
Index

- 21st century, xi, 1, 15, 240
- acceleration, 139
- accuracy calculations, 27, 126, 141
- accuracy formula, 114, 139, 148
- α , 146–148, 250
- apparent error, 216
- atom, 217, 228, 229
- batting averages, 7, 8
- BAX gene, 163, 164
- Bayes, x, 4, 5, 10, 12, 14, 18, 19, 23–25, 28, 55, 57, 69, 82, 90, 95, 97, 192, 205, 217, 220, 223, 231
- hierarchical, 13
 - objective, 13
- Bayes factor, 45, 74
- Bayes intervals, 233
- Bayes methods, 2
- Bayes rule, 2, 3, 7, 13, 18, 53, 189, 211, 212, 230, 232, 233
- Bayes theorem, 2, 27, 80, 188, 199, 217
- Benjamini and Hochberg, x, 18, 23, 28, 46, 48, 52, 69
- Benjamini–Hochberg algorithm, 48, 49, 54, 56, 57, 59, 64, 66, 69, 173, 204
- bin centers, 116, 127, 129
- bin width, 79, 90, 116, 118, 119
- binomial distribution, 24, 31, 48
- bivariate normal density, 117
- block statistic, 150
- block test, 152
- Bonferroni bound, 17, 35, 36, 38, 39, 43, 81, 202
- Bonferroni’s method, 81, 200
- Boole’s inequality, 35, 40
- bootstrap, 113, 115, 138–140
- non-parametric, 115, 138, 139
- bootstrap methods, 112, 138
- brain tumor study, 207
- BRCA data, 152, 162, 167, 168
- canonical parameter, 217, 243
- cardio data, 149–151, 153, 162, 249
- central counts, 98, 160, 161
- central matching, 97, 99–101, 104, 105, 108, 112, 129–131, 226, 227
- estimates, 98–100, 103, 104, 129, 130, 226
- chi-square data, 92, 93, 107, 112, 206, 249
- chi-square distribution, 4, 93
- chromosome, 147, 148, 202–205
- sites, 202, 203
- Clemente, 7–9
- coefficient of variation, 23, 24, 77, 128, 178, 236
- column permutation, 153, 165, 166, 168
- combination, 28, 120, 163, 184–186, 192, 209
- comparability, 94, 185, 206–209
- conditional probability, 2, 188, 197, 228
- confidence interval, 12, 14, 24, 230–233
- control algorithm, 48, 173, 204, 230, 231, 233
- Corbet, Alexander, 233–237, 241
- Cornish–Fisher, 59, 135, 139
- correlation, x, 24, 26, 27, 51, 55, 56, 68, 77, 104, 106, 111, 114, 115, 117, 118, 128, 129, 141, 143, 146–150, 152, 153, 155, 157, 159–162, 165, 167, 174, 176–178, 183, 220, 222, 223
- column, 139, 143–146, 152, 153, 156, 168
 - count, 159
 - ecological, 149
 - negative, 68, 111, 160, 162

