

---

# Subject Index

---

- adaptive MCMC, 308
- adding parameters to a model, 185–186
- adequate summary, 217, 232
- adolescent smoking survey, 148–150
- AECM algorithm, 324, 348
- airline fatalities, 59, 82
- Akaike information criterion (AIC), 172, 177
  - discussion, 182
  - educational testing example, 179
- Alcoholics Anonymous survey, 213–214
- aliasing, 89, 521, 524, 533
- all-at-once Gibbs sampler for hierarchical regression, 393
- alternating conditional sampling, *see* Gibbs sampler
- analysis of variance (Anova), 114, 395–398, 402, 403
  - finite-population and superpopulation models, 396–397
  - fixed and random effects, 396–397
  - for hierarchical logistic regression, 423
  - internet example, 398
  - notation, 395
- ancillary test statistic, 151
- approximate Bayesian computation (ABC), 344
- approximate inference, 263
- approximations based on posterior modes, 311–330
- asymptotic theorems, 87–88, 206, 215–216, 232
  - counterexamples, 89–91
  - proofs, 585–588
- Australia, adolescent smoking survey in, 148–150, 162, 211–212
- auxiliary variables for computation, 297–299, 309
- basis functions, 487–499
  - Gaussian, 487
  - multivariate, 495–498
  - selection and shrinkage, 490–494
  - splines, 488–490
- Bayes factor, 182–184, 193
- Bayes’ rule, 6, 20
  - discrete examples, 8–11, 245
  - original example, 30
- Bayesian data analysis, three steps of, 3
- Bayesian filtering and smoothing, 516
- Behrens–Fisher problem, 80
- belief functions, 98
- Bernoulli distribution, 584
- Bernoulli trials, 29, 147
- beta distribution, 30, 34, 60, 578, 582
- beta-binomial distribution, 578, 584
  - as overdispersed alternative to binomial, 438
- beta-blockers, example of meta-analysis, 124–128
- betting and probability, 13–16
- bias, 94, 99
  - compared to miscalibration, 128
  - difficulties with the notion, 94
  - prediction vs. estimation, 94, 401
- ‘biased-coin’ design, 235
- BIC (‘Bayesian’ information criterion), 175
- bicycle traffic, 81, 136
- binned residual plots, 157–158
- binomial distribution, 578, 583
- binomial model, 29, 37, 80
  - posterior predictive check, 147–148
- bioassay experiment, 74–79, 82
  - normal approximation, 86
- birthdays, example of Gaussian process modeling, 505–510
- births, proportion girls, 37–39
- blockwise Gibbs sampler for hierarchical regression, 392
- bootstrap, 96
- Box-Cox transformation, 188–191, 194–195
- bridge sampling, 347, 348
- Bugs, *see* software
- burn-in for MCMC, why we prefer the term warm-up, 282
- business school grades, hierarchical multivariate regression, 391–392
- calibration, 128
- calibration of probability estimates, 16–19
- cancer maps, 47–51
- capture-recapture sampling, 233

- Cauchy model, 59, 437
- causal inference, 4, 214, 223–224
  - instrumental variables, 224
  - observational studies, 220
    - incumbency example, 358–364
  - principal stratification, 223–224
  - randomized experiments, 214, 231
  - using regression models, 365
- cavity distribution in expectation propagation, 339
- censored data, 61, 224–228
- Census, 422, 466
  - record linkage, 16–19
- central composite design integration (CCD), 344
- chess competition, example of paired comparisons, 427
- $\chi^2$  distribution, 576, 581
- Chinese restaurant process, 550
- Chloride example, 489, 492, 494
- Cholesky factor (matrix square root), 356, 580
- classical methods
  - confidence intervals, 92, 95
  - frequency evaluations, 91
  - hypothesis testing, 145, 150
  - maximum likelihood, 93
  - multiple comparisons, 96
  - nonparametric inference, 96
  - normal-theory inference, 83
  - point estimation, 85, 91
  - standard errors, 85
  - unbiased estimates, 94
  - Wilcoxon rank test, 97
- cluster sampling, 210–212, 232
- coarse data, 230
- cockroach allergen data, 472
- coefficient of variation (CV), 6
- coherence, 13
- coin tossing, 12, 26
- collinearity, 365
- colochos, xiv
- complementary log-log link, 407
- complete data, 199
- complete-data likelihood, 200
- completely randomized experiments, 214–216
- computation, *see* posterior simulation
- computer programs, *see* software
- conditional maximization, *see* posterior modes
- conditional posterior distribution, 122, 325
- confidence intervals, 3, 92, 95
- conjugacy, *see* prior distribution
- conjugate gradient optimization, 313
- consistency, 88, 91
- contingency tables, 428–431
  - with missing data, 462
- continuous models for discrete data, 458
- contour plots, 76, 111, 112
  - and normal approximation, 85
- control variable, 353
- convergence of iterative simulation, 281–286
- covariance matrix, 20, 71
  - for a Gaussian process, 501
  - for a sum of Gaussian processes, 506
  - inverse-Wishart distribution, 72–74, 390, 576, 582
  - literature on models for, 401
  - LKJ distribution, 576, 582
  - scaled inverse-Wishart model, 74, 390
  - Wishart distribution, 576, 582
- covariates, *see* regression models, explanatory variables
- cow feed experiment, 217–218, 379
- cross-validation, 175–177
  - discussion, 182
  - educational testing example, 179
- crude estimation, 76, 263
  - bioassay example, 76
  - educational testing example, 114
  - rat tumor example, 103
  - schizophrenia example, 523, 526
  - Slovenia survey, 463
- curse of dimensionality, 495
- CV, coefficient of variation, 6
- data augmentation, 293
- data collection, 197–236
  - censored and truncated data, 224–228
  - experiments, 214–220
  - formal models, 199–202
  - ignorability, 202–205
  - observational studies, 220–224
  - randomization, 218–220
  - sample surveys, 205–214
- data reduction, 85
- de Finetti’s theorem, 105, 134
- debugging, 270–271
  - comparing inferences from several models, 482
  - EM algorithm, 321
  - in Stan and R, 605–606
- decision analysis, 12, 26, 99, 237–258
  - and Bayesian inference, 237–239
  - medical screening example, 245–246
  - personal and institutional perspectives, 256

