Contents

1 INTRODUCTION 1

2 PROBABILITY AND STATISTICS 6

2.1 Definition of probability 7

2.2 Random variables. Sample space 8

2.3 Calculus of probabilities 9
2.3.1 Definitions 9
2.3.2 Example: Topologies of bubble chamber events (1) 11
2.3.3 Addition rule 11
2.3.4 Conditional probability 11
2.3.5 Example: $K^o p$ scattering cross section 13
2.3.6 Independence; multiplication rule 15
2.3.7 Example: Relay networks 16
2.3.8 Example: Efficiency of a Čerenkov counter 17
2.3.9 Example: π^o detection 18
2.3.10 Example: Beam contamination and δ-rays 19
2.3.11 Example: Scanning efficiency (1) 20
2.3.12 Marginal probability 23
2.3.13 Example: Topologies of bubble chamber events (2) 23

2.4 Bayes' Theorem 24
2.4.1 Statement and proof 24
2.4.2 Example 25
2.4.3 Comments 26
2.4.4 Example: Betting odds 28
2.4.5 Bayes' Postulate 28

3 GENERAL PROPERTIES OF PROBABILITY DISTRIBUTIONS 30

3.1 The probability density function 30

3.2 The cumulative distribution function 31

3.3 Properties of the probability density function 31
3.3.1 Expectation values of a function 32
3.3.2 Mean value and variance of a random variable 33
3.3.3 General moments 34

3.4 The characteristic function 36

3.5 Distributions of more than one random variable 38
3.5.1 The joint probability density function 38
3.5.2 Expectation values 39
3.5.3 The covariance matrix; correlation coefficients 39
3.5.4 Independent variables 41
3.5.5 Marginal and conditional distributions 42
3.5.6 Example: Scatterplots of kinematic variables 43
3.5.7 The joint characteristic function 47

3.6 Linear functions of random variables 48
3.6.1 Example: Arithmetic mean of independent variables with the same mean and variance 50

3.7 Change of variables 50
3.7.1 Example: Dalitz plot variables 52

3.8 Propagation of errors 52
3.8.1 A single function 52
3.8.2 Example: Variance of arithmetic mean 54
3.8.3 Several functions; matrix notation 54

3.9 Discrete probability distributions 57
3.9.1 Modification of formulae 57
3.9.2 The probability generating function 57

3.10 Sampling 58
3.10.1 Universe and sample 58
3.10.2 Sample properties 59
3.10.3 Inferences from the sample 59
3.10.4 The Law of Large Numbers 61

4 SPECIAL PROBABILITY DISTRIBUTIONS 63

4.1 The binomial distribution 64
4.1.1 Definition and properties 64
4.1.2 Example: Histogramming events (1) 67
4.1.3 Example: Scanning efficiency (2) 68

4.2 The multinomial distribution 72
4.2.1 Definition and properties 72
4.2.2 Example: Histogramming events (2) 74

4.3 The Poisson distribution 75
4.3.1 Definition and properties 75
4.3.2 The Poisson assumptions
Example: Bubbles along a track in a bubble chamber 78
4.3.3 Example: Radioactive emissions 81

4.4 Relationships between the Poisson and other probability distributions 83
4.4.1 Example: Distribution of counts from an inefficient counter 83
4.4.2 Example: Subdivision of a counting interval 84
4.4.3 Relation between binomial and Poisson distributions
Example: Forward-backward classification 85
4.4.4 Relation between multinomial and Poisson distributions
Example: Histogramming events (3) 86
4.4.5 The compound Poisson distribution
Example: Droplet formation along tracks in cloud chamber 87

4.5 The uniform distribution 89
4.5.1 The uniform p.d.f. 89
4.5.2 Example: Uniform random number generators 91

4.6 The exponential distribution 92
4.6.1 Definition and properties 92
4.6.2 Derivation of the exponential p.d.f. from the Poisson assumptions 93

4.7 The gamma distribution 95
4.7.1 Definition and properties 95
4.7.2 Derivation of the gamma p.d.f. from the Poisson assumptions 97
4.7.3 Example: On-line processing of batched events 99

4.8 The normal, or Gaussian, distribution 101
4.8.1 Definition and properties of $N(\mu,\sigma^2)$ 101
4.8.2 The standard normal distribution N(0,1) 101
4.8.3 Probability contents of $N(\mu,\sigma^2)$ 103
4.8.4 Central moments; the characteristic function 106
4.8.5 Addition theorem for normally distributed variables 107
4.8.6 Properties of $\bar{x}$ and s^2 for sample from $N(\mu,\sigma^2)$ 109
4.8.7 Example: Position and width of resonance peak 110
4.8.8 The Central Limit Theorem 110
4.8.9 Example: Gaussian random number generator 113

