



## STATISTICS IN PRACTICE

A series of practical books outlining the use of statistical techniques in a wide range of application areas:

- Human and Biological Sciences
- Earth and Environmental Sciences
- Industry, Commerce and Finance

### Statistics and the Evaluation of Evidence for Forensic Scientists

Second Edition

Colin Aitken

University of Edinburgh, UK

Franco Taroni

University of Lausanne, Switzerland

The use of statistical and probabilistic ideas in forensic science is of increasing importance to the administration of justice. This is exemplified by their role in the evaluation and interpretation of trace evidence, such as glass, fibres and DNA. They allow the forensic scientist to evaluate and interpret evidence where there is an element of uncertainty.

For many years, *Statistics and the Evaluation of Evidence for Forensic Scientists* has been considered the leading text in the statistical evaluation of forensic evidence. It is highly regarded amongst forensic scientists. This new edition is fully updated, covering many new topics of interest, and includes coverage of new fields of evidence. It features new examples concerned with up-to-date data sets, and a chapter on evidence evaluation using Bayesian nets.

- Provides a comprehensive introduction to the statistical evaluation of forensic evidence.
- Features new material on glass interpretation, fibres interpretation and Bayesian nets.
- Includes new material on the analysis of multivariate data, useful, for example, in problems involving elemental or chemical compositions.
- Illustrated throughout with examples using real up-to-date data.
- Assumes only a modest mathematical background.
- Includes a foreword by Dennis Lindley.

The clarity of exposition makes this book ideal for all forensic scientists, lawyers and other professionals in related fields interested in the quantitative assessment and evaluation of forensic evidence. It is also suitable for graduate students studying courses in the statistical evaluation of evidence.

***'There can be no doubt that the appreciation of some evidence in a court of law has been greatly enhanced by the sound use of statistical ideas and one can be confident that the next decade will see further developments, during which time this book will admirably serve those who have cause to use statistics in forensic science.'*** — D.V. LINDLEY

 **WILEY**  
wiley.com

ISBN 0-470-84367-5



9 780470 843673

# Contents

<b>List of tables</b>	<b>xiii</b>
<b>List of figures</b>	<b>xix</b>
<b>Foreword</b>	<b>xxii</b>
<b>Preface to the first edition</b>	<b>xxv</b>
<b>Preface to the second edition</b>	<b>xxviii</b>
<b>1 Uncertainty in forensic science</b>	<b>1</b>
1.1 Introduction	1
1.2 Statistics and the law	2
1.3 Uncertainty in scientific evidence	5
1.3.1 The frequentist method	6
1.3.2 Stains of body fluids	7
1.3.3 Glass fragments	9
1.4 Terminology	12
1.5 Types of data	15
1.6 Probability	16
1.6.1 Introduction	16
1.6.2 A standard for uncertainty	18
1.6.3 Events	20
1.6.4 Subjective probability	21
1.6.5 Laws of probability	23
1.6.6 Dependent events and background information	25
1.6.7 Law of total probability	29
1.6.8 Updating of probabilities	32
<b>2 Variation</b>	<b>35</b>
2.1 Populations	35
2.2 Samples and estimates	37
2.3 Counts	40
2.3.1 Probabilities	40
2.3.2 Summary measures	41
2.3.3 Binomial distribution	43
2.3.4 Multinomial distribution	44
2.3.5 Hypergeometric distribution	45
2.3.6 Poisson distribution	48
2.3.7 Beta-binomial distribution	51
2.4 Measurements	52
2.4.1 Summary statistics	52
2.4.2 Normal distribution	53