- radon example, 246–256
- survey incentives example, 239–244
- utility, 238, 245, 248, 256
- decision trees, 238, 245, 252
- delta method, 99
- density regression, 568–571
- dependent Dirichlet process (DDP), 562–564, 572
- derivatives, computation of, 313
- design of surveys, experiments, and observational studies, 197–236
- designs that ‘cheat’, 219
- deviance, 192
- deviance information criterion (DIC), 172–173, 177, 192
  - discussion, 182
  - educational testing example, 179
- differences between data and population, 207, 221, 223, 237, 422
- differential equation model in toxicology, 477–485
- dilution assay, example of a nonlinear model, 471–476, 485
- dimensionality, curse of, 495
- Dirichlet distribution, 69, 578, 583
- Dirichlet process, 545–574
- Dirichlet process mixtures, 549–557
- discrepancy measure, 145
- discrete data
  - adapting continuous models, 458
  - latent-data formulation, 408
  - logistic regression, 406
  - multinomial models, 423–428
  - Poisson regression, 406
  - probit regression, 406
- discrete probability updating, 9, 245
- dispersion parameter for generalized linear models, 405
- distinct parameters and ignorability, 202
- distribution, 575–584
  - Bernoulli, 584
  - beta, 30, 34, 60, 578, 582
  - beta-binomial, 60, 578, 584
  - binomial, 578, 583
  - Cauchy, 98
  - $\chi^2$ , 576, 581
  - Dirichlet, 69, 578, 583
  - double exponential, 368, 493, 576
  - exponential, 576, 581
  - gamma, 45, 576, 581
  - Gaussian, *see* normal distribution
  - inverse- $\chi^2$ , 576, 581
  - inverse-gamma, 43, 576, 581
  - inverse-Wishart, 72, 576, 582
  - Laplace, 368, 493, 576
  - LKJ correlation, 576, 582
  - log-logistic, 578
  - logistic, 578
  - lognormal, 576, 580
  - long-tailed, 435
  - multinomial, 578, 584
  - multivariate normal, 79, 576, 580
    - marginals and conditionals, 580
  - multivariate  $t$ , 319, 578
  - negative binomial, 44, 132, 578, 584
  - normal, 575, 576
  - normal-inverse- $\chi^2$ , 67, 82
  - Pareto, 493
  - Poisson, 578, 583
  - scaled inverse- $\chi^2$ , 43, 65, 576, 581
  - $t$ , 66, 578, 582
  - uniform, 575, 576
  - Weibull, 576, 581
  - Wishart, 576, 582
- divorce rates, 105, 135
- dog metabolism example, 380
- dose-response relation, 74
- double exponential distribution, 368, 493, 576
- $E_{\text{old}}$ , 320
- ECM and ECME algorithms, 323, 348
- educational testing experiments, *see* SAT coaching experiments
- effective number of parameters, 169–182
  - educational testing example, 179
- effective number of simulation draws, 286–288
- efficiency, 91
- eight schools, *see* SAT coaching experiments
- elections
  - forecasting presidential elections, 165–166, 171–172, 383–388
  - incumbency in U.S. Congress, 358–364
  - polling in Slovenia, 463–466
  - polling in U.S., 422–423, 456–462
  - probability of a tie, 27
- EM algorithm, 320–325
  - AECM algorithm, 324
  - as special case of variational inference, 337
  - debugging, 321
  - ECM and ECME algorithms, 323, 348
  - for missing-data models, 452
  - parameter expansion, 325, 348
  - SEM and SECM algorithms, 324–325
- empirical Bayes, why we prefer to avoid the term, 104

