

Index

- actual vs. nominal error probabilities, 308, 458, 585, 616, 625, 644, 685, 688
- actual vs. nominal power, 689
- actual vs. nominal type I error probability, 689, 691
- defective smoke alarm, 690
- admissibility, 486, 487, 507, 554
 - not a minimal property, 486, 487
 - relative to a specific loss function, 487
- algebra and sigma-algebra, *see* field and sigma-field
- “all models are wrong” (Box), 28, 439, 691, 692, 731
- ambien sleep aid real-life example, 21
- analogical reasoning and its perils, 443, 610
- Anscombe’s data, 632
- aspect ratio in graphs, 179
- association (sign) reversal in regression, 671
 - and Simpson’s paradox, 715
 - due to misspecification, 714
- asymptotic properties of estimators, 406, 488
 - asymptotic efficiency, 491, 530
 - asymptotic Normality, 490, 491, 529
 - asymptotic sampling distribution, 505
 - asymptotic unbiasedness, 530
 - consistency (strong), 490
 - consistency (weak), 488
 - consistent and asymptotically Normal (CAN), 457
 - mean square consistency, 489
- autocorrelation, *see also* temporal dependence, 707
 - exchangeable, 458
 - error, 718, 719
- autoskedastic function, 328
- autoregressive function, 191, 328
- autoregressive model [AR(1)], 348
 - exponential, 348
 - LogNormal, 348
 - Normal, *see also* Normal AR(1), 348
 - Pareto, 348
 - Student’s *t*, 348
- auxiliary regressions for dememorizing data, 191, 722
- auxiliary regressions for deseasonalizing data, 191
- auxiliary regressions for detrending data, 190, 191, 712, 722
- auxiliary regressions for M-S testing, *see* misspecification testing
- auxiliary regressions for VIFs, 670
- axiomatic probability theory, 60, 74, 422
- Bahadur and Savage example, 455, 543
 - and non-parametric inference, 453
- Bayes formula, 67
- Bayesian approach to statistics, 432, 440, 441, 685
 - Bayes factor, 595
 - Bayes risk functions, 503
 - credible interval, 504, 505
 - coherence in judgement, 433, 434
 - conjugate priors, 440, 441
 - loss functions, 486, 504, 554
 - posterior distribution, 440, 441, 502
 - predictive distribution, 441
 - primary objective, 502
 - probabilistic ranking by a posterior, 502
 - rule (estimate), 503
 - testing, 594
 - uniform prior, 442
 - vs. frequentist inference, 436
- Behrens–Fisher Problem, 621
- Bernoulli, *see* distributions
- Bernoulli log-linear form, 272
 - log-linear form of the joint density, 271
- Bernoulli regression-like model, 656
 - logit model, 655, 656
 - probit model, 655, 656
- Bernoulli trivariate distribution, 271

- Berry–Esseen bound, 403
- Beta distribution, *see* distributions
- Beta function, 121
- Bhattacharyya lower bound, 496
- Bienayme’s lemma, 101
- binomial distribution, *see* distributions
- bivariate distributions
 - Bernoulli, 133, 525
 - beta distribution, 172, 255, 256, 281, 285
 - binomial (or trinomial), 135, 141, 171, 287
 - Cauchy, 172
 - exponential, 173, 250, 285
 - F, 173, 253, 285
 - gamma, 173, 174, 234, 247, 285, 307
 - logarithmic, 171, 246, 285
 - LogNormal, 285
 - Normal, 135, 136, 140, 174, 242, 280, 283, 285, 309
 - Pareto, 175, 285
 - Pearson II, 175, 244, 285
 - Poisson, 172, 287
 - Student’s t , 175, 241, 242, 281, 285, 294
- Bonferroni rule, 706
- bootstrap method, 461
 - and data-specific patterns, 463
 - and statistical adequacy, 462, 466
 - bootstrap resample, 461
- Borel (measurable) function, 89, 97
- Borel sigma-field, 56, 63, 87
- boy–girl problem, 66
- Brownian bridge, 356
- Brownian motion process, 355, 415
 - as a limit of a random walk, 357
 - geometric Brownian motion, 357
 - integrated, 356
 - standard, 354
- Caratheodory’s extension theorem, 63, 88
- Cardano’s early probability rules, 7, 40
- Cauchy distribution, *see* distributions
- central limit theorem (CLT), 357
 - Chebyshev’s “near” CLT, 397, 401
 - De Moivre–Laplace CLT, 397
 - Hajek–Sidak CLT, 402
 - Lindeberg’s CLT, 400
 - Lindeberg–Feller’s CLT, 400
 - Lindeberg–Levy CLT, 396
 - Lyapunov’s CLT, 399
 - martingale difference (second-order), 402
 - multivariate CLT, 406
 - stationary process (second-order), 403
- chance regularity patterns, 1, 3, 6, 11, 437
 - assessing distribution assumptions, 5, 189
 - assessing t -homogeneity, 185
 - Beta IID data, 198, 199, 214
 - Cauchy IID data, 200, 201, 213
 - data plots and assumptions, 288
 - dememorized data, 191
 - detrended data, 192, 257, 712
 - distribution, dependence and heterogeneity, 12
 - early formalization of, 7
 - exchange rate data, 11
 - exponential IID data, 198
 - histogram of NIID data, 195
 - homogeneity, 184
 - independence, 178
 - irregular cycles, 182, 222
 - LogNormal IID data, 196
 - mean and variance heterogeneity, 189
 - mean heterogeneity, 186, 188
 - negative dependence, 183
 - NIID data, 180, 195
 - NID negatively dependent data, 183
 - NID positively dependent data, 182
 - NI mean heterogeneous data, 186
 - NI mean seasonal and trending data, 190
 - NI mean seasonal data, 190
 - NI mean-shift data, 188
 - NI mean/variance trending data, 189
 - NI mean/variance-shift data, 189
 - NI variance-shift data, 188
 - NI variance-trending data, 187
 - non-Normal IID, 196
 - Normal, dependent, and trending data, 191
 - positive dependence, 21, 182
 - probabilistic assumptions and statistical models, 12, 14
 - seasonality patterns, 187
 - Student’s t IID data, 200
 - Uniform IID data, 200
 - variance heterogeneity, 186, 188
 - Weibull IID data, 199
- Chapman–Kolmogorov equation, 337
- characteristic function (chf), 103
- Chevalier de Mere’s paradox, 10
- coefficient of variation, 118
- combinations, 37
- concordance measure, 265
- conditional cross-product ratios, 272
- conditional density, 143
 - properties, 144
- conditional distribution, 144, 146, 148, 152, 170, 225, 228, 233, 268, 271, 273, 278, 279, 292, 299, 304, 325, 351, 480, 481
 - singular, 350
- conditional expectation, 278, 296, 297, 300, 302, 327, 345, 450
- conditional independence, 268, 269, 330
- conditional moment functions, 278, 280
 - autoregressive function, 328
 - autoskedastic function, 328
 - clitic function, 281
 - kurtic function, 281
 - regression function, 280, 282
 - skedastic function, 280, 282
- conditional moments, 146, 235, 280

