

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/138823> holds various files of this Leiden University dissertation.

Author: Chen, X. (G.)

Title: Prediction sets via parametric and nonparametric Bayes: With applications in pharmaceutical industry

Issue date: 2021-01-05

**Prediction Sets via Parametric and Nonparametric Bayes:
with Applications in Pharmaceutical Industry**

Proefschrift
ter verkrijging van
de graad van Doctor aan de Universiteit Leiden
op gezag van Rector Magnificus Prof. mr. C. J. J. M. Stolker,
volgens besluit van het College voor Promoties
te verdedigen op dinsdag 5 januari 2021
klokke 15:00 uur

door

Xiangyi (Gregory) Chen

geboren te Xiamen, China
in 1987

Promotor: Prof.dr. A.W. van der Vaart

Samenstelling van de promotiecommissie:

Prof.dr. E.R. Eliel

Prof.dr. J.J. Meulman

Prof.dr. E.R. van den Heuvel (Technische Universiteit Eindhoven)

Prof.dr. J.J. Goeman

Dr. T. W. Nagler

To my dear mother, Ms. Luo Yan-Peng

Table of contents

Introduction	1
1 Parametric Prediction Region based on LMM	5
1.1 Introduction	5
1.2 Definitions and setup	7
1.2.1 Frequentist definitions	7
1.2.2 Bayesian definitions	8
1.2.3 Comparison	9
1.2.4 One-sided and two-sided tolerance intervals	10
1.3 Normally distributed future observation	10
1.3.1 Computational strategy	13
1.4 Data from a linear mixed model	17
1.4.1 Approximations to the conditional posterior mean	18
1.4.2 Numerical examples	20
1.5 Frequentist justification of the Bayesian procedure	22
1.5.1 Normal predictions	25
1.6 Annex : Extra Insights on KMW solution in Section 1.3.1	29
1.7 Annex: Gibbs Sampler for one-way Random Effect model	29
1.7.1 Scenario 1: Formulation (a) and Prior (a)	30
1.7.2 Scenario 2: Formulation (a) and Prior (b)	32
1.7.3 Scenario 3: Formulation (b) and Prior (a)	33
1.7.4 Scenario 4: Formulation (b) and Prior (b)	34

1.7.5	Additional Note	35
2	Another Parametric Application: Setting Control Limit at Release for Stability Assurance	37
2.1	Introduction	37
2.2	Problem formulation	38
2.3	Benchmark approach in the industry	39
2.4	Our Population Model	40
2.5	CoT formulation	41
2.5.1	CoT in population model	41
2.5.2	Sample Inference for CoT	42
2.6	CoI formulation	42
2.6.1	CoI in population model	42
2.6.2	Practicality of CoI	44
2.6.3	Sample Inference for CoI	46
2.7	Calculation with Example datasets	49
2.8	Conclusion	50
2.9	Proofs	51
2.10	Annex: Asymptotics of the Bayesian Estimators for CoI Formulation	54
3	Nonparametric Prediction Regions	59
3.1	Introduction	59
3.2	Univariate Nonparametric Prediction Intervals	60
3.3	Data Depth and DD-based Prediction Sets	61
3.3.1	Data Depth	61
3.3.2	DD-based Prediction Sets	63
3.4	Conformity Scores and Conformal Prediction Sets	64
3.5	Bayesian Nonparametric Prediction Sets	65
3.6	Bayesian Nonparametric Prediction Sets via Dirichlet Mixtures	66
3.6.1	Posterior Sampling and SUGS	67

3.6.2	SUGS Drawbacks and Remedies	71
3.6.3	Derive a Prediction Set	72
3.7	Numerical Evaluation	75
3.7.1	Univariate Case	76
3.7.2	Multivariate case	78
3.8	Annex: Stick-Breaking Process	83
3.9	Annex: Generalized Pólya Urn Scheme	83
3.10	Annex: Pitman-Yor and Dirichlet Process Mixture Model	84
3.10.1	Block Gibbs Sampler in Stick-Breaking Scheme	85
3.10.2	Gibbs Sampler in Pólya Urn Scheme	88
4	Nonparametric Tests and Prediction Sets for Functional Data	91
4.1	Introduction	91
4.2	Functional Data Depth (FDD)	92
4.2.1	Definition and Overview	92
4.2.2	ID and WID	93
4.2.3	PBD, RPD and MPD	95
4.2.4	BD and MBD	96
4.2.5	HRD and MHRD	97
4.3	Topic I - Visualization of Outliers	98
4.4	Topic II - Prediction Sets	102
4.5	Topic III - Two Functional Sample Testing	105
4.5.1	Rank Test	105
4.5.2	Permutation Test	106
4.5.3	Bootstrap Test	106
4.5.4	Example Analysis	106
5	An Application: Image-Based Filter for Silicon Oil Particles	109
5.1	Introduction	109
5.2	Industrial Background	110

5.3	Three Benchmark Approaches	111
5.4	Data and Image Pre-processing	112
5.5	Construction of Image-Based (IB) Filter	115
5.5.1	IB-Test 1: General SO Pattern Recognition	115
5.5.2	IB-Test 2: Reference books of SO stereotypes	120
5.6	Performance Evaluation	121
5.7	Conclusion	125
5.8	Discussion	125
5.9	Appendix 1: Gallery on variations of SO/NSO images	128
5.10	Appendix 2: Gallery of the Reference book in Test 2 of IB-Filter	132
5.11	Appendix 3: Gallery of example images from the testing dataset.	134
	References	137
	Summary	143
	Samenvatting	147
	Acknowledgment	151
	Curriculum Vitae	153