

## Index

### a

- Acceptance region 77
- Activation function 169
- Adaptive boost (AdaBoost) 179–180
  - boost weight 180
- Akaike information criterion (AIC) 400
- Alternative hypothesis 76
- Anderson–Darling test 101
- Approximate Bayesian Computation (ABC) 382
- Arithmetic mean 3
- Armijo condition 253
- Artificial neural network (ANN) 168–172
  - activation function 169
  - backpropagation 171
  - batch learning 171
  - feed-forward network 170
  - learning rate 171
  - multi-layer perceptron (MLP) 170
  - neuron 169
  - node 169
  - online learning 171
  - overtraining 172
  - perceptron 170
  - weight 169
- Astrostatistics 382
- Asymptotic distribution 83
- Asymptotic likelihood-ratio test 130
- Averaging data with inconsistencies 63–67
  - outliers 65
  - world average values 65–67
- Averaging of measurements 32–33, 43–46, 63–67

### b

- Backpropagation 171
- Bagging 181
  - bootstrapping 181
- Bayes factor 92, 385

- Bayes' theorem 19, 67
- Bayesian elimination 130
- Bayesian experimental design 404
- Bayesian inference 21–24
- Bayesian information criterion (BIC) 400
- Bayesian interval construction 133–140
- Bayesian parameter estimation 67–69
- Bayesian probability 18
- Bayesian reference analysis approach 135
- Bernoulli distribution 172
- Bernoulli experiment 333
- Best linear unbiased estimator (BLUE) 351
- Bias
  - experimenter's bias 292–293, 354
  - in classification 160, 166
  - in hypothesis testing 81
  - of an estimator 28
- Bias-variance trade-off 159–161
- Binned mass-peak fit: practical considerations 61–63
- Binned maximum likelihood 59–60
- Binning 272
- Binomial distribution 9–10
- Blind analysis 292–293, 351–354
  - hidden signal box method 352
- Blue noise 199
- Boosted decision trees (BDT) 178–179
  - boosting 179
  - Gini index 179
  - pruning 179
- Boosting 179–181
  - adaptive boost (AdaBoost) 179–181
  - base classifiers 180
  - base learner 180
  - gradient boost 180
- Bootstrap method (trigger) 332–333
- Bootstrapping 110, 124–128, 346
  - bootstrap sample 125
  - bootstrap-*t* interval 125–127

- calibration 128
- percentile intervals 127–128
  - automatic percentile bootstrap 127
  - simple percentile interval 127
- Breit–Wigner distribution 10–11
- c**
- Cauchy distribution 10–11
- Central limit theorem (CLT) 6
- Central moments 3
- Characteristic function 5
- $\chi^2$  distribution 14–15, 37
  - number of degrees of freedom 15
- $\chi^2$  test 93–96
- Cholesky decomposition 197, 238, 258
- Classification 153–186
  - all-versus-all approach 153
  - one-versus-all approach 153
  - one-versus-one approach 153
  - pre-processing 182–183
  - Receiver-Operating-Characteristic (ROC) 157–158
  - systematic uncertainties 183–184
- Classifier 154, 162–181
  - artificial neural network (ANN) *see* Artificial neural network (ANN)
  - bias 160
  - boosted decision trees (BDT) *see* Boosted decision trees (BDT)
  - Fisher linear discriminant *see* Fisher linear discriminant
  - generalisation properties 160
  - k-Nearest Neighbour classifier (kNN) *see* k-Nearest Neighbour classifier (kNN)
  - naive Bayes classifier 162–163
  - support vector machine (SVM) *see* Support vector machine (SVM)
  - variance 160
- Clopper–Pearson interval 121
- Closure test 271, 347
- Complementary error function 7
- Composite hypothesis 76
- Confidence belt 111
- Confidence level 108, 363
- Consistency
  - in hypothesis testing 81
  - of an estimator 28
- Constrained fits 227–261, 376
  - Lagrange multipliers *see* Lagrange multipliers
  - propagation of uncertainty 239–244
  - solution by elimination 230–232
  - statistical interpretation 231–232
- Convolution 193–195
- Correlation 4
  - correlation of estimators 351
  - of systematic uncertainties 285–288
- Cost function 82
- Counting experiment 358–361
  - optimisation 360
- Counting method (trigger) 331
- Covariance 4
- Covariance matrix 4, 228
- Coverage 108, 374
  - overcoverage 121
  - undercoverage 116
- Credibility 108
- Critical region 77–79
- Cross-validation 161
- Cumulative distribution function 5, 120
- Curse of dimensionality 163
- Curtosis 3–4
- Cut variations 279–285
- d**
- Data blinding *see* Blind analysis
- Data mining 153
- Data-to-Monte Carlo comparison 277–278
- Deciles 5
- Decision boundary 157
- Deconvolution 193–195
- Degree-of-belief 18
- Dennis–Moré theorem 255
- Deviance information criterion (DIC) 400
- Directional derivative 251
- Discovery 365
- Discrete cosine transformation 193
- Discrete Fourier transform 210
- Distributions 5–16
- Double-blind test 352
- e**
- Educated guess 273–274
- Efficiency
  - of an estimator 28
- Eigenvalue decomposition 198
- Ensemble tests *see* Bootstrapping
- Error function 7
- Errors of first and second kind 79, 155–156
- Estimator 28
  - bias 28
  - consistency 28
  - efficiency 28
- Evidence 386
- Evidence-based prior 386
- Exclusion 365

