor Model

Multinomial distribution $M(\mathbf{n} \mid \alpha)$ for common-cause failures in a k -component system

$$p(\mathbf{n} \mid \alpha) = \prod_{j=1}^k \alpha_j^{n_j}$$

where

- ▶ **alpha-factor** α_j := probability of j of the k components failing due to a common cause given that failure occurs
- ▶ **failure count** n_j := corresponding number of failures observed
- ▶ \mathbf{n} denotes (n_1, \dots, n_k) and α denotes $(\alpha_1, \dots, \alpha_k)$

(the model actually serves to estimate failure *rates*, but the above is all what matters in this talk)



Alpha-Factor Model: Parameter Estimation

maximum likelihood estimator:

$$\hat{\alpha}_j = \frac{n_j}{n}, \quad \text{where } \sum_{j=1}^n n_j = n$$



Alpha-Factor Model: Parameter Estimation

maximum likelihood estimator:

$$\hat{\alpha}_j = \frac{n_j}{n}, \quad \text{where } \sum_{j=1}^n n_j = n$$

The Problem

- ▶ typically, for $j \geq 2$, the n_j are very low with zero being quite common for larger j
- ▶ zero counts = flat likelihoods → $\hat{\alpha}_j = ?$

Alpha-Factor Model: Parameter Estimation

maximum likelihood estimator:

$$\hat{\alpha}_j = \frac{n_j}{n}, \quad \text{where } \sum_{j=1}^n n_j = n$$

The Problem

- ▶ typically, for $j \geq 2$, the n_j are very low with zero being quite common for larger j
- ▶ zero counts = flat likelihoods → $\hat{\alpha}_j = ?$

→ need to rely on **epistemic information**: Bayesian inference



Bayesian Inference: Dirichlet Prior

Dirichlet-Multinomial Model

$$p(\alpha \mid n^{(0)}, \mathbf{y}^{(0)}) \propto \prod_{j=1}^k \alpha_j^{n^{(0)} y_j^{(0)} - 1}$$

where $(n^{(0)}, \mathbf{y}^{(0)})$
 are *hyperparameters*

$$n^{(0)} > 0$$

$$\mathbf{y}^{(0)} \in \Delta = \left\{ (y_1^{(0)}, \dots, y_k^{(0)}) : y_1^{(0)} \geq 0, \dots, y_k^{(0)} \geq 0, \sum_{j=1}^k y_j^{(0)} = 1 \right\}$$

$$p(\alpha \mid n^{(0)}, \mathbf{y}^{(0)}, \mathbf{n}) = p(\alpha \mid n^{(n)}, \mathbf{y}^{(n)}) = \prod_{j=1}^k \alpha_j^{n^{(n)} y_j^{(n)} - 1}$$



Bayesian Inference: Dirichlet Prior

Dirichlet-Multinomial Model

$$p(\alpha \mid n^{(0)}, \mathbf{y}^{(0)}) \propto \prod_{j=1}^k \alpha_j^{n^{(0)} y_j^{(0)} - 1}$$

where $(n^{(0)}, \mathbf{y}^{(0)})$
 are *hyperparameters*

$$n^{(0)} > 0$$

$$\mathbf{y}^{(0)} \in \Delta = \left\{ (y_1^{(0)}, \dots, y_k^{(0)}) : y_1^{(0)} \geq 0, \dots, y_k^{(0)} \geq 0, \sum_{j=1}^k y_j^{(0)} = 1 \right\}$$

$$p(\alpha \mid n^{(0)}, \mathbf{y}^{(0)}, \mathbf{n}) = p(\alpha \mid n^{(n)}, \mathbf{y}^{(n)}) = \prod_{j=1}^k \alpha_j^{n^{(n)} y_j^{(n)} - 1}$$

Interpretation

- ▶ $\mathbf{y}^{(0)}$ = **prior expectation of α** , i.e., a prior guess for $\frac{n_j}{n}$, $j = 1, \dots, n$
- ▶ $n^{(0)}$ = determines **spread and learning speed**

Bayesian Inference: Example

Example (Kelly & Atwood, 2011)

Consider a system with four redundant components ($k = 4$).
The analyst specifies the following prior expectation $\mu_{\text{spec},j}$ for each α_j :

$$\mu_{\text{spec},1} = 0.950 \quad \mu_{\text{spec},2} = 0.030 \quad \mu_{\text{spec},3} = 0.015 \quad \mu_{\text{spec},4} = 0.005$$