- environmental health
  - allergen measurements, 472
  - perchloroethylene, 477
  - radon, 246
- EP, *see* expectation propagation
- estimands, 4, 24, 267
- exchangeable models, 5, 26, 104–108, 230
  - and explanatory variables, 5
  - and ignorability, 230
  - no conflict with robustness, 436
  - objections to, 107, 126
  - universal applicability of, 107
- expectation propagation, 338–343
  - cavity distribution, 339
  - extensions, 343
  - logistic regression example, 340–343
  - moment matching, 339
  - picture of, 342
  - tilted distribution, 339
- experiments, 214–220
  - completely randomized, 214–216
  - definition, 214
  - distinguished from observational studies, 220
  - Latin square, 216
  - randomization, 218–220
  - randomized block, 216
  - sequential, 217, 235
- explanatory variables, *see* regression models
- exponential distribution, 576, 581
- exponential families, 36, 338
- exponential model, 46, 61
- external validation, 142, 167
  - record linkage example, 17
  - toxicology example, 484
- factorial analysis, internet example, 397–398
- Federalist papers, 447
- finite-population inference, 200–203, 205–209, 212, 214–216, 232
  - in Anova, 396–397
- Fisher information, 88
- fixed effects, 383
  - and finite-population models in Anova, 397
- football point spreads, 13–16, 26, 27
- forecasting presidential elections, 142, 383–388
  - hierarchical model, 386
  - problems with ordinary linear regression, 385
- frequency evaluations, 91–92, 98
- frequentist perspective, 91
- functional data analysis, 512–513
- gamma distribution, 45, 576, 581
- Gaussian distribution, *see* normal distribution
- Gaussian processes, 501–517
  - birthdays example, 505–510
  - golf putting, 517
  - latent, 510–512
  - logistic, 513–515
- gay marriage data, 499
- generalized linear models, 405–434
  - computation, 409–412
  - hierarchical, 409
  - hierarchical logistic regression, 422–423
  - hierarchical Poisson regression, 420–422
  - overdispersion, 407, 431, 433
  - prior distribution, 409
  - simple logistic regression example, 74–78
- genetics, 8, 183
  - simple example of Bayesian inference, 8–9, 27
- geometric mean (GM), 6
- geometric standard deviation (GSD), 6
- Gibbs sampler, 276–278, 280–281, 291
  - all-at-once for hierarchical regression, 393
  - assessing convergence, 281–286
  - blockwise for hierarchical regression, 392
  - efficiency, 293–295
  - examples, 289, 440, 465, 528
  - hierarchical linear models, 288–290, 392–394, 396
  - parameter expansion for hierarchical regression, 393, 396
  - picture of, 277
  - programming in R, 596–606
  - special case of Metropolis-Hastings algorithm, 281
- girl births, proportion of, 37–39
- global mode, why it is not special, 311
- GM (geometric mean), 6
- golf putting
  - Gaussian process, 517
  - nonlinear model for, 486, 499
- goodness-of-fit testing, *see* model checking
- graphical models, 133
- graphics
  - examples of use in model checking, 143, 144, 154–158
  - jittering, 14, 15, 27
  - posterior predictive checks, 153–159
- grid approximation, 76–77, 263

- GSD (geometric standard deviation), 6
- Hamiltonian (hybrid) Monte Carlo, 300–307, 601–605  
   hierarchical model example, 305–307, 601–605  
   leapfrog algorithm, 301  
   mass matrix, 301  
   momentum distribution, 301  
   no U-turn sampler, 304  
   programming in R, 601–605  
   tuning, 303
- heteroscedasticity in linear regression, 369–376  
   parametric model for, 372
- hierarchical Dirichlet process (HDP), 564–566
- hierarchical linear regression, 381–404  
   computation, 392–394, 396  
   interpretation as a single linear regression, 389
- hierarchical logistic regression, 422–423
- hierarchical models, 5, 101–137, 381–404  
   analysis of variance (Anova), 395  
   binomial, 109–113, 136  
   bivariate normal, 209–210  
   business school grades, 391–392  
   cluster sampling, 210–212  
   computation, 108–113  
   forecasting elections, 383–388  
   logistic regression, 422–423  
   many batches of random effects  
     election forecasting example, 386  
     polling example, 422–423  
   meta-analysis, 124–128, 423–425  
   multivariate, 390–392, 423–425, 456–462  
   no unique way to set up, 389  
   normal, 113–128, 288–290, 326–330  
   NYPD stops, 420–422  
   pharmacokinetics example, 480–481  
   Poisson, 137, 420–422  
   pre-election polling, 209–210  
   prediction, 108, 118  
   prior distribution, *see* hyperprior distribution  
   radon, 246–256  
   rat tumor example, 109–113  
   SAT coaching, 119–124  
   schizophrenia example, 524–533  
   stratified sampling, 209–210  
   survey incentives, 239–244
- hierarchical Poisson regression, 420–422
- hierarchical regression, 381–404  
   prediction, 387
- highest posterior density interval, 33
- HMC, *see* Hamiltonian Monte Carlo
- horseshoe prior distribution for regression coefficients, 378
- hybrid Monte Carlo, *see* Hamiltonian Monte Carlo
- hyperparameter, 35, 101, 105
- hyperprior distribution, 107–108  
   informative, 480–481  
   noninformative, 108, 110, 111, 115, 117, 135, 424, 526
- hypothesis testing, 145, 150
- identifiability, 365
- ignorability, 202–205, 230, 450  
   and exchangeability, 230  
   incumbency example, 359  
   strong, 203
- ignorable and known designs, 203
- ignorable and known designs given covariates, 203
- ignorable and unknown designs, 204
- iid (independent and identically distributed), 5
- ill-posed systems  
   differential equation model in toxicology, 477–485  
   mixture of exponentials, 486
- importance ratio, 264
- importance resampling (sampling-importance resampling, SIR), 266, 271, 273, 319  
   examples, 441, 442  
   why you should sample without replacement, 266
- importance sampling, 265, 271  
   bridge sampling, 347, 348  
   for marginal posterior densities, 440  
   path sampling, 347–348  
   unreliability of, 265
- improper posterior distribution, *see* posterior distribution
- improper prior distribution, *see* prior distribution
- imputation, *see* multiple imputation
- inclusion indicator, 200, 449
- incumbency advantage, 358–364  
   two variance parameters, 374
- indicator variables, 366  
   for mixture models, 519
- inference  
   discrete examples, 8–11  
   one of the three steps of Bayesian data analysis, 3
- inference, finite-population and superpopulation, 201–202, 212, 214  
   completely randomized experiments, 215–216, 232

