

---

## Subject index

---

- adding parameters to a model, 177–178
- adequate summary, 222, 240
- adolescent smoking survey, 172–174
- AECM algorithm, 322, 331
- airline fatalities, 70, 99
- Alcoholics Anonymous survey, 216–218
- aliasing, 109
- all-at-once Gibbs sampler for
  - hierarchical regression, 402
- alternating conditional sampling, *see* Gibbs sampler
- analysis of variance (ANOVA), 133, 221, 406–412, 414
  - finite-population and superpopulation models, 408–409
  - fixed and random effects, 408–409
  - internet example, 410
  - notation, 406
  - pre-election polling example, 429
- approximate inference, 276
- approximations based on posterior modes, 311
- asymptotic theorems, 106–108, 208–209, 219–220, 241
  - counterexamples, 108–111
  - proofs, 585–589
- Australia, adolescent smoking survey in, 172–174, 190, 214–216
- auxiliary variables for computation, 335–339, 348
  
- Bayes factor, 184–186, 191–192
- Bayes’ rule, 7, 22
  - discrete examples, 10, 554
- Bayes, T., 34
- Bayesian data analysis, three steps of, 3, 259
- Behrens–Fisher problem, 96
- Bernoulli distribution, 583
- Bernoulli trials, 33, 163
- beta distribution, 34, 40, 70, 576, 581
- beta-binomial distribution, 576, 583
  - as overdispersed alternative to binomial, 447
- beta-blockers, example of
  - meta-analysis, 145–150
- betting and probability, 13–17
- bias, 248, 256
  - difficulties with the notion, 248
  - prediction vs. estimation, 249, 411
- ‘biased-coin’ design, 244
- bicycle traffic, 98, 155
- binned residual plots, 170–171
- binomial distribution, 576, 583
- binomial model, 33, 43, 97
  - posterior predictive check, 163–164
- bioassay experiment, 88–93, 95, 99
  - normal approximation, 104
- births, 43–46
- blockwise Gibbs sampler for
  - hierarchical regression, 401
- bootstrap, 251
- Box-Cox transformation, 195–196, 265–269
- bridge sampling, 344, 348
- Bugs, *see* software
- burn-in for MCMC sequences, 295
- business school grades, hierarchical multivariate regression, 486–488
  
- calibration of probability estimates, 17–21
- cancer maps, 55
- capture-recapture sampling, 242
- Cauchy model, 69, 446
- causal inference, 4, 218, 229–231
  - instrumental variables, 230–231
  - observational studies, 226

- incumbency example, 359–367
- principal stratification, 229, 230
- randomized experiments, 218, 240
- using regression models, 368
- censored data, 72, 231–236
- Census
  - recoding, 261–263
  - record linkage, 17–21
- chess competition, example of paired
  - comparisons, 431
- $\chi^2$  distribution, 574, 580
- $\chi^2$  test, 175
- Cholesky factor (matrix square root), 357, 578
- classical methods
  - confidence intervals, 112, 249
  - frequency evaluations, 111
  - hypothesis testing, 162, 175, 250
  - maximum likelihood, 247
  - multiple comparisons, 250
  - nonparametric inference, 251
  - normal-theory inference, 101
  - point estimation, 103, 111
  - standard errors, 103
  - unbiased estimates, 248
  - Wilcoxon rank test, 252
- cluster sampling, 214–216, 241
- coarse data, 238
- cockroach allergen data, 498
- coefficient of variation (CV), 7
- coherence, 13
- coin tossing, 12, 29
- collinearity, 369
- colochos, xxv
- complementary log-log link, 418
- complete data, 199
- complete-data likelihood, 201
- completely randomized experiments, 218
- computation, *see* posterior simulation
- computational tips, 278–282, 607–608
- computer programs, *see* software
- conditional maximization, *see* posterior modes
- conditional posterior means and variances, 142
- confidence intervals, 4, 112, 249
- conjugacy, *see* prior distribution
- consistency, 107, 111
- contingency tables, 433–437
  - with missing data, 533–534
- continuous models for discrete data, 529
- contour plots, 91, 129, 130
  - and normal approximation, 103
- control variable, 353
- convergence of iterative simulation, 294–298
- covariance matrices, 22, 85
  - for longitudinal data, 485
  - modeling standard deviations and correlations separately, 483–485
- covariates, *see* regression models, explanatory variables
- cow feed experiment, 223, 386
- cross-validation, 190, 194, 253
- crude estimation, 91, 276
  - bioassay example, 91
  - educational testing example, 132
  - for hierarchical models, 401
  - rat tumor example, 119
  - schizophrenia example, 467, 471
  - Slovenia survey, 535
- CV, coefficient of variation, 7
- data augmentation, 303
- data collection, 197–245
  - censored and truncated data, 231–236
  - experiments, 218–226
  - formal models, 200–203
  - ignorability, 203–207
  - observational studies, 226–231
  - randomization, 223–226
  - sample surveys, 207–218
- data reduction, 104, 264
- de Finetti’s theorem, 122, 152
- debugging, 279–281
  - comparing inferences from several models, 510
  - EM algorithm, 319
  - in Bugs and R, 607–608
- decision analysis, 541–570
  - and Bayesian inference, 542–544
  - medical screening example, 552–555
  - personal and institutional perspectives, 567–568

