

## Index

- 2 ln L, *see* deviance
- $\chi^2$ , 286
- $\epsilon$ -greedy choice, 400, 403
- ABC, 163–170, 233
  - overview, 164
  - recognition memory example, 166
  - sampling, 164
  - summary statistic, 165
- accumulation rate, *see* drift rate
- accumulator, 37, 123, 387, 389, 410, 411, 419
- activation, 14–17, 20, 21, 41–43, 90, 334
- activation function
  - linear, 337
  - logistic, 357
- AIC, 258–261, 271, 273, 297, 299–302, 315, 319
  - differences, 259
- AIC vs BIC, 299–301
- AIC with small sample correction, *see* AICc
- AICc, 261
- Akaike weights, 260–261
- Akaike's Information Criterion, *see* AIC
- amnesia, 317, 318
- Approximate Bayesian Computation, *see* ABC
- artificial intelligence, 324
- associative memory, 334
- attentional refreshing, 41
- auto-associator, 314, 349–356
- autocorrelations, 147
- auxiliary assumptions, 311, 319, 333
- averaged data, 106–109
- averaged distributions, *see* Vincentized average
- backpropagation network, 356–365
  - in psychology, 364–365
  - internal representations, 363
- backpropagation of error, 361
- backpropagation through time, 364
- bandit task, 398
- Bayes factor, 276–277
- Bayes' theorem, 127, 128, 135, 273
- Bayesian Information Criterion, *see* BIC
- Bayesian parameter estimation, 102, 126–139
- Bernoulli process, 141
- best-fitting parameter estimates, 52
- Beta function, 135
- bias unit, in backpropagation, 359
- bias-variance trade-off, 245–247, 392
- BIC, 117, 271, 297–301, 315, 319
- Bonini's paradox, 323, 324
- bootstrapping, 65–70, 119
- boundary separation, 29, 122, 371, 387, 420
- Brain-State-in-a-Box model, *see* BSB
- BSB, 352–356
- buckets, 92
- burnin, 151, 175, 178, 180, 209
- burst vs buildup cells, 415
- catastrophic interference, in backpropagation
  - learning, 365
- categorization, 7–9, 13–17, 75, 353
  - distance and angle, 14
  - of faces, 7, 13, 14, 16–17, 87
  - of line lengths, 14
- $\chi^2$  distribution, *see* probability distribution
- $\chi^2$ , 249, 255
- choice RT, 24–26, 85, 86, 369–392
- choice task, 24–26, 369–392
- classification, *see* categorization
- cognitive load, 41
- cognitive psychometrics, 121, 122
- coin tossing, 90
- common coding, 324
- conceptual innovation, 313
- conceptual simulation, 313
- conditional distributions, 172, 175, 176
- conditional probability, *see* probability, conditional
- confidence interval, 67
- conjugacy, 136
- continuous variable, 76–78, 82, 204
- convergence problem, 161
- Copernicus, 3–6
- CRAN, 26
- cross-entropy, 256
- cross-modal integration, 244
- cross-validation, 262, 302, 321, 418