- rms, 55, 56, 104, 106, 113, 122, 123, 141, 145–149, 153–157, 159, 176, 250
- row, 111, 143–147, 153, 156, 168
- row and column, 141–143
- sample, 67, 142–144, 222
- simultaneous tests of, 64, 67
- correlation across cases, 106, 108, 111, 113
- correlation across units, 106, 109, 110
- correlation clusters, 68
- correlation corrections, 221
- correlation distribution, 118, 120
- correlation effects, 109, 139, 151, 159, 162, 176
- correlation matrix, 113, 117, 144, 149, 152, 222
- correlation of t -values, 157
- correlation structure, 42, 56, 174, 183, 221
- count vector, 116, 117, 126, 159, 161, 196
- covariance matrix, 113, 116–119, 125, 128, 142, 148, 149, 154, 155, 157, 167, 221, 236, 246
- covariate, 11, 80, 95, 107, 133, 147, 197, 198, 202, 206
- unobserved, 107, 108, 111
- CRAN, 174, 215
- cross-validation, 215, 216, 219, 221–223, 240, 250
- cumulant, 218, 230, 244
- generating function, 243, 244
- cumulative distribution functions (cdf), 16, 18, 19, 32, 39, 41, 43, 51, 53, 54, 60, 68, 72, 91, 93, 105, 111, 113–115, 119, 125, 132–135, 138, 165, 169, 173, 237
- empirical, 53, 55, 124, 234
- non-null, 86
- deconvolution, 218
- delta-method, 23, 115, 126, 127, 130
- demeaning, 154, 155
- deviance function, 244, 246
- dichotomous case, 212
- diffusion tensor imaging, 26
- Dirichlet, 201, 245
- disease status, 202
- DTI data, 26, 28, 32, 34, 54, 57, 59–61, 71–73, 181, 182, 185, 186, 189, 199, 249
- dyslexic, 26, 34, 181, 183
- and normal children, 26, 193
- Ebay, 215, 218–223, 240, 250
- EER, 43, 45
- effect size, 181, 212–214, 217–220, 223–225, 227–231, 233, 250
- estimation, 211, 212, 227, 230, 241
- effective sample size, 153, 155–157
- efficiency, 1, 27, 90, 195
- eigenvalues, 142–144
- eigenvector, 149, 151, 152
- empirical Bayes, ix, x, 1, 4, 5, 7, 10, 12–14, 18, 20–22, 27–29, 37, 46, 51–53, 55, 57, 63, 66, 69, 82, 88–90, 114, 115, 190, 191, 198, 199, 212, 213, 217–220, 223, 229, 231, 233–235, 240, 241, 250
- confidence intervals, 12, 14, 233
- methods, 7, 17, 30, 37, 70, 113, 206, 209, 240
- posterior band, 229
- empirical distribution, 20, 79, 109, 139, 169
- empirical null, 84, 90, 91, 94–98, 106–108, 111, 112, 130, 141, 163, 167, 169, 183, 184, 186, 194, 224–226, 250
- density, 90, 164
- distribution, 56, 89, 160, 226, 227
- estimation, 97, 102, 194, 196
- empirical probability distribution, 29
- enrichment, 82, 163, 164, 167, 170, 171, 173, 174, 178, 181–185, 191, 192, 194
- analysis, 163–165, 184
- estimation, x, xi, 1–3, 10, 13, 14, 20, 27, 28, 57, 58, 60, 71, 75–77, 80, 82, 88–90, 98, 99, 103, 104, 123, 156, 188, 195, 201, 220, 229, 233, 241, 243
- effect size, 211, 212, 227, 230, 241
- empirical Bayes, 1, 4, 13, 63, 113, 240
- empirical null, 97, 102, 194, 196
- James–Stein, 10, 11, 209
- maximum likelihood, x, 1, 74, 75, 238, 245, 246
- estimation accuracy, 113, 148
- exchangeability, 80
- expected false and true positives, 82, 83
- exponential family, xi, 74, 82, 103, 104, 171, 174, 217, 224, 230, 243–247