- radon example, 555–567
  - survey incentives example, 544–552
  - utility, 543, 553, 557, 567
- decision theory, 13, 28, 255
- decision trees, 543, 553, 562
- delta method, 114
- derivatives, computation of, 313
- design of surveys, experiments, and
  - observational studies, 197–245
- designs that ‘cheat’, 225
- deviance, 175, 180–184, 191
- deviance information criterion (DIC),
  - 182–184, 191
  - educational testing example, 183
- differences between data and
  - population, 209, 227, 229, 428, 541
- differential equation model in
  - toxicology, 504–514
- dilution assay, example of a nonlinear
  - model, 498–504, 514
- Dirichlet distribution, 83, 576, 582
- discrepancy measure, 162
- discrete data
  - adapting continuous models, 529
  - latent-data formulation, 419
  - logistic regression, 417
  - multinomial models, 430–433
  - Poisson regression, 417
  - probit regression, 417
- discrete probability updating, 10, 554
- dispersion parameter for generalized
  - linear models, 416
- distinct parameters and ignorability,
  - 204
- distribution, 573–584
  - Bernoulli, 583
  - beta, 34, 40, 70, 576, 581
  - beta-binomial, 70, 576, 583
  - binomial, 576, 583
  - Cauchy, 113
  - $\chi^2$ , 574, 580
  - Dirichlet, 83, 576, 582
  - exponential, 574, 580
  - gamma, 53, 574, 579
  - Gaussian, *see* normal distribution
  - inverse- $\chi^2$ , 574, 580
  - inverse-gamma, 50, 574, 580
  - inverse-Wishart, 87, 574, 581
  - lognormal, 578
  - long-tailed, 444
  - multinomial, 576, 583
  - multivariate normal, 95, 574, 578
    - marginals and conditionals, 578
  - multivariate  $t$ , 315, 576
  - negative binomial, 52, 150, 576, 583
  - normal, 574, 578
  - normal-inverse- $\chi^2$ , 78, 99
  - Poisson, 576, 582
  - scaled inverse- $\chi^2$ , 50, 75, 574, 580
  - standard normal, 578
  - Student- $t$ , 76, 576, 581
  - uniform, 573, 574
  - Weibull, 580
  - Wishart, 574, 581
- divorce rates, 122, 153
- dog metabolism example, 387
- dose-response relation, 89
- $E_{\text{old}}$ , 318
- ECM and ECME algorithms, 321, 322, 331
- educational testing experiments, *see*
  - SAT coaching experiments
- effective number of parameters, 181–182
  - educational testing example, 183
- effective number of simulation draws,
  - 298
- efficiency, 111
- eight schools, *see* SAT coaching
  - experiments
- elections
  - forecasting Presidential elections,
    - 392–399
  - incumbency in U.S. Congress,
    - 359–367
  - pre-election polling in Slovenia,
    - 534–539
  - pre-election polling in U.S.,
    - 428–430, 526–533
  - probability of a tie, 30
- EM algorithm, 317–324
  - AECM algorithm, 322
  - debugging, 319
  - ECM and ECME algorithms, 321, 322, 331
  - for missing-data models, 521–522

- parameter expansion, 324, 331
- SEM and SEM algorithms, 322–324
- empirical Bayes, 121
- environmental health
  - allergen measurements, 498
  - perchloroethylene, 504
  - radon, 555
- estimands, 4, 26, 276
- exchangeability, 28
- exchangeable models, 5, 121–125, 239
  - and explanatory variables, 6
  - no conflict with robustness, 444
  - objections to, 124, 148
  - universal applicability of, 123
- experiments, 218–226
  - completely randomized, 218–220
  - definition, 218
  - distinguished from observational studies, 226
  - Latin square, 220
  - randomization, 223–226
  - randomized block, 220
  - sequential, 221, 244
- explanatory variables, *see* regression models
- exponential distribution, 574, 580
- exponential families, 42
- exponential model, 55, 71
- external validation, 159
  - record linkage example, 18
  - toxicology example, 514
- factorial analysis, internet example, 409–411
- Federalist papers, 458
- finite-population inference, 202–203, 205, 207–209, 211, 212, 216, 218–221, 241
  - in ANOVA, 408–409
- Fisher information, 107
- fixed effects, 391
  - and finite-population models in ANOVA, 409
- football point spreads, 14–17, 29, 30, 51–52, 82
- forecasting Presidential elections, 158, 392–399
  - hierarchical model, 396
  - problems with ordinary linear regression, 394
- frequency evaluations, 111–113
- frequentist perspective, 111
- gamma distribution, 53, 574, 579
- Gaussian distribution, *see* normal distribution
- generalized linear models, 415–442
  - computation, 421–425
  - hierarchical, 421
  - hierarchical logistic regression, 428–430
  - hierarchical Poisson regression, 425–428
  - overdispersion, 418, 439, 441
  - prior distribution, 420–421
  - simple logistic regression example, 88–93
- genetics, 9, 30
- geometric mean (GM), 7
- geometric standard deviation (GSD), 7
- Gibbs sampler, 287–289, 292, 294, 308–309
  - all-at-once for hierarchical regression, 402
  - blockwise for hierarchical regression, 401
  - efficiency, 302–305
  - examples, 300, 380, 449, 474, 538
  - for separation model of covariance matrices, 484
  - hierarchical linear models, 299–302, 400–405, 407–408
  - linear transformation for hierarchical regression, 403
  - parameter expansion for hierarchical regression, 404, 407
  - picture of, 288
  - programming in R, 601–608
  - scalar for hierarchical regression, 402
  - special case of Metropolis-Hastings algorithm, 293
- girl births, proportion of, 43–46
- global mode, why it is not special, 312
- GM, geometric mean, 7
- golf putting, nonlinear model for, 515

