CAMBRIDGE
Cambridge University Press & Assessment
978-1-108-48967-6 — Principles of Statistical Analysis
Ery Arias-Castro
Index
More Information

Index

A/B testing, *see* sequential design
additive models, 347
admissibility, 188
σ-algebra, 8
  Borel, 35
alternative
  hypothesis, 170
  set, 170
Anderson–Darling tests, 252
antitonic regression, *see* isotonic regression
association
  affine, 301
  in causal inference, 366, 367
  in observational studies, 217
  monotonic, 302
average causal effect, 367
average power, 188
average risk, 185

backfitting algorithm, 347
balanced design, 148
balanced incomplete block design, 149
bandwidth, 264, 334, 336
  choice, 264, 266, 343
bar chart, 203
  segmented, 217
  side-by-side, 203
base rate, 16
Base Rate Fallacy, 15, *see also* base rate
Bayes classifier, 333
Bayes estimator, 185
Bayes formula, 15
Bayes risk, 185, *see* average risk
Benjamini–Hochberg multiple test, 319
Bernoulli distribution, 42, *see also* binomial distribution
Bernoulli trials, 21, 105, 164, 199, 224, *see also* Bernoulli distribution
beta distribution, 65
bias, 166
  blinding to control, 147
  bootstrap estimate, 243
  in survey sampling, 140
bias-variance decomposition, 166, 265
binning, 235
binomial coefficient, 24
binomial distribution, 24, 41, 42
  experiment, 164, 192
  normal approximation, 55
  Poisson approximation, 48
Bonferroni correction, 318
Bonferroni's inequalities, 10
Boole's inequality, 10
bootstrap, 208
  empirical, 242, 260
  parametric, 260
  smooth, 260
bootstrap confidence interval, 246, 275
  Studentized, 247, 273
bootstrap distribution, 242
bootstrap p-value, 207, 210, 220, 255, 282, 286, 300
bootstrap world, 241
Borel–Cantelli lemmas, 100
boxplot, 40

case-control study, 156
Catalan numbers, 30
categorical variable, 70
Cauchy distribution, 66
causal inference, 366
  natural experiment, 158
  randomization, 147
Central Limit Theorem, 105
  Lindeberg, 107
  Lyapunov, 107
Chain Rule, 14

© in this web service Cambridge University Press & Assessment www.cambridge.org

CAMBRIDGE
Cambridge University Press & Assessment
978-1-108-48967-6 — Principles of Statistical Analysis
Ery Arias-Castro
Index
More Information

characteristic function, 89, 106
  goodness-of-fit tests, 269, 287, 306
Chebyshev's inequality, 90
  confidence interval, 194
Chernoff's inequality, 91
chi-squared distribution, 66
classification, 330, 331
  additive, 347
  linear, 342
  local, 336
classification boundary, 350
clinical trial, 144
cluster sampling, 143
Cochran–Mantel–Haenszel test, 322
cohort study, 155, 321
combination tests, 312
Combinatorics, 20
completely randomized design, 148
compound Poisson distribution, 93
compound sum, 93
concentration inequalities, 90–92
conditional
  distribution, 75
  expectation, 87
  probability, 12
  variance, 87
confidence band, 253
confidence interval, 168
confidence region, 168
conformal prediction, *see* prediction interval
confounding, 146
consistency, 182
  estimator, 183
  test, 184
contingency table, 211, 219, 323
continuous distribution, 57, 72
  absolutely, 58, 72
  random variable, 60
  random vector, 72
Continuous Mapping Theorem, 109
convenience sampling, 141
convergence
  distribution, 102
  probability, 101
convolution, 77
correlation, 86
  Kendall, 304
  Pearson, 301
  Spearman, 303
correlation analysis, 299
counterfactual model, 367
counts, 48, 71, 209, 235, *see also* contingency table
  estimated expected, 210
  expected, 206
  observed, 206
covariance, 85
Cramér–von Mises test, 252
critical value, 175
cross-sectional study, 157
cross-validation, 266, 344
crossover design, 152
cumulative distribution function, *see* distribution function
curse of dimensionality, 338