We have 36 observations, in which 35 showed one component failing,
and 1 showed two components failing:

$$n_1 = 35$$

$$n_2 = 1$$

$$n_3 = 0$$

$$n_4 = 0$$



Bayesian Inference: Example

Example (Kelly & Atwood, 2011)

Consider a system with four redundant components ($k = 4$).
The analyst specifies the following prior expectation $\mu_{\text{spec},j}$ for each α_j :

$$\mu_{\text{spec},1} = 0.950 \quad \mu_{\text{spec},2} = 0.030 \quad \mu_{\text{spec},3} = 0.015 \quad \mu_{\text{spec},4} = 0.005$$

We have 36 observations, in which 35 showed one component failing,
and 1 showed two components failing:

$$n_1 = 35 \quad n_2 = 1 \quad n_3 = 0 \quad n_4 = 0$$

we focus on $E[\alpha_j | n^{(n)}, \mathbf{y}^{(n)}] = y_j^{(n)} = \frac{n^{(0)}}{n^{(0)} + n} \cdot y_j^{(0)} + \frac{n}{n^{(0)} + n} \cdot \frac{n_j}{n}$

(in a decision context, this expectation would typically end up
in expressions for expected utility)



Example: Non-Informative Priors

large variation in posterior under different non-informative priors

- ▶ with constrained maximum entropy prior
(Atwood, 1996; Kelly & Atwood, 2011):

$$E[\alpha_1 | n^{(n)}, \mathbf{y}^{(n)}] = 0.967$$

$$E[\alpha_2 | n^{(n)}, \mathbf{y}^{(n)}] = 0.028$$

$$E[\alpha_3 | n^{(n)}, \mathbf{y}^{(n)}] = 0.003$$

$$E[\alpha_4 | n^{(n)}, \mathbf{y}^{(n)}] = 0.001$$

Example: Non-Informative Priors

large variation in posterior under different non-informative priors

- ▶ with constrained maximum entropy prior
(Atwood, 1996; Kelly & Atwood, 2011):

$$E[\alpha_1 | n^{(n)}, \mathbf{y}^{(n)}] = 0.967$$

$$E[\alpha_2 | n^{(n)}, \mathbf{y}^{(n)}] = 0.028$$

$$E[\alpha_3 | n^{(n)}, \mathbf{y}^{(n)}] = 0.003$$

$$E[\alpha_4 | n^{(n)}, \mathbf{y}^{(n)}] = 0.001$$

- ▶ with uniform prior ($y_j^{(0)} = 0.25$ and $n^{(0)} = 4$):

$$E[\alpha_1 | n^{(n)}, \mathbf{y}^{(n)}] = 0.9$$

$$E[\alpha_2 | n^{(n)}, \mathbf{y}^{(n)}] = 0.05$$

$$E[\alpha_3 | n^{(n)}, \mathbf{y}^{(n)}] = 0.025$$

$$E[\alpha_4 | n^{(n)}, \mathbf{y}^{(n)}] = 0.025$$



Example: Non-Informative Priors

large variation in posterior under different non-informative priors

- ▶ with constrained maximum entropy prior (Atwood, 1996; Kelly & Atwood, 2011):

$$E[\alpha_1 | n^{(n)}, \mathbf{y}^{(n)}] = 0.967$$

$$E[\alpha_2 | n^{(n)}, \mathbf{y}^{(n)}] = 0.028$$

$$E[\alpha_3 | n^{(n)}, \mathbf{y}^{(n)}] = 0.003$$

$$E[\alpha_4 | n^{(n)}, \mathbf{y}^{(n)}] = 0.001$$

- ▶ with uniform prior ($y_j^{(0)} = 0.25$ and $n^{(0)} = 4$):

$$E[\alpha_1 | n^{(n)}, \mathbf{y}^{(n)}] = 0.9$$

$$E[\alpha_2 | n^{(n)}, \mathbf{y}^{(n)}] = 0.05$$

$$E[\alpha_3 | n^{(n)}, \mathbf{y}^{(n)}] = 0.025$$

$$E[\alpha_4 | n^{(n)}, \mathbf{y}^{(n)}] = 0.025$$

- ▶ with Jeffrey's prior ($y_j^{(0)} = 0.25$ and $n^{(0)} = 2$):