- goodness-of-fit testing, *see also* model checking, 250
- graphical models, 151
- graphical posterior predictive checks, 165–172
- graphics
  - examples of use in model checking, 159, 160, 166–170
  - jittering, 15, 16, 29
  - posterior predictive checks, 165–172
- grid approximation, 91–92, 284
- GSD, geometric standard deviation, 7
- Hamiltonian (hybrid) MCMC, 335–336, 348
- heteroscedasticity in linear regression, 372–382
  - parametric model for, 376
- hierarchical linear regression, 389–414
  - computation, 400–405, 407–408
  - interpretation as a single linear regression, 399
- hierarchical logistic regression, 428–430
- hierarchical models, 6, 117–156, 389–414
  - analysis of variance (ANOVA), 406
  - binomial, 127–131, 155
  - bivariate normal, 212–214
  - business school grades, 486–488
  - cluster sampling, 214–216
  - computation, 125–131
  - forecasting elections, 392–399
  - logistic regression, 428–430
  - many batches of random effects
    - election forecasting example, 396
    - polling example, 428–430
  - meta-analysis, 145–150, 488–491
  - multivariate, 486–491, 526–533
  - no unique way to set up, 400
  - normal, 131–150, 299–302, 325–331
  - NYPD stops, 425–428
  - pharmacokinetics example, 260
  - Poisson, 155, 425–428
  - pre-election polling, 212–214
  - prediction, 125, 137
  - prior distribution, *see* hyperprior distribution
  - radon, 555–567
  - rat tumor example, 127–131
  - SAT coaching, 138–145
  - schizophrenia example, 468–479
  - stratified sampling, 212–214
  - survey incentives, 544–552
- hierarchical Poisson regression, 425–428
- hierarchical regression, 389–414
  - an alternative to selecting regression variables, 405–406
  - prediction, 397
- highest posterior density interval, 38
- hybrid MCMC, 335–336, 348
- hyperparameter, 40, 117, 122
- hyperprior distribution, 124–125
  - informative, 260
  - noninformative, 125, 127–129, 134, 136, 137, 154, 470, 489
- hypothesis testing, 162, 175, 180, 250
- identifiability, 369
- ignorability, 203–207, 239, 518
  - incumbency example, 361
  - strong, 205
- ignorable and known designs, 205
- ignorable and known designs given covariates, 205
- ignorable and unknown designs, 206
- iid (independent and identically distributed), 6
- ill-posed systems
  - differential equation model in toxicology, 504–514
  - mixture of exponentials, 516
- importance ratio, 284
- importance resampling (sampling-importance resampling, SIR), 316, 331, 332
  - examples, 450, 452
  - unreliability of, 316
  - why you should sample without replacement, 316
- importance sampling, 342, 348
  - bridge sampling, 344, 348
  - for marginal posterior densities, 343, 450
  - path sampling, 344–345, 348
  - unreliability of, 343
- improper posterior distributions, *see* posterior distribution