754 Index

- and dependence, 232
- conditional covariance, 269
- conditional mean, 147
- conditional variance, 147, 245
- stochastic, 297
- conditional probability, 65, 74, 143
- conditioning information set, 301, 302
- conditioning on a sigma-field, 295–298
 - conditioning on the sigma-field $\sigma(X)$, 295, 297
 - corset property, 298
 - law of iterated expectations (LIE) property, 298
 - least-squares prediction property, 298
 - stochastic moment functions, 295, 297
 - taking out what is known property, 298
- conditioning on values of random variables, 143, 144, 288
 - the role of the quantifier, 151, 153
 - vs. on events, 151, 152
- conditioning on events, 65, 66
- confidence intervals (CIs), 365, 396, 448, 455, 498, 499, 501, 554, 575, 610, 611
 - and factual reasoning, 611
 - coverage error, 499, 501, 506
 - coverage probability, 499, 612–614, 644
 - expected length, 501, 613
 - fallacious interpretation, 614, 615
 - long-run metaphor, 499, 500
 - vs. Bayesian credible interval, 505
 - uniformly most accurate, 612, 636
 - vs. hypothesis testing, 613
 - vs. p-values, 614
 - vs. severity, 614
- confounding variables/factors, 564, 648, 665, 693, 717, 730
- consistent and asymptotically Normal (CAN) estimators, 491, 506, 507, 692
- consistent and asymptotically Normal (CAN) estimators, 457
- consistent estimators, 488, 529
 - consistency as a minimal property, 456, 489, 505
 - consistent but useless estimator, 489
- contemporaneous (synchronic) dependence, 223, 288
- continuous mapping theorem, 418
- convergence, probabilistic
 - almost surely (with probability one), 382, 425
 - almost surely vs. in probability, 382
 - convergence in distribution, 382, 396
 - convergence in probability, 381, 383
 - uniform convergence, 409
- convolution formula, 163
- Cook's D, 674
- copulas and dependence, 258, 259
 - Archimedean, 260
 - modeling and inference perspective, 261
 - Normal (Gaussian), 260
 - Sklar's theorem, 261
 - Student's t , 261
- correlation coefficient, 231
 - exchangeable, 458
 - linear (first-order) dependence, 231
 - nonsense, 710
 - partial, 269
 - properties, 232
- count (point) process, 360, 361
- countable additivity, 60, 63, 85
- countable set, 47
- covariance, 136
 - partial, 268
 - properties, 137
- covariance ratio, 675
- coverage probability, *see* confidence intervals
- Cramer's theorem, 407
- Cramer–Rao lower bound, 478
 - asymptotic, 491
- cross-product (odds) ratio, 264, 271, 272, 526
- cumulants, 103
- cumulative distribution function (cdf), 89
 - empirical (ecdf), 193, 207, 450
 - joint, 134
 - properties, 90
- curve-fitting as empirical modeling, 2, 261, 262, 305, 306, 308, 311, 365, 649, 650, 675, 692, 716
 - conflating statistical and substantive models, 308, 365
 - driven by goodness-of-fit, 2, 262, 655, 676
 - theory-driven modeling, 676
 - vs. statistical modeling, 675
- data classification, 15
 - cross-section data, 15
 - cross-section vs. time series data, 20
 - panel (longitudinal) data, 15
 - time series, 15
- de Morgan's laws, 50
- decision making under uncertainty, 443
- decision-theoretic framing of inference, 464, 502
- delta method, 408
- demand schedule, 26, 305
- density function, 35, 83, 91
 - (continuous) properties, 92
 - (discrete) properties, 93
 - support, 96
- dependence concepts (assumptions)
 - asymptotic average non-correlation, 331
 - asymptotic independence, 330, 369
 - asymptotic non-correlation, 331, 370
 - correlation, 231
 - ergodicity, 370
 - linear dependence, 232, 235, 250, 263, 269, 276, 330, 389
 - Markov dependence, 227, 268
 - martingale, 331
 - martingale difference, 331
 - of order m , 330

