

---

Luca Lista

# Statistical Methods for Data Analysis

With Applications in Particle Physics

Third Edition

 Springer

# Contents

<b>1</b>	<b>Introduction to Probability and Inference</b> .....	1
1.1	Why Probability Matters to a Physicist .....	1
1.2	Random Processes and Probability .....	2
1.3	Different Approaches to Probability .....	2
1.4	Classical Probability .....	4
1.5	Problems with the Generalization to the Continuum .....	6
1.6	The Bertrand's Paradox .....	7
1.7	Axiomatic Probability Definition .....	8
1.8	Conditional Probability .....	9
1.9	Independent Events .....	9
1.10	Law of Total Probability .....	10
1.11	Inference .....	12
1.12	Measurements and Their Uncertainties .....	13
1.13	Statistical and Systematic Uncertainties .....	14
1.14	Frequentist vs Bayesian Inference .....	15
	References .....	15
<b>2</b>	<b>Discrete Probability Distributions</b> .....	17
2.1	Introduction .....	17
2.2	Joint and Marginal Probability Distributions .....	18
2.3	Conditional Distributions and Chain Rule .....	18
2.4	Independent Random Variables .....	19
2.5	Statistical Indicators: Average, Variance, and Covariance .....	20
2.6	Statistical Indicators for Finite Samples .....	23
2.7	Transformations of Variables .....	25
2.8	The Bernoulli Distribution .....	26
2.9	The Binomial Distribution .....	27
2.10	The Multinomial Distribution .....	28
2.11	The Poisson Distribution .....	30
2.12	The Law of Large Numbers .....	33
2.13	Law of Large Numbers and Frequentist Probability .....	34
	References .....	35

<b>3</b>	<b>Probability Density Functions</b> .....	37
3.1	Introduction .....	37
3.2	Definition of Probability Density Function .....	37
3.3	Statistical Indicators in the Continuous Case .....	39
3.4	Cumulative Distribution .....	41
3.5	Continuous Transformations of Variables .....	41
3.6	Marginal Distributions .....	43
3.7	Uniform Distribution .....	44
3.8	Gaussian Distribution .....	45
3.9	$\chi^2$ Distribution .....	47
3.10	Log Normal Distribution .....	48
3.11	Exponential Distribution .....	49
3.12	Gamma Distribution .....	50
3.13	Beta Distribution .....	51
3.14	Breit–Wigner Distribution .....	54
3.15	Relativistic Breit–Wigner Distribution .....	55
3.16	Argus Distribution .....	57
3.17	Crystal Ball Function .....	58
3.18	Landau Distribution .....	59
3.19	Mixture of PDFs .....	60
3.20	Central Limit Theorem .....	61
3.21	Probability Distributions in Multiple Dimension .....	63
3.22	Independent Variables .....	63
3.23	Covariance, Correlation, and Independence .....	65
3.24	Conditional Distributions .....	67
3.25	Gaussian Distributions in Two or More Dimensions .....	68
	References .....	73
<b>4</b>	<b>Random Numbers and Monte Carlo Methods</b> .....	75
4.1	Pseudorandom Numbers .....	75
4.2	Properties of Pseudorandom Generators .....	76
4.3	Uniform Random Number Generators .....	78
4.4	Inversion of the Cumulative Distribution .....	79
4.5	Random Numbers Following a Finite Discrete Distribution .....	80
4.6	Gaussian Generator Using the Central Limit Theorem .....	85
4.7	Gaussian Generator with the Box–Muller Method .....	86
4.8	Hit-or-Miss Monte Carlo .....	87
4.9	Importance Sampling .....	88
4.10	Numerical Integration with Monte Carlo Methods .....	90
4.11	Markov Chain Monte Carlo .....	91
	References .....	93
<b>5</b>	<b>Bayesian Probability and Inference</b> .....	95
5.1	Introduction .....	95
5.2	Bayes’ Theorem .....	95

5.3	Bayesian Probability Definition .....	101
5.4	Decomposing the Denominator in Bayes' Formula .....	102
5.5	Bayesian Probability Density and Likelihood Functions .....	104
5.6	Bayesian Inference .....	105
5.7	Repeated Observations of a Gaussian Variable .....	109
5.8	Bayesian Inference as Learning Process .....	111
5.9	Parameters of Interest and Nuisance Parameters .....	112
5.10	Credible Intervals .....	113
5.11	Bayes Factors .....	117
5.12	Subjectiveness and Prior Choice .....	119
5.13	Jeffreys' Prior .....	120
5.14	Reference Priors .....	121
5.15	Improper Priors .....	121
5.16	Transformations of Variables and Error Propagation .....	124
	References .....	125
<b>6</b>	<b>Frequentist Probability and Inference .....</b>	<b>127</b>
6.1	Frequentist Definition of Probability .....	127
6.2	Estimators .....	128
6.3	Properties of Estimators .....	129
6.4	Robust Estimators .....	131
6.5	Maximum Likelihood Method .....	131
6.6	Likelihood Function .....	132
6.7	Binned and Unbinned Fits .....	133
6.8	Numerical Implementations .....	133
6.9	Likelihood Function for Gaussian Distribution .....	134
6.10	Errors of Maximum Likelihood Estimates .....	136
6.11	Covariance Matrix Estimate .....	136
6.12	Likelihood Scan .....	139
6.13	Properties of Maximum Likelihood Estimators .....	141
6.14	Extended Likelihood Function .....	142
6.15	Minimum $\chi^2$ and Least Squares Methods .....	145
6.16	Linear Regression .....	146
6.17	Goodness of Fit and $p$ -Value .....	151
6.18	Minimum $\chi^2$ Method for Binned Histograms .....	152
6.19	Binned Poissonian Fits .....	153
6.20	Constrained Fitting .....	154
6.21	Error Propagation .....	155
6.22	Simple Cases of Error Propagation .....	156
6.23	Propagation of Asymmetric Errors .....	158
6.24	Asymmetric Error Combination with a Linear Model .....	159
	References .....	162
<b>7</b>	<b>Combining Measurements .....</b>	<b>163</b>
7.1	Introduction .....	163
7.2	Control Regions .....	164