$$E[\alpha_1 | n^{(n)}, \mathbf{y}^{(n)}] = 0.9342$$

$$E[\alpha_2 | n^{(n)}, \mathbf{y}^{(n)}] = 0.0395$$

$$E[\alpha_3 | n^{(n)}, \mathbf{y}^{(n)}] = 0.0132$$

$$E[\alpha_4 | n^{(n)}, \mathbf{y}^{(n)}] = 0.0132$$

Imprecise Dirichlet Model: Definition

Troffaes, Walter & Kelly (2013): model vague prior info more cautiously

Imprecise Dirichlet Model (IDM) for Common-Cause Failure

use a **set of hyperparameters** (Walley 1991, 1996):

$$= \left\{ (n^{(0)}, \mathbf{y}^{(0)}) : n^{(0)} \in [\underline{n}^{(0)}, \bar{n}^{(0)}], \mathbf{y}^{(0)} \in \Delta, y_j^{(0)} \in [\underline{y}_j^{(0)}, \bar{y}_j^{(0)}] \right\}$$

Imprecise Dirichlet Model: Definition

Troffaes, Walter & Kelly (2013): model vague prior info more cautiously

Imprecise Dirichlet Model (IDM) for Common-Cause Failure

use a **set of hyperparameters** (Walley 1991, 1996):

$$= \left\{ (n^{(0)}, \mathbf{y}^{(0)}) : n^{(0)} \in [\underline{n}^{(0)}, \bar{n}^{(0)}], \mathbf{y}^{(0)} \in \Delta, y_j^{(0)} \in [\underline{y}_j^{(0)}, \bar{y}_j^{(0)}] \right\}$$

Interpretation

- ▶ we are doing a **sensitivity analysis** (à la robust Bayes) over $(n^{(0)}, \mathbf{y}^{(0)}) \in$
- ▶ we take a **set of priors** based on as model for prior information



Imprecise Dirichlet Model: Definition

Troffaes, Walter & Kelly (2013): model vague prior info more cautiously

Imprecise Dirichlet Model (IDM) for Common-Cause Failure

use a **set of hyperparameters** (Walley 1991, 1996):

$$= \left\{ (n^{(0)}, \mathbf{y}^{(0)}) : n^{(0)} \in [\underline{n}^{(0)}, \bar{n}^{(0)}], \mathbf{y}^{(0)} \in \Delta, y_j^{(0)} \in [\underline{y}_j^{(0)}, \bar{y}_j^{(0)}] \right\}$$

Interpretation

- ▶ we are doing a **sensitivity analysis** (à la robust Bayes) over $(n^{(0)}, \mathbf{y}^{(0)}) \in$
- ▶ we take a **set of priors** based on as model for prior information

Analyst has to specify ('elicit')

bounds $[\underline{n}^{(0)}, \bar{n}^{(0)}]$ and bounds $[\underline{y}_j^{(0)}, \bar{y}_j^{(0)}]$ for each $j \in \{1, \dots, k\}$

Imprecise Dirichlet Model: Elicitation

- ▶ $[\underline{y}_j^{(0)}, \bar{y}_j^{(0)}]$: Cautious interpretation of prior specifications $\mu_{\text{spec},j}$:

$$[\underline{y}_1^{(0)}, \bar{y}_1^{(0)}] = [0.950, 1]$$

$$[\underline{y}_2^{(0)}, \bar{y}_2^{(0)}] = [0, 0.030]$$

$$[\underline{y}_3^{(0)}, \bar{y}_3^{(0)}] = [0, 0.015]$$

$$[\underline{y}_4^{(0)}, \bar{y}_4^{(0)}] = [0, 0.005]$$

Imprecise Dirichlet Model: Elicitation

- ▶ $[\underline{y}_j^{(0)}, \bar{y}_j^{(0)}]$: Cautious interpretation of prior specifications $\mu_{\text{spec},j}$:

$$[\underline{y}_1^{(0)}, \bar{y}_1^{(0)}] = [0.950, 1]$$

$$[\underline{y}_2^{(0)}, \bar{y}_2^{(0)}] = [0, 0.030]$$

$$[\underline{y}_3^{(0)}, \bar{y}_3^{(0)}] = [0, 0.015]$$

$$[\underline{y}_4^{(0)}, \bar{y}_4^{(0)}] = [0, 0.005]$$

- ▶ $[\underline{n}^{(0)}, \bar{n}^{(0)}]$: Good (1965):

