

Contents

	Preface: A Great Promise	xi
	Acknowledgments	xv
1	Overview	1
1.1	Two Examples	1
1.2	Key Concepts	5
1.2.1	Probabilities and their Meaning in Databases	5
1.2.2	Possible Worlds Semantics	5
1.2.3	Types of Uncertainty	6
1.2.4	Types of Probabilistic Databases	6
1.2.5	Query Semantics	6
1.2.6	Lineage	7
1.2.7	Probabilistic Databases v.s. Graphical Models	8
1.2.8	Safe Queries, Safe Query Plans, and the Dichotomy	9
1.3	Applications of Probabilistic Databases	10
1.4	Bibliographic and Historical Notes	13
2	Data and Query Model	17
2.1	Background of the Relational Data Model	17
2.2	The Probabilistic Data Model	19
2.3	Query Semantics	21
2.3.1	Views: Possible Answer Sets Semantics	22
2.3.2	Queries: Possible Answers Semantics	22
2.4	C-Tables and PC-Tables	23
2.5	Lineage	27
2.6	Properties of a Representation System	29
2.7	Simple Probabilistic Database Design	30
2.7.1	Tuple-independent Databases	31
2.7.2	BID Databases	35
2.7.3	U-Databases	37
2.8	Bibliographic and Historical Notes	41

3	The Query Evaluation Problem	45
3.1	The Complexity of $P(\Phi)$	45
3.2	The Complexity of $P(Q)$	48
3.3	Bibliographic and Historical Notes	51
4	Extensional Query Evaluation	53
4.1	Query Evaluation Using Rules	55
4.1.1	Query Independence	55
4.1.2	Six Simple Rules for $P(Q)$	56
4.1.3	Examples of Unsafe (Intractable) Queries	61
4.1.4	Examples of Safe (Tractable) Queries	62
4.1.5	The Möbius Function	65
4.1.6	Completeness	69
4.2	Query Evaluation using Extensional Plans	75
4.2.1	Extensional Operators	75
4.2.2	An Algorithm for Safe Plans	80
4.2.3	Extensional Plans for Unsafe Queries	81
4.3	Extensions	84
4.3.1	BID Tables	84
4.3.2	Deterministic Tables	86
4.3.3	Keys in the Representation	87
4.4	Bibliographic and Historical Notes	87
5	Intensional Query Evaluation	91
5.1	Probability Computation using Rules	92
5.1.1	Five Simple Rules for $P(\Phi)$	92
5.1.2	An Algorithm for $P(\Phi)$	96
5.1.3	Read-Once Formulas	98
5.2	Compiling $P(\Phi)$	99
5.2.1	d-DNNF ⁺	100
5.2.2	FBDD	101
5.2.3	OBDD	101
5.2.4	Read-Once Formulas	102
5.3	Approximating $P(\Phi)$	102
5.3.1	A deterministic approximation algorithm	102
5.3.2	Monte Carlo Approximation	104
5.4	Query Compilation	108

5.4.1	Conjunctive Queries without Self-Joins	109
5.4.2	Unions of Conjunctive Queries	110
5.5	Discussion	119
5.6	Bibliographic and Historical Notes	120
6	Advanced Techniques	123
6.1	Top- k Query Answering	123
6.1.1	Computing the Set Top_k	124
6.1.2	Ranking the Set Top_k	129
6.2	Sequential Probabilistic Databases	129
6.3	Monte Carlo Databases	134
6.3.1	The MCDB Data Model	134
6.3.2	Query Evaluation in MCDB	135
6.4	Indexes and Materialized Views	137
6.4.1	Indexes for Probabilistic data	137
6.4.2	Materialized Views for Relational Probabilistic Databases	140
	Conclusion	143
	Bibliography	145
	Authors' Biographies	163