- in Anova, 396–397
  - pre-election polling, 208–209
  - simple random sampling, 205–206
- information criteria, 169–182
- information matrix, 84, 88
- informative prior distribution
  - alternative to selecting regression variables, 367–369
  - spell checking example, 10
  - toxicology example, 480
- institutional decision analysis, 256
- instrumental variables, 224
- integrated nested Laplace approximation (INLA), 343
- intention-to-treat effect, 224
- interactions
  - in basis-function models, 497
  - in Gaussian processes, 504, 511
  - in loglinear models, 429
  - in regression models, 242, 367
- internet connect times, 397–398
- intraclass correlation, 382
- inverse probability, 56
- inverse- $\chi^2$  distribution, 576, 581
- inverse-gamma distribution, 43, 576, 581
- inverse-Wishart distribution, 72, 576, 582
- iterative proportional fitting (IPF), 430–431
- iterative simulation, *see* Markov chain Monte Carlo, 293–310
- iterative weighted least squares (EM for robust regression), 444
  
- jackknife, 96
- Jacobian, 22
- Jeffreys' rule for noninformative prior distributions, 52–53, 57, 59
- jittering, 14, 15, 27
- joint posterior distribution, 63
  
- Kullback-Leibler divergence, 88, 331–336, 585–587
  - connection to deviance, 192
  
- label switching in mixture models, 533
- Laplace distribution, 368, 493, 576
- Laplace's method for numerical integration, 318, 348
- large-sample inference, 83–92
- lasso (regularized regression), 368–369, 379
- latent continuous models for discrete data, 408
- latent-variable regression, 515
- Latin square experiment, 216–217
- LD50, 77–78
  
- leapfrog algorithm for Hamiltonian Monte Carlo, 301
- leave-one-out cross-validation, 175–177
  - discussion, 182
  - educational testing example, 179
- life expectancy, quality-adjusted, 245
- likelihood, 7–10
  - complete-data, 200
  - observed-data, 201
- likelihood principle, 8, 26
  - misplaced appeal to, 198
- linear regression, 353–380, *see also* regression models
  - $t$  errors, 444–445
  - analysis of residuals, 361
  - classical, 354
  - conjugate prior distribution, 376–378
    - as augmented data, 377
  - correlated errors, 369–376
  - errors in  $x$  and  $y$ , 379, 380
  - heteroscedasticity, 369–376
    - parametric model for, 372
  - hierarchical, 381–404
    - interpretation as a single linear regression, 389
  - incumbency example, 358–364
  - known covariance matrix, 370
  - model checking, 361
  - posterior simulation, 356
  - prediction, 357, 364
    - with correlations, 371
  - residuals, 358, 362
  - robust, 444–445
  - several variance parameters, 369–376
    - weighted, 372
- link function, 405, 407
- LKJ correlation distribution, 576, 582
- location and scale parameters, 54
- log densities, 261
- log-logistic distribution, 578
- logistic distribution, 578
- logistic regression, 74–78, 406
  - for multinomial data, 423
  - hierarchical, 422–423
  - latent-data interpretation, 408
- logit (logistic, log-odds) transformation, 22, 125
- loglinear models, 428–431
  - prior distributions, 429
- lognormal distribution, 576, 580
- longitudinal data
  - survey of adolescent smoking, 211–212
  
- maps

- artifacts in, 47–51, 57
- cancer rates, 47–51
- for model checking, 143
- MAR (missing at random), 202, 450
  - a more reasonable assumption than MCAR, 450
- marginal and conditional means and variances, 21
- marginal posterior distribution, 63, 110, 111, 122, 261
  - approximation, 325–326
  - computation for the educational testing example, 594–595
  - computation using importance sampling, 440
  - EM algorithm, 320–325
- Markov chain, 275
- Markov chain Monte Carlo (MCMC), 275–310
  - adaptive algorithms, 297
  - assessing convergence, 281–286
    - between/within variances, 283
    - simple example, 285
  - auxiliary variables, 297–299, 309
  - burn-in, why we prefer the term warm-up, 282
  - data augmentation, 293
  - effective number of simulation draws, 286–288
  - efficiency, 280, 293–296
  - Gibbs sampler, 276–278, 280–281, 291
    - assessing convergence, 281–286
    - efficiency, 293–295
    - examples, 277, 289, 392, 440, 465, 528
    - picture of, 277
    - programming in R, 596–606
  - Hamiltonian (hybrid) Monte Carlo, 300–307, 309
    - hierarchical model example, 305–307
    - leapfrog algorithm, 301
    - mass matrix, 301
    - momentum distribution, 301
    - no U-turn sampler, 304
    - tuning, 303
  - inference, 281–286
  - Metropolis algorithm, 278–280, 291
    - efficient jumping rules, 295–297
    - examples, 278, 290
    - generalizations, 293–300
    - picture of, 276
    - programming in R, 598–599
    - relation to optimization, 279
  - Metropolis-Hastings algorithm, 279, 291
    - generalizations, 293–300
  - minimal analysis, 217
  - missing at random (MAR), 202, 450
    - a more reasonable assumption than MCAR, 450
    - a slightly misleading phrase, 202
  - missing completely at random (MCAR), 450
  - missing data, 449–467
    - and EM algorithm, 452, 454
    - intentional, 198
    - monotone pattern, 453, 455, 459–462
    - multinomial model, 462
  - Metropolis-Hastings algorithm, 279, 291
    - generalizations, 293–300
    - multiple sequences, 282
    - output analysis, 281–288
    - overdispersed starting points, 283
    - parallel tempering, 299–300
    - perfect simulation, 309
    - regeneration, 309
    - reversible jump sampling, 297–299, 309
    - simulated tempering, 309
    - slice sampling, 297, 309
    - thinning, 282
    - trans-dimensional, 297–299, 309
    - warm-up, 282
  - matrix and vector notation, 4
  - maximum entropy, 57
  - maximum likelihood, 93
  - MCAR (missing completely at random), 450
  - measurement error models
    - hierarchical, 133
    - linear regression with errors in  $x$  and  $y$ , 380
    - nonlinear, 471–476
  - medical screening, example of decision analysis, 245–246
  - meta-analysis, 133, 137
    - beta-blockers study, 124–128, 423–425
    - bivariate model, 423–425
    - goals of, 125
    - survey incentives study, 239–242
  - Metropolis algorithm, 278–280, 291
    - efficient jumping rules, 295–297
    - examples, 278, 290
    - generalizations, 293–300
    - picture of, 276
    - programming in R, 598–599
    - relation to optimization, 279
  - Metropolis-Hastings algorithm, 279, 291
    - generalizations, 293–300
  - minimal analysis, 217
  - missing at random (MAR), 202, 450
    - a more reasonable assumption than MCAR, 450
    - a slightly misleading phrase, 202
  - missing completely at random (MCAR), 450
  - missing data, 449–467
    - and EM algorithm, 452, 454
    - intentional, 198
    - monotone pattern, 453, 455, 459–462
    - multinomial model, 462