- cued recall, 334, 336, 342  
cumulative distribution function, 77, 79, 98, 168, 183, 220, 376  
cumulative prospect theory, *see* prospect theory
- Darwin's theory of evolution, *see* evolution  
data averaging, *see* averaged data  
data description, 12  
data model, 91  
data vector, 82  
debugging, 327–328  
deep belief networks, 365  
delay discounting, 218, 219, 222, 418  
derivative, 79, 263, 268, 295, 296, 361  
determinant, 272, 296, 297  
deviance, 96, 249, 256, 258, 297, 298  
DIC, Deviance Information Criterion, 302, 419  
differential evolution, ABCDE, 170  
diffusion model, 121, 316, 369, 412, 414, 419  
    Bayesian approaches, 392  
    drift rate, 370  
    EZ, 392  
    falsifiability, 385  
    fitting, 371  
    interpretation, 383  
    quantile probability functions, 371  
diminishing utility, 250  
discount factor, 402  
discounting of the future, future discounting,  
    *see* delay discounting  
discrepancy function, 47–49, 62  
discrete variable, 75, 84, 204, 256  
distance, 14–15  
distributed representation, 335  
distribution  
    posterior, 137  
dopamine neurons, 404, 405, 409  
drift diffusion model, 121, 122  
drift rate, 26, 28, 31, 32, 34, 236, 316, 370, 376,  
    384–387, 411–415, 420
- EEG, 414, 418–420  
eigenvalue, 351  
eigenvector, 351, 353  
eligibility traces, 404, 409  
EM, *see* Expectation-Maximization  
error surface, 51, 59  
errors  
    fast, 33, 35  
    slow, 33, 36  
Euclidian distance, 89, 90, 339  
event, 73  
evidence, 19, 24–26, 28, 37, 72, 128, 134, 147, 172,  
    256, 273, 274, 276, 300, 319, 386, 387  
evolution, 18  
exemplar model, 8, 9, 13, 18, 75, 320, 397
- Expectation-Maximization (EM), 115–117  
exponential function  
    relationship to logarithm, 95, 249  
exponential learning, 10–11  
express saccades, 114  
EZ diffusion model, 392
- falsifiability, 18–20, 264, 317, 385  
FEF, 410, 411, 413  
Fisher information, 263, 272, 295, 297, 299  
fMRI, 409, 411–415, 418  
forgetting, 12, 211, 320  
    exponential function, 212, 278, 287, 292  
    interference-based, 320  
    power function, 212, 278, 287, 292  
    time-based, 320  
free recall, 118, 316  
frontal eye field, *see* FEF  
functional magnetic resonance imaging,  
    *see* fMRI
- $G^2$ , 249, 257  
Gaussian mixture model, 113–117  
Gaussian quadrature, 278  
GCM, 8, 13–18, 75–76, 87–91, 97–101, 271, 307,  
    316, 320, 397  
generalization, 345  
generalized linear mixed models, 228  
Gibbs sampling, 172–177  
    ABC, 234  
    bivariate example, 173  
    vs. Metropolis-Hastings, 176  
git, 328–330  
    collaboration, 329  
global minimum, 60  
goodness-of-fit, 258, 260  
gradient descent, 361, 365  
graphical models, 204–206  
    signal-detection, 204  
grid search, 52, 59, 70  
grouping, in memory, 119
- harmonic mean estimator, 281  
Hebbian associator  
    limitations, 356  
Hebbian learning, 337, 338, 349  
    limitations, 356  
heliocentric model, 3, 5, 6  
Hessian matrix, 271, 272, 295, 296  
hierarchical model, 22, 23, 103, 122, 203–234, 287,  
    301–303  
hierarchical modeling  
    advantages, 233  
    concepts, 203  
    forgetting, 211  
    inter-temporal preferences, 218

- maximum likelihood, 228
- shrinkage, 211
- signal detection, 207
  - Bayesian vs. maximum likelihood, 233
  - maximum likelihood, 228
  - Stein's paradox, 211
- high-threshold model, 186, 269, 281, 284
- hyper-hyperparameters, 203
- hyperparameters, 152, 168, 417, 418
  
- identifiability, 102, 264–269, 315
- identifiability problem, 321
- importance sampling, 280–284
- indicator variable, 287, 294
- individual differences, 94, 121–122, 206, 209, 211, 226, 233, 234, 316, 385
- information, 144, 256, 297–299
- integration, 79, 263, 273–303
- inter-temporal preference, 218, 226, 418
- IPS, 413
- irrelevant specification problem, 320
  
- Jacobian matrix, 268, 269, 315
- JAGS, 163, 172
  - convergence diagnostics, 184
  - declarative language, 179
  - forgetting, 212
  - installation, 177
  - inter-temporal preferences, 221
  - signal-detection model, signal-detection model, 182
  - specify prior distribution, 180
- joint probability, *see* probability, joint
  
- K-means clustering, 118–120
  - gap statistic, 119
- Kepler's model, 6
- knitr, 330
- Kullback-Leibler distance, 256–258
  - expected, 258, 259, 261
  
- Laplace approximation to marginal likelihood, 294–297
- latency, 102, 103, 121
- LATER model, Linear Approach to Threshold with Ergodic Rate, 410
- lateral interparietal cortex, *see* LIP
- LBA, 316, 386–392, 412–414, 417–419
  - fitting, 388
- learning rate, 337, 357, 400
- likelihood, 80–128
  - vs. probability, 83
  - colloquial usage, 72
  - definition, 82
  - fitting, *see* maximum likelihood estimation
  - joint, 93
  - likelihood surface, 95
  - multiple parameters, 94
  - multiple participants, 93, 94
- likelihood function, 82, 95, 128, 130, 131
  - regularity, 101
- likelihood ratio, 249, 260
- likelihood ratio test, 249, 255
- Likert scale, 48, 75
- Linear Ballistic Accumulator, *see* LBA
- linear ballistic accumulator, *see* LBA
- linear regression, 50
- LIP, 411, 419
- local minimum, 60
- local necessity, 319–321
- localist representation, 335
- log-likelihood, 95–258
  - additive properties, 95
  - joint, 96, 252
  - negative, 97
  - simplifying, 96–261
- log-likelihood function
  - normal, 96
- logarithm, 95
- logarithm function, 260
  - natural log, 95
  - relationship to exponential, 95
- loss aversion, 251, 255
- LRT, *see* likelihood ratio test
- Luce's choice rule, 16, 17, 75
  