reason about posterior expectations for hypothetical data



Imprecise Dirichlet Model: Elicitation

- ▶ $[\underline{y}_j^{(0)}, \bar{y}_j^{(0)}]$: Cautious interpretation of prior specifications $\mu_{\text{spec},j}$:

$$[\underline{y}_1^{(0)}, \bar{y}_1^{(0)}] = [0.950, 1]$$

$$[\underline{y}_2^{(0)}, \bar{y}_2^{(0)}] = [0, 0.030]$$

$$[\underline{y}_3^{(0)}, \bar{y}_3^{(0)}] = [0, 0.015]$$

$$[\underline{y}_4^{(0)}, \bar{y}_4^{(0)}] = [0, 0.005]$$

- ▶ $[\underline{n}^{(0)}, \bar{n}^{(0)}]$: Good (1965):

reason about posterior expectations for hypothetical data

$\bar{n}^{(0)}$ = number of one-component failures required
 to reduce the upper probabilities of multi-component failure by half

$\underline{n}^{(0)}$ = number of multi-component failures required
 to reduce the lower probability of one-component failure by half



Imprecise Dirichlet Model: Elicitation

$\bar{n}^{(0)}$ = number of one-component failures required
to reduce the upper probabilities of multi-component failure by half

$\underline{n}^{(0)}$ = number of multi-component failures required
to reduce the lower probability of one-component failure by half

Imprecise Dirichlet Model: Elicitation

$\bar{n}^{(0)}$ = number of one-component failures required
to reduce the upper probabilities of multi-component failure by half

$\underline{n}^{(0)}$ = number of multi-component failures required
to reduce the lower probability of one-component failure by half

Reasonable values in example:

- ▶ $\bar{n}^{(0)} = 10$: after observing 10 one-component failures
➔ halve upper probabilities of multi-component failures
- ▶ $\underline{n}^{(0)} = 1$: immediate multi-component failure
➔ keen to reduce lower probability for one-component failure



Imprecise Dirichlet Model: Elicitation

$\bar{n}^{(0)}$ = number of one-component failures required
to reduce the upper probabilities of multi-component failure by half

$\underline{n}^{(0)}$ = number of multi-component failures required
to reduce the lower probability of one-component failure by half

Reasonable values in example:

- ▶ $\bar{n}^{(0)} = 10$: after observing 10 one-component failures
→ halve upper probabilities of multi-component failures
- ▶ $\underline{n}^{(0)} = 1$: immediate multi-component failure
→ keen to reduce lower probability for one-component failure

Difference between $\underline{n}^{(0)}$ and $\bar{n}^{(0)}$ reflects a **level of caution**:

The rate at which we reduce upper probabilities
is less than the rate at which we reduce lower probabilities

Imprecise Dirichlet Model: Inference

With $[\underline{n}^{(0)}, \bar{n}^{(0)}] = [1, 10]$, we get...

prior bounds + data \rightarrow posterior bounds

j	$\underline{y}_j^{(0)}$	$\bar{y}_j^{(0)}$	n_j	$\underline{E}[\alpha_j \cdot]$	$\bar{E}[\alpha_j \cdot]$
1	0.950	1	35	0.967	0.978
2	0	0.030	1	0.0270	0.0283
3	0	0.015	0	0	0.00326
4	0	0.005	0	0	0.00109

Imprecise Dirichlet Model: Inference

With $[\underline{n}^{(0)}, \bar{n}^{(0)}] = [1, 10]$, we get...

prior bounds + data \rightarrow posterior bounds

j	$\underline{y}_j^{(0)}$	$\bar{y}_j^{(0)}$	n_j	$\underline{E}[\alpha_j \cdot]$	$\bar{E}[\alpha_j \cdot]$
1	0.950	1	35	0.967	0.978
2	0	0.030	1	0.0270	0.0283
3	0	0.015	0	0	0.00326
4	0	0.005	0	0	0.00109

- **Bounds**, rather than precise values, are desirable due to inferences being strongly sensitive to the prior particularly when faced with zero counts

Imprecise Dirichlet Model: Inference

With $[\underline{n}^{(0)}, \bar{n}^{(0)}] = [1, 10]$, we get...

prior bounds + data \rightarrow posterior bounds

j	$\underline{y}_j^{(0)}$	$\bar{y}_j^{(0)}$	n_j	$\underline{E}[\alpha_j \]$	$\bar{E}[\alpha_j \]$
1	0.950	1	35	0.967	0.978
2	0	0.030	1	0.0270	0.0283
3	0	0.015	0	0	0.00326
4	0	0.005	0	0	0.00109