- multivariate normal model, 454–456
  - multivariate  $t$  model, 456
  - notation, 199, 449–452
  - paradigm for data collection, 199
  - Slovenia survey, 463–466
  - unintentional, 198, 204, 449
- mixed-effects model, 382
- mixture models, 17, 20, 105, 135, 519–543
  - computation, 523–524
  - continuous, 520
  - de Finetti's theorem and, 105
  - Dirichlet process, 549–557
  - discrete, 519
  - exponential distributions, 486
  - hierarchical, 525
  - label switching, 533
  - model checking, 531, 532
  - prediction, 530
  - schizophrenia example, 524–533
- mixture of exponentials, as example of an ill-posed system, 478, 486
- model, *see also* hierarchical models, regression models, etc.
  - beta-binomial, 438
  - binomial, 29, 37, 80, 147
  - Cauchy, 59, 437
  - Dirichlet process, 545
  - exponential, 46, 61
  - lognormal, 188
  - multinomial, 69, 79, 423–428
  - multivariate normal, 70
  - negative binomial, 437, 446
  - nonlinear, 471–486
  - normal, 39, 41, 42, 60, 64–69
  - overdispersed, 437–439
  - Poisson, 43, 45, 59, 61
  - Polya urn, 549
  - robit, 438
  - robust or nonrobust, 438–439
  - $t$ , 293, 437, 441–445
  - underidentified, 89
- model averaging, 193, 297
- model building, one of the three steps of Bayesian data analysis, 3
- model checking, 141–164, 187–195
  - adolescent smoking example, 148–150
  - election forecasting example, 142, 386
  - incumbency example, 361
  - one of the three steps of Bayesian data analysis, 3
  - power transformation example, 189
  - pre-election polling, 210
  - psychology examples, 154–157
  - residual plots, 158, 476, 484
  - SAT coaching, 159–161
  - schizophrenia example, 531, 532
  - speed of light example, 143, 146
  - spelling correction example, 11
  - toxicology example, 483
- model comparison, 178–184
- model complexity, *see* effective number of parameters
- model expansion, 184–192
  - continuous, 184, 372, 439
  - schizophrenia example, 531–532
- model selection
  - bias induced by, 181
  - why we reluctantly do it, 178, 183–184, 367
- moment matching in expectation propagation, 339
- momentum distribution for Hamiltonian Monte Carlo, 301
- monitoring convergence of iterative simulation, 281–286
- monotone missing data pattern, 453, 455, 459–462
- Monte Carlo error, 267, 268, 272
- Monte Carlo simulation, 267–310
- multilevel models, *see* hierarchical models
- multimodal posterior distribution, 299, 319
- multinomial distribution, 578, 584
- multinomial logistic regression, 426
- multinomial model, 69, 79
  - for missing data, 462
- multinomial probit model, 432
- multinomial regression, 408, 423–428
  - parameterization as a Poisson regression, 427
- multiparameter models, 63–82
- multiple comparisons, 96, 134, 150, 186
- multiple imputation, 201, 451–454
  - combining inferences, 453
  - pre-election polling, 456–462
  - Slovenia survey, 463–466
- multiple modes, 311, 321
- multivariate models
  - for nonnormal data, 423–425
  - hierarchical, 390–392
  - prior distributions
    - noninformative, 458
- multivariate normal distribution, 576, 580
- multivariate  $t$  distribution, 319, 578
- natural parameter for an exponential family, 36
- negative binomial distribution, 44, 132, 578, 584