7.3	Simultaneous Fits .....	164
7.4	Weighted Average .....	165
7.5	Combinations with $\chi^2$ in $n$ Dimensions .....	167
7.6	The Best Linear Unbiased Estimator (BLUE) .....	168
7.7	Quantifying the Importance of Individual Measurements .....	171
7.8	Negative Weights .....	173
7.9	Iterative Application of the BLUE Method .....	175
	References .....	175
<b>8</b>	<b>Confidence Intervals</b> .....	<b>177</b>
8.1	Introduction .....	177
8.2	Neyman Confidence Intervals .....	177
8.3	Construction of the Confidence Belt .....	178
8.4	Inversion of the Confidence Belt .....	180
8.5	Binomial Intervals .....	181
8.6	The Flip-Flopping Problem .....	183
8.7	The Unified Feldman–Cousins Approach .....	185
	References .....	188
<b>9</b>	<b>Convolution and Unfolding</b> .....	<b>189</b>
9.1	Introduction .....	189
9.2	Convolution .....	189
9.3	Convolution and Fourier Transform .....	191
9.4	Discrete Convolution and Response Matrix .....	192
9.5	Effect of Efficiency and Background .....	192
9.6	Unfolding by Inversion of the Response Matrix .....	194
9.7	Bin-by-Bin Correction Factors .....	196
9.8	Regularized Unfolding .....	197
9.9	Tikhonov Regularization .....	198
9.10	$L$ -Curve Scan .....	199
9.11	Iterative Unfolding .....	202
9.12	Adding Background in Iterative Unfolding .....	204
9.13	Other Unfolding Methods .....	206
9.14	Software Implementations .....	206
9.15	Unfolding in More Dimensions .....	206
	References .....	206
<b>10</b>	<b>Hypothesis Testing</b> .....	<b>209</b>
10.1	Introduction .....	209
10.2	Test Statistic .....	210
10.3	Type I and Type II Errors .....	211
10.4	Fisher’s Linear Discriminant .....	213
10.5	The Neyman–Pearson Lemma .....	215
10.6	Projective Likelihood Ratio Discriminant .....	216
10.7	Kolmogorov–Smirnov Test .....	217
10.8	Wilks’ Theorem .....	219

10.9	Likelihood Ratio in the Search for New Signals .....	220
	References .....	223
<b>11</b>	<b>Machine Learning</b> .....	<b>225</b>
11.1	Supervised and Unsupervised Learning .....	226
11.2	Machine-Learning Classification from a Statistics Perspective... ..	227
11.3	Terminology .....	229
11.4	Curve Fit as Machine-Learning Problem .....	232
11.5	Undertraining and Overtraining .....	233
11.6	Bias-Variance Trade-Off .....	235
11.7	Logistic Regression .....	238
11.8	Softmax Regression .....	241
11.9	Support Vector Machines .....	243
11.10	Artificial Neural Networks .....	243
11.11	Deep Learning .....	245
11.12	Deep Learning in Particle Physics .....	247
11.13	Convolutional Neural Networks .....	248
11.14	Recursive Neural Networks .....	250
11.15	Graph Neural Networks .....	251
11.16	Random Forest and Boosted Decision Trees .....	252
11.17	Unsupervised Learning .....	268
11.18	Clustering .....	268
11.19	Anomaly Detection .....	268
11.20	Autoencoders .....	269
11.21	Reinforcement Learning .....	270
11.22	Generative Adversarial Network .....	271
11.23	No Free Lunch Theorem .....	273
	References .....	273
<b>12</b>	<b>Discoveries and Limits</b> .....	<b>277</b>
12.1	Searches for New Phenomena: Discovery and Limits .....	277
12.2	$p$ -Values and Discoveries .....	278
12.3	Significance Level .....	279
12.4	Signal Significance and Discovery .....	280
12.5	Significance for Poissonian Counting Experiments .....	281
12.6	Significance with Likelihood Ratio .....	282
12.7	Significance Evaluation with Toy Monte Carlo .....	283
12.8	Excluding a Signal Hypothesis .....	283
12.9	Combined Measurements and Likelihood Ratio .....	284
12.10	Definitions of Upper Limit .....	284
12.11	Bayesian Approach .....	285
12.12	Bayesian Upper Limits for Poissonian Counting .....	285
12.13	Limitations of the Bayesian Approach .....	288
12.14	Frequentist Upper Limits .....	288
12.15	Frequentist Upper Limits for Counting Experiments .....	289
12.16	Frequentist Limits in Case of Discrete Variables .....	290