- improper prior distribution, *see* prior distribution
- imputation, *see* multiple imputation
- inclusion indicator, 201, 518
- incumbency advantage, 359–367
  - two variance parameters, 377, 381
- indicator variables, 370
  - for mixture models, 463
- inference, finite-population and superpopulation, 203, 216, 218
  - completely randomized experiments, 219–221, 241
  - in ANOVA, 408–409
  - pre-election polling, 211–212
  - simple random sampling, 208–209
- information matrix, 102, 107
- informative prior distributions, *see* prior distribution
- institutional decision analysis, 567–568
- instrumental variables, 230–231
- intention-to-treat effect, 230, 231
- interactions
  - in loglinear models, 434
  - in regression models, 370, 549
- internet connect times, 409–411
- intraclass correlation, 391
- inverse probability, 65
- inverse- $\chi^2$  distribution, 574, 580
- inverse-gamma distribution, 50, 574, 580
- inverse-Wishart distribution, 87, 574, 581
  - other models for covariance matrices, 483–486
- iterative proportional fitting (IPF), 435–437
- iterative simulation, 283–310, 335–350
- iterative weighted least squares (EM for robust regression), 455
- jackknife, 251
- Jacobian, 24
- Jeffreys' rule for noninformative prior distributions, 62–63, 66, 69
- jittering, 15, 16, 29
- joint posterior distribution, 73
- Kullback-Leibler information, 107, 586–588
  - connection to deviance, 181, 191
- Langevin updating for MCMC, 336
- Laplace's method for numerical integration, 341–342, 348
- Laplace, P., 34
- large-sample inference, 101–114
- latent continuous models for discrete data, 419
- Latin square experiment, 220–221
- LD50, 92–93
- life expectancy, quality-adjusted, 553
- likelihood, 9–11
  - complete-data, 201
  - observed-data, 202
- likelihood principle, 9, 28
  - misplaced appeal to, 198
- linear regression, 353–387, *see also* regression models
  - analysis of residuals, 364
  - classical, 355
  - conjugate prior distribution, 382–385
    - as augmented data, 383
  - correlated errors, 372–375
  - errors in  $x$  and  $y$ , 386, 387
  - fitting two variance parameters, 377
  - heteroscedasticity, 372–382
    - parametric model for, 376
  - hierarchical, 389–414
    - interpretation as a single linear regression, 399
  - incumbency example, 359–367
  - known covariance matrix, 373
  - model checking, 364
  - multivariate, 481–482
    - prior distributions, 483–486
  - posterior simulation, 357
  - prediction, 358, 367
    - with correlations, 374
  - residuals, 359, 365
  - robust, 455–457
  - several variance parameters, 382
  - Student- $t$  errors, 455–457
  - weighted, 376
- linear transformation, with Gibbs sampler for hierarchical regression, 403

- link function, 416, 418
- location and scale parameters, 64
- logistic regression, 88–93, 417
  - for multinomial data, 430
  - hierarchical, 428–430
  - latent-data interpretation, 419
- logit (logistic, log-odds) transformation, 24, 146
- loglinear models, 433–437
  - prior distributions, 434
- lognormal distribution, 578
- longitudinal data
  - modeling covariance matrices, 485
  - survey of adolescent smoking, 214–216
  
- map, as used in model checking, 159
- maps of cancer rates, 55
- maps, artifacts in, 55–66
- MAR (missing at random), 204, 518
  - a more reasonable assumption than MCAR, 519
- marginal and conditional means and variances, 23, 142
- marginal posterior distribution, 73, 128, 129, 275
  - approximation, 324–325
  - computation for the educational testing example, 600–601
  - computation for the survey incentives example, 548
  - computation using importance sampling, 343, 450
  - EM algorithm, 317–324
- Markov chain, 286
- Markov chain Monte Carlo (MCMC), 285–310, 335–350
  - adaptive algorithms, 307
  - assessing convergence, 294–298
    - between/within variances, 296
    - simple example, 297
  - auxiliary variables, 335–339, 348
  - burn-in, 295
  - data augmentation, 303
  - efficiency, 291, 302–307
  - Gibbs sampler, 287–289, 292–294, 308–309
    - efficiency, 302–305
      - examples, 288, 300, 380, 400, 449, 474, 538
      - picture of, 288
      - programming in R, 601–608
  - hybrid (Hamiltonian) Monte Carlo, 335–336, 348
  - inference, 294–298
  - Langevin updating, 336
  - Metropolis algorithm, 289–292, 308–309
    - efficient jumping rules, 305–307
    - examples, 290, 301–302
    - generalizations, 335–340
    - picture of, 286
    - programming in R, 604–606
    - relation to optimization, 290
  - Metropolis-Hastings algorithm, 291, 308–310
    - generalizations, 335–340
  - multiple sequences, 294
  - output analysis, 294–299
  - overdispersed starting points, 295
  - perfect simulation, 340, 348
  - recommended strategy, 307, 308
  - regeneration, 340, 348
  - restarting methods, 340, 348
  - reversible jump sampling, 338–339, 348
  - simulated tempering, 337–338, 348
  - slice sampling, 336, 348
  - thinning, 295
  - trans-dimensional, 338–339, 348
- matrix and vector notation, 5
- maximum entropy, 66, 494
- maximum likelihood, 247
- MCAR (missing completely at random), 518
- measurement error models
  - hierarchical, 151
  - linear regression with errors in  $x$  and  $y$ , 387
  - nonlinear, 498–504
- medical screening, example of decision analysis, 552–555
- meta-analysis, 151, 156
  - beta-blockers study, 145–150, 488–491
  - bivariate model, 488–491
  - goals of, 147