- Expectation value 2
- Expected significance (level) 359
- Expected-posterior prior construction 388
- Experimenter's bias *see* Blind analysis
- Extended maximum likelihood 55–59, 370
  - fitting rates of processes 56–57
  - improving fitted parameters 57–59
- f**
- Factorisation 298–308
  - factorisation scale 300, 305–307
  - scope and limitations 300–301
- Feature vector 154
- Feed-forward network 170
- Feldman–Cousins intervals 123
- Fisher information 23, 38
  - information matrix 28
- Fisher linear discriminant 165–168
  - between-class matrix 167
  - bias 166
  - weight vector 166
  - within-class matrix 167
- Fletcher's augmented Lagrangian 255
- Flip-flopping 143
- Fredholm integral equation 187, 391
- Frequentist inference 20–21
- Frequentist interval construction 110–133
- Full width at half maximum (FWHM) 11
- g**
- Garwood intervals 123
- Gaussian distribution 6–8
- Gibbs phenomenon 203, 205
- Gini index 179
- Goldstein conditions 254
- Goodness-of-fit (GoF) test 84, 92–102
  - Kolmogorov–Smirnov test 99–101
  - maximum-likelihood-based test 98–99
  - Pearson's  $\chi^2$  test 93–96
  - run test 96–98
  - Smirnov–Cramér–von Mises test 101
  - two-sample tests 101–102
  - unbinned  $\chi^2$  test 98
- Gradient boost 180
- h**
- Hessian matrix 35, 239, 319–321
- Hidden signal box method 352
- Hodges–Lehmann estimator 109
- Hypothesis testing 75–104
  - Bayesian approach 92
  - bias 81
  - combination of tests 87–88
  - consistency 81
- discovery 365
- exclusion 365
- frequentist approach 76
- goodness-of-fit (GoF) test *see* Goodness-of-fit (GoF) test
- hybrid method 86
- look-elsewhere effect (LEE) *see* Look-elsewhere effect (LEE)
- marginalization approach 86
- power of a test 78
- profile likelihood *see* Profile likelihood
- size of a test 78
- systematic uncertainties 86
- test inversion 89–90
- test statistic *see* Test statistic
- i**
- Image reconstruction 390–394
- Improper prior 22, 386
- Inference 2, 20–24, 381
  - Bayesian inference 21–24
  - frequentist inference 20–21
- Information *see* Fisher information
- Interval construction 108–110
  - asymptotic likelihood-ratio test 130
  - automatic percentile 131
  - Bayesian elimination 130
  - Bayesian interval construction 133–140
  - equal-tailed intervals 133
  - frequentist interval construction 110–133
  - generality 110
  - highest posterior density 109, 133
  - interval length 108
  - likelihood regions 134
  - likelihood-ratio bootstrap 131
  - likelihood-ratio profile bootstrap 131
  - likelihood-ratio test inversion 130
  - lowest posterior loss regions 134
  - Mandelkern–Schulz intervals 141
  - naive method 130
  - Neyman's construction 110–116
  - ordering rule 113–115
  - parameter transformation 109
  - physical boundaries 109
  - relation to point estimate 109
  - simple percentile 131
  - systematic uncertainties 109
- Interval estimation 107–146
- Intrinsic discrepancy loss 134
- Intrinsic prior construction 388
- Inverse problem 187–196
  - direct processes 187–189