4.9 The binormal distribution 114
4.9.1 Definition and properties 114
4.9.2 Example: Construction of a binormal random number generator 119

4.10 The multinormal distribution 120
4.10.1 Definition and properties 120
4.10.2 The quadratic form Q 122

4.11 The Cauchy, or Breit-Wigner, distribution 123

5 SAMPLING DISTRIBUTIONS 127

5.1 The chi-square distribution 127
5.1.1 Definition 127
5.1.2 Proof for the chi-square p.d.f. 129
5.1.3 Properties of the chi-square distribution 130
5.1.4 Probability contents of the chi-square distribution 135
5.1.5 Addition theorem for chi-square distributed variables 135
5.1.6 Proof that $(n-1)s^2/\sigma^2$ for sample from $N(\mu,\sigma^2)$ is $\chi^2(n-1)$ 136

5.2 The Student's t-distribution 140
5.2.1 Definition 140
5.2.2 Proof for the Student's t p.d.f. 141
5.2.3 Properties of the Student's t-distribution 142
5.2.4 Probability contents of the Student's t-distribution 143

5.3 The F-distribution 145
5.3.1 Definition 145
5.3.2 Proof for the F p.d.f. 146
5.3.3 Properties of the F-distribution 146
5.3.4 Probability contents of the F-distribution 148

5.4 Limiting properties - connection between probability distributions 149

6 COMPARISON OF EXPERIMENTAL DATA WITH THEORY 151

6.1 Rejection of bad measurements 151
6.2 Experimental errors on measurements. The resolution function 152
6.2.1 Example: Gaussian resolution function and exponential p.d.f. 153
6.2.2 Example: Gaussian resolution function and Gaussian p.d.f. 155
6.2.3 Example: Breit-Wigner resolution function and Breit-Wigner p.d.f. 156
6.2.4 Example: Width of a resonance 156
6.2.5 Experimental determination of resolution function; ideogram 157

6.3 Systematic effects. Detection efficiency 158
6.3.1 Example: Truncation of an exponential distribution 161
6.3.2 Example: Truncation of a Breit-Wigner distribution 161
6.3.3 Correcting for finite geometry - modifying the p.d.f. 162
6.3.4 Correcting for unobservable events - weighting of the events 163

6.4 Superimposed probability densities 163
6.4.1 Example: Particle beam with background 164
6.4.2 Example: Resonance peaks in an effective-mass spectrum 164

7 STATISTICAL INFERENCE FROM NORMAL SAMPLES 166

7.1 Definitions 166

7.2 Confidence intervals for the mean 169
7.2.1 Case with σ^2 known 169
7.2.2 Case with σ^2 unknown 171

7.3 Confidence intervals for the variance 174
7.3.1 Case with μ known 174
7.3.2 Case with μ unknown 176

7.4 Confidence regions for the mean and variance 177

8 ESTIMATION OF PARAMETERS 179

8.1 Definitions 180

8.2 Properties of estimators 180

8.3 Consistency 181

8.4 Unbiassedness 182
8.4.1 Example: s^2 as an estimator of σ^2 183
8.4.2 Example: Estimator of the third central moment 184

8.5 Minimum variance and efficiency 185
8.5.1 Example: Estimator of the mean in the Poisson distribution 187
8.5.2 Example: Estimators of the mean in the normal p.d.f. 188
8.5.3 Example: Estimators of σ^2 and σ in the normal p.d.f. 189

8.6 Sufficiency 190
8.6.1 One-parameter case 190
8.6.2 Example: Single sufficient statistics for the normal p.d.f. 191
8.6.3 Extension to several parameters 193
8.6.4 Example: Jointly sufficient estimators for μ and σ^2 in $N(\mu,\sigma^2)$ 193

9 THE MAXIMUM-LIKELIHOOD METHOD 195

9.1 The Maximum-Likelihood Principle 196
9.1.1 Example: Estimate of mean lifetime 198

9.2 Estimation of parameters in the normal distribution 199
9.2.1 Estimation of μ; measurements with common error 199
9.2.2 Estimation of μ; measurements with different errors (weighted mean) 199
9.2.3 Simultaneous estimation of mean and variance 200