- as overdispersed alternative to Poisson, 437, 446
- nested Dirichlet process (NDP), 566–568
- neural networks, 485
- New York population, 188–191
- Newcomb’s speed of light experiment, 66, 79
- Newton’s method for optimization, 312
- no interference between units, 200
- no U-turn sampler for Hamiltonian Monte Carlo, 304
- non-Bayesian methods, 92–97, 100
  - difficulties for SAT coaching experiments, 119
- nonconjugate prior distributions, *see* prior distribution
- nonidentified parameters, 89
- nonignorable and known designs, 204
- nonignorable and unknown designs, 204
- noninformative prior distribution, 51–57
  - binomial model, 37, 53
  - difficulties, 54
  - for hyperparameters, 108, 110, 111, 115, 117, 526
  - in Stan, 594
  - Jeffreys’ rule, 52–53, 57, 59
  - multivariate normal model, 73
  - normal model, 64
  - pivotal quantities, 54, 57
- nonlinear models, 471–486
  - Gaussian processes, 501–517
  - golf putting, 486, 499
  - mixture of exponentials, 486
  - serial dilution assay, 471–476
  - splines, 487–499
  - toxicology, 477–485
- nonparametric methods, 96
- nonparametric models, 501–517, 545–574
- nonparametric regression, 487–499
- nonrandomized studies, 220
- normal approximation, 83–87, 318–319
  - bioassay experiment, 86
  - for generalized linear models, 409
  - lower-dimensional, 85
  - meta-analysis example, 125
  - multimodal, 319
- normal distribution, 575, 576
- normal model, 39, 41, 60, 64–69, *see also*
  - linear regression *and* hierarchical models
  - multivariate, 70, 454–462
  - power-transformed, 188–191, 194–195
- normalizing factors, 7, 345–349
- notation for data collection, 199
- notation for observed and missing data, 199, 449, 452
- nuisance parameters, 63
- numerical integration, 271, 318–319, 345–348
  - Laplace’s method, 318, 348
- numerical posterior predictive checks, 143–152
- NYPD stops example, 420–422
- objective assignment of probability distributions
  - football example, 13–16
  - record linkage example, 16–19
- objectivity of Bayesian inference, 13, 24
- observational studies, 220–224
  - difficulties with, 222
  - distinguished from experiments, 220
  - incumbency example, 358–364
- observed at random, 450
- observed data, *see* missing data
- observed information, 84
- odds ratio, 8, 80, 125
- offsets for generalized linear models, 407
  - chess example, 428
  - police example, 420
- optimization and the Metropolis algorithm, 279
- ordered logit and probit models, 408, 426
- outcome variable, 353
- outliers, models for, 435
- output analysis for iterative simulation, 281–288
- overdispersed models, 407, 431, 433, 437–439
- overfitting, 101, 367, 409
- p*-values, 98, 145, *see also* model checking
  - Bayesian (posterior predictive), 146
  - classical, 145
  - interpretation of, 150
- packages, *see* software
- paired comparisons with ties, 432
  - multinomial model for, 427
- parallel tempering for MCMC, 299–300
- parameter expansion
  - election forecasting example, 393
  - for Anova computation, 396
  - for EM algorithm, 325, 348
  - for hierarchical regression, 393, 396
  - programming in R, 600–601
- parameters, 4
  - frequentist distinction between parameters and predictions, 94, 401
- Pareto distribution, 493

- partial pooling, *see* shrinkage
- path sampling, 347–348
- perchloroethylene, 477
- perfect simulation for MCMC, 309
- permutation tests, 96
- personal (subjective) probability, 13, 256
- pharmacokinetics, 480–481
- philosophy, references to discussions of, 26
- pivotal quantities, 54, 57, 66
- pivotal quantity, 151
- point estimation, 85, 91, 99
- Poisson distribution, 578, 583
- Poisson model, 43, 59, 61
  - parameterized in terms of rate and exposure, 45
- Poisson regression, 82, 406, 433
  - for multinomial data, 426
  - hierarchical, 420–422
- police stops, example of hierarchical Poisson regression, 420–422
- Polya urn model, 549
- pooling, partial, 25, 115
- population distribution, 101
- posterior distribution, 3, 7, 10
  - as compromise, 32, 40, 58
  - improper, 54, 90, 135
  - joint, 63
  - marginal, 63
  - normal approximation, *see* normal approximation
  - predictive, 7
  - summaries of, 32
  - use as prior distribution when new data arrive, 9, 251
- posterior intervals, 3, 33, 267
- posterior modes, 311–330
  - approximate conditional posterior density using marginal modes, 325
  - conditional maximization (stepwise ascent), 312
  - EM algorithm for marginal posterior modes, 320–325, 348
  - ECM and ECME algorithms, 323, 456, 526
  - examples, 322, 329, 444, 465, 526
  - generalized EM algorithm, 321
  - marginal posterior density increases at each step, 329
  - missing data, 452, 454
  - SEM algorithm, 465
  - SEM and SECM algorithms, 324–325
  - joint mode, problems with, 350
  - Newton’s method, 312
- posterior predictive checks, 143–161, *see also* model checking
  - graphical, 153–159
  - numerical, 143–152
- posterior predictive distribution, 7
  - hierarchical models, 108, 118
  - linear regression, 357
  - missing data, 202
  - mixture model, 530
  - multivariate normal model, 72
  - normal model, 66
  - speed of light example, 144
- posterior simulation, 22–24, 267–310, *see also* Markov chain Monte Carlo (MCMC)
  - computation in R and Stan, 589–606
  - direct, 263–264
  - grid approximation, 76–77, 263
  - hierarchical models, 112
  - how many draws are needed, 267, 268, 272
  - rejection sampling, 264
  - simple problems, 78
  - two-dimensional, 76, 82
  - using inverse cdf, 23
- poststratification, 222, 422–423, 460
- potential scale reduction factor, 285
- power transformations, 188–191, 194–195
- pre-election polling, 69, 79, 233–234, 422–423
  - in Slovenia, 463–466
  - missing data, 456–466
  - state-level opinions from national polls, 422–423
  - stratified sampling, 207–210
- precision (inverse of variance), 40
- prediction, *see* posterior predictive distribution
- predictive simulation, 28
- predictor variables, *see* regression models, explanatory variables
- predictors
  - including even if not ‘statistically significant’, 241–244
  - selecting, 186
- principal stratification, 223–224
- prior distribution, 6, 10
  - boundary-avoiding, 313–318
  - conjugate, 35–37, 56
    - binomial model, 34–35, 38
    - exponential model, 46
    - generalized linear models, 409
    - linear regression, 376–378
    - multinomial model, 69, 429, 462
    - multivariate normal model, 71, 72