- discretisation 189–192
  - inverse processes 187–189
  - Iterative unfolding 213–215
    - implicit regularisation 213
    - Landweber iteration 214
    - Lucy–Richardson convolution 213
- j**
- Jacobian determinant 232
  - Jacobian matrix 239, 258
  - Jeffreys prior 23–24, 135, 386
- k**
- k-Nearest Neighbour classifier (kNN)
    - 163–165
      - metric 165
      - smoothing 165
  - Kernel density estimation 340
  - Kernel function 188
  - Kolmogorov axioms 17
  - Kolmogorov–Smirnov test 99–101
  - Kullback–Leibler divergence 134, 399
  - Kurtosis 3–4
- l**
- Lagrange multipliers 176, 232–237, 321
    - feasible direction 233
    - feasible points 233
    - feasible set 233
    - iterative solution 244–259
      - choice of direction 245–250
      - convergence 256–258
      - initial values 258–259
      - propagation of uncertainty 259
      - step length 250–256
    - Lagrange function 232
    - linear independence constraint qualification (LICQ) 234, 258
    - unmeasured parameters 236–237
  - Landau distribution 11–12
  - Landweber iteration 214
  - Least squares 40–52
    - binned fit 60–61
    - design matrix 42
    - least-squares estimate 40
      - variance 41
    - linear least squares 42–48
      - averaging of correlated measurements 44–46
      - averaging of measurements 43–44
      - best linear unbiased estimator (BLUE) 42
      - normal equations 42
      - straight-line fit 46–48
        - variance 42
        - non-linear least squares 48–52
          - mass-peak fit (signal position) 50–52
- Leptokurtic 4
- Likelihood 20
- Likelihood principle 20
- Likelihood-free analysis 382
- Likelihood-ratio bootstrap 131
- Likelihood-ratio profile bootstrap 131
- Likelihood-ratio test inversion 130
- Line search 252–253
  - Armijo condition 253
  - cubic interpolation 254
  - Goldstein conditions 254
  - sufficient-decrease condition 253
  - Wolfe conditions 253
- Log-normal distribution 15
- Look-elsewhere effect (LEE) 52, 88–89, 361
- Lorentzian distribution 10–11
- Loss function 159
- Lucy–Richardson convolution 213
- m**
- Machine learning 159
  - Mahalanobis distance 165
  - Mandelkern–Schulz intervals 141
  - Manhattan norm 251
  - Maratos effect 255–256
  - Marginal distribution 4, 339
  - Markov chain Monte Carlo (MCMC) 401
  - Mass-peak fit 50–52, 61–63
  - Matrix
    - Cholesky decomposition 197, 238, 258
    - column rank 239
    - condition of a matrix 197
    - correlation matrix 4
    - covariance matrix 4, 228
    - design matrix 42
    - effective rank of a matrix 200
    - Fisher information matrix 28
    - Hessian matrix 35, 239, 319–321
    - Jacobian matrix 239, 258
    - Moore–Penrose pseudo-inverse 191, 248–249, 256
    - resolution matrix 203
    - response matrix 189, 217–218
    - unfolding matrix 215
  - Matrix method 334–337
  - Maximum entropy method 391, 393–394
  - Maximum-likelihood method 29–40
    - averaging of measurements 32–33
    - binned maximum likelihood 59–60
    - extended maximum likelihood 55–59, 370