---

12.17	Feldman–Cousins Unified Approach .....	291
12.18	Modified Frequentist Approach: The $CL_s$ Method.....	295
12.19	Presenting Upper Limits: The Brazil Plot .....	298
12.20	Nuisance Parameters and Systematic Uncertainties .....	299
12.21	Nuisance Parameters with the Bayesian Approach.....	300
12.22	Hybrid Treatment of Nuisance Parameters .....	300
12.23	Event Counting Uncertainties .....	301
12.24	Upper Limits Using the Profile Likelihood .....	301
12.25	Variations of the Profile Likelihood Test Statistic .....	303
12.26	Test Statistic for Positive Signal Strength .....	303
12.27	Test Statistic for Discovery .....	303
12.28	Test Statistic for Upper Limits .....	304
12.29	Higgs Test Statistic .....	304
12.30	Asymptotic Approximations .....	304
12.31	Asimov Data Sets .....	305
12.32	The Look Elsewhere Effect.....	316
12.33	Trial Factors .....	318
12.34	Look Elsewhere Effect in More Dimensions .....	321
	References.....	323
<b>Index</b> .....		<b>325</b>

# List of Figures

Fig. 1.1	Probabilities of the possible outcomes of the sum of two dice, $d_1$ and $d_2$ .....	6
Fig. 1.2	Illustration of Bertrand's paradox: three different choices of <i>random extraction</i> of a chord in a circle lead apparently to probabilities that the cord is longer than the inscribed triangle's side of $1/2$ (left), $1/3$ (center), and $1/4$ (right), respectively. Red solid lines and blue dashed lines represent chords longer and shorter than the triangle's side, respectively .....	8
Fig. 1.3	Illustration of conditional probability. Two events, $A$ (blue) and $B$ (red), are represented as sets of elementary outcomes. The conditional probability $P(A   B)$ is equal to the area of the intersection, $A \cap B$ , divided by the area of $B$ ....	9
Fig. 1.4	Illustration of the partition of a set $E_0$ into the union of the disjoint sets $E_1, E_2, E_3, \dots, E_N$ .....	10
Fig. 1.5	Illustration of the law of total probability. The sample space $\Omega$ is partitioned into the sets $A_1, A_2, A_3, \dots, A_N$ , and the set $E_0$ is correspondingly partitioned into $E_1, E_2, E_3, \dots, E_N$ , where each $E_i$ is $E_0 \cap A_i$ and has probability equal to $P(E_0   A_i) P(A_i)$ .....	11
Fig. 1.6	Illustration of the complementary role of data fluctuation and measurement uncertainty .....	12
Fig. 2.1	A set of $R = 3$ red balls plus $W = 7$ white balls considered in a Bernoulli process. The probability to randomly extract a red ball is $p = R/(R + W) = 3/10 = 30\%$ .....	26
Fig. 2.2	Binomial process represented as subsequent random extractions of a single red or white ball (Bernoulli process). The tree shows all the possible combinations at each extraction step. Each branching has a corresponding probability equal to $p$ or $1 - p$ for a red or white ball, respectively. The number of paths corresponding to each possible combination is shown in parentheses and is equal to the binomial coefficient in Eq. (2.60) .....	28

Fig. 2.3	The Yang Hui triangle showing the construction of binomial coefficients. Reprinted from [2], public domain	29
Fig. 2.4	Binomial distributions for $N = 15$ and for $p = 0.2, 0.5$ and $0.8$	30
Fig. 2.5	Poisson distributions with different value of the rate parameter $\nu$	31
Fig. 2.6	Poisson distributions with different value of the parameter $\nu$ compared with Gaussian distributions with $\mu = \nu$ and $\sigma = \sqrt{\nu}$	32
Fig. 2.7	An illustration of the law of large numbers using a computer simulation of die rolls. The average of the first $N$ out of 1000 random extraction is reported as a function of $N$ . 1000 extractions have been repeated twice (red and blue lines) with independent random extractions in order to show that two different lines converge to the same value	34
Fig. 3.1	Uniform distributions with different values of the extreme values $a$ and $b$	44
Fig. 3.2	Gaussian distributions with different values of average and standard deviation parameters $\mu$ and $\sigma$	46
Fig. 3.3	$\chi^2$ distributions with different numbers of degrees of freedom $k$	48
Fig. 3.4	Log normal distributions with different values of the parameters $\mu$ and $\sigma$	49
Fig. 3.5	Exponential distributions with different values of the parameter $\lambda$	50
Fig. 3.6	Examples of gamma distributions for parameter values $\alpha = \kappa = 1, 2$ and $4$ and for $\beta = 1/\theta = 1$	51
Fig. 3.7	Examples of beta distributions for different values of $\alpha$ and $\beta$ . The sum $\alpha + \beta$ is equal to 6 for all distributions. The distribution with $\alpha = \beta$ is symmetric	52
Fig. 3.8	Example of uniform random extractions of a variable $\xi$ . Two intervals are shown of sizes $x$ and $X$ . We consider $x \ll X$	52
Fig. 3.9	Occurrence times (dots) of events uniformly distributed in time, represented along a horizontal axis. The time origin ( $t_0 = 0$ ) is marked as a cross. The occurrence time of the first event is marked as $t_1$	53
Fig. 3.10	Breit–Wigner distributions centered around zero for different values of the width parameter $\gamma$	55
Fig. 3.11	Relativistic Breit–Wigner distributions with mass parameter $m = 100$ and different values of the width parameter $\Gamma$	56
Fig. 3.12	Argus distributions with different values of the parameters $\theta$ and $\xi$	57