- Makefile, 331
- marginal likelihood, 130, 134–135, 147, 172, 273–274, 276, 277, 277–303
- marginal probability, 129
- Markov Chain, 147
- Marr's levels of analysis, 396, 422
- matrix algebra, 339–342, 348–349
  - addition and subtraction, 339
  - distributive property, 348
  - inner product, 341, 345
  - multiplication, 340, 344
  - non-commutative, 340
  - outer product, 341
- maximum likelihood estimation, 95–103, 126
  - asymptotic normality, 102
  - biased estimators, 102
  - bounded parameters, 99
  - consistency of estimators, 102
  - efficiency of estimators, 102
  - fitting individuals, 111–112
  - minimization vs. maximization, 95
  - normalized, *see* normalized maximum likelihood
  - overdispersion of estimates, 103
  - parameterization invariance, 102
  - properties of estimators, 101

- MCMC, 146  
 autocorrelations, 147, 162  
 chain, 149, 158, 175  
 chain, MCMC  
   convergence, 183  
 convergence diagnostics, 162, 184  
 convergence problem, 161  
 fit mixture model, 156  
 JAGS, 163, 172  
 Metropolis-Hastings, 148  
 multiple chains, MCMC  
   starting values, 161  
 problems, 161  
 proposal distribution, 148  
 scale reduction factor, 186  
 shrink factor, 186  
 target distribution, 149  
 thinning, 162  
 tuning parameter, 149  
 measurement model, 265, 270, 315  
 mental rotation, 419  
 mercurial, 328  
 Metropolis-Hastings algorithm, 148  
 minimum description length, 262–263, 272, 296  
 mixture model, 113–117, 154–160, 180, 235  
   Gaussian, *see* Gaussian mixture model  
 model, 6  
   choice task, 24  
   cognitive aid, 20–22  
   descriptive, 9–13  
   multinomial processing tree, 22  
   process explanation, 17  
   quantitative, 5  
   random-walk, 25–37  
     boundary, 28  
     drift rate, 28, 32  
     Trial-to-trial variability, 33  
   sequential-sampling, 33, 37, 369  
   toolkit  
     discrepancy function, 47–49  
     vs. intuition, 21  
 model comparison, 6, 9, 103, 248, 276, 319, 320, 332  
 model complexity, 243–245, 247, 258, 274–276, 300, 315  
   out of set prediction, 247  
   polynomial, 243–244  
 model exploration, 312–314, 324  
 model flexibility, *see* model complexity, *see* flexibility  
 model weights, 261  
 model-based cognitive neuroscience, 396  
 modeling  
   toolkit, 38  
   parameters, 38  
   momentum, in learning by backpropagation, 361  
 Monte Carlo integration, 280  
 Monte Carlo methods, 146  
 MPT, 22, 186  
   eyewitness testimony, 190  
   no-conflict model, JAGS  
   no-conflict model, 192  
 multidimensional scaling, 88, 363  
 multilevel modeling, 203  
   advantages, 233  
   concepts, 203  
   forgetting, 211  
   inter-temporal preferences, 218  
   maximum likelihood, 228  
   parent distribution, 203  
   shrinkage, 211  
   signal detection, 207  
     Bayesian vs. maximum likelihood, 233  
     maximum likelihood, 228  
   Stein's paradox, 211  
 multinomial distribution, *see* probability distribution, multinomial  
 multinomial processing tree, *see* MPT  
 multiple participants  
   likelihood, *see* likelihood, multiple participants  
 mutual exclusivity, 73  
  
