CAMBRIDGE

Cambridge University Press & Assessment
978-1-107-17891-5 — Foundations of Agnostic Statistics
Peter M. Aronow, Benjamin T. Miller
Index
More Information

# Index

absolute convergence, 45n2, 55n7
Addition Rule, 9–10
almost sure convergence, 117n29
asymptotic estimation theory, 111–116
asymptotic mean squared error, 113
asymptotic normality, 112, 125
  of plug-in estimators, 120
asymptotic relative efficiency, 113
asymptotic standard error, 113
asymptotic unbiasedness, 112
atomic events, 6
average marginal causal effect, 255
average partial derivative, 166–168, 171, 255
average treatment effect (ATE), 237–238
  conditional, 249–250
  decomposition, 238
  plug-in estimation of, 256–257
  sharp bounds for, 240–243
  under random assignment, 245
  under strong ignorability, 248–249, 253, 255
average treatment effect among the treated (ATT), 238, 247, 273

balance, 218, 251–252
balance test, 252, 270n14
bandwidth, 121
Bayes' Rule, 11–12
  alternative form of, 13
Bayesian probability, 4n2
Bernoulli distribution, 19, 46, 58, 67, 70, 135, 182
  generalized, 180
Bernoulli random variable, *see* Bernoulli distribution
best linear predictor (BLP), 76–84
  and independence, 82–84
  and the CEF, 78–79
  coefficients of, 87
  estimating, 144–147
  multivariate case, 86–87
  properties of deviations from, 79, 87
  under nonlinearity, 81
bias, 103
binary choice model, 182–185, 226, 233, 264, 279
  CEF for, 183, 198
bivariate normal distribution, 66n13
bivariate relationships, 31–39
bootstrap, 130–132, 141–142, 153–154, 203, 234
bounds, *see* sharp bounds

cardinality, 7, 284
categorical distribution, 180
Cauchy–Schwarz Inequality, 63n12
causal inference, 235–281
  as a missing data problem, 238–239
  Double Robustness Theorem for, 267–269
  doubly robust estimator for, 267, 269–270
  propensity score for, 250–252
  regression estimator for, 260, 261
Central Limit Theorem (CLT)
  for cluster samples, 139
  for sample means, 108–110



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

CAMBRIDGE

Cambridge University Press & Assessment
978-1-107-17891-5 — Foundations of Agnostic Statistics
Peter M. Aronow, Benjamin T. Miller
Index
More Information



Chebyshev's Inequality, 55, 105
  for the sample mean, 98
classical linear model, 180–182
  CEF for, 181, 198
  maximum likelihood estimation of, 189–192
cluster average, 136–139
cluster sample mean, 136–137
  sampling variance of, 137, 140
cluster samples, 135–141
  CLT for, 139
  estimation with, 136–139
  inference with, 140–141
  WLLN for, 138
collinearity, 132n43, 149
complement, 6, 22, 24, 283
complement rule, 8, 9
conditional average treatment effect, 249–250
  under strong ignorability, 249
conditional expectation, 67–69
  linearity of, 69, 249
  multivariate case, 85
  of a function of a random variable, 68
conditional expectation function (CEF), 69–76
  and independence, 82–84
  and the BLP, 78–79
  for binary choice models, 183, 198
  for the classical linear model, 181, 198
  kernel plug-in estimation of, 124
  multivariate case, 85
  partial derivatives of, 165–168
  polynomial approximation of, 158, 165n11, 169
  properties of deviations from, 74–75
  under MAR, 217
conditional independence assumption, 216, 248, 254, 255, 270–272, 274
conditional PDF, 37–38
  multivariate case, 42
conditional PMF, 34
  multivariate case, 42
conditional probability, 10–12, 67, 182
  and independence, 14–15, 39
conditional variance, 68, 74
conditional variance function, 70, 85
confidence interval, 124–127, 134–135, 212, 231
  asymptotic validity of, 125
  for regression, 154
  for the population mean, 127
  normal approximation-based, 125, 128, 131, 133–135, 140
  validity of, 124
confidence level, 124
consistency, 105
  of maximum likelihood estimation, 193
  of sampling variance and standard error estimators, 113
  of the sample mean, 105
Continuous Mapping Theorem (CMT), 100, 107, 108, 110, 114, 116, 125–126, 140
continuous random variable, 24
  event probabilities for, 25–27
convergence
  absolute, 45n2, 55n7
  almost sure, 117n29
  in distribution, 108
  in probability, 99–100
correlation, 62–63
  and independence, 64–66
  and linear dependence, 62
  properties of, 63
covariance, 59–60
  alternative formula for, 60
  and independence, 64
  properties of, 61–62
covariance matrix, 84
cumulative distribution function (CDF), 21–24, 101–102
  empirical, 116–120
  estimating, 102, 117
  joint, 32–33, 40
  properties of, 22