- normal model, 40, 43, 67
- Poisson model, 44
- estimation from past data, 102
- for covariance matrices
  - noninformative, 458
- hierarchical, *see* hierarchical models *and* hyperprior distribution
- improper, 52, 82
  - and Bayes factors, 194
- informative, 34–46, 480–481
- nonconjugate, 36, 38, 75
- noninformative, 51–57, 93
  - $t$  model, 443
  - binomial model, 37, 53
  - difficulties, 54
  - for hyperparameters, 108, 110, 111, 115, 117, 526
  - generalized linear models, 409
  - in Stan, 590, 594
  - Jeffreys' rule, 52–53, 57, 59
  - linear regression, 355
  - multinomial model, 464
  - multivariate normal model, 73
  - normal model, 64
  - pivotal quantities, 54, 57
  - warnings, *see* posterior distribution, improper
- predictive, 7
- proper, 52
- weakly informative, 55–57, 128–132, 313–318
  - in Stan, 594
- prior predictive checks, 162, 164
- prior predictive distribution, 7
  - normal model, 41
- probability, 19–22, 26
  - assignment, 13–19, 26, 27
  - foundations, 11–13, 25
  - notation, 6
- probability model, 3
- probit regression, 406
  - for multinomial data, 426, 432
  - Gibbs sampler, 408
  - latent-data interpretation, 408
- probit transformation, 22
- programming tips, 270–271, 605–606
- propensity scores, 204, 221, 222, 230
- proper prior distribution, *see* prior distribution
- proportion of female births, 29, 37–39
- psychological data, 154–157, 524–533
- PX-EM algorithm, 325, 348, *see also* parameter expansion
- QR decomposition, 356, 378
- quality-adjusted life expectancy, 245
- quasi-Newton optimization, 313
- R, *see* software
- $\hat{R}$  for monitoring convergence of iterative simulation, 285
- radial basis functions, 487
- radon decision problem, 194, 246–256, 378
- random probability measure (RPM), 545–574
- random-effects model, 382–388
  - analysis of variance (Anova), 395
  - and superpopulation model in Anova, 397
  - election forecasting example, 386
  - non-nested example, 422–423
  - several batches, 383
- randomization, 218–220
  - and ignorability, 220, 230
  - complete, 218
  - given covariates, 219
- randomized blocks, 231
- rank test, 97
- rat tumors, 102–103, 109–113, 133
- ratio estimation, 93, 98
- record linkage, 16–19
- record-breaking data, 230
- reference prior distributions, *see* noninformative prior distribution
- regeneration for MCMC, 309
- regression models, 353–380, *see also* linear regression
  - Bayesian justification, 354
  - explanatory variables, 5, 200, 353, 365–367
  - exchangeability, 5
  - exclude when irrelevant, 367
  - ignorable models, 203
  - goals of, 364–365
  - hierarchical, 381–404
  - variable selection, 367
  - why we prefer to use informative prior distributions, 367–369
- regression to the mean, 95
- regression trees, 485
- regularization, 52, 113–124, 368–369, 493
- rejection sampling, 264, 273
  - picture of, 264
- replications, 145
- residual plots, 162, 358
  - binned, 157–158
  - dilution example, 476
  - incumbency example, 362
  - nonlinear models, 476, 484
  - pain relief example, 158
  - toxicology example, 484