- ▶ **Bounds**, rather than precise values, are desirable due to inferences being strongly sensitive to the prior particularly when faced with zero counts
- ▶ Simple ways to elicit the parameters of the model by **reasoning on hypothetical data**

Conclusion

- ▶ Conjugate priors are a convenient tool for Bayesian inference but there are some pitfalls
 - ▶ Hyperparameters $n^{(0)}$, $y^{(0)}$ are easy to interpret and elicit
 - ▶ Averaging property makes calculations simple, but leads to inadequate model behaviour in case of prior-data conflict



Conclusion


- ▶ Conjugate priors are a convenient tool for Bayesian inference but there are some pitfalls
 - ▶ Hyperparameters $n^{(0)}, y^{(0)}$ are easy to interpret and elicit
 - ▶ Averaging property makes calculations simple, but leads to inadequate model behaviour in case of prior-data conflict
- ▶ Sets of conjugate priors maintain advantages & mitigate issues
 - ▶ Hyperparameter set shape is important
 - ▶ Reasonable choice: *rectangular* $= [n^{(0)}, \bar{n}^{(0)}] \times [y^{(0)}, \bar{y}^{(0)}]$
 (Walter & Augustin 2009: *generalised iLUCK-models*, luck)
 - ▶ Bounds for hyperparameters are easy to interpret and elicit
 - ▶ Additional imprecision in case of prior-data conflict leads to **cautious inferences if, and only if, caution is needed**
 - ▶ Shape for more precision in case of strong prior-data agreement is in development (joint work with Frank Coolen and Miķ Bickis)



Works the Thesis is Based on

- Augustin, Thomas, Gero Walter, and Frank Coolen (2013). “Statistical Inference”. In: *Introduction to Imprecise Probabilities*. Ed. by Frank Coolen et al. In preparation. Wiley.
- Troffaes, Matthias, Gero Walter, and Dana Kelly (2013). “A Robust Bayesian Approach to Modelling Epistemic Uncertainty in Common-Cause Failure Models”. In: *Reliability Engineering & System Safety*. doi: 10.1016/j.res.2013.05.022.
- Walter, Gero (2012). *A Technical Note on the Dirichlet-Multinomial Model — The Dirichlet Distribution as the Canonically Constructed Conjugate Prior*. Tech. rep. 131. Department of Statistics, LMU Munich. URL: <http://epub.ub.uni-muenchen.de/14068/>.
- Walter, Gero and Thomas Augustin (2009). “Imprecision and Prior-data Conflict in Generalized Bayesian Inference”. In: *Journal of Statistical Theory and Practice* 3, pp. 255–271.
- Walter, Gero and Thomas Augustin (2010). “Bayesian linear regression — different conjugate models and their (in)sensitivity to prior-data conflict”. In: *Statistical Modelling and Regression Structures*. Ed. by Thomas Kneib and Gerhard Tutz. Festschrift in Honour of Ludwig Fahrmeir. Springer, pp. 59–78. Extended version published as Tech. Rep. 69, Department of Statistics, LMU Munich. URL: <http://epub.ub.uni-muenchen.de/11050/1/tr069.pdf>.
- Walter, Gero, Thomas Augustin, and Frank P.A. Coolen (2011). “On Prior-Data Conflict in Predictive Bernoulli Inferences”. In: *ISIPTA’11: Proceedings of the Seventh International Symposium on Imprecise Probabilities: Theories and Applications*. Ed. by Frank Coolen et al. SIPTA, pp. 391–400. URL: <http://www.sipta.org/isipta11/proceedings/046.html>.
- Walter, Gero and Norbert Krautenbacher (2013). *lucK: R package for Generalized iLUCK-models*. URL: <http://lucK.r-forge.r-project.org/>.