- orthogonality between random variables, 232
- second-order dependence, 245
- dependence measures, 274
 - and joint raw moments, 230
 - and nominal variables, 266
 - and ordinal variables, 263
 - and the elliptically symmetric family, 240
 - and the Normal distribution, 234
 - between two random variables, 228
 - gamma coefficient, 265
 - Goodman and Kruskal tau, 267
 - negative correlation, 252
 - Pearson vs. Yule on discrete/continuous measures, 263
 - Theil's uncertainty coefficient, 267
 - Yule's Q, 264
- descriptive vs. inferential statistics, 30, 32
- deterministic regularity, 3
- deterministic trends, 731
- different approaches to inference, 436
- dimensionality problem, 225, 233, 277
 - and reduction, 224
- discordance measure, 265
- distribution of the sample, 167, 168, 224, 301, 422, 437, 444, 476, 481, 493, 511, 512, 516, 526, 531
 - reinterpreted in Bayesian statistics, 440
 - vs. the likelihood function, 437, 447
- distributions (univariate)
 - Bernoulli, 36, 83, 121
 - beta, 96, 104, 113, 116, 123, 208, 440, 503, 545
 - binomial, 36, 84, 122
 - Cauchy, 100, 110, 112, 113, 115–117, 123, 163, 325, 454
 - chi-square, 124
 - exponential, 91, 124, 134, 198, 199
 - extreme value (Gumbel), 124
 - F, 125
 - gamma, 96, 125
 - generalized gamma, 125
 - geometric, 86, 122
 - hypergeometric, 122
 - Laplace (double exponential), 106, 126
 - logarithmic series distribution, 122
 - logistic, 116, 126
 - LogNormal, 126, 196
 - negative binomial, 122
 - non-central chi-square, 126
 - non-central Student's t , 127
 - Normal (Gaussian), 36, 93, 94, 127
 - Pareto, 128
 - Poisson, 94, 123
 - power exponential (error), 128
 - Student's t , 107, 128, 497
 - uniform (continuous), 94, 129, 165
 - uniform (discrete), 123
 - Weibull, 129
- distribution-free, *see* non-parametric
- Doob martingale, 341
- Doob–Meyer decomposition, 364
- double truncation, 149
- Duhem's conundrum, 2
- dummy variable, 191
- duration (hazard-based) models, 363
- Durbin–Watson (DW) test, 719
- Dutch book argument, 433
- dynamic linear regression (DLR(1)) model, 720, 728
- Edgeworth expansion, 404
- Edgeworth's testing, key concepts, 556
- elastic net regression estimation, 650
 - as curve-fitting, 650
- elementary outcomes, 7, 8, 48, 49
- elliptically symmetric family of distributions, 233
- empirical cumulative distribution function (ecdf), 207
- empirical modeling, 1
 - as curve-fitting guided by goodness-of-fit, 2, 262
 - crucial features, 1
 - guided by statistical and substantive information, 1
 - modelling vs. inference, 446
 - primary aim of, 28, 301
 - statistical modeling vs. curve-fitting, 675
- empirical modeling examples with data, 1
 - antique grandfather clock, 630, 635, 636, 638, 646
 - capital asset pricing model (CAPM), 643, 644
 - dice data, 701
 - effect of education on income data, 714
 - exam scores data, 699
 - exchange rate data, 215
 - Keynes' absolute income hypothesis data, 716
 - Yule's nonsense correlations data, 710
- equality of random variables, 89
- equal-probability contours, 235, 236, 238
- error probabilities, 425, 438, 563, 685
 - are not conditional, 579
 - actual vs. nominal, 458
 - pre-data, 578, 579, 616
 - post-data, 570, 579
 - post-data severity evaluation, 600
 - pre-data vs. post-data, 579
 - type I error probability, 455, 573, 576, 578, 588
 - type II error probabilities, 455, 573, 577, 588
- error–statistical perspective, 585, 596
- error term assumptions, 305, 348, 627, 720
- error term, statistical vs. structural, 647
- errors of measurement, 648
- estimation (point)
 - admissibility, 486
 - bias of an estimator, 474
 - crystal ball estimator and admissibility, 486, 487
 - finite sample properties, 474
 - fixed-window (rolling), 707

756 Index

- inferential claim for, 506
- minimal sufficient estimator, 481
- primary objective, 471, 512
- recursive, 707
- sampling distribution, 374, 472, 505
- strong consistency, 490
- sufficient estimator, 480
- vs. estimate vs. parameter, 448
- event in probability, 48
 - events vs. elementary outcomes, 51
- event space, 51
- existence of moments, 110, 336, 387, 413, 459
- experimental data, 37, 306, 564
 - sampling procedure probabilities, 37
 - sampling with replacement, 37
 - sampling without replacement, 37
 - vs. observational data, 14
- exponential distribution, *see* distributions
- exponential family of distributions, 681
 - one-parameter, 574, 589, 680
 - two-parameter, 680, 681
- F distribution, *see* distributions
- factor analysis, 351
- factorization theorem, 481
- factual reasoning, 438, 472, 500, 559, 611, 614
 - vs. Bayesian reasoning, 441
 - vs. hypothetical, 560, 576, 613, 614
- false negative/positive in probability, 67
 - in medical detection devices, 609
 - vs. error probabilities, 610
- feasible GLS, 719
- field (algebra), 54, 55
- filtration of sigma-fields, 341
- finite additivity, 434
- finite sample properties of estimators, 474
 - bias, 474
 - full efficiency, 476
 - median unbiasedness, 492
 - minimum MSE, 485
 - minimal sufficiency and completeness, 484
 - mode unbiasedness, 492
 - parameterization invariance, 525, 546
 - Pitman closeness, 492
 - relative efficiency, 476
 - relative vs. full efficiency, 475
 - sufficiency, 480, 496, 528
 - sufficiency and unbiasedness, 482
 - unbiasedness (mean), 474
- Fisher and experimental design
 - blocking, 564
 - material experiment, 565
 - randomization, 564
 - replication, 564
- Fisher information, 476, 477, 494, 514, 517, 519, 630
 - asymptotic Fisher information, 491
 - for a single observation, 477, 530
 - observed, 521
- Fisher–Neyman factorization, 482
- frequentist approach to statistical inference, 436, 437, 512
 - model-based, 24, 30, 424–427, 429, 430, 444, 547
 - primary objective, 471, 561
 - reliability/unreliability of inference, 462, 548, 625, 721, 729, 731
 - untrustworthy evidence, 258, 453, 686, 714, 730
 - vs. Bayesian inference, 436
- Fisher's recasting of Pearson's approach to statistics, 444, 511, 543
 - parametric model-based, 424–426, 444, 511, 512
- Fisher's significance testing, 558, 579
 - falsificationist stance, 562
 - hypothetical reasoning, 560
 - key elements, 568
 - p-value, 560
 - null hypothesis, 559
 - vs. N-P Testing, 598
- for Logit/Probit Models, 710
- frequentist interpretation of probability, 391, 423, 426, 438, 464
 - circularity charge, 427
 - long-run criticism, 429
 - repeatability in principle, 429
 - stipulated provisions, 426
 - vs. the propensity interpretation, 431
 - vs. Kolmogorov's algorithmic complexity, 430
- frequentist testing, 424
 - fallacy of acceptance, 577, 599
 - fallacy of rejection, 578, 599, 605, 704, 720
 - generic capacity (power) of tests, 577, 579, 582, 603
 - hypothetical reasoning, 576, 614
 - primary objective, 561
 - statistically vs. substantively significant, 607
- frequentist vs. Bayesian inference, 436
- full efficiency, 476, 479, 529
- function, 58
 - bijection, 59
 - co-domain, 58
 - composite, 88
 - domain, 59
 - range of values, 59
 - relation, 58
- functional (Donsker) CLT, 418
 - invariance principle, 414
- functions of random variables
 - one random variable, 97, 161
 - two random variables, 162
 - several random variables, 162
- functions of the sample moments, 542
- fundamental theorem of calculus, 91
- Galileo's three-dice puzzle, 38
- Galton on concentric ellipses, 240
 - from a scatterplot to elliptical contours, 238

