Index

AC, *see* Algorithmic complexity
Afterlife theory, 397
AIC (Akaike Information Criterion), 303
Akaike, H., 35
Akaike Information Criterion (AIC), 35, 303
Algorithmic complexity (AC), 100–110, 311
– explanations and, 118–121
– Shannon information versus, 107–110
Algorithmic Probability (ALP), 406
Allison, L., 209, 274
ALP (Algorithmic Probability), 406
Alphabet, 58
Alternative priors, 350–353
Ancillary statistic, 242
Approximate theories, 14
Arcs, 305–306
Arithmetic coding, 73–76
Arrow of Time, Thermodynamic, 337–384
Assertion, 152
– imprecise, of discrete parameters, 284–286
– length of, 235
– *see also* Explanation, first part of
Assertion code for probabilistic finite-state machines, 308–309
Asymmetry, reasons for, 367–369
Atom count difference, 383

Bakus–Naur form (BNF), 306–307
Barron, A.R., 311
Baxter, R.A., 274
Bayesian decision theory, 40–45
Bayesian inference, 1, 35–40
– relation to, 116–118
Bayes Information Criterion (BIC), 303
Bayes posterior density, 182
Bayes' theorem, 22
Bernardo, J.M., 47
Bernoulli sequence, 147
Best explanation of data, 1
Beta density, 246
Beta function, 246
Beta prior density, 47
Bhansali, R.J., 35
Bibby, J.M., 300
BIC (Bayes Information Criterion), 303
Binary codes, 59–63
Binary digits, 59
Binary sequence segmentation problem, 321–326
Binary tree codes, 95–96
Binomial distributions, 24
– irregularities in, 248
– negative, 253
Binomial example, 157–160
– using sufficient statistics, 163–164
Bit, 66
Blurred images, 208–209
BNF (Bakus–Naur form), 306–307
Boltzmann, Stefan, 338–339
Boltzmann's constant, 88
Boulton, D.M., 81, 401
Boundary rule
– for growing data groups, 171–173
– for ideal data groups, 198–199

Categories, 315
– regression in, 320
Cauchy density, 32
Causal explanations, 365–367
Causal nets, 326–336
Cause, effect and, 365–367
Carnap, R., 5
Chaitin, G.J., 3-4, 57, 102, 109, 275, 401
Chickering, D.M., 330
ChiSquared form, 34

Class distributions at leaves, coding, 317–318
Classification trees and nets, 314–320
Classified models
– summary of, 293–294
– unclassified models versus, 295–297
Class labels, 276
– Fisher Information with, 291–293
– surrogate estimate, 288–289
Class proportions, 294
Cleary, J.G., 73
Clocks, example of, 353–355
Closed systems, 339–340
Coarse data, normal distribution with, 265–266
Codeable estimates, 213–215
Code length of imprecise discrete estimates, 286–288
Codes, 59
– binary tree, 95–96
– feasible, for infinite sets, 96, 98
– for infinite sets, 91–92
– non-binary, 77–78
– optimal, *see* Optimal codes
– universal, 98–100
–– in theory descriptions, 115–116
Code tree, 61
Code word in Huffman code, 282
Coding
– arithmetic, 73–76
– class distributions at leaves, 317–318
– of data, 146–147
– of inferences, 148–150
– multi-word messages, 72–73
– random, of estimates, 210
– tree structure, 316–317
Coding probability, 149–150
– prior probability density and, 222
Coding scheme, MML, 222–228
Coding transitions, 312–313
Coding trick, 281-284, 293
Collision table, 370–375
Communication of information, 57–58
Complexity Approximation Principle, 405
Computer process, past of a, 375
Concatenation, 119
Concentration parameter, 259
Conditional probability, 21
Confidence, 30
Confidence interval, 30
Conjugate priors, 46–48
– invariant, 52–53, 261
– for multivariate Normal distribution, 261–264
– for Normal distribution, 258–264
Consequential propositions, 7
Continuous data, SMML explanation for, 166–187
Continuous distributions, 24–25
Continuous random variables, 20
Conway, J.H., 178, 257
Cost function, 54
Counter-instances, 10
Cover, T.M., 98, 311
Cross-Validation (CV) criterion, 321–323
Curved-prior message length, MML, 236–237
Cut points, 321–322
CV (Cross-Validation) criterion, 321–322