9.3 Estimation of the location parameter in the Cauchy p.d.f. 201

9.4 Properties of Maximum-Likelihood estimators 202
9.4.1 Invariance under parameter transformation 202
9.4.2 Consistency 203
9.4.3 Unbiassedness 203
9.4.4 Sufficiency 204
9.4.5 Efficiency 205
9.4.6 Uniqueness 206
9.4.7 Asymptotic normality of ML estimators 207
9.4.8 Example: Asymptotic normality of the ML estimator of the mean lifetime 208

9.5 Variance of Maximum-Likelihood estimators 210
9.5.1 General methods for variance estimation 210
9.5.2 Example: Variance of the lifetime estimate 212
9.5.3 Variance of sufficient ML estimators 213
9.5.4 Example: Variance of the weighted mean 215
9.5.5 Example: Errors in the ML estimated of μ and σ^2 in $N(\mu,\sigma^2)$ 215
9.5.6 Variance of large-sample ML estimators 217
9.5.7 Example: Planning of an experiment; polarization (1) 218
9.5.8 Example: Planning of an experiment; density matrix elements (1) 219

9.6 Graphical determination of the Maximum-Likelihood estimate and its error 221
9.6.1 The one-parameter case 221
9.6.2 Example: Scanning efficiency (3) 222
9.6.3 The two-parameter case; the covariance ellipse 225

9.7 Interval estimation from the likelihood function 229
9.7.1 Likelihood intervals, the one-parameter case 232
9.7.2 Confidence intervals from the Bartlett functions 236
9.7.3 Example: Confidence intervals for the mean lifetime 237
9.7.4 Likelihood regions, the two-parameter case 239
9.7.5 Example: Likelihood region for μ and σ^2 in $N(\mu,\sigma^2)$ 245
9.7.6 Likelihood regions, the multi-parameter case 246

9.8 Generalized likelihood function 249

9.9 Application of the Maximum-Likelihood method to classified data 249

9.10 Combining experiments by the Maximum-Likelihood method 251

9.11 Application of the Maximum-Likelihood method to weighted events 252
9.12 A case study: an ill-behaved likelihood function 255

10 THE LEAST-SQUARES METHOD 259

10.1 Basis for the Least-Squares method 259
10.1.1 The Least-Squares Principle 259
10.1.2 Connection between the LS and the ML estimation methods 261

10.2 The linear Least-Squares model 262
10.2.1 Example: Fitting a straight line (1) 262
10.2.2 The normal equations 263
10.2.3 Matrix notation 265
10.2.4 Properties of the linear LS estimator 267
10.2.5 Example: Fitting a parabola 268
10.2.6 Example: Combining two experiments 271
10.2.7 General polynomial fitting 272
10.2.8 Orthogonal polynomials 273
10.2.9 Example: Fitting a straight line (2) 274

10.3 The non-linear Least-Squares model 275
10.3.1 Newton's method 276
10.3.2 Example: Helix parameters in track reconstruction 278

10.4 Least-Squares fit 282
10.4.1 "Improved measurements" (fitted variables) and residuals 282
10.4.2 Estimating σ^2 in the linear model 283
10.4.3 The normality assumption; degrees of freedom 285
10.4.4 Goodness-of-fit 288
10.4.5 Stretch functions, or "pulls" 289

10.5 Application of the Least-Squares method to classified data 290
10.5.1 Construction of X^2 290
10.5.2 Choice of classes 294
10.5.3 Example: Polarization of antiprotons (2) 295
10.5.4 Example: Angular momentum analysis (1) 296

10.6 Application of the Least-Squares method to weighted events 298

10.7 Linear Least-Squares estimation with linear constraints 301
10.7.1 Example: Angles in a triangle 301
10.7.2 Linear LS model with linear constraints; Lagrangian multipliers 304

10.8 General Least-Squares estimation with constraints 307
10.8.1 The iteration procedure 308
10.8.2 Example: Kinematic analysis of a V^0 event (1) 312
10.8.3 Calculation of errors 314

10.9 Confidence intervals and errors from the X^2 function 316
10.9.1 Basis for the determination of LS confidence intervals 316
10.9.2 LS errors and confidence intervals, the one-parameter case 317
10.9.3 LS errors and confidence regions, the multi-parameter case 319

11 THE METHOD OF MOMENTS 321

11.1 Basis for the simple moments method 321

11.2 Generalized moments method 323
11.2.1 One-parameter case 323
11.2.2 Multi-parameter case 324
11.2.3 Example: Density matrix elements (2) 324

11.3 Moments method with orthonormal functions 327
11.3.1 Example: Polarization of antiprotons (3) 328
11.3.2 Example: Angular momentum analysis (2) 330
11.3.3 Confidence intervals for MM estimates 330