data splitting, 343
de Moivre–Laplace Theorem, 55, *see also* Central Limit Theorem
deconvolution, *see* measurement error
Defendant's Fallacy, 17
density function, 59, 72
  conditional, 75
  convolution, 77
  estimation, 263, 266, 336, 350
  goodness-of-fit tests, 253
  independence testing, 306
  likelihood, 179
differential entropy, 98
discovery, *see* rejection
discrete distribution, 69
  random variable, 41
discrete distributions, 41
discriminant analysis, 350
  linear, 350
disjunct design, 151
distance covariance, 307, 308, *see also* energy statistics
distribution function, 36, 69
double-blind experiment, 147
Dvoretzky–Kiefer–Wolfowitz Theorem, 233

empirical bootstrap, 260
empirical distribution, 231, *see also* Glivenko–Cantelli Theorem, Dvoretzky–Kiefer–Wolfowitz Theorem, empirical bootstrap
  distribution function, 231, 250, 253, 276, 305
  quantile function, 233, *see also* order statistics
Empirical Process Theory, 259

© in this web service Cambridge University Press & Assessment www.cambridge.org

CAMBRIDGE
Cambridge University Press & Assessment
978-1-108-48967-6 — Principles of Statistical Analysis
Ery Arias-Castro
Index
More Information

empirical risk, 339
  minimization, 340
energy statistics, 287
entropy, 97
estimate, 165, *see also* estimator
estimation, 94
estimator, 165
events, 7
exchangeability, 110, 223, 295
expectation, 78
expected loss, *see* risk
expected risk, 332
experiment, 6
experimental design, elements of, 144
experimental design, examples of, 148–153
exponential distribution, 57, 63

F-distribution, *see* Fisher distribution
F-test, 281, 287
factor, *see* categorical variable
factorial, 23
factorization criterion, 181
fail-safe number, 325
false discovery rate (FDR), 316
  marginal (mFDR), 328
  procedures, 319–321
false negative, *see* Type II error
false non-discovery rate, 317
false positive, *see* Type I error
family-wise error rate
  procedures, 317
family-wise error rate (FWER), 316
  procedures, 319
file drawer problem, 324
Fisher distribution, 66
Fisher's Exact Test, 215
Fisher's exact test, 217
fitting, 332
Fourier transform, 89
Fréchet distribution, 108
frequencies, *see* binning, counts
Friedman test, 297
funnel plot, 322

Galton–Watson process, 121
Gambler's Fallacy, 12, 22
Gambler's Ruin, 120, *see also* random walk
gamma distribution, 65
Gaussian distribution, *see* normal distribution
general exponential family, 191
General Law of Multiplication, 14
generalization error, *see* expected risk
generalized likelihood ratio, *see* likelihood ratio
geometric distribution, 44
  experiment, 199
Glivenko–Cantelli Theorem, 232
global null hypothesis, 312
global testing, 281, *see also* global null hypothesis
goodness-of-fit, testing for, 204, 208, 250, 276, 308, 359
graph, 117
Grenander estimator, 267
group testing, 150
Gumbel distribution, 108

hazard rate, 257
higher criticism, 313, *see also* Anderson–Darling tests
histogram, *see also* kernel density estimation, 234
Hochberg multiple test, 319
Hodges–Lehmann estimator, 298
Hoeffding test, 306
Holm multiple test, 318
Hommel multiple test, 319
homogeneity, testing for, 208
hypergeometric distribution, 44
  experiment, 197
hypothesis testing, 170, *see also* test, multiple hypothesis testing

identifiability, *see also* factorization criterion, 164
inclusion-exclusion formula, 11
independence, 13
  events, 13
  mutual, 14
  pairwise, 14
  random variables, 42, 74
  testing for, 217, 299
independent and identically distributed (iid), 103
interpolation, 340
Ising model, 135, 363
isotonic regression, 348

joint distribution, 69
joint independence, *see* mutual independence

Kaplan–Meier estimator, 258
Kendall correlation, 304

© in this web service Cambridge University Press & Assessment www.cambridge.org

CAMBRIDGE
Cambridge University Press & Assessment
978-1-108-48967-6 — Principles of Statistical Analysis
Ery Arias-Castro
Index
More Information