DAG (directed acyclic graph), 326–327
Data
– coarse, normal distribution with, 265–266
– coding of, 146–147
– continuous, SMML explanation for, 166–187
– discrete, SMML explanation for, 153–166
– explanation of, 359
– perturbed, normal distribution with, 264–265
– possible, set X of, 144–145
– probabilistic model of, 146
Data acquisition, 397
Data groups, 222
– growing, boundary rule for, 171–173
– ideal, 198–199
Data representation invariance, 187–188
Decimal numbers, 59
Decision graphs, 318–320
Decision tree explanation, 315–316
Decision trees, 315
Deduction
– with deterministic laws, 346–348
– with non-deterministic laws, 348–350
– of past disorder, 345–355
– uses of, 361–362
Deductive reasoning, 5
Defining propositions, 7
Density, 25
Descriptive MML theory, 385–399
Detail, 152
– *see also* Explanation, second part of
Detail length, 235, 325–326
Deterministic laws, 343–344
– deduction with, 346–348
– induction of the past with, 363–365
Devolution, 346
Diatomic molecules, 374–375
Dirac delta function, 193
Directed acyclic graph (DAG), 326–327
Disc collision model, 351
Discrete data, SMML explanation for, 153–166
Discrete distributions, 23–24
Discrete estimates, imprecise, code length of, 286–288
Discrete hypothesis sets, 156
Discrete parameters, imprecise assertion of, 284–286
Discrete random variables, 20
Disorder, 337
– entropy as measure of, 341–343
– past, deduction of, 345–355
Dissipation, 361
Dissipative laws, 345
Distributions, 23
– binomial, *see* Binomial distributions
– entropy of, 89
– infinite entropy, 94
– multinomial, *see* Multinomial distributions
–– information content of, 81–87
– Normal, *see* Normal distributions
– predictive, 206
– probability, *see* Probability distributions
– uniform, of known range, mean of, 183–187
Dowe, D.L., 203, 209–216, 219, 252, 268–269, 274, 323–326
Dowe's approximation to message length, 209–213
Dowe's construction, uncertainty regions via, 216
Downham, D.Y., 35

Educated Turing machines (ETM), 130–131
Effect, cause and, 365–367
EM (Expectation Maximization) algorithm, 276–279
Empirical Fisher Information, 240–245
– transformation of, 243–244
Empirical information matrix, 241
Entropy, 51,
– of distribu
– increasing,
– as measur
– time rever
Equilibrium,
Equivalence
– partial orc
– structural,
Equivalence
Escape sequ
Estimate(s),
– class label
– codeable,
– imprecise 286–288
– random c
– Schou, 26
Estimate sp
Estimation
Normal p
Estimator, 3
Estimator f
ETM (Educ 130–131
Euclidean d
Euler's cons
Evolutionar
Evolution p
Exceptions,
Expectatio
Expectatio algorithm
Expected st
Expected v
Experiment
Explanatio
– algorithm
– of data, 3
– first part
– second p
Explanatio
Explanatio
– shortest,
Explanatio
Explanator

Factor ana
– defining
– MML, 3
Factor loa
Factor sco
Falsifiable

Falsifying data, 12
Family of models, 28
Fano, R.M., 70–72
Farr, G.E., 159, 164–165, 172
Farr's algorithm, 164–165
Feasible codes for infinite sets, 96, 98
Fermi–Dirac statistics, 378
Finite-state machines (FSMs), 127–128, 305–314
– less-redundant code for, 309–310
– transparency and redundancy in, 310–312
Fisher determinant, 227
Fisher Information, 48, 225, 411–412
– alternative expression for, 228–229
– with class labels, 291–293
– empirical, *see* Empirical Fisher Information
– for mixtures, 290–291
– safer empirical approximation to, 244–245
Fisher matrix, 232
Fitzgibbon, L., 209, 274
Formula I1A, 226
– for many parameters, 240–241
Formula I1B, 226
Formula I1C, 243
Freeman, P.R., 226–227, 300
FSMs, *see* Finite-state machines
Function approximation, 272–275
Future, 368

Gaines, B.R., 313
Gammerman, A., 405
Gas simulations, 370–375
Geometric constants, 257–258
Glymour, C., 327
God theory, 397
Grammars, regular, *see* Regular grammars
Group, 50
Growing data groups, boundary rule for, 171–173
Grünwald, P.D., 227

Hexagonal Voronoi regions, 181
Huffman, D.A., 70–73, 78, 103, 107, 282, 284–285
Huffman code, 70–71
– code word in, 282
Human induction, 394–396
Hutter, M., 403–405
Hypothesis space
– partitions of, 213–215
– uncertainty regions in, 214–215