11.4 Combining MM estimates from different experiments 331

12 A SIMPLE CASE STUDY WITH APPLICATION OF DIFFERENT PARAMETER ESTIMATION METHODS 332

12.1 Simulation of a polarization experiment 333

12.2 Application of different estimation methods 335
12.2.1 The method of moments 335
12.2.2 The Maximum-Likelihood method 336
12.2.3 The Maximum-Likelihood method for classified data 336
12.2.4 The Least-Squares method 338
12.2.5 The simplified Least-Squares method 340

12.3 Discussion 340
12.3.1 The estimated parameter values and their errors 340
12.3.2 Goodness-of-fit 342

13 MINIMIZATION PROCEDURES 346

13.1 General remarks 347

13.2 Step methods 350
13.2.1 Grid search and random search 350
13.2.2 Minimum along a line; the success-failure method 352
13.2.3 The coordinate variation method 353
13.2.4 The Rosenbrock method 354
13.2.5 The simplex method 356

13.3 Gradient methods 359
13.3.1 Numerical calculation of derivatives 360
13.3.2 Method of steepest descent 362
13.3.3 The Davidon variance algorithm 363

13.4 Multiple minima 365

13.5 Evaluation of errors 366

13.6 Minimization with constraints 367
13.6.1 Elimination of constraints by change of variables 370
13.6.2 Penalty functions; Carroll's response surface technique 372
13.6.3 Example: Determination of resonance production 373

13.7 Concluding remarks 374

14 HYPOTHESIS TESTING 376

14.1 Introductory remarks 376

14.2 Outline of general methods 378
14.2.1 Example: Separation of one-π^0 and multi-π^0 events 381
14.2.2 The Neyman-Pearson test for simple hypotheses 384
14.2.3 Example: Neyman-Pearson test on the Ξ^0 mean lifetime 385
14.2.4 The likelihood-ratio test for composite hypotheses 388
14.2.5 Example: Likelihood-ratio test on the mean of a normal p.d.f. 390

14.3 Parametric tests for normal variables 395
14.3.1 Tests of mean and variance in $N(\mu,\sigma^2)$ 395
14.3.2 Comparison of means in two normal distributions 397
14.3.3 Comparison of variances in two normal distributions 401
14.3.4 Summary table 403
14.3.5 Example: Comparison of results from two different measuring machines 404
14.3.6 Example: Significance of signal above background 406
14.3.7 Comparison of means in N normal distributions; scale factor 411

14.4 Goodness-of-fit tests 414
14.4.1 Pearson's χ^2 test 415
14.4.2 Choice of classes for Pearson's χ^2 test 418
14.4.3 Degrees of freedom in Pearson's χ^2 test 420
14.4.4 General χ^2 tests for goodness-of-fit 421
14.4.5 Example: Kinematic analysis of a V^0 event (2) 423
14.4.6 The Kolmogorov-Smirnov test 424
14.4.7 Example: Goodness-of-fit in a small sample 427

14.5 Tests of independence 429
14.5.1 Two-way classification; contingency tables 429
14.5.2 Example: Independence of momentum components 433

14.6 Tests of consistency and randomness 433
14.6.1 Sign test 436
14.6.2 Run test for comparison of two samples 438
14.6.3 Example: Consistency between two effective-mass samples 440
14.6.4 Run test for checking randomness within one sample 442
14.6.5 Example: Time variation of beam momentum 443
14.6.6 Run test as a supplement to Pearson's χ^2 test 444
14.6.7 Example: Comparison of experimental histogram and theoretical distribution 446
14.6.8 Kolmogorov-Smirnov test for comparison of two samples 448
14.6.9 Wilcoxon's rank sum test for comparison of two samples 450
14.6.10 Example: Consistency test for two sets of measurements of the π^0 lifetime 452
14.6.11 Kruskal-Wallis rank test for comparison of several samples 453
14.6.12 The χ^2 test for comparison of histograms 455

APPENDIX STATISTICAL TABLES 459

Table A1 The binomial distribution 461
Table A2 The cumulative binomial distribution 465
Table A3 The Poisson distribution 469
Table A4 The cumulative Poisson distribution 473
Table A5 The standard normal probability density function 477
Table A6 The cumulative standard normal distribution 478
Table A7 Percentage points of the Student's t-distribution 479
Table A8 Percentage points of the chi-square distribution 480
Table A9 Percentage points of the F-distribution 481
Table A10 Percentage points of the Kolmogorov-Smirnov statistic 486
Table A11 Critical values of the run statistic 487
Table A12 Critical values of the Wilcoxon rank sum statistic 488

BIBLIOGRAPHY 493

INDEX 497