degenerate distribution, 47
difference-in-means estimator, 245, 256, 257, 259
discrete random variable, 18–21
  event probabilities for, 21
distribution, 3–4, 20, 21
Double Robustness Theorem
  for causal inference, 267–269
  for missing data, 229–230
doubly robust (DR) estimator
  for causal inference, 267, 269–270
  for missing data, 228, 230
draw, 4

efficiency, *see* relative efficiency
elementary event, 6
empirical CDF, 116–117, 120
estimation theory, 96–116
  asymptotic, 111–116
estimator, 103–105
  bias of, 103
  consistency of, 105

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

CAMBRIDGE
Cambridge University Press & Assessment
978-1-107-17891-5 — Foundations of Agnostic Statistics
Peter M. Aronow, Benjamin T. Miller
Index
More Information



doubly robust, *see* doubly robust (DR) estimator
inverse probability–weighted (IPW), *see* inverse probability–weighted estimators
kernel density, 122
kernel plug-in, 123
lasso, 169–170, 174, 202
MSE of, 104–105
plug in, *see* plug-in estimator
post-stratification, 219, 256
regression, *see* regression
relative efficiency of, 105
sampling variance of, 104
sieve, 168–169
standard error of, 104
unbiased, 103
event, *see* random event
event space, 5–8
expectation, 45–50
conditional, 67–69
linearity of, 49–50
of a function of a random variable, 47
of a function of two random variables, 49
of a random vector, 48
of the sample mean, 97
properties of, 47–50
under MAR, 216
under MCAR, 213–215
expectation operator, 45, 47
expected value, *see* expectation
experiment, 4
explanatory variable, 79

finite population, 93–94
frequentist probability, 4
function of a random variable, 16–17
conditional expectation of, 68
equality of, 32, 70
expectation of, 47, 49, 68
functional, *see* statistical functional
Fundamental Problem of Causal Inference, 236–239, 270

geometric distribution, 19
Glivenko–Cantelli Theorem, 102n14, 117n27, 120n31

heteroskedasticity, 156
homoskedasticity, 155, 156, 261
hot deck imputation, 224–225, 233
hypothesis testing, 128–130

i.i.d., 91–92, 94–98
identification
with missing data, 208–219
with potential outcomes, 236–256
ignorability, *see* missing at random (MAR)
imputation
hot deck, 224–226, 233
under MAR, 220
under MCAR, 214–215
under random assignment, 246
under strong ignorability, 258
independence
CEFs and BLPs under, 82–84
conditional, 216, 248, 251, 254, 255, 270–272, 274
conditional probability and, 14–15, 39
correlation and, 64–66
covariance and, 59–60
implications of, 39, 64–65, 82
of events, 14
of random variables, 38–39
independent and identically distributed, *see* i.i.d.
inference, 124–135
with clustering, 140–141
with maximum likelihood estimation, 202–203
with regression, 151–156
interactions, 164–165
inverse probability–weighted (IPW) estimators, 226–228, 231, 234, 264, 266–267, 270, 275, 279
for causal inference, 266–267
for missing data, 227–228
stabilized, 228, 266–267

joint CDF, 32–33
multivariate case, 40
joint PDF, 36
multivariate case, 41
joint PMF, 32
multivariate case, 40
joint probability, 9–10
jointly continuous random variables, 36–38, 41, 42, 49, 50, 64–65, 67–69, 72–73, 85

kernel, 121
kernel density estimator, 121
kernel plug-in estimator, 123
Kolmogorov Axioms, 6–8, 20, 183
Kullback–Leibler Divergence, 195–196

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

CAMBRIDGE
Cambridge University Press & Assessment
978-1-107-17891-5 — Foundations of Agnostic Statistics
Peter M. Aronow, Benjamin T. Miller
Index
More Information



lasso estimator, 169–170, 174, 202
Law of Iterated Expectations, 71–76, 155, 167, 198, 211, 212, 216, 218, 220, 227, 229, 241–243, 248, 250, 251, 258, 265, 268
Law of Total Probability, 12–13, 15, 33
Law of Total Variance, 73–75
leave-one-out cross-validation, 170, 174
likelihood function, 186, 189
linearity of expectations, 49–50
link function, 183n3
log-likelihood, 187–190, 192, 194, 202
logit model, 183–184, 233