2.4.3	Student's <i>t</i> -distribution	60
2.4.4	Beta distribution	62
2.4.5	Dirichlet distribution	63
2.4.6	Multivariate Normal and correlation	64
<b>3</b>	<b>The evaluation of evidence</b>	<b>69</b>
3.1	Odds	69
3.1.1	Complementary events	69
3.1.2	Examples	70
3.1.3	Definition	70
3.2	Bayes' Theorem	72
3.2.1	Statement of the theorem	72
3.2.2	Examples	73
3.3	Errors in interpretation	78
3.3.1	Fallacy of the transposed conditional	79
3.3.2	Source probability error	81
3.3.3	Ultimate issue error	82
3.3.4	Defender's fallacy	82
3.3.5	Probability (another match) error	83
3.3.6	Numerical conversion error	84
3.3.7	False positive fallacy	85
3.3.8	Uniqueness	86
3.3.9	Other difficulties	87
3.3.10	Empirical evidence of errors in interpretation	89
3.4	The odds form of Bayes' theorem	95
3.4.1	Likelihood ratio	95
3.4.2	Logarithm of the likelihood ratio	99
3.5	The value of evidence	101
3.5.1	Evaluation of forensic evidence	101
3.5.2	Summary of competing propositions	105
3.5.3	Qualitative scale for the value of the evidence	107
3.5.4	Misinterpretations	111
3.5.5	Explanation of transposed conditional and defence fallacies	112
3.5.6	The probability of guilt	116
3.6	Summary	118
<b>4</b>	<b>Historical review</b>	<b>119</b>
4.1	Early history	119
4.2	The Dreyfus case	122
4.3	Statistical arguments by early twentieth-century forensic scientists	125
4.4	People v. Collins	126
4.5	Discriminating power	129
4.5.1	Derivation	129
4.5.2	Evaluation of evidence by discriminating power	130
4.5.3	Finite samples	133
4.5.4	Combination of independent systems	135
4.5.5	Correlated attributes	136
4.6	Significance probabilities	141
4.6.1	Calculation of significance probabilities	141
4.6.2	Relationship to likelihood ratio	144
4.6.3	Combination of significance probabilities	147

4.7	Coincidence probabilities	149
4.7.1	Introduction	149
4.7.2	Comparison stage	151
4.7.3	Significance stage	151
4.8	Likelihood ratio	153
<b>5</b>	<b>Bayesian inference</b>	<b>157</b>
5.1	Introduction	157
5.2	Bayesian inference for a Bernoulli probability	160
5.3	Estimation with zero occurrences in a sample	162
5.4	Estimation of products in forensic identification	165
5.5	Bayesian inference for a Normal mean	166
5.6	Interval estimation	170
5.6.1	Confidence intervals	170
5.6.2	Highest posterior density intervals	172
5.6.3	Bootstrap intervals	172
5.6.4	Likelihood intervals	173
5.7	Odds ratios	175
<b>6</b>	<b>Sampling</b>	<b>179</b>
6.1	Introduction	179
6.2	Choice of sample size	182
6.2.1	Large consignments	182
6.2.2	Small consignments	186
6.3	Quantity estimation	190
6.3.1	Frequentist approach	190
6.3.2	Bayesian approach	191
6.4	Misleading evidence	197
<b>7</b>	<b>Interpretation</b>	<b>205</b>
7.1	Concepts and court cases	205
7.1.1	Relevant population	205
7.1.2	Consideration of odds	206
7.1.3	Combination of evidence	208
7.1.4	Specific cases	208
7.2	Pre-assessment and relevant propositions	214
7.2.1	Levels of proposition	214
7.2.2	Pre-assessment of the case	217
7.2.3	Pre-assessment of the evidence	220
7.3	Assessment of value of various evidential types	221
7.3.1	Earprints	221
7.3.2	Firearms and toolmarks	223
7.3.3	Fingerprints	226
7.3.4	Speaker recognition	228
7.3.5	Hair	229
7.3.6	Documents	231
7.3.7	Envelopes	233
7.3.8	Handwriting	235
7.3.9	Paint	239
7.4	Pre-data and post-data questions	239



<b>8</b>	<b>Transfer evidence</b>	<b>245</b>
8.1	The likelihood ratio	245
8.1.1	Probability of guilt	245
8.1.2	Justification	246
8.1.3	Combination of evidence and comparison of more than two propositions	248
8.2	Correspondence probabilities	254
8.3	Direction of transfer	255
8.3.1	Transfer of evidence from the criminal to the scene	255
8.3.2	Transfer of evidence from the scene to the criminal	260
8.3.3	Transfer probabilities	261
8.3.4	Two-way transfer	270
8.3.5	Presence of non-matching evidence	271
8.4	Grouping	271
8.5	Relevant populations	274
<b>9</b>	<b>Discrete data</b>	<b>283</b>
9.1	Notation	283
9.2	Single sample	283
9.2.1	Introduction	283
9.2.2	General population	286
9.2.3	Particular population	286
9.2.4	Examples	286
9.3	Two samples	288
9.3.1	Two stains, two offenders	288
9.3.2	DNA profiling	291
9.4	Many samples	292
9.4.1	Many different profiles	292
9.4.2	General cases	293
9.5	Relevance of evidence and relevant material	295
9.5.1	Introduction	295
9.5.2	Subjective probabilities	296
9.5.3	Association propositions	296
9.5.4	Intermediate association propositions	297
9.5.5	Examples	298
9.5.6	Two stains, one offender	302
9.6	Summary	304
9.6.1	Stain known to have been left by offenders	304
9.6.2	Relevance: stain may not have been left by offenders	305
9.6.3	Relevance and the crime level	307
9.7	Missing persons	308
9.7.1	Case 1 (Kuo, 1982)	309
9.7.2	Case 2 (Ogino and Gregonis, 1981)	309
9.7.3	Calculation of the likelihood ratio	310
9.8	Paternity: combination of likelihood ratios	312
9.8.1	Likelihood of paternity	314
9.8.2	Probability of exclusion in paternity	317