- on reverse regression, 308
 - regression toward the mean, 307
- gamma distribution, *see* distributions
- gamma function, 121
- gamma regression-like model, 658
 - and generalized linear models, 658
- garbage in garbage out, 23
- Gauss linear model, 306, 649, 655, 657
 - best linear unbiased estimator (BLUE), 536, 651, 665
 - vs. the Linear Regression model, 650, 677
- Gauss–Markov theorem, 536, 651, 665, 719
 - and different distributions, 652
 - limited value in inference, 536
- generalized least squares (GLS), 719
 - feasible GLS, 719
- generalized linear (GL) models, 655, 657, 682
 - common features, 681
 - deviance, 683
 - deviance residuals, 683
 - linear predictor, 681
 - link function, 657, 658, 681
 - Pearson residuals, 683
 - response residuals, 683
- Glivenko–Cantelli theorem, 451
- Gnedenko’s theorem, 405
- Goodman and Kruskal tau, 267
- goodness-of-fit, 655
- graphical analysis, 28, 176
- graphical causal modeling, 270, 272
- hazard function, 150, 363
 - cumulative, 363
- hazard-based statistical models, 363
 - Weibull, 365
- Heaviside function, 452
- heterogeneity concepts (assumptions)
 - first order stationarity, 333
 - Markov homogeneity, 332
 - second-order stationarity, 333
 - separable heterogeneity, 310, 334, 349, 367
 - spatially homogeneous, 339
 - stationarity of order $m \geq 1$, 334
 - strict, 332
- high leverage points, 675
- histogram, 4, 176, 203
 - and non-IID data, 206
 - smoothed histogram, 204, 205
- hypothetical reasoning, 438, 553, 560, 561, 576, 613–615
 - vs. Bayesian reasoning, 441
 - vs. factual reasoning, 438
- ideal estimator, 471, 488
- idempotent matrix, 663
- identical distributions (ID), 37, 71, 158, 331
- impossible event, 51
- incidental parameter problem, 225, 227, 328, 343, 346, 533, 655, 693, 707, 719
- increasing conditioning information set, 227
- independence, 37, 70, 303, 329
 - among events, 69
 - among random variables, 155, 156
- indicator function, 203
- individual decision making, 443
- induction by enumeration, 427
- inductive inference, 24
 - premises of inference, 23
 - ampliative dimension, 24, 32
 - vs. deduction, 121, 457
 - model-based induction, 24, 427
- inequalities (probabilistic), 412
 - Bernstein inequality, 413
 - Boole inequality, 64
 - Bonferroni inequality, 65
 - Cauchy–Schwarz inequality, 413
 - Chebyshev inequality, 102, 385, 412, 488
 - Hoeffding inequality, 229, 413
 - Holder inequality, 414
 - Jensen inequality, 414
 - Lyapunov inequality, 414
 - Markov inequality, 412
 - Mill’s inequality, 413
 - Minkowski inequality, 414
 - Schwarz inequality, 232
- inference vs. modeling facet, *see* modeling vs. inference
- infinitely divisible distributions, 377, 406
- initial conditions, 337
- innovation process, 344
- interpretations of probability, 421
 - degrees of belief, 432, 441, 464
 - frequentist (model-based), 424, 426
 - frequentist (von Mises), 426, 427
 - Kolmogorov’s algorithmic complexity, 430
 - logical, 435
 - Popper’s propensity, 431
- interval estimation, *see* confidence intervals
- intervals on the real line, 46
 - closed interval, 46
 - half-closed, 46
 - open interval, 46
 - singleton, 46
- Ito’s Lemma, 359
- Itô process, 358
- joint and conditional probability formulae, 152
- joint density function, 132–134, 139, 144, 150
 - properties, 135
- joint independence, 70
- Karl Pearson approach to statistics, 263, 445, 497, 547, 551, 556
 - vs. Fisher, 547
- Karl Pearson’s chi-square test, 557, 567
 - the initial misspecification test, 558, 686
- Karl Pearson’s method of moments, 543