Fig. 3.13	Crystal ball distributions with $\mu = 0, \sigma = 1, n = 2$ and different values of the parameter $\alpha$ .....	59
Fig. 3.14	Landau distributions with $\mu = 0$ , and different values of $\sigma$ .....	60
Fig. 3.15	Illustration of the central limit theorem using a Monte Carlo extraction. Random variables $x_i$ are generated uniformly in the interval $[-\sqrt{3}, \sqrt{3}[$ to have average value $\mu = 0$ and standard deviation $\sigma = 1$ . All plots correspond to $10^5$ random extractions. The top-left plot shows the distribution of a single variable $x_1$ ; the other plots show the distribution of $(x_1 + x_2)/\sqrt{2}$ , $(x_1 + x_2 + x_3)/\sqrt{3}$ and $(x_1 + x_2 + x_3 + x_4)/\sqrt{4}$ , respectively, where all $x_i$ are extracted from the same uniform distribution. A Gaussian curve with $\mu = 0$ and $\sigma = 1$ , with proper normalization, in order to match the sample size, is superimposed for comparison. The Gaussian approximation is visually more and more stringent as a larger number of variables is added .....	61
Fig. 3.16	Same as Fig. 3.15, using a PDF that is uniformly distributed in two disjoint intervals, $[-3/2, -1/2[$ and $[1/2, 3/2[$ , in order to have average value $\mu = 0$ and standard deviation $\sigma = 1$ . The individual distribution and the sum of $N = 2, 3, 4, 6$ and $10$ independent random extractions of such a variable, divided by $\sqrt{N}$ , are shown in the six plots. Gaussian distributions with $\mu = 0$ and $\sigma = 1$ are superimposed for comparison .....	62
Fig. 3.17	Illustration of relation between marginal distributions and joint distribution for independent variables. In the $xy$ plane, a slice in $x$ of width $\delta x$ corresponds to a probability $\delta P(x) = f_x(x) \delta x$ , and a slice in $y$ with width $\delta y$ corresponds to a probability $\delta P(y) = f_y(y) \delta y$ . Their intersection corresponds to a probability $\delta P(x, y) = f(x, y) \delta x \delta y$ .....	64
Fig. 3.18	Example of a PDF of two variables $x$ and $y$ that are uncorrelated but not independent .....	66
Fig. 3.19	Illustration of conditional PDF in two dimensions. The PDF $f(y x_0)$ can be obtained by rescaling the red curve that shows $f(x_0, y)$ by a proper factor, in order to ensure normalization .....	67
Fig. 3.20	One-sigma contour for a two-dimensional Gaussian PDF. The two ellipse axes have length equal to $\sigma_{x'}$ and $\sigma_{y'}$ . The $x'$ axis is rotated by an angle $\phi$ with respect to the $x$ axis. The lines tangent to the ellipse parallel to the $x$ and $y$ axes, shown in gray, have a distance with respect to the corresponding axes equal to $\sigma_y$ and $\sigma_x$ . In this example, the correlation coefficient $\rho_{xy}$ is positive .....	71

Fig. 3.21	Illustration of two-dimensional $1\sigma$ and $2\sigma$ Gaussian contours. Each one-dimensional projection of the $1\sigma$ or $2\sigma$ contour corresponds to a band which has a 68.27% or 95.45% probability content, respectively. As an example, three possible projections are shown: a vertical, a horizontal and an oblique one. The probability content of the areas inside the ellipses are smaller than the corresponding one-dimensional projected intervals .....	72
Fig. 4.1	Density of 100,000 consecutive values in the logistic sequence for $\lambda = 4$ (blue histogram) compared to the function $f(x) = 1/\pi\sqrt{x(1-x)}$ (red line) .....	77
Fig. 4.2	Logistic map .....	78
Fig. 4.3	Simulated cosmic rays generated in all direction (blue) intersecting a pair of detector planes (red).....	84
Fig. 4.4	Angular distribution of simulated cosmic rays. The blue histogram represents all generated cosmic rays, the red histogram represents only the ones that cross the detector geometry. The histograms are normalized to the total number of generated cosmic rays .....	85
Fig. 4.5	Illustration of the hit-or-miss Monte Carlo method. Points are uniformly generated within a rectangle that contains the PDF curve $f(x)$ . Green points are accepted, while red points are rejected. The $x$ values of the accepted points are distributed according to $f(x)$ .....	88
Fig. 4.6	Variation of the hit-or-miss Monte Carlo using the importance sampling. Points are uniformly distributed within rectangles whose union contains $f(x)$ . This extraction has a greater efficiency compared to Fig. 4.5 .....	89
Fig. 4.7	Hit-or-miss Monte Carlo using the importance sampling with an exponential envelope, shown as the orange curve .....	90
Fig. 4.8	Example of random generation with the Metropolis–Hastings method. The sampled PDF is the sum of two two-dimensional Gaussian distributions with relative weights equal to 0.45 and 0.55. The PDF is plotted as a white-to-red-to-black color map in the background. The generated points are connected by a line in order to show the sequence. The first 20 generated points are shown in purple, all the subsequent ones are in blue. The proposal function is a two-dimensional Gaussian with $\sigma = 0.5$ . The total number of generated points is 100 (top, left), 1000 (top, right), 2000 (bottom, left), and 5000 (bottom, right) .....	93