<b>10</b>	<b>Continuous data</b>	<b>319</b>
10.1	The likelihood ratio	319
10.2	Normal distribution for between-source data	321
10.2.1	Sources of variation	322
10.2.2	Derivation of the marginal distribution	322
10.2.3	Approximate derivation of the likelihood ratio	324
10.2.4	Lindley's approach	326
10.2.5	Interpretation of result	327
10.2.6	Examples	328
10.3	Estimation of a probability density function	330
10.4	Kernel density estimation for between-source data	337
10.4.1	Application to medullary widths of cat hairs	339
10.4.2	Refractive index of glass	339
10.5	Probabilities of transfer	342
10.5.1	Introduction	342
10.5.2	Single fragment	342
10.5.3	Two fragments	345
10.5.4	A practical approach to glass evaluation	349
10.5.5	Graphical models for the assessment of transfer probabilities	352
10.6	Approach based on $t$ -distribution	353
10.7	Appendix Derivation of $V$ when the between-source measurements are assumed normally distributed	357
<b>11</b>	<b>Multivariate analysis</b>	<b>359</b>
11.1	Introduction	359
11.2	Description of example	360
11.3	Univariate $t$ -tests	362
11.4	Hotelling's $T^2$	363
11.5	Univariate Normality, two sources of variation	365
11.6	Multivariate Normality, two sources of variation	366
11.7	Caveat lector	371
11.8	Summary	372
11.9	Appendix	373
11.9.1	Matrix terminology	373
11.9.2	Determination of a likelihood ratio with an assumption of Normality	377
<b>12</b>	<b>Fibres</b>	<b>381</b>
12.1	Introduction	381
12.2	Likelihood ratios in scenarios involving fibres	381
12.2.1	Fibres evidence left by an offender	382
12.2.2	Comments on the fibres scenario	387
12.2.3	Fibres evidence not left by the offender	388
12.2.4	Cross-transfer	389
12.3	Pre-assessment in fibres scenarios	392
12.3.1	The case history	392
12.3.2	Formulation of the pairs of propositions and events	392
12.3.3	Assessment of the expected value of the likelihood ratio	394
12.4	The relevant population of fibres	396



<b>13</b>	<b>DNA profiling</b>	<b>399</b>
13.1	Introduction	399
13.2	Hardy–Weinberg equilibrium	401
13.3	DNA likelihood ratio	404
13.4	Uncertainty	405
13.5	Variation in sub-population allele frequencies	405
13.6	Related individuals	409
13.7	More than two propositions	412
13.8	Database searching	414
	13.8.1 Search and selection effect (double counting error)	418
13.9	Island problem	419
13.10	Mixtures	421
13.11	Error rate	424
<b>14</b>	<b>Bayesian networks</b>	<b>429</b>
14.1	Introduction	429
14.2	Bayesian networks	430
	14.2.1 The construction of Bayesian networks	431
14.3	Evidence at the crime level	437
	14.3.1 Preliminaries	437
	14.3.2 Description of probabilities required	437
14.4	Missing evidence	439
	14.4.1 Preliminaries	439
	14.4.2 Determination of a structure for a Bayesian network	440
	14.4.3 Comments	442
14.5	Error rates	442
	14.5.1 Preliminaries	442
	14.5.2 Determination of a structure for a Bayesian network	443
14.6	Transfer evidence	444
	14.6.1 Preliminaries	444
	14.6.2 Determination of a structure for a Bayesian network	444
	14.6.3 Comment on the <i>transfer</i> node	446
14.7	Combination of evidence	447
14.8	Cross-transfer evidence	449
	14.8.1 Description of nodes	451
	14.8.2 Probabilities for nodes	452
14.9	Factors to consider	452
	14.9.1 Parameter choice	453
14.10	Summary	453
	<b>References</b>	<b>455</b>
	<b>Notation</b>	<b>485</b>
	<b>Cases</b>	<b>489</b>
	<b>Author index</b>	<b>491</b>
	<b>Subject index</b>	<b>499</b>