- multiparameter, 105, 137, 245, 247
 - one-parameter, 137, 241, 243
- exponential tilting, 243, 244
- expression matrix, 47, 61, 133, 144, 146, 149
- false coverage rate, 230–232
- false discoveries, 21, 22, 47, 51, 59, 232
- false discovery proportion, 22, 23, 29, 47–49, 51, 53, 55, 56, 106, 107, 130, 160, 201
- false discovery rate, *x*, *xi*, 18, 20, 22, 23, 26, 28, 30, 37, 42, 43, 45–47, 50, 53, 57, 58, 60, 62, 64, 69, 70, 72, 73, 81–83, 97, 102, 105, 106, 128, 159, 160, 164, 184–186, 188, 194, 197–201, 205, 208, 212, 223, 225
 - Bayes, 18, 19, 23–25, 29, 55, 200
 - local, 19, 28, 29, 57, 70, 73, 77, 78, 82, 83, 86–88, 90, 97, 183, 185, 187, 197, 223, 226, 229, 250
 - positive, 51
- false negative rate, 57, 69
- FCR control, 230, 231, 233
- FDR control, 30, 37, 45, 46, 48, 50, 51, 57, 58, 69, 70, 173, 204, 231, 233
- filtration, 109
- Fisher, R.A., *ix*, *x*, 14, 15, 31, 32, 36, 45, 48, 65, 111, 112, 234, 237–239, 241
 - Fisher’s linear discriminant function, 211, 212, 214
 - Fisher’s prediction formula, 237
 - Fisher’s scale of evidence, 31, 46
- fold-change, 207–209
- frequentist, *ix*, *x*, 1, 2, 5, 10, 14, 15, 18, 30, 37, 48, 54, 115, 199, 200, 205, 230, 233
 - methods, 205, 233
- FWER, 30, 34, 35, 43, 44, 47, 49, 50, 81, 83, 185, 199, 200, 205
 - control, 35, 37, 40, 43
- gamma density, 237
- gamma distribution, 134
- gene expression, 15, 141, 149, 163, 167, 212
- gene set, 163–171, 173, 174, 176, 181, 183, 184
- generalized linear model (GLM), 75, 127, 247
 - Poisson, 131, 247, 248
- GSEA statistic, 165, 172–174
- Hermite polynomial, 118, 119
- hierarchical model, 241
 - Bayes, 27, 28, 96, 182
 - normal, 217, 223–226, 228, 241
 - Poisson, 240
- high-dimensional statistical inference, 1
- higher criticism, 59, 69, 205
- hippocampus, 197
- HIV data, 94, 96, 99, 105, 110–112, 148, 249
- Holm’s procedure, 36–38, 40, 81
- homogeneity assumption, 123–125
- hyperparameter, 13
- independence, 24–26, 36, 39, 50, 55, 56, 59, 66, 76, 84, 92, 106, 111, 113, 124, 125, 128, 130, 141, 152, 153, 221, 222, 235, 236
 - assumption, 24, 26, 55, 60, 106
 - column-wise, 141, 149, 153, 162
- indirect evidence, 7, 229, 240
- influence function, 127, 129, 130, 139
- James–Stein estimator, 1, 5–7, 14, 27
- Jensen’s inequality, 22, 23, 155
- k*-FWER, 43–45
- kidney data, 11, 14, 249
- Kolmogorov–Smirnov test statistic, 165, 172
- Kronecker product, 154
- Kullback–Leibler distance, 244
- kurtosis, 132, 135
- large-scale inference, 3, 9, 34, 163
- learning from the experience of others, 10, 27, 184, 206, 229, 240
- least squares, 10, 12, 110, 129, 131, 245
- Lehmann alternatives, 20
- leukemia data, 63, 64, 98, 111–114, 124, 132, 138, 146, 147, 224, 225, 249
- likelihood, *x*, 1–3, 5, 13, 74, 75, 102–104, 171, 238, 245, 246
- limma, 184
- Lindsey’s method, 74–76, 88, 247
- linear logistic model, 194, 195
- linear regression, 5, 11, 34, 150, 152, 240
- local averaging, 181–183
- local true discovery rate, 84
- locfdr, 34, 75, 76, 82, 86, 101, 103, 104, 127, 153, 226, 250
- logistic regression estimate, 189
- Malaysian butterfly data, 233, 234, 236
- marginal density, 18, 19, 123, 187, 188, 199, 217, 218, 224, 233

- marginal distribution, 2, 4, 177
max- T , 40–43
maximum likelihood estimator (MLE), 3, 5, 102
maxmean statistic, 171–173, 184
mean, 2, 8, 9, 24, 28, 56, 68, 77, 87, 98, 104, 107, 115–117, 121, 125, 126, 132–135, 137, 141–144, 147, 149, 154, 156–159, 161, 165–168, 172, 173, 177, 178, 184, 211, 214, 250
and variance, 6, 23, 60, 82, 114, 144, 158, 159, 161, 180, 228, 230, 244
posterior, 13, 218, 223
Mehler's identity, 118, 121
Michigan lung cancer study, 222, 223, 241
microarray, ix, 1, 15, 27, 28, 43, 45, 61, 81, 91, 96, 106, 107, 109, 121, 141, 143, 145, 148–150, 152, 163, 164, 167, 181, 206–208, 212, 215
microarray experiment, 61, 107, 112, 133, 143, 152, 163, 223
microarray study, 27, 61, 64, 91, 109, 113, 148, 163
Miller, Rupert, xi, 30, 45, 186, 205
min- p , 40, 41
missing mass problem, 239
missing species problem, 233, 237, 241
mixture density, 18, 21, 29, 71, 74, 88, 96, 101, 114, 187, 250
MLE method, 1, 90, 99, 101, 102, 104, 105, 108, 112, 153, 183, 189
multi-class model, 117, 119, 187
multinomial, 118, 125, 161, 236, 245, 247
multivariate normal, 2, 6, 28, 36, 146, 211, 245
calculations, 153, 180
natural parameter, 137, 217, 243, 244, 247
natural spline basis, 76, 90
nearest shrunken centroid, 215, 240
negative binomial, 237, 238
new species, 234, 235, 237, 238
non-central χ^2 , 178, 180
non-central t , 132, 137, 221
distribution, 133
non-null counts, 84, 85
smoothed, 84, 85, 96, 124
normal distribution, 3, 16, 84, 139, 153, 213
matrix normal, 154
multivariate normal, 211
null and alternative densities, 29
null hypothesis, x, 16, 26, 31, 32, 48, 49, 51, 55, 58, 65, 74, 83, 89, 96, 97, 105, 111, 113, 132, 135–137, 141, 149–151, 164, 165, 170–172, 174, 194, 195, 201, 224, 227, 250
blurry, 97
complete, 40–42
enrichment, 194
null proportion, 20, 28, 49, 54, 60, 66, 90, 98
odds ratio, 2, 211, 227
optimal discovery procedure, 45