Fig. 5.1	Illustration of the conditional probabilities, $P(A   B)$ and $P(B   A)$ . The events $A$ and $B$ are represented as subsets of a sample space $\Omega$ . This Illustration was originally presented by Robert Cousins, courtesy of the author. All rights reserved .....	96
Fig. 5.2	Illustration of the Bayes' theorem. The areas of events $A$ and $B$ , equal to $P(A)$ and $P(B)$ , respectively, simplify when $P(A   B) P(B)$ and $P(B   A) P(A)$ are multiplied. This illustration was originally presented by Robert Cousins, courtesy of the author. All rights reserved .....	97
Fig. 5.3	Illustration Bayes' theorem applied to the computation of the probability that a person with a positive test is really ill. The two red rectangular areas correspond to the cases of a true positive diagnosis for an ill person ( $P(+   \text{ill})$ , vertical red area) and a false positive diagnosis for a healthy person ( $P(+   \text{healthy})$ , horizontal red area). The probability that a person with a positive test is really ill, $P(\text{ill}   +)$ , is equal to the ratio of the vertical red area to the total red area .....	100
Fig. 5.4	Posterior for the binomial parameter $p$ for two cases of positive feedbacks: $8/8 = 100\%$ (red) and $97/100 = 97\%$ (blue) of positive feedbacks. The most probable values of $p$ are 1.0 and 0.97, respectively. The first case has a large probability that $p$ is below 0.8 (0.13), while this probability is negligible ( $4.8 \times 10^{-7}$ ) for the second case .....	108
Fig. 5.5	Repeated application of Bayes' theorem to the same sample of randomly generated Bernoulli processes. Two different priors $\pi(p) = p_0(p)$ are considered: uniform (top) and exponential, $\pi(p) \propto e^{-4p}$ (bottom). As $i$ increases, the posteriors $p_i(p)$ become less sensitive on the assumed prior $\pi(p)$ . The two red bars show on each plot the fractions of the 18 extracted values that are equal to 0 and 1. 13 out of 18 outcomes are equal to 1 in this example.....	113
Fig. 5.6	Different interval choices, shown as shaded areas, corresponding to a 68.27% (top) or 90% (bottom) probability. Top left: central interval; the left and right tails have equal probability. Top right: symmetric interval; the most probable value lies at the center of the interval. Bottom left: fully asymmetric intervals for an upper limit. Bottom right: fully asymmetric interval for a lower limit .....	115
Fig. 5.7	Poisson posterior PDFs for $n = 5$ with a central 68.27% probability interval (left) and for $n = 0$ with a fully asymmetric 90% probability interval (right). Credible intervals are shown as shaded areas .....	117

Fig. 5.8	Posterior distribution for $\tau$ using a uniform prior (dashed line), a uniform prior on $\lambda = 1/\tau$ (dotted line) and Jeffreys' prior (solid line). Data values $t_1, \dots, t_N$ are shown as blue histogram. Numbers of measurements $N = 5$ (top), 10 (middle) and 50 (bottom) have been considered. ....	123
Fig. 6.1	Scan of $-2 \log L$ as a function of a parameter $\theta$ . The bounds of the error interval are determined as the values for which $-2 \log L$ increases by one unit with respect to its minimum, corresponding to $\theta = \hat{\theta}$ .....	140
Fig. 6.2	Example of unbinned extended maximum likelihood fit of a simulated data set. The fit curve (blue solid line) is superimposed to the data points (black dots with error bars) ....	144
Fig. 6.3	Two-dimensional $1\sigma$ contour plot surrounding the uncertainty region for the parameters $s$ and $b$ , equal to the expected number of signal and background events, respectively. The plot corresponds to the fit in Fig. 6.2. The contour is determined from the equation $-2 \log L(x_1, \dots, x_N; s, b, \vec{\theta}) = -2 \log L_{\max} - 1$ .....	144
Fig. 6.4	Example of minimum $\chi^2$ fit of a computer-generated data set. The points with the error bars are used to fit a function model of the type $y = f(x) = a x e^{-bx}$ , where $a$ and $b$ are unknown parameters determined by the fit. The fit curve is superimposed as solid blue line. Residuals $\varepsilon_i$ are shown in the bottom section of the plot .....	146
Fig. 6.5	Transformation of variable $\eta = \eta(\theta)$ and visualization of the procedure of error propagation using local linear approximation .....	156
Fig. 6.6	Transformation of a variable $x$ into a variable $x'$ through a piece-wise linear transformation characterized by two different slopes. If $x$ follows a Gaussian distribution with standard deviation $\sigma$ , $x'$ follows a bifurcated Gaussian, made of two Gaussian halves having different standard deviation parameters, $\sigma'_+$ and $\sigma'_-$ .....	160
Fig. 7.1	Contours at the 68% and 95% confidence level in the plane mass of the W boson vs mass of the top quark, $(m_W, m_t)$ , from global electroweak fits including (blue) or excluding (gray) the measurement of the Higgs boson mass $m_H$ . The direct measurements of $m_W$ and $m_t$ (green) are shown as comparison. The fit was performed by the GFitter group [1], and the figure Reprinted from [2], $\Gamma$ The LEP Electroweak Working Group. Reproduced under CC-BY-4.0 license .....	168