- maximum-likelihood estimate (MLE) 29
- properties 31
- transformation invariance 31
- variance 33–35
- profile likelihood *see* Profile likelihood
- solution 30–31
- unbinned maximum likelihood 52
- Mean 3
- Median 5
- Minimum-variance bound (MVB) 28, 38–40
- Mode 5
- Model complexity 399–400
- Moments 3
- Monte Carlo generators 313–314
- Moore–Penrose pseudo-inverse 191, 248–249, 256
- Morozov discrepancy principle 207
- Multi-layer perceptron (MLP) 170
- MultiNest 404
- Multinomial distribution 10, 338
- Multivariate analysis 381
- Multivariate classification 153
- n**
- Naive Bayes classifier 162–163
- Negative binomial distribution 12
- Nested sampling 401–404
- Neuron 169
- Newton–Raphson method 48, 246
- Neyman–Pearson lemma 155, 158–159
- Neyman’s  $\chi^2$  344
- Neyman’s construction 110–116
- Normal distribution 6
- Nuisance parameters 86, 128–133, 372
- Null hypothesis 75, 155
- o**
- Objective function 228
- Objective prior 24, 69, 108
- Observed significance (level) 83, 359
- On/off problem 383–390
- Overtraining 160, 172
- p**
- p*-value 7, 78, 83–89, 358
- Parameter estimation 27–72
  - Bayesian approach 67–69
  - least squares *see* Least squares
  - maximum-likelihood method *see* Maximum-likelihood method
- Parametrised unfolding 195–196
- Parton distribution functions 299, 314–324
  - comparison of PDF sets 322–324
  - DGLAP evolution equation 315
  - NNPDF approach 321–322
  - parametric uncertainties 318–322
  - uncertainties 316–322
- Pattern recognition 153
- Pearson’s  $\chi^2$  344
- Pearson’s  $\chi^2$  test 93–96
- Penalty function  $\ell_1$  251
- Percentiles 5
- Perceptron 170
- Philipps regularisation 205
- Pivoting 118–123
- Pixon method 391
- Platykurtic 4
- Plug-in principle 124
- Point spread function (PSF) 188, 213, 390
- Poisson distribution 8–9
- Posterior probability 19, 338
  - marginal distribution 339
- Power of a test 78
- Power-constrained limits 90, 366
- Prediction 2
  - retrospective prediction 15
- Principal component analysis (PCA) 182
- Prior probability 19, 22–24, 69, 338
  - evidence-based prior 386
  - expected-posterior prior construction 388
  - improper prior 22, 386
  - intrinsic prior construction 388
  - Jeffreys prior 23–24, 135, 386
  - objective prior 24, 69, 108
  - proper prior 108, 386
  - reference prior 24
  - subjective prior 386
  - uniform prior 22
- Probability 2, 16–19
  - Bayesian definition 18
  - classical definition 17
  - frequentist definition 17–18
  - mathematical definition 17
- Probability density function 2–5
  - marginalisation 4
  - projection 4
- Profile likelihood 37–38, 362–369, 372
  - profile likelihood ratio 362
- Projection 4
- Projection methods
  - discrete cosine transformation 210–211
- Proper prior 108, 386
- Pruning 179