M-estimators, 187n7
Manski bounds, *see* sharp bounds
marginal effect, 165
marginal PDF, 37
  multivariate case, 41–42
marginal PMF, 33–34
  multivariate case, 41–42
matching, 262–264
maximum likelihood (ML) estimation, 185–203, 233
  consistency of, 193
  inference with , 202–203
  logic of, 185–193
  of mixture models, 198–201
  of the classical linear model , 189–192
  of the probit model, 192–193
  penalized, 201–202
  plug-in estimator, 197–198
  under misspecification, 194
mean function, 183–184, 226, 264
mean squared error (MSE), 56–58
  alternative formula for, 57
  asymptotic, 113
  expected value minimizes, 57–58, 151
  of an estimator, 104
  root, 57
micronumerosity, 132–135, 149, 161, 162, 168, 274
missing at random (MAR), 215–217
  and the propensity score, 217–219
  CEF under, 217
  expected value under, 216
  imputation under, 220–221
  reweighting under, 226–227
missing completely at random (MCAR), 213–215
  expected value under, 213–214
  imputation under, 214–215
missing data, 208–219
  Double Robustness Theorem for, 230, 267
  doubly robust estimator for, 228, 230
  plug-in estimator for, 220–222
  propensity score for, 217–219
  regression estimator for, 224
  with binary outcomes, 209, 212, 214, 220, 222, 239
mixture model, 198–201
models, *see* parametric models
moments
  central, 51
  raw, 51
multicollinearity, *see* collinearity
multiplicative law for PDFs, 38
multiplicative law for PMFs, 35
Multiplicative Law of Probability, 10, 11, 13, 14, 35, 38
multivariate normal distribution, 152

negative control, *see* placebo testing
Newton–Raphson method, 191n11
Neyman–Rubin Causal Model (NRCM), 236, 239, 254
normal distribution, 55–56
  bivariate, 66n13
  mean and standard deviation of, 56
  properties of, 56
  standard, 28–30
null hypothesis, 128–130, 278

operator, 17, 45, 47, 67, 118
ordinary least squares (OLS) regression estimator, 145–146, *see also* regression
outcome, 4–5, *see also* potential outcomes model
overfitting, 161–164
overlap
  empirical, 273–276
  population, 271–273

*p*-value, 128–133, 140, 152, 154, 278, 279
parameter space, 179, 180, 183
parameter vector, 179, 182, 185
parametric models, 178–203
  as approximations, 203
  binary choice models, 182–185
  classical linear model, 180–182
  maximum likelihood estimation of, *see* maximum likelihood estimation
partial derivative at the average, 166–168, 171

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

CAMBRIDGE
Cambridge University Press & Assessment
978-1-107-17891-5 — Foundations of Agnostic Statistics
Peter M. Aronow, Benjamin T. Miller
Index
More Information



partition, 12–13
penalized maximum likelihood estimation, 201–202
penalized regression, 169–170
penalty parameter, 169, 202
placebo testing, 270–271
plug-in estimator, 120
  asymptotic normality of, 120
  for the ATE under strong ignorability, 256–257
  for the expected value under MAR, 219–220
  maximum likelihood, 197–198
  plug-in sample variance as, 119
  regression as, 143–147
  regularity conditions, 120
  sample mean as, 118, 210, 240
plug-in principle, 116–124
plug-in sample variance, 105–108
  as plug-in estimator, 119
  properties of, 106
point identification, 207
polynomial approximation of the CEF, 158, 165n11, 169
population, 91, 95–96
  finite, 93–94
population mean, 93, 94, 97, 101, 103, 105, 106, 110, 115, 124, 127, 132, 133, 143, 151, 212, 214, 219, 224, 243
  confidence interval for, 127
  consistency of the sample mean for, 105
population regression function, 145
population variance, 94, 105, 116
positivity, 271–276
post-stratification estimator, 219, 256
post-treatment variables, 252–254
potential outcomes model, 236–239, 245, 248, 262, 265, 277, 278
  generalized, 254–256
probability density function (PDF), 24–25
  conditional, 37–38, 42
  joint, 36, 40
  marginal, 37, 42
  multiplicative law for, 38
  properties of, 25
probability limit, 99
probability mass function (PMF), 18–21
  conditional, 42
  joint, 32, 40
  marginal, 33–34, 41–42
  multiplicative law for, 35
  properties of, 20
probability measure, 6, 7
probability space, 6, 8
probit model, 183
  maximum likelihood estimation of , 192–193
propensity score
  for causal inference, 250–252
  for missing data, 217–219
  imputation with, 224–225
  MAR and, 217–219
  matching on, 262–264
  maximum likelihood plug-in estimation of, 225, 264
  strong ignorability and, 251–252
  Winsorization, 275
pseudopopulation, 95