Fig. 7.2	BLUE coefficient for the second measurement $w_2$ (top) and combined BLUE variance $\sigma_x^2$ divided by $\sigma_1^2$ (bottom) as a function of the correlation $\rho$ between the two measurements for various fixed values of the ratio $\sigma_2/\sigma_1$ .....	173
Fig. 7.3	The most conservative value of an unknown correlation coefficient $\rho(\text{cor})$ of the uncertainties $\sigma_1(\text{cor})$ and $\sigma_2(\text{cor})$ as a function of $\sigma_1(\text{cor})/\sigma_1$ , for different possible values of $\sigma_2(\text{cor})/\sigma_1(\text{cor}) \geq 1$ .....	174
Fig. 8.1	Illustration of Neyman belt construction (left) and inversion (right) .....	178
Fig. 8.2	Three possible choices of ordering rule: central interval (top) and fully asymmetric intervals (bottom left, right).....	179
Fig. 8.3	Neyman belt for the parameter $\mu$ of a Gaussian with $\sigma = 1$ at the 68.27% confidence level .....	180
Fig. 8.4	Neyman belt at the 68.27% confidence level for the parameter $p$ of a binomial distribution with $N = 10$ .....	182
Fig. 8.5	Coverage of Clopper–Pearson intervals as a function of $p$ for the cases with $N = 10$ (top) and $N = 100$ (bottom).....	184
Fig. 8.6	Illustration of the flip-flopping problem. The plot shows the quoted point estimate of $\mu$ as a function of the measured $x$ (dashed line), and the 90% confidence interval corresponding to the choice of quoting a central interval for $x/\sigma \geq 3$ and an upper limit for $x/\sigma < 3$ . The coverage decreases from 90 to 85% for a value of $\mu$ corresponding to the horizontal lines with arrows .....	185
Fig. 8.7	Ordering rule in the Feldman–Cousins approach. The interval of the random variable $x$ is determined on the basis of the likelihood ratio $\lambda(x   \theta_0) = f(x   \theta_0)/f(x   \hat{\theta}(x))$ ....	186
Fig. 8.8	Neyman confidence belt constructed using the Feldman–Cousins ordering rule for a Gaussian PDF .....	188
Fig. 9.1	Examples of convolution of two PDFs $f(x)$ (solid light blue lines) with Gaussian kernel functions $r(x, x')$ (dashed black lines). Top: $f(x)$ is the superposition of three Breit–Wigner functions (see Sect. 3.14); bottom: $f(x)$ is a piece-wise constant PDF. The result of the convolution is the solid dark blue line. The kernel function is scaled down by a factor of 2 and 20 in the top and bottom plots, respectively, for display convenience .....	190
Fig. 9.2	Effect of bin migration. Top: response matrix $R_{ij}$ , and bottom: the original distribution $\mu_j$ (light blue line) with superimposed the convoluted distribution $\nu_i = \sum_{j=1}^M R_{ij} \mu_j$ (dark blue line) .....	193

Fig. 9.3	Top: Monte Carlo generated histogram with 10,000 entries randomly extracted according to the convoluted distribution $\vec{v}$ in Fig. 9.2 (bottom) superimposed to the original convoluted distribution, and bottom: the result of response matrix inversion, or the maximum likelihood estimate of the unconvoluted distribution $\hat{\hat{\mu}}$ , with error bars. Note that the vertical scale on the bottom plot extends far beyond the range of the top plot .....	196
Fig. 9.4	Top: response matrix $R_{ij}$ , and bottom: the original distribution $\mu_j$ is shown in light blue, while the convoluted distribution $v_i = \sum R_{ij} \mu_j$ is shown in dark blue .....	200
Fig. 9.5	Top: data sample $n_i$ (dashed histogram with error bars) randomly generated according to the convoluted distribution $v_i$ (solid histogram, superimposed), and bottom: $L$ -curve scan (black line). The optimal $L$ -curve point is shown as blue cross .....	201
Fig. 9.6	Unfolded distribution using Tikhonov regularization, $\hat{\mu}_j$ , shown as points with error bars, superimposed to the original distribution $\mu_i$ , shown as dashed line .....	201
Fig. 9.7	Unfolded distribution using the iterative unfolding, shown as points with error bars, superimposed to the original distribution, shown as dashed line. 100 iterations have been used .....	204
Fig. 9.8	Unfolded distribution as in Fig. 9.7 using the iterative unfolding with 1000 (top), 10,000 (middle), and 100,000 (bottom) iterations .....	205
Fig. 10.1	Distribution of a test statistic $t$ according to two possible PDFs for the signal (blue) and background (red) hypotheses under test .....	211
Fig. 10.2	Signal efficiency versus background misidentification probability (receiver operating characteristic or ROC curve).....	212
Fig. 10.3	Examples of two-dimensional selections of a signal (blue dots) against a background (red dots). A linear cut is chosen on the left plot, while a box cut is chosen on the right plot.....	212
Fig. 10.4	Illustration of the projection of two-dimensional distributions along two different axes (top plots). The red and blue distributions are projected along the black dashed lines; the normal to that line is shown as a green line with an arrow. The bottom plots show the one-dimensional projections of the corresponding top plots .....	214
Fig. 10.5	Illustration of the Kolmogorov–Smirnov test.....	218