Ideal data groups, 198–199
Ideal group (IG) estimator, 197–200
– for Neyman–Scott problem, 201–202
IG, *see* Ideal group estimator
Images, blurred, 208–209
Imprecise assertion of discrete parameters, 284–286
Imprecise discrete estimates, code length of, 286–288
Independence, 22
Induction, 1
– evolutionary, 396–397
– human, 394–396
– MML, practice of, 391–399
– of the past, 356–365
–– with deterministic laws, 363–365
–– by maximum likelihood, 357–358
–– by MML, 358–361
Inductive conclusions, 6
Inductive inference, 1–55
– introduction to, 5–11
– primitive, 387–388
Inductive process, 6
Inexplicable present views, 362–363
Inference(s)
– Bayesian, *see* Bayesian inference
– coding of, 148–150
– inductive, *see* Inductive inference
– Non-Bayesian, 28–35
– possible, set of, 147–148
– of regular grammars, 305–314
– statistical, *see* Statistical inference
Inference problem, specified, 153
Infinite entropy distributions, 94
Infinite sets
– codes for, 91–92
– feasible codes for, 96, 98
Information, 57–141
– communication of, 57–58
– defined, 57
– Fisher, *see* Fisher Information
– measurement of, 66–68
– pattern and noise, separation of, 190–191
– Shannon, *see* Shannon information
– subjective nature of, 79–81
Information content of multinomial distribution, 81–87
Information matrix, empirical, 241
Input tape, Turing machines, 101
Instructions, Turing machines, 101
Integers, optimal codes for, 93–96
Invariance
– data representation, 187–188
– model representation, 188–189
Invariant conjugate priors, 52–53, 261
Invariants, 340
Irregular likelihood functions, 242–243
ISO-7 code, 61

Jeffreys prior, 48–49, 410, 412
J-factor model, 300–303
Joreskog, K.G., 299

Kearns *et al.* MDL criterion (KMDL), 322–325
Kepler's laws, 13
Kent, J.T., 300
KLD (Kullback–Leibler distance), 204–205, 287–288
KMDL (Kearns *et al.* MDL criterion), 322–325
Knowledge, 5–6
Kolmogorov, A.N., 4, 57, 102, 275, 401
Korb, K.B., 323–326, 335
Kraft's inequality, 65
Kuhn, T., 385
Kullback–Leibler distance (KLD), 204–205, 287–288

Langdon, G.G., 73
Language(s)
– choice of, 132
– natural, *see* Natural languages
– scientific, 391
Large-*D* message length, MML, 237
Latent factor model, 297–303
– defining equations for, 299
– MML, 300–303
Lattice constant, 180
Leaf nodes, 315
Lee, T.C.M., 319
Lengths of explanations, 143
Leung Yan Cheong, S.K., 98
Levin, L.A., 392, 396–397
Levin search, 396–397
Likelihood, 30
– maximum, *see* Maximum likelihood
Likelihood functions, irregular, 242–243
Likelihood principle, minimum Message Length approach and, 254–255
Linear regression, 270–272
log* code, 99–100, 409
log likelihood ratio, 33–35

Loss funct
– expecte

Machine l
Mansour,
Mardia, K
Marginal
203–20
Marginal
Maximur
Maximur
Maximur
– inducti
– margin
– for Ney
– normal
Maximun
299–30
Maximun
Maxwell–
MDL (M
criteric
MDL (M
princip
Mealey n
305–30
Mean, 27
– discrim
– of mul
– Norma
estima
– sample
– of unif
range,
Measure
Meek, C
MEKL,
distan
Memorie
Message
Message
Message
– averag
– curvec
– Dowe'
– large-
– small-
Message
Message
Metropc
Micro-st
Ming Li
Minimal