Further References

- Atwood, C. (1996). “Constrained noninformative priors in risk assessment”. In: *Reliability Engineering and System Safety* 53, pp. 37–46.
- Benavoli, A. and M. Zaffalon (2012). “A model of prior ignorance for inferences in the one-parameter exponential family”. In: *Journal of Statistical Planning and Inference* 142, pp. 1960–1979. doi: 10.1016/j.jspi.2012.01.023.
- Berger, J. et al. (1994). “An overview of robust Bayesian analysis”. In: *TEST* 3, pp. 5–124.
- Evans, M. and H. Moshonov (2006). “Checking for Prior-Data Conflict”. In: *Bayesian Analysis* 1, pp. 893–914.
- Good, I. (1965). *The estimation of probabilities*. Cambridge (MA): MIT Press.
- Høyland, Arnljot and Marvin Rausand (1994). *System reliability theory: models and statistical methods*. A Wiley interscience publication. New York, NY: Wiley. ISBN: 0-471-59397-4.
- Kelly, Dana and Corwin Atwood (2011). “Finding a minimally informative Dirichlet prior distribution using least squares”. In: *Reliability Engineering and System Safety* 96.3, pp. 398–402. ISSN: 0951-8320. doi: 10.1016/j.res.2010.11.008.
- Quaeghebeur, E. and G. de Cooman (2005). “Imprecise probability models for inference in exponential families”. In: *ISIPTA '05. Proceedings of the Fourth International Symposium on Imprecise Probabilities and Their Applications*. Ed. by F. Cozman, R. Nau, and T. Seidenfeld. Manno: SIPTA, pp. 287–296.
- Walley, Peter (1991). *Statistical Reasoning with Imprecise Probabilities*. London: Chapman and Hall. ISBN: 0-412-28660-2.
- Walley, Peter (1996). “Inferences from multinomial data: Learning about a bag of marbles”. In: *Journal of the Royal Statistical Society, Series B* 58.1, pp. 3–34. 

Arguments for Imprecise Bayesian Inference ◀

- ▶ The formal arguments for the Bayesian approach as a coherent way of inferential reasoning hold only if one can make “*arbitrarily fine discriminations in judgement about unknowns and utilities*” (Berger et al. 1994, p. 303). (Your subjective prior must *exactly* express your preferences, anticipating *anything* that may happen.)
- ▶ Walley’s (1991) framework for coherent inference using imprecise probabilities instead allows for incomplete and imprecise prior specifications.
 - ▶ more realistic description of uncertainties that are often hidden by spuriously precise models (through ‘arbitrary’ modeling decisions)
 - ▶ ‘overprecision’ is often compensated by ‘taking models not too seriously’ (“all models are wrong, but some are useful”)
 - ▶ why not use sensible, reliable models in the first place?

Hierarchical Modelling ◀

- ▶ Usually: space of parametric priors indexed by parameter ϕ .
- ▶ Very useful if ϕ has a clear interpretation (e.g., as a global mean), such that a prior on ϕ can be meaningfully elicited.
- ▶ Noninformative priors on ϕ can be problematic (incoherence, improper posteriors).

Example (e.g., Walley 1991, p. 232)

$$x_i \stackrel{iid}{\sim} N(\mu_i, 1), \mu_i \stackrel{iid}{\sim} N(\mu, \sigma^2), i = 1, \dots, n.$$

For $p(\mu, \sigma^2) \propto \sigma^{-1}$, the posterior is improper.

- ▶ High-variance priors on ϕ do not express ignorance about ϕ , but a strong prior belief that $|\phi|$ is large (Walley 1991, p. 233).



Updating and Mixture Commute ◀

Let

$$p_m(\vartheta \mid n_1^{(0)}, y_1^{(0)}, n_2^{(0)}, y_2^{(0)}, \kappa) := \kappa p(\vartheta \mid n_1^{(0)}, y_1^{(0)}) + (1 - \kappa) p(\vartheta \mid n_2^{(0)}, y_2^{(0)}),$$

with marginals

$$f_1(\mathbf{x}) = \int_{\Theta} f(\mathbf{x} \mid \vartheta) p(\vartheta \mid n_1^{(0)}, y_1^{(0)}) d\vartheta,$$

$$f_2(\mathbf{x}) = \int_{\Theta} f(\mathbf{x} \mid \vartheta) p(\vartheta \mid n_2^{(0)}, y_2^{(0)}) d\vartheta,$$

$$\begin{aligned} f_m(\mathbf{x}) &= \int_{\Theta} f(\mathbf{x} \mid \vartheta) p_m(\vartheta \mid n_1^{(0)}, y_1^{(0)}, n_2^{(0)}, y_2^{(0)}, \kappa) d\vartheta \\ &= \kappa \int_{\Theta} f(\mathbf{x} \mid \vartheta) p(\vartheta \mid n_1^{(0)}, y_1^{(0)}) d\vartheta + (1 - \kappa) \int_{\Theta} f(\mathbf{x} \mid \vartheta) p(\vartheta \mid n_2^{(0)}, y_2^{(0)}) d\vartheta \\ &= \kappa f_1(\mathbf{x}) + (1 - \kappa) f_2(\mathbf{x}). \end{aligned}$$