758 Index

- Karl Pearson's testing, 556
 - key concepts, 558
- kernel smoothing, 204, 219, 454, 460
 - bandwidth, 205
 - biweight kernel, 205
 - Epanechnikov, 205
 - Normal, 204
 - rolling histogram, 203
 - uniform, 204
- Kolmogorov algorithmic complexity, 430
- Kolmogorov axioms of probability, 60, 422
- Kolmogorov's distance theorem, 451
- Kolmogorov's extension theorem, 322, 366
- Kullback–Liebler distance, 229
 - and likelihood, 524
- kurtosis coefficient, 105, 138
- large n problem, 563, 598, 603, 616
 - decreasing α for as sample size increases, 599
- LASSO regression estimation, 650
 - and curve-fitting, 650
- law of iterated logarithm (LIL), 391, 395
 - Khinchin's LIL, 395
- learning from data, 383, 437, 438, 441, 466, 572, 577
- least squares as a statistical method, 535
 - best linear unbiased estimator (BLUE), 536, 651, 665
 - least-squares estimator, 535
 - ordinary least squares, 305
- least-squares mathematical approximation, 534
 - approximating function, 648
 - Boscovitch, 534
 - curve-fitting, 534, 648
 - Gauss, 536
 - Laplace, 534
 - Legendre, 534, 535
 - principle of least squares, 648
 - theory of errors, 41, 510
- Lehmann–Scheffé theorems, 483, 484
- leptokurtic distributions, 106, 241, 286
- likelihood function, 437, 447, 511, 513
 - log-likelihood function, 515
- likelihood ratio test, 571, 591, 594, 595
 - and monotonicity, 588
 - asymptotic, 594
- limit theorems in probability, 375, 410
 - main assertions, 375
 - with non-testable assumptions, 409, 410
- linear regression, traditional formulation, 659, 718
 - ad hoc M-S testing, 718, 725
 - asymptotic sampling distributions, 666
 - autocorrelation-consistent standard errors (ACSE), 721
 - computational perspective, 660
 - curve-fitting problem, 716
 - error-fixing strategies, 720, 721
 - Gauss–Markov theorem, 665
 - hat matrix, 673
 - heteroskedasticity-consistent standard errors (HCSE), 721
 - Huber condition, 666
 - matrix formulation, 661, 667
 - Normality and the reliability of inference, 643
 - robustness claims, 457, 693, 731
 - traditional specification, 627
 - vs. the Gauss linear model, 649
- logit model, 655, 656, 710
 - misspecification testing, 710
- logit transformation, 657
- long-run frequency, 429
 - vs. probability, 429
- long-run metaphor, 425, 427, 437, 466, 500
 - conflating probability with relative frequencies, 429
 - no temporal dimension, 427
 - repeatability in principle, 429
- loss function, 113, 486, 504, 554
- Mann and Wald theorem, 407
- marginal density function, 140
- marginal distributions, 139
- marginalization, 139, 140
- marginalization vs. conditioning, 150
- Markov chains, 337
- Markov dependence, 227, 268, 277, 329
 - of order p , 330
- Markov process, 325, 336, 337
- martingale difference process, 327, 342, 344, 394
 - dependence, 331
 - second-order, 344
- martingale process, 327, 340
 - dependence, 331
 - sub-martingale, 340
 - super-martingale, 340
- mathematical deduction, 64
- mathematical duality between testing and CIs, 610, 612
 - not an inferential duality, 614
- mathematization of chance regularities, 7
- matrix norm, 668
- maximal ancillary statistic, 705
- maximum likelihood method, 547
 - asymptotic variance, 490
 - criticisms of, 532
 - estimator (MLE), 514
 - inconsistent estimators, 533
 - Neyman and Scott model, 533
 - optimal properties of MLEs, 532
- max-stable family of distributions, 405
- mean of a distribution, 99
 - properties, 100
- mean ergodic process, 371
- mean square error (MSE), 300, 485, 503
 - and biased estimators, 487
 - frequentist vs. Bayesian definition, 486

- measure theory, 60, 73, 422, 428
 - Lebesgue measure, 63, 428
- median of a distribution, 113
- Mendel's cross-breeding experiments, 557
 - substantive model, 557
- method of least squares, 547
- methods of estimation, 510
- mid-range, 515
- minimal sufficient statistic, 480
- misinterpreting the p-value, 563
- misspecification (M-S) testing, 2, 23, 26, 177, 289, 426, 461, 631, 686
 - Anderson–Darling test for Normality, 696
 - custom tailoring, 704
 - data snooping/mining charge, 693
 - directions of departure, 697
 - double use of data charge, 693
 - encompassing alternative model, 697, 720
 - for Logit/Probit Models, 710
 - formalization of, 316
 - independence and mean constancy, 698
 - independence and variance constancy, 700
 - infinite regress/circularity charge, 427, 693
 - joint tests based on auxiliary regressions, 645, 670, 699–701, 704, 705, 707, 709, 724
 - Kolmogorov's test for Normality, 695
 - large enough sample size (n), 25, 379, 506
 - legitimate conditioning information set, 708
 - multiple testing charge, 693
 - non-parametric (omnibus), 694
 - Normality test, 646
 - pre-test bias charge, 693
 - runs test, 694
 - skedastic function tests, 724
 - skewness-kurtosis test of Normality, 700
 - testing outside the statistical model, 687, 703
 - testing within vs. testing outside, 567, 687, 703
 - t -invariance of the parameters, 707
 - too many tests charge, 706
 - vs. N-P testing, 704
- mixing conditions, 369
 - non-testable, 403
- mixingale, 370
- mode of a distribution, 112
- model-based frequentist interpretation of
 - probability, 424, 426
 - single-event probability problem, 429
 - statistical induction, 24, 74, 427, 444, 457, 466, 686
- modeling vs. inference facet, 2, 119, 365, 374, 446, 447, 465, 507, 510, 642, 687, 729
- moment generating function (mgf), 103
- moment matching principle, 111, 536, 537, 548
- moments of distributions, 99
 - central moments and dependence, 231
 - higher central moments, 103
 - higher raw moments, 102
 - joint central moments, 136, 231
 - joint product moments, 136
 - problem of moments, 110
- Monty Hall puzzle, 68
- moving average [MA(q)] process, 352
- multiplication counting rule, 37
- multivariate distributions, 134
 - Bernoulli distribution, 271
 - density properties, 138
 - elliptically symmetric, 241
 - Normal distribution, 139, 269
 - Student's t , 344
- mutually exclusive events, 51, 70, 422
- near-collinearity problem, 666, 667, 677
 - and ordering of observations, 669
 - determinant of $(\mathbf{X}^T \mathbf{X})$, 668
 - high correlations among regressors, 667
 - ill-conditioning of $(\mathbf{X}^T \mathbf{X})$, 667, 668
 - numerical perspective, 667
 - statistical perspective, 670
- negative binomial, 122, 589, 680
 - bivariate, 171, 287
- nesting restrictions, 643
- Newton–Raphson algorithm, 522
- Neyman and Scott model, 532
- Neyman–Pearson (N-P) testing, 449
 - acceptance region, 449, 572, 611–613
 - alternative hypothesis, 570
 - arbitrariness of the significance level, 608
 - archetypal framing of hypotheses, 572, 575, 603
 - behavioristic interpretation of accept/reject, 573
 - coarseness of accept/reject rules, 606
 - consistent N-P test, 582
 - convexity of the alternative parameter space, 589
 - default alternative hypothesis, 576
 - Egon Pearson's retrosperspective, 569, 570
 - key elements, 586
 - modifications to Fisher testing, 570, 573
 - monotone likelihood ratio, 587
 - multiple testing (comparisons), 609, 706
 - N-P lemma, 586
 - non-centrality parameter, 586
 - null hypothesis, 559
 - origins of the Fisher vs. Neyman prolonged dispute, 574
 - pre-designation of error probabilities, 598
 - randomization, 590
 - rejection region, 450, 572, 576, 586
 - simple vs. composite hypotheses, 571
 - testing within, 572, 687, 703
 - testing within vs. testing outside, 567, 687, 703
 - the large n problem, 585
 - trade-off between the type I and II, 580, 598
 - trade-off rationale, 580
 - type I and II errors, *see also* error probabilities, 573
 - unbiased N-P test, 581