 NaN, not a number, 328  
 necessity, 318–321  
   local, *see* local necessity  
 nested models, 249, 255, 284, 300, 320  
 neural network, 314, 316  
   activation, 335, 336  
   graceful degradation, 346  
   input layer, 335  
   lesioning, 345  
   output layer, 335  
   pattern of activation, 335  
 neuroscience, criticisms of, 395  
 normalized maximum likelihood, 263  
 normative behavior, normative expectations, 12  
 null hypothesis, 186, 249, 255, 256, 284, 285, 301, 304  
 numerical integration, 278–279  
 numerical overflow, 299  
  
 Occam's razor, 244, 274  
 outcome, 73, 80  
 outcome space, 19  
 output interference, 320  
 over-fitting, 243, 245  
  
 parameter estimates  
   best-fitting estimates, 52  
   variability, 65

- parameter estimation, 50–64, 257
- bounded parameters, 99
  - by eye, 321
  - error surface, 51
  - general limitation, 60
  - global minimum, 60
  - grid search, 52
  - histogram of estimates, 67
  - least-squares, 50
  - local minimum, 60
  - maximum likelihood, 50, *see* maximum likelihood estimation, 126
  - parameter estimates
    - variability, 65
  - simulated annealing, 61
  - starting values, 56
- parameter recovery, 315
- parameter vector, 80
- parameters, 10, 38, 39
- best-fitting, 39
  - continuous, 59
  - contribution to model behavior, 314–315
  - discrete, 59
  - fixed, 39
  - free, 39
- parametric resampling, 65
- parent distribution, 203, 207
- parsimony, 39, 258, 274, 276, 300, 322, 323
- phasic bursts, in dopamine neurons, 405, 408
- plain text, vs rich text, 326
- planetary motion, retrograde motion, 3
- posterior distribution, 135, 137
- posterior predictive distribution, 215
- posterior probability, 129
- power law, 10–12, 108
- pre-SMA, 412, 414, 417, 420
- pre-supplementary motor area, *see* pre-SMA
- precision (vs. standard deviation), 159
- prediction error, 397, 400, 405, 408, 409
- present subjective value, 219
- primacy gradient, 42
- prior, 411
- conjugate, 136
  - data-informed, 299, 304
  - determining, 139
  - Haldane, 142, 303
  - hyperparameters, 152
  - hyperparameters, ABC
    - hyperparameters, 168
  - improper, 303
  - informative, 304, 305
  - Jeffrey's prior, 263
  - Jeffreys, prior
    - non-informative, 159
  - non-informative, 303, 305
  - non-uniform, 151
  - pseudoprior, *see* pseudoprior reference, 142
  - sensitivity analysis, *see* sensitivity analysis
  - specification of, 299, 303–305
  - subjective, 303, 304
  - unit information, 297–299
- prior distribution, 102, 131
- uniform, 140
- prior predictive, 304
- prior predictive distribution, 304
- prior probability, 128
- priority heuristic, 258–259, 298
- probability, 72, 80, 92
- axioms, 73
  - conditional, 74, 126
  - definition, 72
  - joint, 73, 74, 82, 127, 130
  - marginal, 129
  - of observed data, 128
  - posterior, 129
  - prior, 128
  - properties of, 73–75
- probability density function, 78–80, 85, 95, 256
- probability distribution, 75–80
- $\chi^2$ , 249
  - Bernoulli, 94, 130, 131, 222
  - beta, 132–135, 141, 144, 146, 284, 288, 292, 301
  - bimodal, 113
  - binomial, 67, 75, 84, 85, 90, 91, 93, 95, 131, 205, 211, 214, 286, 288
  - ex-Gaussian, 102
  - exponential, 95
  - multinomial, 92–93, 95, 261
  - multivariate Gaussian, 296
  - multivariate normal, 173, 298
  - normal, 35, 95, 96, 98, 102, 115, 151, 166, 168, 189, 200, 201, 220, 242, 244, 274, 302, 387, 398
  - von Mises, 157
  - Wald, 77, 78, 86, 94
  - Weibull, 109
- probability distribution function, *see* cumulative distribution function
- probability functions, 75–80
- probability integral transform, 169
- probability mass function, 75–76, 80, 85, 92, 95
- probit regression, 230
- process models, 13
- product space method, 287
- proposal distribution, 148, 149
- prospect theory, 249–255, 298
- cumulative, 251
  - probability weighting function, 251
  - value function, 250
- prototype model, 9, 271
- pseudoprior, 291–292