Fig. 11.1	Random data sample (red) extracted from a non-linear function model (blue) used in the example described in the text .....	233
Fig. 11.2	Fit of the data sample from Fig. 11.1 (red) using a third-order polynomial (orange). The function model used for the random generation is superimposed (blue) .....	233
Fig. 11.3	Fits of the data sample from Fig. 11.1 (red) using a polynomial (orange) of degrees from 0 to 11 (from top left to bottom right). The function model used for the random generation is superimposed (blue) .....	234
Fig. 11.4	Illustration of overtraining of a machine-learning classification algorithm that exploits two features, represented as the $x$ and $y$ variables corresponding to the horizontal and vertical axes, respectively. Blue points are signal observations, and red points are background in the training sample. The multivariate boundary of the classifier selection is represented as a dashed curve. Left: a smooth selection is based on a regular algorithm output; right: a selection from an overtrained algorithm picks artifacts following statistical fluctuations of the finite-size training sample .....	235
Fig. 11.5	Mean square error evaluated over the train sample (blue) and average of the mean square errors evaluated over 100 independently generated test samples (orange) .....	236
Fig. 11.6	Top: a sample of 50 fourth-degree polynomial fit curves (gray) from randomly generated data sets compared to the originating curve (blue); bottom: bias squared (blue) and variance (orange) contributions to the total square error for all measurement points .....	239
Fig. 11.7	Average over the $N$ observed points of the bias squared (blue) and variance (orange) and total (sum of the two, red) as a function of the polynomial degree $p$ .....	239
Fig. 11.8	Structure of a multilayer artificial neural network .....	243
Fig. 11.9	Construction of a feature map for a $3 \times 3$ filter applied on a $10 \times 10$ pixel image at different positions. The resulting feature map has size $8 \times 8$ . A grayscale image has been assumed, with one color code instead of the three RGB, reducing the dimensionality of the example .....	249
Fig. 11.10	Construction of a feature map for a $3 \times 3 \times 3$ filter applied on a $10 \times 10 \times 3$ colored pixel image at different positions. The resulting feature map has size $8 \times 8$ .....	250
Fig. 11.11	Simplified structure of a convolutional neural network [26] .....	250

Fig. 11.12	Example of a decision tree. Each node represented as an ellipse contains a different number of signal (left number) and background (right number) observations. The relative amount of signal and background in each node is represented as blue and red areas, respectively. The applied requirements are represented as rectangular boxes .....	253
Fig. 11.13	Distribution in two dimensions for signal (blue) and background (red) test samples. Only a fraction of the data is displayed to avoid an overcrowded scatter plot .....	257
Fig. 11.14	ROC curve for the six multivariate classifiers trained on the distribution in Fig. 11.13 .....	258
Fig. 11.15	Distribution of different multivariate classifier outputs (color map plus contour levels) compared to the test samples for signal (blue) and background (red) .....	259
Fig. 11.16	Distribution of the outputs of the different classifiers for signal (blue) and background (red). Solid histograms are the training samples, while dots with error bars are the test samples. The reasonable level of agreement gives indication that no substantial overtraining is present .....	260
Fig. 11.17	Distribution in two dimensions for signal (blue) and background (red) test samples. Only a fraction of the data is displayed to avoid an overcrowded scatter plot .....	261
Fig. 11.18	ROC curve for the four multivariate classifiers.....	261
Fig. 11.19	Distribution of different multivariate classifier outputs (color map plus contour levels) compared to the test samples for signal (blue) and background (red) .....	262
Fig. 11.20	Distribution of the outputs of the different classifiers for signal (blue) and background (red). Solid histograms are the training samples, while dots with error bars are the test samples. The reasonable level of agreement gives indication that no substantial overtraining is present .....	263
Fig. 11.21	Distribution of different multivariate classifier outputs (color map plus contour levels) compared to the test samples for signal (blue) and background (red) .....	264
Fig. 11.22	ROC curve for the four multivariate classifiers .....	265
Fig. 11.23	Distribution of different multivariate classifiers compared to the test samples for signal (blue) and background (red) .....	266
Fig. 11.24	Distribution of the outputs of the different classifiers for signal (blue) and background (red). Solid histograms are the training samples, while dots with error bars are the test samples. The reasonable level of agreement gives indication that no substantial overtraining is present .....	267
Fig. 11.25	Illustration of the $k$ -means clustering algorithm applied to a data sample generated from five two-dimensional Gaussian distributions .....	269