- Pseudo-experiment 345
- Pull distribution 12, 347, 374
- q**
- QR decomposition 248
- Quantiles 5
- Quartiles 5
- r**
- Random process 2
- Random variable 2, 5
- Receiver-Operating-Characteristic (ROC) 157–158
- Rectangular distribution 10
- Reference prior 24
- Regularisation 203–209
  - derivative regularisation 206–207
  - low-pass regularisation 212–213
  - norm regularisation 204–206
  - Philipps regularisation 205
  - presentation of results 220–221
  - regularisation methods 203
  - regularisation parameter 203, 207–208
  - regularisation schemes 203
  - Tikhonov regularisation 205
- Resampling 124, 346
- Response function 188
- Response matrix 189, 217–218
- ROOT 222
- Root mean square 257
- Run test 96–98
- s**
- Schwarz information criterion (SIC) 400
- Semi-convergence 213
- Sideband fit 360–361, 383
- Sideband subtraction 274
- Significance (level) 78, 84–85
  - expected significance (level) 359
  - observed significance (level) 83, 359
- Significance tests 84
- Simple hypothesis 76
- Single-blind test 352
- Singular value decomposition (SVD) 196–198, 258
  - Fourier coefficients 199
  - pre-scaling 197
  - pre-whitening 197
  - singular values 197
  - thin SVD 197
  - truncated SVD 199–202
- Size of a test 78
- Skew 3–4
- Smirnov–Cramér–von Mises test 101
- Software
  - APLCON 260
  - BAT 69, 337
  - GURU 222
  - MINUIT 31, 38, 124, 195
  - R 7
  - ROOFIT 357, 378
  - ROOSTATS 69, 357
  - ROOT 7, 37
  - ROOUNFOLD 222
  - RUN 217
  - TUNFOLD 222
  - for constrained fits 260
  - for unfolding 221–222
- Standard deviation 3
- Standard normal distribution 6
- Statistical hypothesis 75–76
  - alternative hypothesis 76
  - composite hypothesis 76
  - null hypothesis 75
  - simple hypothesis 76
- Statistical inference 381
- Straight-line fit 46–48
- Student's *t* distribution 12–14
- Subjective prior 386
- Subjective probability 18
- Supervised machine learning 159
- Support vector machine (SVM) 172–178
  - margin 173
  - soft-margin approach 174
  - support vector 173
- Systematic uncertainties 263–295, 375
  - binning 272
  - combination 285–288
  - correlation 285–288
  - cut variations 279–285
  - data-to-Monte Carlo comparison 277–278
  - detector acceptance 268–269
  - educated guesses 273–274
  - estimation 272–288
  - experimenter's bias 292–293
  - in hypothesis testing 86
  - in interval construction 109
  - sources 265
  - theory uncertainties *see* Theory uncertainties
  - tolerance 273
- t**
- Tag-and-probe method 331–332
- Taxicab norm 251
- Template method 337–345
- Template morphing 342, 371

Test inversion 89–90, 116–117  
 – likelihood-ratio test inversion 130  
 Test statistic 76–77, 155, 362  
 – sufficiency 80  
 Theory uncertainties 276–277, 297–325  
 – colour reconnection 314  
 – hadronisation 311–313  
 – corrections 312  
 – fragmentation functions 312  
 – multi-parton interactions 310–311  
 – parton distribution functions 316–322  
 – perturbative expansion 301–308  
 – combining different orders 307  
 – factorisation scale 305–307  
 – multi-scale problems 307–308  
 – renormalisation scale 301–305  
 – resummation methods 307–308  
 – power corrections 300, 308–310  
 – higher-twist corrections 308  
 – operator product expansion 308–309  
 – sources 297  
 – underlying event 310–311  
 Tikhonov regularisation 205  
 Times series analysis 382  
 Tolerance 273  
 Top-hat distribution 10  
 Transfer function 213  
 Two-sample tests 101–102  
 Type I and type II errors 79, 155–156

**u**

Unbinned  $\chi^2$  test 98  
 Unbinned maximum likelihood 52–59,  
 370–371  
 – fitting fractions of processes 54–55  
 Unfolding 187–223, 391  
 – bin migration 188  
 – bin-by-bin correction method 208, 214

– convolution 193–195  
 – deconvolution 193–195  
 – eigenvalue decomposition 198  
 – folding vs. unfolding 202  
 – iterative unfolding *see* Iterative  
 unfolding  
 – kernel function 188  
 – Landweber iteration 214  
 – migration problems 218  
 – Morozov discrepancy principle 207  
 – parameterised unfolding 195–196, 202  
 – Poisson-distributed data 192–193  
 – projection methods 211–212  
 – regularisation *see* Regularisation  
 – resolution matrix 203  
 – response function 188  
 – response matrix 189  
 – singular value decomposition (SVD)  
*see* Singular value decomposition (SVD)  
 – transformation problems 218  
 – unfolding matrix 215

Uniform distribution 10

Uniform prior 22

Unit Gaussian distribution 6

Unsupervised machine learning 159

**v**

Variance 3  
 – in classification 160

**w**

Weibull distribution 15–16  
 White noise 211  
 Wilks' theorem 123  
 Wolfe conditions 253

**z**

Z-value 78, 84