- survey incentives study, 544–550
- Metropolis algorithm, 289–292, 308–309
  - efficient jumping rules, 305–307
  - examples, 290, 301–302
  - generalizations, 335–340
  - picture of, 286
  - programming in R, 604–606
  - relation to optimization, 290
- Metropolis-Hastings algorithm, 291, 308–310
  - generalizations, 335–340
- minimal analysis, 222
- missing at random (MAR), 204, 518
  - a more reasonable assumption than MCAR, 519
  - a slightly misleading phrase, 204
- missing completely at random (MCAR), 518
- missing data, 517–540
  - and EM algorithm, 521, 523
  - intentional, 199
  - monotone pattern, 522, 524–525, 530–534
  - multinomial model, 533–534
  - multivariate normal model, 523–526
  - multivariate  $t$  model, 525
  - notation, 200, 517–519, 521
  - paradigm for data collection, 199
  - Slovenia survey, 534–539
  - unintentional, 199, 207, 517
- mixed-effects model, 391
- mixture models, 23, 153, 463–480
  - computation, 467–468
  - continuous, 464
  - discrete, 463
  - exponential distributions, 516
  - hierarchical, 470
  - model checking, 477, 479
  - prediction, 476
  - schizophrenia example, 468–479
- mixture of exponentials, as example of an ill-posed system, 506, 516
- model, *see also* hierarchical models, regression models, etc.
  - beta-binomial, 447
  - binomial, 33, 43, 97, 163
  - Cauchy, 69, 446
  - exponential, 55, 71
  - lognormal, 266
  - multinomial, 83, 95, 430–433
  - multivariate, 481–495
  - multivariate normal, 85
  - negative binomial, 446, 458
  - nonlinear, 497–516
  - normal, 46, 48, 50, 70, 74–83
  - overdispersed, 445–448
  - Poisson, 51, 53, 70, 71
  - robit, 447
  - robust or nonrobust, 448
  - space-time, 494
  - spatial, 493, 494
  - Student- $t$ , 303, 446, 451–457
  - time series, 491–494
  - underidentified, 108
- model averaging, 191, 338
- model checking, 157–196, 265–271
  - adolescent smoking example, 172–174
  - election forecasting example, 158, 394
  - incumbency example, 364
  - power transformation example, 267
  - pre-election polling, 214
  - psychology examples, 165–170
  - residual plots, 170, 503, 513
  - SAT coaching, 186–190
  - schizophrenia example, 477, 479
  - speed of light example, 160, 164
  - toxicology example, 512
- model comparison, 179–186
- model complexity, 181–182
  - educational testing example, 183
- model expansion, 177–179
  - continuous, 186, 376, 449
  - schizophrenia example, 478–479
- model selection, why we avoid it, 180, 185–186, 371, 405
- monitoring convergence of iterative simulation, 294–298
- monotone missing data pattern, 522, 524–525, 530–534
- Monte Carlo error, 277, 278, 282
- Monte Carlo simulation, 276–350
- multilevel models, *see* hierarchical models
- multimodal posterior distribution, 314, 337

- multinomial distribution, 576, 583
- multinomial logistic regression, 430
- multinomial model, 83, 95
  - for missing data, 533–534
- multinomial probit model, 440
- multinomial regression, 420, 430–433
  - parameterization as a Poisson regression, 432
- multiparameter models, 73–100
- multiple comparisons, 152, 174, 250, 253
- multiple imputation, 203, 519–523
  - combining inferences, 522
  - pre-election polling, 526–533
  - recoding Census data, 261–263
  - Slovenia survey, 534–539
- multiple modes, 312, 319
- multivariate models, 481–495
  - for nonnormal data, 488–491
  - hierarchical, 486–488
  - prior distributions, 483–486
    - noninformative, 481, 483, 529
- multivariate normal distribution, 574, 578
- multivariate  $t$  distribution, 315, 576
  
- natural parameter for an exponential family, 42
- negative binomial distribution, 52, 150, 576, 583
  - as overdispersed alternative to Poisson, 446, 458
- neural networks, 515
- New York population, 265–269
- Newcomb, S., 77
- Newton’s method for optimization, 313
- no interference between units, 201
- non-Bayesian methods, 247–252, 256, 257
  - cross-validation, 253
  - difficulties for SAT coaching experiments, 139
- nonconjugate prior distributions, *see* prior distribution
- nonidentified parameters, 108
- nonignorable and known designs, 206
- nonignorable and unknown designs, 206
- noninformative prior distribution, 61–66, 252, 597
  - binomial model, 43, 63
  - Bugs software, 596
  - difficulties, 64
  - for hyperparameters, 125, 127–129, 134, 136, 137, 470
  - for many parameters, 253
  - Jeffreys’ rule, 62–63, 66, 69
  - multivariate normal model, 88
  - normal model, 74
  - pivotal quantities, 64, 66
- nonlinear models, 497–516
  - golf putting, 515
  - mixture of exponentials, 516
  - serial dilution assay, 498–504
  - toxicology, 504–514
- nonparametric methods, 251
- nonrandomized studies, 225
- normal approximation, 101–106, 314–316
  - bioassay experiment, 104
  - for generalized linear models, 422
  - lower-dimensional, 104
  - meta-analysis example, 146–147
  - multimodal, 314
- normal distribution, 574, 578
- normal model, 46, 48, 70, 74–83, *see also* linear regression *and* hierarchical models
  - multivariate, 85, 483–486, 523–533
  - power-transformed, 195–196, 265–269
- normalizing factors, 8, 345–349
- notation for data collection, 199
- notation for observed and missing data, 200, 517, 521
- nuisance parameters, 73
- numerical integration, 340–345, 348
  - Laplace’s method, 341–342, 348
- numerical posterior predictive checks, 172–177
- NYPD stops example, 425–428
  