Kendall's τ, *see* Kendall correlation
kernel density estimation, 263
kernel function, 263
kernel regression, 335
Kolmogorov distribution, 251
Kolmogorov–Smirnov test, 251
Kruskal–Wallis test, 283
Kullback–Leibler divergence, 98, 254

Laplace transform, 88
Law of Addition, 9
Law of Large Numbers, 103
Law of Multiplication, 13
Law of Small Numbers, 48
Law of Total Probability, 9
leave-q-out cross-validation, 345
leave-one-out cross-validation, 266, 346
Lebesgue integral, 59
level of a test, *see* significance level
likelihood function, 167, 179
likelihood ratio, 171
linear classification, 342
linear models, 341, *see also* linear regression
linear regression, 341
  least squares, 342
  polynomial regression, 341
Literary Digest Poll, 142
local average, 334, *see also* kernel regression
local linear regression, 335
local methods for regression, 334
location family of distributions, 60
location-scale family of distributions, 60
LOESS, *see* local linear regression
log-concave density, 269
log-rank test, 288
logarithmic distribution, 93
logistic regression, 343
longest run test, 224
loss, 166, 331
  surrogate, 342

many-to-one comparison, 281
margin of error, 196
marginal distribution, 69
Markov chain, 113
  ergodic theorem, 134
  irreducible, 116
  positive recurrent, 116
  reversible, 117
  stationary distribution, 115
  transition matrix, 114
Markov chain Monte Carlo (MCMC), 132
Markov's inequality, 90
mass function, 19, 69, *see also* density function
matched-pairs design, 153, 213
matching, 157
maximum likelihood estimator, 167
maximum risk, 184
mean, 78, *see also* expectation
mean absolute error, 166, *see also* risk
mean integrated squared error, 266, 352
mean squared error, 166, *see also* risk
measurable function, 36
measurable space, 8
median, 40, *see also* quartile, quantile
  inference, 237
median test, 284, *see also* sign test
Meta-analysis, 321
method of least squares, 342
method of moments, 260, 261
Metropolis–Hastings Algorithm, 135
minimax estimator, 184
minimax risk, 184, *see also* maximum risk
minimum power, 188
moment, 82
  central, 84
moment generating function, 87
monotone likelihood ratio, 190
Monte Carlo integration, 128
Monte Carlo simulation, 127, 128
Monty Hall Problem, 12
Moran urn model, 28
multinomial distribution, 71
multiple test, 314
multiple testing, 310
multivariate hypergeometric distribution, 72
mutual independence, 14, *see also* independence
  events, 14
  random variables, 74

Nadaraya–Watson estimator, *see* kernel regression
naive Bayes, 351
natural experiment, 158
nearest neighbor classifier, 337
negative binomial distribution, 46
  experiment, 199
negative hypergeometric distribution, 47
  experiment, 199
neighborhood

© in this web service Cambridge University Press & Assessment www.cambridge.org

CAMBRIDGE
Cambridge University Press & Assessment
978-1-108-48967-6 — Principles of Statistical Analysis
Ery Arias-Castro
Index
More Information

ball neighbors, 333
nearest neighbors, 333
network sampling, 143
Neyman–Pearson Lemma, 189
non-response bias, 140
normal approximation to the binomial distribution, *see* de Moivre–Laplace Theorem
normal distribution, 62
standard, 56, 62
testing for, 255
normal sequence model, 311
null distribution, 250
null hypothesis, 170
null set, 170
number of runs test, 224

observational study, 153
optional stopping, 201
order statistics, 230
outcomes, 6
overfitting, 341

p-value, 173
adjusted, 327
permutation, 212, 215, 296, 300
Monte Carlo, 213
randomization, 356, 357
valid, 174
Pólya urn model, 27
pairwise independence, 14
Paley–Zygmund inequality, 97
parameter space, 164
parametric bootstrap, 260
pattern, 280
rank, 303
sign, 294
Pearson correlation, 301
permutation distribution, 277
rank tests, 296
pie chart, 203
pivot, 246
placebo, 147
Poisson approximation to the binomial, *see* Law of Small Numbers
Poisson distribution, 47
pooled adjacent violators algorithm, 348
power calculations, 146
power of a test, *see also* Type II error, average power, 177
power set, 8
prediction, 331
prediction error, *see* risk
prediction interval, 261
regression, 348
prediction residual error sum-of-squares, 346, *see also* leave-one-out cross-validation
predictor variable, 329
prior, 185
probabilistic modeling, 163
probability axioms, 8
probability distribution, 8
probability generating function, 88
probability space, 9
proportion test
one-sample, 201
two-sample, 228
proposal distribution, 130
Prosecutor's Fallacy, 16
pseudo-random number generator, 137
publication bias, 322