random assignment, 244–247
  ATE under, 245
  imputation under, 246
random event, 4, 15
  independence of, 14
random generative process, 3–8, 14–16, 93
random sampling, 92–94
  finite population, 93–94
  with replacement, 94
  without replacement, 94
random variable, 15–31
  continuous, 24–30
  degenerate, 47, 108, 110, 113
  discrete, 18–21
  equality of, 31–32
  function of, 16
  independence of, 38–39
  independent and identically distributed, 91–92
  operator on, 17, 45
random vector, 39–40
  covariance matrix of, 84
  expectation of, 48
range, 16n15, 18, 284
realization, 4
regression, 143–151
  as plug-in estimator, 143–147
  bivariate case, 144–145
  classical standard errors for, 155–156
  confidence intervals for, 154
  estimator for causal inference, 258–261
  estimator for missing data, 222–224
  penalized, 169–170
  properties of residuals from, 149–151
  robust sampling variance estimator for, 153
  robust standard errors for, 152–154
  with matrix algebra, 147–151

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

CAMBRIDGE

Cambridge University Press & Assessment
978-1-107-17891-5 — Foundations of Agnostic Statistics
Peter M. Aronow, Benjamin T. Miller
Index
More Information



regressor matrix, 147
relative efficiency, 105
  asymptotic, 113
residual, 147–149
response propensity function, 217, 225, 230
reweighting
  survey, 220–222
  under MAR, 226–227
  under strong ignorability, 265–266
root mean squared error, 57

sample mean
  as plug-in estimator, 118, 210, 240
  Chebyshev's Inequality for, 98–99
  consistency of, 105
  consistency of the standard error of, 115
  estimating the sampling variance of, 114
  expected value of, 97
  sampling variance of, 97–98
  standard error of, 115
  standardized, 109, 139
sample points, 5
sample space, 5
sample statistic, 96
sample variance
  plug-in, 105–107
  unbiased, 107
sampling, *see* random sampling
sampling variance, 104
  estimator of, 114–115
  of the cluster sample mean, 137, 140
  of the sample mean, 97–98
sandwich estimator, 153
sharp bounds, 209–213, 231, 240–244
  estimating, 211–212, 243–244
  for the ATE, 240–243
  for the expected value, 209–211
sieve estimator, 168–169
sigma field, 6n6
simple event, 6
Slutsky's Theorem, 109–110, 125, 126
stabilized IPW estimator
  for causal inference, 266–267
  for missing data, 228
stable outcomes model, 208
stable unit treatment value assumption, 237
standard deviation, 53
  of the normal distribution, 56
  properties of, 53–54
standard error, 104
  asymptotic, 113
  estimator of, 113–115
  of the regression estimator, *see* regression
  of the sample mean, 115
standard normal distribution, 28–30, 55
  expected value of, 46
standard uniform distribution, 28
standardized sample mean, 109, 139
statistical functional, 118
strong ignorability, 247–258, 261, 262
  and the propensity score, 250–252
  ATE under, 248–249, 253, 255
  conditional ATE under, 249
  imputation under, 257–258
  reweighting under, 265–266
superpopulation, 94
support, 30–31
survey reweighting, 221

*t*-statistic, 129
test statistic, 128n37
treatment effect, *see* unit-level treatment effect
treatment propensity function, 250, 264, 269, 273

unbiased sample variance, 107–108
  properties of, 108
unbiasedness, 103–105
  asymptotic, 112
uniform distribution, 28
unit-level treatment effect, 237

variance, 51–53
  alternative formula for, 52, 57
  conditional, 68
  population, 94, 105, 116
  properties of, 52–53
  sample, 105–108
  sampling, 104
variance rule, 60, 65
  multivariate, 84, 137

Weak Law of Large Numbers (WLLN), 99–102, 105, 107, 189, 193, 212, 243
  for cluster samples, 138
Weierstrass Approximation Theorem, 158
weighting estimator, 226–228, 264–267, 273–276
Winsorization, 275

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