Minimum-cost estimation, 189
Minimum Description Length (MDL) criterion, 321–325
Minimum Description Length (MDL) principle, 401, 408–415
Minimum expected K–L distance (MEKL), 205–208
– for Neyman–Scott problem, 206–208
Minimum Message Length (MML) approach, 117–118
– binomial example, 246–248
– coding scheme, 222–228
– curved-prior message length, 236–237
– as descriptive theory, 385–399
– details in specific cases, 257–303
– efficiency of, 230–231
– extension to Neyman–Scott problem, 252–253
– induction of the past by, 358–361
– large-D message length, 237
– likelihood principle and, 254–255
– limitations of, 249–250
– message length formulae, 235
– model invariance, 229–230
– multi-parameter properties of, 234–235
– multiple latent factors, 300–303
– multiple parameters in, 232–233
– negative binomial distribution, 253
– normal distribution, 250–253
– practice of induction, 391–399
– precision of estimate spacing, 238–240
– properties of estimator, 228–240
– quadratic, assumptions of, 226–227
– quadratic approximations to SMML, 221–255
– related work and, 401–415
– singularities in priors, 237
– small-sample message length, 235–236
– standard formulae, 235
Minimum Variance Unbiased estimators, 31–32
Mixture models, 275–297
Mixtures
– Fisher Information for, 290–291
– message format for, 279–280
ML* estimator, 299–300
ML (Maximum Likelihood) method, 30
MML, *see* Minimum Message Length approach
Model, 28
Model classes, 408
Model density, 177
Model family, 28
Model invariance, MML, 229–230
Model representation invariance, 188–189
Multinomial distributions, 247
– information content of, 81–87
– irregularities in, 248
Multivariate mean estimator, summary of, 183
Multivariate Normal, mean of, 177–183
Multivariate Normal distribution,
– conjugate priors for, 261–264
Multi-word messages, coding, 72–73

Nat, 78
Natural languages
– efficiencies of, 389–390
– hypotheses of, 388–390
– inefficiencies of, 390
Neal, R., 73
Negative binomial distribution, 253
Nested union, 192
Neyman–Pearson testing, 413
Neyman–Scott problem, 200–204, 218
– ideal group estimator for, 201–202
– maximum likelihood for, 203–204
– minimum expected K–L distance for, 206–208
– MML extension to, 252–253
– other estimators for, 202–203
Ng, A.Y., 321–325
Nit, 78
NML (normalized maximum likelihood), 410–415
Nodes, 61
– children, 64
– dummy, 318–319
Noise, 190
Noise level, 273
Non-adiabatic experiment, 382–384
Non-Bayesian estimation, 30–32
Non-Bayesian inference, 28–35
Non-Bayesian model selection, 32–35
Non-binary codes, 77–78
Non-deterministic laws, 344
– deduction with, 348–350
Non-random strings, 109
Normal, multivariate, mean of, 177–183
Normal density, 28
Normal distribution
– with coarse data, 265–266
– conjugate priors for, 258–264
– multivariate, conjugate priors for, 261–264
– with perturbed data, 264–265
Normal distribution function, 29
Normalization, 407–408
Normalized maximum likelihood (NML), 410–415
Normalized probability measure, 403
Normal mean, with Normal prior, estimation of, 173–177
Normal prior, estimation of Normal mean with, 173–177
Notation, confusion of, 144–145
Noun, 389
Nuisance parameters, 285
Null hypothesis, 19, 32–33

Occam's Razor, 1
Observations, 8
Oliver, J.J., 268, 319
Optimal codes, 63–66
– construction of, 69–72
– for integers, 93–96
– properties of, 76–77
Optimum quantizing lattice, 178, 285
Order, 337
Organisms, 7–8
Output tape, Turing machines, 101

Parameters, discrete, imprecise assertion of, 284–286
Parameter values, 150
Partial order equivalence, 330
Particle-state probability distribution, 342
Partitions of hypothesis space, 213–215
Past, 368–369
– of a computer process, 375
– induction of the, *see* Induction of the past
– regained, 369–370
– simulation of the, 381–382
Past disorder, deduction of, 345–355
Patrick, J.D., 275, 314–315
Pattern, 190
Perturbed data, normal distribution with, 264–265
PFSM, *see* Probabilistic finite-state machine
Philosophy of science, 1–2
Pipe model, 356
Poisson distribution, 269–270
Popper, K
Populatio
Possible c
Possible i
Posterior,
Posterior
Posterior
Posterior
Precision
Precision
Predictiv
Prefix co
Prefix pr
Present,
Present v
Previous
Price, H.,
Primitive
Primitive (*UTM*
Prior den
Prior dist
Prior esti
Prior pre
Prior pro
Prior pro
– coding
Priors, 3
– alterna
– conjug
– evoluti
– invaria
– maxim
– Norma estima
– origins
– singul
– theory
– Unifor
– uninfo
– Univer 124–1
Probabil (PFSM
– asserti
Probabil
Probabil
Probabil
– coding
– condit
Probabil
Probabil
– define

– particle-state, 342
– Turing, 103–104
Probability line, 74
Punctuated binary codes, 92–93

Quantization noise, 181
Quantizing lattices, 178, 285
Quantum mechanics, 340
Quinlan, J.R., 315