Updating and Mixture Commute ◀

$$p_m(\vartheta \mid n_1^{(0)}, y_1^{(0)}, n_2^{(0)}, y_2^{(0)}, \kappa, \mathbf{x})$$

$$\begin{aligned} &= \frac{f(\mathbf{x} \mid \vartheta)}{f_m(\mathbf{x})} \left(\kappa p(\vartheta \mid n_1^{(0)}, y_1^{(0)}) + (1 - \kappa) p(\vartheta \mid n_2^{(0)}, y_2^{(0)}) \right) \\ &= \kappa \frac{f(\mathbf{x} \mid \vartheta) p(\vartheta \mid n_1^{(0)}, y_1^{(0)})}{f_m(\mathbf{x})} + (1 - \kappa) \frac{f(\mathbf{x} \mid \vartheta) p(\vartheta \mid n_2^{(0)}, y_2^{(0)})}{f_m(\mathbf{x})} \\ &= \kappa \frac{f_1(\mathbf{x}) p(\vartheta \mid n_1^{(n)}, y_1^{(n)})}{f_m(\mathbf{x})} + (1 - \kappa) \frac{f_2(\mathbf{x}) p(\vartheta \mid n_2^{(n)}, y_2^{(n)})}{f_m(\mathbf{x})} \\ &= p_m(\vartheta \mid n_1^{(n)}, y_1^{(n)}, n_2^{(n)}, y_2^{(n)}, \kappa^*), \end{aligned}$$

where
$$\kappa^* = \kappa \frac{f_1(\mathbf{x})}{f_m(\mathbf{x})} = \frac{\kappa f_1(\mathbf{x})}{\kappa f_1(\mathbf{x}) + (1 - \kappa) f_2(\mathbf{x})}.$$

Updating and Mixture Commute

- ▶ updated mixture distribution is a mixture of the updated components with mixture parameter κ^* instead of κ .
- ▶ convex hull of prior components
= set of prior mixture distributions with $\kappa \in [0, 1]$.
- ▶ for any $\kappa \in [0, 1]$, the corresponding $\kappa^* \in [0, 1]$.
- ▶ in fact, $\{\kappa^* \mid \kappa \in [0, 1]\} = [0, 1]$.
- ▶ set of updated mixture distributions with $\kappa \in [0, 1]$
= convex hull of updated components ($\kappa^* \in [0, 1]$).
- ▶ arbitrary number of components by complete induction.

Other Models Based on Sets of Priors ◀ 3 ◀ 10

- ▶ Neighbourhood models
 - ▶ set of distributions 'close to' a central distribution P_0
 - ▶ common in robust Bayesian approaches
 - ▶ example: ε -contamination class: $\{P : P = (1 - \varepsilon) P_0 + \varepsilon Q, Q \in \mathcal{Q}\}$
 - ▶ not necessarily closed under Bayesian updating
- ▶ Density ratio class / interval of measures
 - ▶ set of distributions by bounds for the density function $p(\vartheta)$:

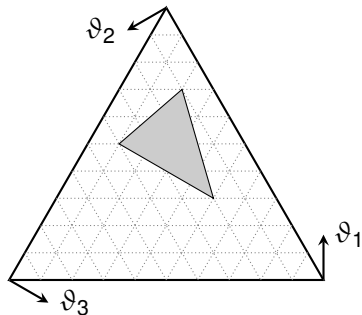
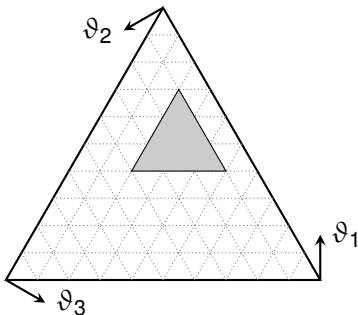
$$\mathcal{M}_{l,u} = \{p(\vartheta) : \exists c \in \mathbb{R}_{>0} : l(\vartheta) \leq cp(\vartheta) \leq u(\vartheta)\}$$

- ▶ posterior set is bounded by updated $l(\vartheta)$ and $u(\vartheta)$
- ▶ $u(\vartheta)/l(\vartheta)$ is constant under updating
 - ➔ size of the set does not decrease with n
 - ➔ too vague posterior inferences