760 Index

- uniformly most powerful (UMP) test, 580, 581, 586
- non-correlation, 330
 - of order m , 331
- non-parametric (distribution free) inference, 453
 - and its reliance on asymptotics, 453
 - indirect distribution assumptions, 112, 453
 - inference based on asymptotic bounds, 458
 - non-testable smoothness restrictions, 453
 - perils of ‘as n tends to infinity’, 374, 409, 458, 692, 731
 - vs. parametric, 453
 - weak probabilistic assumptions, 454, 456, 466
- non-parametric tests
 - and claims of robustness, 454
 - low power, 696
 - Normality of bust, 454
 - t test vs. Wilcoxon test, 454
 - Wilcoxon-type tests, 453
- non-parametric (omnibus), 696
- non-stationarity vs. separable heterogeneity, 367
- non-typical observation, *see* outlier
- norm condition number, 667
- Normal autoregressive [AR(1)] model, 294, 347, 348, 456, 457, 642
 - AR(p) model, 345, 353
 - AR(p) with mean heterogeneity model, 348
 - Normal, Markov, stationary process, 346
 - statistical GM, 347
 - statistical parameterization, 347
 - testable, internally consistent model assumptions, 347
 - underlying joint distribution, 346
 - vs. the unit root [UR(1)] model, 350, 351
 - with second-order separable heterogeneity, 351
- Normal autoregressive, moving average [ARMA(p , q)] model, 353, 354, 367
- Normal distribution, *see* distributions
- Normal dynamic linear regression (DLR) model, 728
- Normal equations, 534
- Normal linear regression (LR) model, 538, 625, 627, 631, 658, 690, 705
 - and mean heterogeneity, 310, 691
 - fitted values and residuals, 633
 - heterogeneous conditional variance, 310, 351
 - heteroskedasticity, 286, 304, 310, 718, 725
 - heteroskedasticity vs. heterogeneity, 310, 718
 - linearity assumption, 284, 298, 303, 649
 - mean heterogeneity, 716
 - Normality and the LR model, 642
 - role of the constant term, 664
 - t -invariance assumption, 303
 - variance decomposition, 664
- Normal martingale difference process, 708
- Normal moving average [MA(q)] process, 352
- Normal, unit root [UR(1)] model, 349, 350
 - AR(1) vs. UR(1) model, 350
 - statistical GM, 350
 - statistical parameterization, 350
 - underlying joint distribution, 349
 - unit root testing, 350, 626
 - Wiener process, 349
- Normal vs. Pearson type II, 108
- Normal vs. Student’s t distribution, 107, 108, 201, 202
- Normal white-noise process, 336
- nuisance parameters, 533
- null hypothesis, 559
- number sets, 46
 - integers, 46
 - natural numbers, 46
 - positive integers, 46
 - rational numbers, 46
 - real numbers, 46
- numerical evaluation of optimization, 522
- objectivity in statistical inference, 444, 686
- observational data, 14, 26, 37, 40, 304, 305, 443, 730
 - vs. experimental data, 14, 306
- observed confidence intervals and severity, 614
- odds ratio, 264, 433
 - test, 568, 577, 581, 582
- omitted variables, 307, 646, 648, 692
 - vs. discarded variables, 709
- one-way analysis of variance (ANOVA), 526, 624, 680
- operational statistical model, 343, 386
- ordered sample, 208
 - and its distributions, 165
- ordering of interest, 20, 21, 178, 317, 365, 710, 715
 - cross-section data, 21, 179, 274, 307, 317, 365
 - of the sample, 20
- Ornstein–Uhlenbeck process, 356
- orthogonal decomposition, 302
- orthogonal polynomials, 673
- orthogonal projectors, 663
- outcomes Set, 45
 - countable, 61
 - finite, 60
 - uncountable, 62
- outlier (non-typical observation), testing for, 675
- overidentifying restrictions, 693, 729
- paired sample tests, 622
- parameter space, 95, 446
- parameters and moments, 97
- parameters vs. estimators vs. estimates, 537
- parametric (directional) M-S testing, 697
- parametric inference, 548
- parametric method of moments (PMM), 544, 548
- parametric vs. non-parametric inference, 453
- parametrically nested, 24, 350
- partial sums process, 324
- partition of a set, 51, 62, 67, 74, 449, 571, 576