p -values, 31–41, 45, 48–53, 57, 66, 70, 136, 141, 167, 169, 173, 174, 200, 204
adjusted, 34–37, 39–42, 202
p53 data, 163, 166, 173, 184, 194, 250
P53_UP, 194, 195
pamr, 215–219, 240
PCER, 43
penalty, 5, 50, 76, 201, 204, 205
correlation, 114, 118, 119, 122, 123, 128
permutation algorithms, 39
permutation calculations, 40, 97, 109, 110, 177–180, 183
permutation distribution, 41, 62, 64, 89, 111, 169, 180
permutation methods, 41, 42, 110–112, 149, 171, 174
permutation null, 97, 109, 111, 141, 149
distribution, 64, 109, 110, 184
point estimate, 22
Poisson assumption, 25, 170
Poisson-independence, 24, 25, 58, 200
Poisson deviance, 76, 248
Poisson process, 234
Poisson regression, 71, 75, 218, 219, 245
estimate, 74, 114, 123, 127, 196, 233
Poisson selection model, 170, 174
police data, 93, 95, 112, 226, 227, 249
positive regression dependence, 50, 69
posterior distribution, 2, 3, 12, 229
posterior interval, 12, 13, 232
posterior probability, 18, 53, 57, 74, 95
power, 34, 48, 57, 74, 83, 85–89, 104, 153, 163, 170–173
diagnostics, 83, 87, 88

- prediction, xi, 1, 7–10, 28, 211–217, 223, 228, 229, 234, 235, 237–241
 computer-intensive, 212
 prediction error, 7, 214, 219–221
 prediction rule, 97, 211–213, 216, 217, 219–221, 223
 ideal, 213
 prior density, 2, 27, 95–97, 228
 prior knowledge, 80, 88
 prior probability, 17, 70, 83, 187, 211, 213
 prostate data, 15, 17, 21, 24, 27, 41, 42, 44, 47, 50, 57, 60–63, 68, 71–73, 78–80, 82, 84, 85, 88, 148, 151, 184, 185, 207, 208, 212, 214–216, 219, 221, 224, 225, 229, 249, 250
 randomization, 164–167, 169, 174, 176, 179
 row, 165, 167–171, 174, 177, 183, 184
 ras pathway, 164, 166, 171
 relevance, x, 81, 82, 94, 185, 186, 196–198, 205, 206, 209
 relevance function, 14, 197, 198, 209
 restandardization, 168, 169, 173, 174, 176, 178, 179, 183, 184
 restandardized, 168, 173, 184
 risk function, 3
 rms approximation, 121–123, 125, 161
 Robbins, Herbert, x, 1, 14, 88, 240, 241
 SAM, 61–64, 94, 110, 174, 209
 scaling properties, 80, 81
 score function, 171, 246
 scoring function, 165, 169, 170, 174
 secant, 19
 second derivative matrix, 246
 selection bias, 215, 230
 separate analyses, 185, 186, 199, 205
 separation, 184–186, 196, 199, 200, 205, 209
 Shakespeare data, 238, 239, 241, 250
 Shakespeare, Wm., 238–241
 canon, 238, 239
 Šidák procedure, 36, 37
 significance level, 34, 81, 178, 200
 Simes' inequality, 39
 simulation, 7, 8, 55, 56, 76, 87, 104, 106, 121, 146, 148, 149, 155, 173, 196, 217, 232
 experiment, 9, 56, 107, 125, 126
 simz, 148, 159, 162, 250
 single nucleotide polymorphism (snp), 81, 147, 162, 202–206, 210
 singular value decomposition (svd), 118, 142
 skewness, 132, 135
 slope, 19, 34, 110, 194, 225, 233
 statistic, 194
 test, 152, 205
 small subclasses, 192
 smoothed cdf estimate, 72
 soft thresholding rule, 215, 218, 219
 spherical cap, 65–67
 spike-in, 207
 squared error loss, 3, 8
 standard deviation, 24, 55, 56, 76, 77, 87, 104, 109, 113–117, 122–126, 128–130, 132–135, 137–139, 143–146, 160, 165–168, 184, 195, 196, 217, 236
 binomial, 114, 115, 122
 standard error, 16, 96, 130, 138, 156, 194, 202, 208, 219, 221
 standard normal, 16, 18, 32, 43, 53, 59, 60, 71, 72, 91, 116, 133, 136, 173, 244
 standardization, 91, 121, 144, 146, 148, 149, 151, 169
 column, 144, 145, 152, 250
 double, 142–147, 149, 151–153, 155, 162, 209, 250
 quantile, 153
 row, 144, 145, 169
 row and column, 144, 145
 standardized, 132, 135, 146, 169, 182, 183, 213, 221, 222, 249
 Stein, Charles, x, 1, 5, 7, 14, 28, 241
 step-down procedure, 37, 39
 step-up procedure, 39, 42, 49, 50, 81
 stepwise algorithms, 37
 strata, 200–202
 strong control, 35
 Student-*t*, 16, 33, 51, 65, 68, 105, 108, 112, 132
 effects, 219, 221
 sufficient statistic, 243–245, 247
t-statistic, 64, 98, 110, 111, 133, 157, 206–209, 219, 221
 two-sample, 16, 34, 41, 94, 105, 106, 108, 111, 125, 176, 214
t-test, x, 28, 32, 64–66, 69, 91, 96, 112, 149, 163, 168, 249