Fig. 12.1	Poisson distribution of the number of collision events $n$ with an expected background yield of $\nu = b = 4.5$ . The probability corresponding to $n \geq 8$ (light blue area) is 0.087 and is equal to the $p$ -value, taking the event counting as test statistic .....	279
Fig. 12.2	Upper limits at the 90% CL (top) and 95% CL (bottom) to the signal yield $s$ for a Poissonian process using the Bayesian approach as a function of the expected background $b$ and for a number of observed counts $n^*$ , ranging from 0 to 10 .....	287
Fig. 12.3	Illustration of Neyman belt construction for upper limits determination .....	288
Fig. 12.4	Poisson distribution of the total number of counts $n = s + b$ for $s = 4$ and $b = 0$ . The white bins show the smallest possible fully asymmetric confidence interval, $\{2, 3, 4, \dots\}$ in this case, that gives at least the required coverage of 90% .....	290
Fig. 12.5	Confidence belt at the 90% CL for a Poissonian process using the Feldman–Cousins approach for $b = 3$ .....	292
Fig. 12.6	Upper limits at 90% CL to the signal $s$ for a Poissonian process using the Feldman–Cousins method as a function of the expected background $b$ and for a number of observed events $n^*$ ranging from 0 to 10 .....	293
Fig. 12.7	Example of evaluation of $CL_s$ from pseudo-experiments. The distribution of the test statistic $-2 \log \lambda$ is shown in blue assuming the signal-plus-background hypothesis and in red assuming the background-only hypothesis. The black line shows the value of the test statistic measured in data, $\lambda^*$ , and the shaded areas represent $p_{s+b}$ (blue) and $p_b$ (red). $CL_s$ is determined as $p_{s+b}/(1 - p_b)$ .....	296
Fig. 12.8	Illustration of the application of the $CL_s$ method in cases of well separated distributions of the test statistic $-2 \log \lambda$ for the $s + b$ and $b$ hypotheses (top) and in case of largely overlapping distributions (bottom) where the experiment has poor sensitivity to the signal .....	297
Fig. 12.9	Upper limit reported as Brazil plot in the context of the search for the Higgs boson at the LHC by the ATLAS experiment. The expected limit to the signal strength $\mu = \sigma/\sigma_{SM}$ is shown as a dashed line, surrounded by the $\pm 1\sigma$ (green) and $\pm 2\sigma$ (yellow) bands. The observed limit is shown as a solid line. All mass values corresponding to a limit below $\mu = 1$ (dashed horizontal line) are excluded at the 95% confidence level. Reprinted from [9], © ATLAS Collaboration. Reproduced under CC-BY-4.0 license .....	299

Fig. 12.10	Monte Carlo generated data sample superimposed to an exponential background model (top) and to an exponential background model plus a Gaussian signal (bottom) .....	308
Fig. 12.11	Distribution of the test statistic $q$ for the background-only hypothesis (blue) and the signal-plus-background hypothesis (red). The value determined with the considered data sample (black arrow) is superimposed. $p$ -values can be determined from the shaded areas of the two PDFs .....	309
Fig. 12.12	Scan of the test statistic $q$ as a function of the parameter of interest $\mu$ .....	310
Fig. 12.13	The considered data sample superimposed to an exponential background model (top) and to an exponential background model plus a Gaussian signal (bottom) adding a $\pm 10\%$ uncertainty to the background normalization. Background excursions of $\pm 10\%$ correspond to the histograms shown in purple and green .....	311
Fig. 12.14	Scan of the test statistic $q$ as a function of the parameter of interest $\mu$ including systematic uncertainty on $\beta$ (red) compared with the case with no uncertainty (blue) .....	313
Fig. 12.15	Example of pseudo-experiment generated with a Gaussian signal on top of an exponential background. The assumed distribution for the background is shown as a red dashed line, while the distribution for signal plus background is shown as a blue solid line .....	315
Fig. 12.16	Negative logarithm of the profile likelihood as a function of the signal yield $s$ . The blue curve is computed assuming a negligible uncertainty on the background yield $b$ , while the red curve is computed assuming a 30% uncertainty. The intersection of the curves with the green line at $-\log \lambda(s) = 0.5$ determines the uncertainty intervals on $s$ .....	316
Fig. 12.17	Illustration of upcrossing of a test statistic $q(m)$ counted in order to determine $\langle N_{u_0} \rangle$ . In this example, we have $N_{\mu} = 3$ .....	319
Fig. 12.18	Local $p$ -value as a function of the Higgs boson mass in the search for the Higgs boson at the LHC performed by ATLAS. The solid line shows the observed $p$ -value, while the dashed line shows the median expected $p$ -value according to the prediction of a Standard Model Higgs boson corresponding to a given mass value $m_H$ . Reprinted from [9], $\copyright$ ATLAS Collaboration. Reproduced under CC-BY-4.0 license .....	320

---

Fig. 12.19 Examples of sets with different values of the Euler characteristic  $\varphi$ , defined as the number of disconnected components minus the number of holes. Top left:  $\varphi = 1$  (one component, no holes), top right:  $\varphi = 0$  (one component, one hole), bottom left:  $\varphi = 2$  (two components, no holes), bottom right:  $\varphi = 1$  (two components, one hole) ..... 322

---

## List of Tables

Table 1.1	Possible values of the sum of two dice with all possible pair combinations and their corresponding probabilities .....	5
Table 3.1	Probabilities corresponding to $Z\sigma$ one-dimensional intervals and corresponding two-dimensional areas inside the elliptic contours for different values of $Z$ .....	73
Table 5.1	Threshold values for Bayes factors proposed to assess the evidence of an alternative hypothesis $H_1$ against a null hypothesis $H_0$ according to [4] .....	119
Table 5.2	Jeffreys' priors corresponding to the parameters of some of the most frequently used PDFs .....	120
Table 7.1	Properties of the different definitions of the importance of a measurement within a BLUE combination .....	172
Table 12.1	Significance levels expressed as $Z\sigma$ and the corresponding $p$ -values for typical cases .....	280
Table 12.2	Upper limits in the presence of negligible background evaluated under the Bayesian approach for a different number of observed counts $n^*$ .....	286
Table 12.3	Upper and lower limits in the presence of negligible background ( $b = 0$ ) with the Feldman–Cousins approach .....	292