- Pearson family of frequency curves, 445, 551
- percentiles of a distribution, 114
- permutation invariance, 322
- permutations, 38
- pivotal quantity, 499, 559, 613
- platykurtic distributions, 105, 108
- Poisson distribution, *see* distributions
- Poisson process, 361, 362
- Poisson regression-like model, 657
- Poisson's law of small numbers, 379
- popular misconceptions about limit theorems, 376
- positive dependence, 21, 182, 188, 193, 222, 253
- positive predictive value (PPV), 609
 - blameworthiness by association, 610
 - misinterpreting error probabilities, 610
- power of the test, 573
 - evaluation, 582
 - increases as discrepancies increase, 583
 - increases as n increases, 583
- power curve, 583
- pre-data role, 584
- power set of an outcomes set, 52, 53
- P–P plot, 207, 209
 - Normal IID data, 210
 - standardized, 211
 - standardized Normal, 211
 - Student's t , 217
 - Uniform IID data, 209
 - vs. Q–Q Plots, 214
- prediction (forecasting), 450
- prediction intervals, 636
- pre-image of a function, 80
 - and set-theoretic operations, 88
- preliminary data analysis, 677
- primary objective of frequentist testing, 561
- primitive notions, 73
- principal component analysis, 351
- prior distribution, 440, 441, 464
 - conjugate, 441
 - Jeffreys, 435, 443, 504
 - logistic, 442
 - objective (default, reference), 442, 443
 - reparameterization invariant, 442
 - uniform, 442
 - vs. prior substantive information, 444
- probabilistic assumptions underlying statistical models, 329
 - three broad categories, 12, 28, 177, 179, 316, 318, 386, 454
- probabilistic reduction, 278, 282, 328, 628, 655, 721
 - reduction and model assumptions, 628, 631, 642, 644
- probability integral transformation, 98, 212, 259
 - converse, 116
- probability model, 33, 95
- probability set function, 60, 63
- probability space, 63, 302
 - induced by a random variable, 89
- probit model, 655, 656, 710
- product rule for conditional probability, 66
- propensity interpretation of probability, 431
 - Humphrey's paradox, 432
 - vs. the frequentist interpretation, 431
- pseudo-random numbers, 178
- p-value, 482, 560, 570, 578, 579, 616
 - a post-data perspective, 579
 - a pre-data perspective, 578
 - misinterpretations, 554, 562, 563
 - vs. type I error probability, 578, 579
- Q–Q plot, 207, 213
- quantifier, universal, 151, 486, 507
- quantile function, 114
 - properties, 114
- quantile transformation, 210
- quantiles of a distribution, 114
- quartiles of a distribution, 114
 - interquartile range, 117
 - quartile deviation, 117
- R square, 635
- random experiment, 42, 78
- random sample, 36, 39, 74, 160
 - non-random sample, 222
- random trials, 70, 72
- random variable, 34, 79–81, 87, 169, 422
 - a naive view, 34
 - continuous, 35, 87
 - continuous/discrete, 146
 - degenerate, 82
 - discrete, 35, 80
 - neither random nor a variable, 81, 119
 - range of values, 116
- random vector, 132, 133
- random walk process, 325
 - second-order, 339
- Rao–Blackwell theorem, 482
- regression function, 280, 282
 - characterization, 298, 708
- regression models, 281, 285, 287, 311
 - and heterogeneity, 308
 - and homoskedasticity, 285, 303
- beta, 285
- Binomial, 287
- Exponential, 285, 289
- F, 285
- gamma (Cherian), 285
- gamma (Kibble), 285
- logistic, 285
- LogNormal, 285
- modeling strategy, 289
- Negative Binomial, 287
- Normal, linear regression, 303, 351, 627, 658, 690, 705
- Pareto, 285
- Pearson type II linear regression, 285

762 Index

- Poisson, 287
- selecting a regression model, 312
- statistical GM, 303
- Student's t , linear regression model, 285, 304
- regression-like statistical models, 655
 - generalized linear models, 655, 657
- reluctance to validate model assumptions, 691
- reparameterization, 293, 294, 351, 442
- reparameterization invariance, 442, 507
- replication crises, 609
 - and severity, 609
 - trustworthiness of the evidence, 28
- rescaling variables, 673
- residuals, 629, 632, 634, 644, 646, 647, 660
 - small vs. non-systematic, 650, 655
- residual sum of squares, 527, 634, 663, 707
 - explained sum of squares, 634
 - restricted, 640, 641, 699
 - total sum of squares, 634
 - unrestricted, 640, 664, 699
- reverse regression, 282, 307, 308
 - problem, 307
- ridge regression estimation, 650
 - and curve-fitting, 650
- rules of total variance, 299
- runs up and down, 181
 - test, 694
- sample correlation coefficient, 541
- sample moments and their properties, 539
 - central moments and their first two moments, 540
 - mean and its first four moments, 540
 - raw moments and their first two moments, 539
 - variance and its first two moments, 541
- sample space, 446
- sample vs. sample realization, 319, 446
- sampling distribution, 110, 374, 409, 429, 439, 445, 447, 448, 455, 460, 461, 472, 473, 476
 - actual vs. assumed sampling distribution, 689
- sampling model, 33, 130, 161, 169
- sampling space, 70, 72
- sampling survey methods, 39
 - cluster sampling, 40
 - quota sampling, 40
 - simple random sampling, 39
 - stratified sampling, 39
- scales of measurement for data, 16, 20, 262, 273
 - and mathematical operations, 17
 - and ordering of data, 17
 - interval scale, 17
 - nominal scale, 17
 - ordinal scale, 17
 - ratio scale, 17
- scatterplot (or cross-plot), 237
- score function, 517
 - properties, 518
- sinusoidal polynomials, 309
- sequential conditioning, 225, 226, 277, 328
- set theoretic vs. probabilistic terminology, 57
- set theory and operations, 45, 48
 - cardinality of a set, 47
 - complementation, 49
 - difference of two sets, 50
 - empty set, 49, 82
 - equality of sets, 50
 - intersection, 49
 - subset, 48
 - union of sets, 49
 - universal set, 45, 49
- severity (post-data) evaluation, 578, 598, 600, 601
 - addressing abuses of N-P testing, 601
 - addressing abuses of significance testing, 609
 - addressing the fallacy of acceptance, 605
 - addressing the fallacy of rejection, 605
 - and observed confidence intervals, 614
 - and the large n problem, 603
 - and the p-value, 605
 - error-statistical account of evidence, 602, 603, 609
 - interchanging the null and alternative hypotheses, 601
 - manipulating the level of significance, 601, 608
 - post-data direction of departure, 605
 - statistical vs. substantive significance, 602
 - warranted discrepancy from null, 565
- severity curve, 602, 608
 - accept the null, 604
 - reject the null, 602
- severity principle, 609
 - strong, 609
 - weak, 609
- sigma-field (σ -field), 56, 62
- significance level, *see* type I error probability
- significance level vs. the p-value, 578, 579
- simple random walk, 338, 339
- simple statistical models, 33, 166
 - Bernoulli model, 167, 423, 438, 448, 470, 512, 566
 - bivariate Bernoulli model, 567
 - bivariate Normal model, 711
 - Cauchy model, 552, 591
 - exponential model, 475, 518, 591
 - gamma model, 521
 - Laplace model, 516
 - logistic model, 523
 - Normal model, 167, 177, 225, 303, 423, 438, 493, 519, 558
 - Normal (1-parameter) model, 471, 568, 570, 576, 585, 613
 - Poisson model, 528, 590
 - uniform model, 478, 514
- Simpson's paradox, 715
- skewed distributions, 104
- skewness coefficient, 104, 138
- Spearman's rank coefficient, 229
- square contingency coefficient, 229