Cambridge University Press

978-1-107-61967-8 - Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction

Bradley Efron

Index

[More information](#)*Index*

263

- tail counts, 160, 161
- tangent, 19, 23, 226, 227
- testing, x, xi, 1, 14–17, 34–36, 46, 57, 68, 88–90, 126, 134, 150, 152, 170, 172, 173, 194, 198, 208, 209, 240, 243
- hypothesis, x, xi, 1, 2, 10, 14, 15, 27, 28, 30, 32, 58, 63, 70, 97, 110, 113, 141, 225
- large-scale, 1, 15, 18, 25, 26, 34, 37, 42, 46, 48, 51, 70, 89, 90, 105, 109, 198, 225, 229
- multiple, xi, 17, 30, 43, 45, 46, 48, 51, 57, 58, 70, 205
- significance, 30, 37, 46, 48, 51, 70, 89, 227, 229
- simultaneous, 15, 27, 28, 30, 64, 67, 164, 227
- single-case, 47, 51, 55, 57, 70
- testing algorithm, 30, 34, 37, 45, 48, 51
- theoretical null, 17, 20, 21, 51, 60, 61, 75, 84, 89–97, 105–108, 110–112, 130, 141, 149, 169, 194, 202, 223, 225–227, 250
- distribution, 16, 20, 32, 33, 81, 89, 90, 105, 106, 168, 185, 190, 223
- trace, 143
- training set, 211–213, 216
- transcription factor, 163
- trapping period, 234, 235, 237
- true and false discoveries, 46
- true discoveries, 47, 55, 59, 232
- Turing, Alan, 14, 241
- two-groups model, 17, 20, 21, 24, 27–29, 48, 53, 60, 70, 76, 78, 81–84, 98, 103, 187, 197, 211, 217, 227, 228
- Type I error rate, 30, 47
- voxels, 26, 34, 54, 57, 73, 181, 183, 185, 186, 192–194, 197
- weak control, 35
- Westfall and Young, 39, 40, 45
- Wilcoxon test, 28, 98
- wing-shaped function, 161
- Wishart distribution, 155, 156
- z-values, 16–18, 20–22, 26, 29, 31–34, 43, 44, 47, 51, 55, 56, 59, 60, 62–64, 69, 71, 88, 90, 91, 96, 99, 103, 104, 106, 110, 111, 114, 115, 125, 132, 134, 145, 149, 159, 160, 163–165, 167–169, 171, 173, 174, 177, 181, 184, 185, 190, 192–194, 196, 198, 200, 202, 205–207, 217–219, 245, 247, 249
- correlated, 82, 148, 174, 250
- histogram of, 34, 64, 71, 72, 78, 84, 90–94, 105, 106, 139, 164, 167, 182, 183, 193, 202, 204–206
- non-null distribution of, 115, 132, 137
- vector of, 61, 109, 113, 148, 159, 175, 219, 249, 250
- zero assumption, 60, 87, 98–100, 102, 227