quantile, 40
function, 39
quartile, *see also* quantile

Rademacher distribution, 294
random matrix, 77
random variable, 34
random vector, 68
random walk, 124
simple, 119
randomization, 146
p-value, 357
randomized complete block design, 149
randomness, testing for, 222
range, 36
rank
pattern, 303
statistic, 278
test, 278
Rasch model, 360
ratio of uniforms, 132
re-randomization, 355
p-value, 356
test, 355
regression analysis, 329
regression discontinuity design, 160
regression estimator, 331
regression function, 332
rejection, 170
region, 175
rejection sampling, 130
relative entropy, *see* Kullback–Leibler divergence

© in this web service Cambridge University Press & Assessment www.cambridge.org

CAMBRIDGE
Cambridge University Press & Assessment
978-1-108-48967-6 — Principles of Statistical Analysis
Ery Arias-Castro
Index
More Information

repeated measures design, 151
response bias, 140
response variable, 329
Riemann integral, 59
risk, 166, 331, *see also* average risk, expected risk
risk unbiased estimator, 186
run, 223

Saint Petersburg Paradox, 94
sample mean, 244
sample median, 237
sample space, 6
sample variance, 245
sampling
  cluster, 143
  network, 143
  stratified, 143
  systematic, 142
  with replacement, 25
  without replacement, 25
scale family of distributions, 60
self-selection bias, 140
sensitivity, 16, *see also* Type I error
sequential design, 149
sequential probability ratio test, 200
Set Theory, 3
shape constraint, 266, 348
sign test, 239
significance level, *see also* Type II error, 175
simple random sampling, 140
Simpson's paradox, 221
Simpson's reversal, 32
size of a test, *see also* significance level, 175, *see also* level
Slutky's theorem, 110
small study effect, 322
Smirnov test for symmetry, 297
smooth bootstrap, 260
spacing, 359
Spearman correlation, 303
Spearman's ρ, *see* Spearman correlation
species co-occurrence, 362
specificity, 16, *see also* Type II error
split plot design, 150
standard deviation, 84
statistic, 165, *see also* estimator, test
statistical inference, 163
statistical model, 163
statistical procedure, 182
Stirling's formula, 29
stochastic dominance, 284
stratification, 143
stratified sampling, 143
Student distribution, 66
Student test, 269, 286
sufficiency, 181
support, 36, 69
support vector machines, 343
surrogate loss, 342
Survey Sampling, 139
Survival Analysis, 256
survival function, 38, *see also* distribution function
symmetric distribution, 291
syndromic surveillance, 309
systematic sampling, 142

t-distribution, *see* Student distribution
t-test, *see* Student test
test, 174, *see also* multiple test
test set, 344
Tippett–Šidák multiple test, 317
total variation distance, 254
Tracy–Widom distribution, 112
training, *see* fitting
training set, 343
Two Envelopes Problem, 30
Type I error, 175
Type II error, 175

unbiased estimator, 187, *see* risk unbiased estimator
unbiased test, 191
uniform distribution, 55, 61
  discrete, 10, 20, 46
  testing for, 250
uniformly most powerful, 189
  unbiased, 191
union bound, *see* Boole's inequality
urn model, 6, 24, 27–29

validation set, 343
variance, 84
  bootstrap estimate, 243
Venn diagram, 4

weak convergence, *see* convergence in distribution
Weibull distribution, 108
Wilcoxon rank-sum test, 279
Wilcoxon signed-rank test, 293
Wright–Fisher urn model, 28

zero-one laws, 100

© in this web service Cambridge University Press & Assessment www.cambridge.org