Random coding of estimates, 210
Random variables, 19–21
Range, 20
– known, uniform distribution of, mean of, 183–187
Raman, A., 314
Reality, 2
Records, 355–356
Redundant code, 79
Regression in categories, 320
Regular expressions, 127
Regular grammars
– inferences of, 305–314
– probabilistic, 308
Retrace equality, 348–349
Reversible laws, 340–341
Rissanen, J.J., 4–5, 73, 99–100, 401, 408–415
Rivest, R.L., 315
Ron, D, 321–325
Rounding-off quantum, 168
Roundoff error, 224
Roundoff vector, 241
Rumantir, G.W., 272

Sample mean, 260
Sample variance, 260
Scheines, R., 327
Schou, G., 268
Schou estimate, 268
Schwarz, G., 35
Science, 1
– Grand Theories of, 386–387
– philosophy of, 1–2
Scientific languages, 391
Second law of thermodynamics, 88
Set(s)
– discrete hypothesis, 156
– of possible inferences, 147–148
Set X of possible data, 144–145
Shannon, C.E., 4, 57–100, 103–110, 118, 121, 225
Shannon–Fano code, 70
Shannon information, 57–100
– algorithmic complexity versus, 107–110
Shannon measure, 104
Simplicity, 12
Singularities in priors, 237
Sloane, N.J.A., 178, 257
Small-sample message length, MML, 235–236
Smith, A.F.M., 47
Smith, B., 125
SMML, *see* Strict minimum message length estimators
Solomonoff, R.J., 3, 5, 401–408
Solomonoff prediction, 396, 404, 405
Spacing, 170–171
Spacing functions, 223–226
– family of, 242
Speech, 59
Spirtes, P., 327
Spherical von Mises–Fisher distribution, 268–269
SRM (Structural Risk Minimization), 274, 321–322
Standard deviation, 201
State space, 340
State variables, 134
Statistical critique, summary of, 54–55
Statistical inference, 3
Statistics, sufficient, *see* Sufficient statistics
Stick collision model, 352
Stirling's approximation, 86, 179
Stochastic Complexity, 409
Stop and start conditions, Turing machines, 102–103
Strict minimum message length (SMML) estimators, 143–196
– approximations to, summary, 218–219
– approximation to sequence problem, 325–326
– binomial example, 157–160
– defined, 143
– discrimination of, 192–195
– efficiency of, 189–192
– explanation for continuous data, 166–187
– explanation for discrete data, 153–166
– generality of, 189
– general properties of, 187–195
– minimizing relations for, 156–157
– problem definition, 144–153
– quadratic approximations to, 221–255
String length, expected, 64
Strings, non-random, 109
Structural equivalence, 330–331
Structural models, 305–336
Structural Risk Minimization (SRM), 274, 321–322
Subjective nature of information, 79–81
Success count ranges, 159
Sufficient statistics, 29, 161–163
– binomial example using, 163–164
– minimal, 163
Surrogate class label estimate, 288–289
Symbols, 58
Symmetry, 50

Target functions, 274
Teleological explanation, 367
Test statistic, 33
TETRAD II, 327, 335–336
Theories, 10
– approximate, 14
– demise of, 11–14
– Grand, of Science, 386–387
Theory description codes, 124
– as priors, 114–115
Theory descriptions, universal codes in, 115–116
Thom, A., 275
Time, Thermodynamic Arrow of, 337–384
Time reversal, 350
Time reversibility, 337, 340–341
TM, *see* Turing machines
TOM (Totally ordered model), 328–333
Totally ordered model (TOM), 328–333
Transformed parameter, 241
Transition arcs, 305
Tree
– coding structure, 316–318
– coding class distributions at leaves, 317–318
– leaves, 61
– levels, 64
Trinomial distribution, solution for, 165–166
Turing machines (TM), 101–102
– choice of, 103
– Educated (ETM), 130–131
– Universal, *see* Universal Turing Machines
Turing probability distributions, 103–104
Two-cl

Unary
Uncert
– in h
– mea
– via
– via
Uncla
vers
Unifor
mea
Unifor
Unifor
Uninf
Unive
– in t
Unive
105
– diff
– pri
– as
UTM
UTM
Ma

Vapn
Varia
– sar
V-C
Vecto
va
Verb
Visw
Vita
von
– cir
– sp
Voro
– he
Vovk
V'yu

Wisl
Witl
Witt
Woc
Wor
Wor
Wor

Zad
Zad
Zero