- residuals, 157
- response surface, 126
- response variable, 353
- reversible jump sampling for MCMC, 297–299, 309
- ridge regression, 401
- robit regression (robust alternative to logit and probit), 438
- robust inference, 162, 185, 192, 435–447
  - for regression, 444–445
  - SAT coaching, 441–444
  - various estimands, 191
- rounded data, 80, 234
- sampling, 205–214, *see also* surveys
  - capture-recapture, 233
  - cluster, 210–212, 232
  - poststratification, 222, 422–423
  - ratio estimation, 93, 98
  - stratified, 206–210
  - unequal selection probabilities, 212–214, 233–234
- SAT coaching experiments, 119–124
  - difficulties with natural non-Bayesian methods, 119
  - information criteria and effective number of parameters, 179
  - model checking for, 159–161
  - robust inference for, 441–444
- scale parameter, 43
- scaled inverse- $\chi^2$  distribution, 43, 576, 581
- scaled inverse-Wishart model, 74, 390
- schizophrenia reaction times, example of mixture modeling, 524–533
- selection of predictors, 186
- SEM and SECM algorithms, 324–325, 348
- sensitivity analysis, 160–161, 184–185, 435–447
  - and data collection, 191
  - and realistic models, 191
  - balanced and unbalanced data, 221
  - cannot be avoided by setting up a super-model, 141
  - estimating a population total, 188–191
  - incumbency example, 363
  - SAT coaching, 441–444
  - using  $t$  models, 443–444
  - various estimands, 191
- sequential designs, 217, 235
- serial dilution assay, example of a non-linear model, 471–476, 485
- sex ratio, 29, 37–39
- shrinkage, 32, 40, 45, 113–124, 132, 368–369, 490, 493
  - graphs of, 113, 122
- simple random sampling, 205–206
  - difficulties of estimating a population total, 188
- simulated tempering for MCMC, 309
- simulation, *see* posterior simulation
- single-parameter models, 29–62
- SIR, *see* importance resampling
- slice sampling for MCMC, 297, 309
- Slovenia survey, 463–466
- small-area estimation, 133
- software, 589–606
  - Bugs, 27, 133, 269, 272
  - debugging, 270–271, 605–606
  - extended example using Stan and R, 589–605
  - programming tips, 270–271, 605–606
  - R, 22, 27, 589–606
  - R programming, 594–606
  - running Stan from R, 589
  - setting up, 589
  - Stan, 22, 269, 307–308, 589–594
- speed of light example, 66, 143
  - posterior predictive checks, 146
- spelling correction, simple example of Bayesian inference, 9–11
- splines, 487–499
  - gay marriage, 499
  - golf putting, 499
  - multivariate, 495–498
- sports
  - football, 13–16, 26
  - golf, 486, 499, 517
- stability, 200
- stable estimation, 91
- stable unit treatment value assumption, 200, 231
- Stan, 307–308, 589–594
- standard errors, 85
- state-level opinions from national polls, 422–423
- statistical packages, *see* software
- statistically significant but not practically significant, 151
  - regression example, 363
- stepwise ascent, 312
- stepwise regression, Bayesian interpretation of, 367
- stratified sampling, 206–210
  - hierarchical model, 209–210, 292
  - pre-election polling, 207–210
- strong ignorability, 203
- Student- $t$  model, *see*  $t$  model
- subjectivity, 12, 13, 26, 28, 100, 248, 256
- sufficient statistics, 36, 93, 338
- summary statistics, 85

- superpopulation inference, 200–203, 205–206, 208, 209, 212, 214–216, 232
  - in Anova, 396–397
- supplemented EM (SEM) algorithm, 324–325
- survey incentives, example of meta-analysis and decision analysis, 239–244
- surveys, 205–214, 454–466, *see also* sampling
  - adolescent smoking, 148–150
  - Alcoholics Anonymous, 213–214
  - incentives to increase response rates, 239–244
  - pre-election polling, 207–210, 422–423, 456–466
  - telephone, unequal sampling probabilities, 233–234
- $t$  approximation, 319
- $t$  distribution, 66, 578, 582
- $t$  model, 437, 441–445
  - computation using data augmentation, 293–294
  - computation using parameter expansion, 295
  - interpretation as mixture, 437
- tail-area probabilities, *see p-values*
- target distribution (for iterative simulation), 261
- test statistics and test quantities, 145
  - choosing, 147
  - examples, *see* model checking
  - graphical, 153–159
  - numerical, 143–152
- thinning of MCMC sequences, 282
- three steps of Bayesian data analysis, 3
- tilted distribution in expectation propagation, 339
- toxicology model, as example of an ill-posed system, 477–485
- trans-dimensional MCMC, 297–299, 309
- transformations, 21, 99
  - logarithmic, 380
  - logistic (logit, log-odds), 22, 125
  - not needed for incentives example, 241
  - not needed for incumbency example, 360
  - power, 188–191, 194–195
  - probit, 22
  - rat tumor example, 110
  - to improve MCMC efficiency, 293–295
  - to reduce correlations in hierarchical models, 480
  - useful in setting up a multivariate model, 424
- treatment variable, 353
- truncated data, 224–228
- $2 \times 2$  tables, 80, 125, 423–425
- type I errors, why we do not care about, 150
- U.S. House of Representatives, 358
- unbiasedness, *see* bias
- unbounded likelihoods, 90
- underidentified models, 89
- uniform distribution, 575, 576
- units, 353
- unnormalized densities, 7, 261
- unseen species, estimating the number of, 349
- utility in decision analysis, 238, 245, 248, 256
- variable selection, why we prefer to use informative prior distributions, 367
- variance matrix, *see* covariance matrix
- variational inference, 331–338
  - EM as special case, 337
  - hierarchical model example, 332–335
  - model checking for, 336
  - picture of, 334, 335, 342
  - variational lower bound, 336
- varying intercepts and slopes, 390–392
- vector and matrix notation, 4
- warm-up for MCMC sequences, 282
- Watanabe-Akaike or widely available information criterion (WAIC), 173–174, 177
  - discussion, 182
  - educational testing example, 179
- weakly informative prior distribution, 55–57, 128–132, 313–318
  - in Stan, 594
- Weibull distribution, 576, 581
- Wilcoxon rank test, 97
- Wishart distribution, 576, 582
- $y^{\text{rep}}$ , 145
- $\tilde{y}$ , 4, 7, 145