Other Models Based on Sets of Priors ◀ 3 ◀ 10

► Discrete models

- discretize the parameter space as $\Theta = \{\theta_j\}_{j \in \{1, \dots, m\}}$
- set of distributions by bounds for $p(\vartheta_j)$ (or for expectations)
- posterior bounds determined by linear programming algorithm
- more flexibility at the cost of computational complexity
- no clear measure for weight of prior information as compared to n





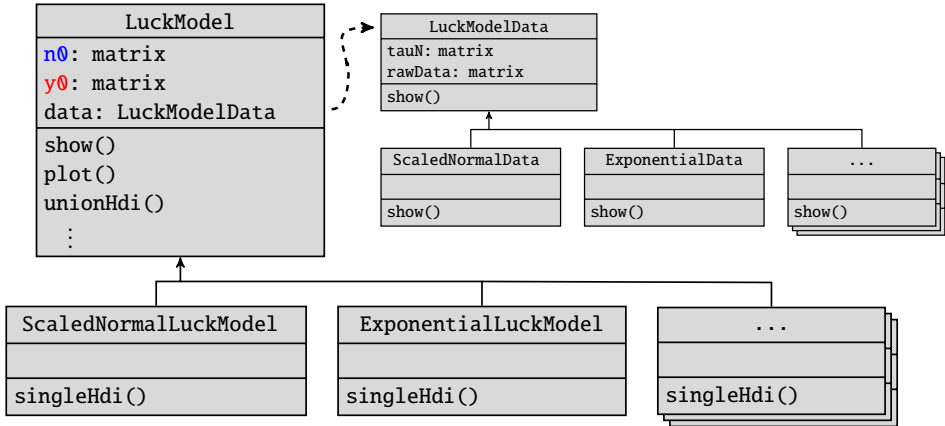
R package luck ◀

- ▶ S4 implementation of the general canonical prior parameter structure with rectangular sets $= [\underline{n}^{(0)}, \bar{n}^{(0)}] \times [\underline{y}^{(0)}, \bar{y}^{(0)}]$
- ▶ lean subclasses for concrete sample distributions (currently implemented: scaled normal, exponential)
- ▶ currently on R-Forge, to be submitted to CRAN

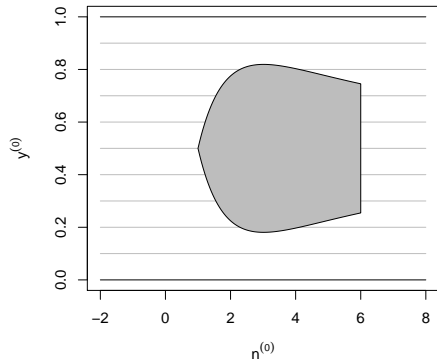
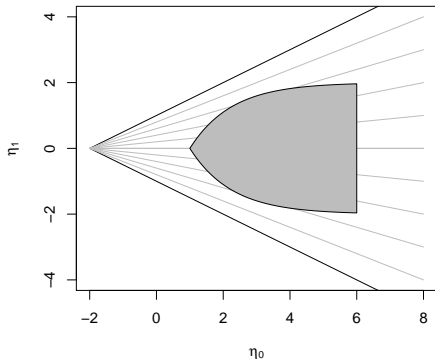
```
install.packages("luck", repos="http://R-Forge.R-project.org")
```



R package luck ◀



Strong Prior-Data Agreement Modelling ◀





'Noninformative' Priors ◀

- ▶ noninformative priors can be useful as a technical device when they hardly influence the posterior
- ▶ they are often improper: problematic for testing and model selection
- ▶ they give a (precise) probability for any subset A of the parameter space, which seems incompatible with the notion of ignorance
- ▶ they express *indifference* instead of *ignorance*
- ▶ a set of priors with $P(A) = [0, 1]$ is noninformative
- ▶ model framework allows for near-noninformative sets of priors
 - ▶ IDM (Walley 1996): range of $y_j^{(0)} = (0, 1) \forall j$
 - ▶ Benavoli & Zaffalon (2012): range of $y^{(0)} = (-\infty, +\infty)$, while $\bar{n}^{(0)}$ decreases with $y^{(0)}$ (to avoid $n^{(0)}|y^{(0)}| = \infty$, i.e. vacuous posterior inferences)