- objective assignment of probability distributions
  - football example, 14–17
  - record linkage example, 17–21
- objectivity of Bayesian inference, 14, 271
- observational studies, 226–231
  - difficulties with, 228

- distinguished from experiments, 226
- incumbency example, 359–367
- observed at random, 518
- observed data, *see* missing data
- observed information, 102
- odds ratio, 9, 96, 145
- offsets for generalized linear models, 418
  - chess example, 433
  - police example, 426
- optimization and the Metropolis algorithm, 290
- ordered logit and probit models, 420, 430
- outcome variable, 353
- outliers, models for, 443
- output analysis for iterative simulation, 294–299
- overdispersed models, 418, 439, 441, 445–448
- overfitting, 117, 371, 421
  
- p*-values, 162, 255, *see also* model checking
  - Bayesian (posterior predictive), 162
  - classical, 162
  - interpretation of, 175
- packages, *see* software
- paired comparisons with ties, 440
  - multinomial model for, 431
- parameter expansion
  - election forecasting example, 404
  - for ANOVA computation, 407
  - for EM algorithm, 324, 331
  - for hierarchical regression, 404, 407
  - programming in Bugs, 597–598
  - programming in R, 606–607
- parameters, 5
  - effective number of, 181
  - frequentist distinction between parameters and predictions, 249, 411
- partial pooling, 133
- path sampling, 344–345, 348
- perchloroethylene, 504
- perfect simulation for MCMC, 340, 348
- permutation tests, 251
  
- personal (subjective) probability, 14, 567–568
- pharmacokinetics, 260
- pivotal quantities, 64, 66, 77
- point estimation, 103, 111, 256
- Poisson distribution, 576, 582
- Poisson model, 51, 70, 71
  - parameterized in terms of rate and exposure, 53
- Poisson regression, 99, 417, 441
  - for multinomial data, 431
  - hierarchical, 425–428
- police stops, example of hierarchical Poisson regression, 425–428
- pooling, partial, 133, 271
- population distribution, 117
- posterior distribution, 3, 8
  - as compromise, 36–37, 47, 67
  - calculating marginal, 81
  - improper, 64, 109, 154
  - joint, 73
  - marginal, 73
  - normal approximation, *see* normal approximation
  - predictive, 8
  - summaries of, 37
  - use as prior distribution when new data arrive, 11, 561
- posterior intervals, 4, 38, 276
- posterior modes, 312–314
  - approximate conditional posterior density using marginal modes, 324
  - conditional maximization (stepwise ascent), 312
  - EM algorithm for marginal posterior modes, 317–324, 331
  - ECM and ECME algorithms, 321, 471, 526
  - examples, 319, 328, 379, 455, 471, 537
  - generalized EM algorithm, 318
  - marginal posterior density increases at each step, 329
  - missing data, 521, 523
  - SEM algorithm, 538
  - SEM and SECM algorithms, 322–324
- joint mode, problems with, 333