- stable family of distributions, 404
- standard deviation, 102
- stationarity, 331
 - first order, 333
 - of order $m \geq 1$, 334
 - second order, 333
 - strict, 332
- statistical adequacy, 2, 23, 24, 28, 308, 312, 365, 439, 445, 465, 469, 565, 572, 616, 635, 647, 655, 693
 - and the reliability/unreliability of inference, 25, 462, 548, 625, 721, 729, 731
 - vs. substantive adequacy, 647, 648, 721
- statistical curve-fitting, 648
- statistical generating mechanism (GM), 278, 301, 311, 312, 366, 429, 728
 - hazard-based models, 364
 - vs. actual generating mechanism, 7, 431, 648, 649
- statistical inference (induction), 32, 74
 - different approaches, 436
- statistical information set, 709
- statistical learning methods, 650
 - and curve-fitting, 650
- statistical misspecification, 21, 262, 453, 456, 462, 465, 466, 610, 685, 730
 - nominal vs. actual error probabilities, 458, 585, 616, 644, 685, 688, 693, 717, 729
 - unreliable inferences, 365, 462, 548, 729
- statistical models, 23, 28, 167, 315, 366, 425, 437, 559, 643, 716
 - and completeness, 484
 - and substantive information, 11
 - data has 'a life of its own', 26
 - identification of parameters, 168
 - internally consistent and testable assumptions, 13, 306, 318
 - misspecified, 616, 632, 646, 690, 691, 713, 715, 724
 - model validation vs. model selection, 685, 687, 703
 - non-regular probability models, 444, 479
 - non-systematic component, 302, 364, 647, 663
 - non-testable model assumptions, 409, 454, 457
 - population vs. stochastic generating mechanism, 40
 - probabilistic reduction, 628, 693, 721
 - probabilistic structure vs. parameterization, 350
 - reduction and model assumptions, 628, 642
 - regularity conditions, 477, 514
 - respecification, 316, 704, 713, 716
 - rigged models, 466
 - set of all possible probability models, 168
 - specification, 13, 177, 316, 721
 - statistical meaningfulness of parameters, 626, 655
 - statistical vs. substantive models, 304
 - systematic component, 364, 647, 663
 - testable probabilistic assumptions, 312, 350, 427, 454, 457, 693
 - what is a statistical model?, 625
- statistical space, 73, 78, 79, 118, 130, 166
- statistical systematic information, 13, 315, 436, 648, 699, 701
 - vs. substantive information, 1, 25, 575
- statistical vs. substantive significance, 554, 599, 603, 607
- statistical vs. structural error term, 647, 648
- statistically meaningful parameterizations, 300, 304, 307, 310, 312, 315, 366, 626, 627, 648, 650, 655
- stereogram, 238
 - smoothed, 238
- stochastic calculus, 358
- stochastic difference equation, 625
- stochastic differential equation, 358
- stochastic integral, 358
- stochastic phenomena of interest, 2, 3, 8, 11, 30, 34, 78, 160, 167, 176, 323, 383, 421, 424, 426, 436, 443, 451
- stochastic processes, 318
 - Bernoulli IID process, 335
 - building block processes, 335
 - classification, 320
 - dependence assumptions, 329
 - ensemble of a stochastic process, 320
 - exponential IID process, 335
 - functional perspective, 323
 - heterogeneity assumptions, 331
 - independent increments homogeneity, 332
 - IID process, 365
 - independent increments process, 326
 - index set (ordering), 318
 - joint distribution perspective, 323
 - Normal IID process, 336
 - sample path viewing angle, 319
 - taming a wild process, 342
- stochastic regression model, 278
- strong law of large numbers (SLLN), 375, 393, 395, 425
 - Borel's SLLN, 391
 - Kolmogorov's first SLLN, 391
 - Kolmogorov's second SLLN, 392
 - uniform SLLN, 409
- structural model, *see* substantive model
- Student's t distribution, *see* distributions
- Student's t martingale difference process, 344
- Student's t test, 454, 455, 565, 593, 637, 638, 640, 690, 698
 - vs. non-parametric tests, 453
- substantive adequacy, 439, 565, 643, 647, 693, 728
 - confounding variables/factors, 564, 648, 665
 - errors of measurement, 648
 - external shocks, 648
 - omitted variables, 307
 - probing for, 647, 693
- substantive (structural) model, 1, 643, 676, 692, 693, 730

764 Index

- structural error term, 648
- vs. statistical model, 365, 675
- substantive information, 1, 7, 11, 24, 26, 301, 308, 312, 315, 432, 437, 444, 605
 - non-systematic, 648
- sufficiency principle (SP), 483
- sure event, 51
- symmetric distribution, 103, 302, 663
- Taylor's series expansion, 542
- temporal dependence, 20, 223, 243, 288, 339, 354, 719, 722
- test statistic, 355, 556, 560, 561, 566, 568, 576, 578
- testing linear restrictions, 639, 641, 663, 729
 - F-test, 640, 643, 664, 665, 675, 701
- testing the difference of two means, 620
- testing the difference of two proportions, 623
- total probability rule, 67
- t -plot, 3, 4, 11, 178, 179, 219
- transition probabilities, 338
- trend polynomials, 219, 701
- trigonometric polynomials, 191
- trivial event space, 52
- trivial field, 296
- truncation, 148, 149
- two-parameter exponential family, 680, 681
- unbiased estimator, 474
 - and consistent, 489
 - and full efficient, 527
 - vs. parameterization invariant, 475
- uniformly asymptotic negligibility, 398
- untrustworthy evidence, 365, 453, 644, 717
- variance, 100
 - heterogeneity, 186, 725
 - properties, 101
- variance inflation factors (VIF), 670
- Venn diagrams, 49
- von Mises' frequentist interpretation of probability, 426
 - von Mises collective, 427
 - vs. model based frequentist interpretation, 427
- weak exogeneity, 292–294, 297
- variation freeness, 293
- weak law of large numbers (WLLN), 375
 - Bernoulli's WLLN, 378, 383
 - Bernstein's WLLN, 388
 - Chebyshev's WLLN, 387
 - Khinchin's WLLN, 390
 - Markov's WLLN, 388
 - Poisson's WLLN, 386
- Weibull hazard-based model, 365
- white-noise process, 336, 344, 352, 367
 - vs. innovation process, 345
- Wiener process, 349
- Yule's Q, 264
- Yule's reverse engineering, 711
- zero probability paradox, 578