- psychophysics, 241  
   polynomial law, 241–243  
 Ptolemaic system, 3, 5, 6  
 Pythagorean theorem, 15
- Q values, 399–401  
 quantile, 67–68  
 quantile probability functions, 371  
 quantile probability plots, 372  
 quasi-identifiability, 266
- R, 26  
   apply, 89, 111  
   drawing random samples, 28  
   lapply, 90  
   log, 95  
   loops, 28  
   matrix, 29  
   packages, 331  
   packrat, 332  
   pnorm, 98, 286  
   sample, 400  
   sourcing, 87  
 random walk, 25–37  
 re-parameterization, 266–267  
 reading, 316–317  
 reasoning, 121  
 recency effect, 314  
 recognition, 7–9, 166, 264, 317, 318  
   familiarity, 166, 318  
 reduced model, *see* nested models  
 reference priors, 142  
 regression, 50  
 regularity effect, 317, 357, 362  
 reinforcement learning, 398–410  
   action value learning, 398–401  
   future expected reward, 402  
   state-action value learning, 401–404  
 rejection sampling, 170  
 relative strength of evidence, *see* strength of evidence  
   evidence  
 remember-know, 317  
 replicability, *see* reproducibility  
 repository, 26, 329, 330  
 reproducibility, 20  
 resampling procedures, *see* bootstrapping  
 response boundary, 28  
 response suppression, 314–315  
 response time, *see* latency  
 RMSD, 48, 50, 169  
 RNG, random number generator, 331  
 rolling dice, 18  
 RStudio, 26
- saccadic decision-making, 410–411  
 sample, 73  
 sample space, 73  
 sampling distribution, 65, 67, 255  
 sampling variability, 79, 90  
 SARSA, 403–404  
 SARSA( $\lambda$ ), 403  
 Savage-Dickey ratio, 284, 301  
 scientific reasoning, 20  
 scope, 18–20  
 sensitivity analysis, 304  
 sequential-sampling models, 33, 37, 369  
 serial recall, 314  
 sharing code, 325–326, 329  
 short-term memory, 320, 321  
 shrinkage, 211  
 signal detection theory, 166, 269, 281, 285, 302, 315, 317  
 similarity, 14–15  
 SIMPLE, 320  
 Simplex, 57–60, 96  
   limitations, 59  
 Simpson's paradox, 106  
 simulated annealing, 61  
   acceptance function, 61, 62  
   candidate function, 61  
   cooling schedule, 63  
   temperature, 62  
 single cell recording, 410, 411, 415–417  
 skill acquisition, 10–12  
 slow errors, 36  
 softmax, 252  
 speed-accuracy trade-off, 412, 417  
 speeded-choice task, 24–26, 369–392  
 spike train, *see* single cell recording  
 spreading activation theory, 20, 21  
 stage-like behavior, 317  
 Stan, 172  
 starting point, 31, 34, 35, 370, 371, 375, 386–388, 392, 411–413  
 starting values, 52, 56, 100, 161  
 Stein's paradox, 211  
 Sternberg task, 264  
 strength of evidence, 256, 260, 261, 319  
 striatum, 409, 412, 417, 420  
 strong inference, 318  
 structural equation modeling (SEM), 121, 316  
 Subplex, 60  
 sufficiency, 316–318  
 superior colliculus, 411
- taxation, 10  
 Taylor series expansion, 295  
 TBRS, 41  
 TBRS\*, 41  
 temporal difference learning, 403  
 testability, *see* falsifiability, 269–270, 315  
 thinning, 162

- 
- thought experiment, *see* conceptual simulation  
threshold, 25, 26, 42, 77, 86, 97, 123, 187, 236, 318, 359, 370, 387, 388, 410–415  
time-based resource-sharing model, *see* TBRS  
toolkit, 38  
    discrepancy function, 47–49  
    parameters, 38  
tract strength, 417  
Transdimensional MCMC, 287–294  
Trial-to-trial variability, 33, 370, 371, 392  
    drift rate, 36, 37, 370  
    starting point, 34, 370  
truth, 256, 257, 322  
tuning parameter, 149  
  
uncertainty, 248, 256  
unlearning, 314  
utility, 252  
  
value function, 250  
vector cosine, 338  
vector length, *see* vector norm  
vector norm, 339, 341  
vector normalization, 339  
verbal theorizing, 6, 9, 41, 311  
verisimilitude, 323  
version control, 328–329  
Vincentized average, 109–111  
visual working memory, 154  
  
WAIC, widely applicable information criterion, 302  
Walsh matrix, 349  
Widrow-Hoff learning, 360  
WinBUGS, 172  
word frequency, 317  
word naming, 316, 317, 356  
working memory, 121, 154