- Newton's method, 313
- posterior predictive checks, 159–177,
  - 253, *see also* model checking
  - graphical, 165–172
  - numerical, 172–177
- posterior predictive distribution, 8
  - hierarchical models, 125, 137
  - linear regression, 358
  - missing data, 203
  - mixture model, 476
  - multivariate normal model, 86
  - normal model, 77
  - speed of light example, 160
- posterior simulation, 25–26, 276–350,
  - see also* Markov chain Monte Carlo (MCMC)
  - computation in R and Bugs,
    - 591–608
  - direct, 283–285
  - grid approximation, 91–92, 284
  - hierarchical models, 130
  - how many draws are needed, 277, 278, 282
  - rejection sampling, 284
  - simple problems, 93–94
  - two-dimensional, 82, 92, 99
  - using inverse cdf, 25–26
- poststratification, 228, 428–430
- potential scale reduction factor, 297
- power transformations, 195–196, 265–269
- pre-election polling, 83, 95, 242–244, 428–430
  - in Slovenia, 534–539
  - missing data, 526–539
  - state-level opinions from national polls, 428–430
  - stratified sampling, 210–214
- precision (inverse of variance), 47
  - used by Bugs software, 592
- prediction, *see* posterior predictive distribution
- predictive simulation, 31
- predictor variables, *see* regression models, explanatory variables
- predictors
  - including even if not ‘statistically significant’, 263, 546–552
  - selecting, 254
- principal stratification, 229–231
- prior distribution, 7
  - conjugate, 40–43, 66
    - binomial model, 40–41, 45
    - exponential model, 55
    - generalized linear models, 421
    - linear regression, 382–385
    - multinomial model, 83, 434, 533
    - multivariate normal model, 85, 87
    - normal model, 46, 50, 78
    - Poisson model, 52
  - estimation from past data, 118
  - for covariance matrices, 483–486
    - noninformative, 481, 483, 529
  - for multivariate models, 483–486
  - hierarchical, *see* hierarchical models *and* hyperprior distribution
  - improper, 61, 100
    - and Bayes factors, 194
  - informative, 39–55, 260
  - nonconjugate, 41, 45, 90
  - noninformative, 61–66, 247, 252
    - binomial model, 43, 63
    - Bugs software, 593, 596, 597
    - difficulties, 64
    - for hyperparameters, 125, 127–129, 134, 136, 137, 470
    - for many parameters, 253
    - generalized linear models, 420
    - Jeffreys' rule, 62–63, 66, 69
    - linear regression, 355
    - multinomial model, 536
    - multivariate normal model, 88
    - normal model, 74
    - pivotal quantities, 64, 66
    - Student-*t* model, 454
    - warnings, *see* posterior distribution, improper
- predictive, 8
  - proper, 61
  - semi-conjugate, 81, 134
- prior predictive checks, 190, 193
- prior predictive distribution, 8
  - normal model, 47
- probability, 22–25, 29
  - assignment, 14–21, 29, 30
  - foundations, 11–14, 28

- notation, 7
- probability model, 3
- probit regression, 417
  - for multinomial data, 430, 440
  - Gibbs sampler, 419
  - latent-data interpretation, 419
- probit transformation, 25
- propensity scores, 206–207, 227, 228, 238
- proper prior distribution, *see* prior distribution
- proportion of female births, 33, 43–46
- psychological data, 165–170, 468–479
- PX-EM algorithm, 324, 331, *see also* parameter expansion
  
- QR decomposition, 357, 385
- quality-adjusted life expectancy, 553
  
- R, *see* software
- $\hat{R}$  for monitoring convergence of
  - iterative simulation, 297
- radon decision problem, 195, 385, 555–567
- random-effects model, 390–392
  - analysis of variance (ANOVA), 406
  - and superpopulation model in ANOVA, 409
  - election forecasting example, 396
  - non-nested example, 428–430
  - several batches, 391
- random-effects models, *see also* hierarchical models *and* analysis of variance
- randomization, 223–226
  - and ignorability, 226, 239
  - complete, 223
  - given covariates, 224
- randomized blocks, 240
- rank test, 252
- rat tumors, 118–120, 127–131, 151
- ratio estimation, 260, 272
- record linkage, 17–21
- record-breaking data, 238
- reference prior distributions, *see* noninformative prior distribution
- reference set for predictive checking, 165
- regeneration for MCMC, 340, 348
  
- regression models, 353–387, *see also* linear regression
  - Bayesian justification, 354
  - explanatory variables, 6, 201, 353, 369–371
  - exchangeability, 6
  - exclude when irrelevant, 371
  - ignorable models, 205
  - include even when nonidentified, 261–263
  - goals of, 367–369
  - hierarchical, 389–414
  - variable selection, 371
  - why we prefer hierarchical models, 405–406
- regression to the mean, 249
- regression trees, 515
- rejection sampling, 284, 310
  - picture of, 285
- replications, 161
- residual plots, 192, 359
  - binned, 170–171
  - dilution example, 503
  - incumbency example, 365
  - nonlinear models, 503, 513
  - pain relief example, 170
  - toxicology example, 513
- residuals, 170
- response surface, 148
- response variable, 353
- restarting for MCMC, 340, 348
- reversible jump sampling for MCMC, 338–339, 348
- ridge regression, 412
- robit regression (robust alternative to logit and probit), 447
- robust inference, 177, 191, 270, 443–459
  - for regression, 455–457
  - SAT coaching, 451–455
  - various estimands, 269
- rounded data, 96, 244
  
- S, S-Plus, and R, 591
- sampling, 207–218, *see also* surveys
  - capture-recapture, 242
  - cluster, 214–216, 241
  - poststratification, 228, 428–430
  - ratio estimation, 260, 272
  - stratified, 209–214

- unequal selection probabilities, 216–218, 242–244
- SAT coaching experiments, 138–145
  - difficulties with natural non-Bayesian methods, 139
  - model checking for, 186–190
  - robust inference for, 451–455
- scalar Gibbs sampler for hierarchical regression, 402
- scale parameter, 50
- scaled inverse- $\chi^2$  distribution, 50, 574, 580
- schizophrenia reaction times, example of mixture modeling, 468–479
- selection of predictors, 254
- SEM and SECM algorithms, 322–324, 331
- semi-conjugate prior distribution, 81, 134
- sensitivity analysis, 177, 189–190, 443–459
  - and data collection, 270
  - and realistic models, 270
  - balanced and unbalanced data, 227
  - cannot be avoided by setting up a super-model, 158
  - Census recoding, 261–263
  - estimating a population total, 265–269
  - incumbency example, 366
  - SAT coaching, 451–455
  - using  $t$  models, 454–455
  - various estimands, 269
- separation models for covariance matrices, 483–485
- sequential designs, 221, 244
- serial dilution assay, example of a nonlinear model, 498–504, 514
- sex ratio, 33, 43–46
- shrinkage, 36–37, 47, 54, 133, 150
  - graphs of, 131, 142
- simple random sampling, 207–209
  - difficulties of estimating a population total, 265
- simulated tempering for MCMC, 337–338, 348
- simulation, *see* posterior simulation
- single-parameter models, 33–72
- SIR, *see* importance resampling
- slice sampling for MCMC, 336, 348
- Slovenia survey, 534–539
- small-area estimation, 151
- software, 591–609
  - Bugs, 27, 29, 151, 592–600
  - debugging, 607–608
  - extended example using Bugs and R, 592–607
  - programming tips, 278–282, 607–608
  - R, 27, 29, 600–607
  - running Bugs from R, 592
  - setting up, 591
  - WinBugs and Bugs, 592
- space-time models, 494
- spatial models, 493, 494
- speed of light example, 77, 160
  - posterior predictive checks, 164
- spline models, 515
- sports
  - football, 14–17, 29
  - golf, 515
- stability, 200
- stable estimation, 111
- stable unit treatment value assumption, 201, 239
- standard errors, 103
- standard normal distribution, 578
- state-level opinions from national polls, 428–430
- statistical packages, *see* software
- statistically significant but not practically significant, 176
  - regression example, 366
- stepwise ascent, 312
- stepwise regression, Bayesian interpretation of, 405
- stratified sampling, 209–214
  - hierarchical model, 212–214, 310
  - pre-election polling, 210–214
- strong ignorability, 205
- Student- $t$  approximation, 315
- Student- $t$  distribution, 76, 576, 581
- Student- $t$  model, 446, 451–457
  - computation using data augmentation, 303–304
  - computation using parameter expansion, 304–305
  - interpretation as mixture, 446

- subjectivity, 12, 14, 28, 31, 257, 557, 567
- sufficient statistics, 42, 247
- summary statistics, 104
- superpopulation inference, 202–203, 205, 208–209, 211, 212, 216, 218–221, 241
  - in ANOVA, 408–409
- supplemented EM (SEM) algorithm, 322–324
- survey incentives, example of
  - meta-analysis and decision analysis, 544–552
- surveys, 207–218, *see also* sampling
  - adolescent smoking, 172–174
  - Alcoholics Anonymous, 216–218
  - pre-election polling, 210–214, 428–430, 526–539
  - telephone, unequal sampling probabilities, 242–244
- t* model, *see* Student-*t* model
- tail-area probabilities, *see p*-values
- target distribution (for iterative simulation), 275
- test statistics and test quantities, 161
  - choosing, 172
  - examples, *see* model checking
  - graphical, 165–172
  - numerical, 172–177
- thinning of MCMC sequences, 295
- three steps of Bayesian data analysis, 3, 259
- time series models, 491–494
  - frequency domain, 492
  - time domain, 491
- tolerance intervals, 71
- toxicology model, as example of an ill-posed system, 504–514
- trans-dimensional MCMC, 338–339, 348
- transformations, 24, 114
  - logarithmic, 387
  - logistic (logit, log-odds), 24, 146
  - not needed for incentives example, 547
  - not needed for incumbency example, 363
  - power, 195–196, 265–269
  - probit, 25
  - rat tumor example, 128
  - to improve MCMC efficiency, 302–305
  - to reduce correlations in hierarchical models, 509
  - useful in setting up a multivariate model, 489
- treatment variable, 353
- truncated data, 231–236
- $2 \times 2$  tables, 96, 145–147, 488–491
- type I errors, why we do not care about, 175
- U.S. Census, 261, 428, 539
- U.S. House of Representatives, 359
- unbiasedness, *see* bias
- unbounded likelihoods, 109
- underidentified models, 108
- uniform distribution, 573, 574
- units, 354
- unnormalized densities, 8, 275
- unseen species, estimating the number of, 332
- utility in decision analysis, 543, 553, 557, 567
- validation, 159, 253
- variable selection, why we prefer to use hierarchical models, 405
- variance as parameterized in Bugs, 592
- variance matrices, *see* covariance matrices
- vector and matrix notation, 5
- wavelet models, 494
- Weibull distribution, 580
- Wilcoxon rank test, 252
- WinBugs and Bugs, 592
- Wishart distribution, 574, 581
  - other models for inverse covariance matrices, 483–486
- $y^{\text{rep}}$ , 161
- $\tilde{y}$ , 5, 8, 161