

## Index

All references are to pages

- Abrahams, Phyllis; Baxter, Johnson and, refs., *Medieval Latin Word-List*, 250
- Accident distributions; analogy with word-distributions, 4–5, 41: examples, 42, 43: respects in which such distributions differ from word-distributions, 44: theory, elementary exposition, 44–48: rule for Poisson distribution  $\sigma^2 = M$ , 45: for actual distributions  $\sigma^2$  exceeds  $M$ , 45: derivation of the characteristic  $K$ , 45–47: more detailed discussion of theory, 48–54: refs., 56. *See also* Word-distributions
- Adjectives; distribution in *De Imitatione Christi*, 19
- à Kempis, Thomas; noun-distribution for samples from, 15, 236; characteristic  $K$ , 57, 120, 129, 237, 284: his transcription of Vulgate, 51: attempt to estimate his total vocabulary of nouns, 71–73; and thence  $v_\lambda$ , 76–77: effect of similes on vocabulary, 74–76: nouns not in Lewis and Short, 101–102, 241–242: authorship controversy in relation to *Imitatio* and Gerson, generally, 221–280: method of sampling, 227–228: list of samples, 229: subsequent procedure and some results, 230–231: characteristics  $K$  at successive stages of work, 231: comments on the methods of sampling used, 232–233: statistical comparison of the three noun-distributions, 235–238: concordance and dictionaries used for obtaining special classes of nouns, 238: non-Vulgate nouns, 239–241: nouns not in Lewis and Short, 241–242: nouns not in Vulgate nor in dictionaries used, 242–244: non-religious, non-Vulgate, non-classical nouns borrowed from the Greek, 244–245: summary of evidence to this point, 245–246: the *Scutum Kempense* of Amort, criticism, 247–249: correlation and contingency tables and coefficients, 251–255: linearity of regression, 255–258: ‘frequent’ nouns in, general discussion, 258–268: table, 262; summary, 267–268: association tables, 268–270: alphabetical distribution, 271–272: supplementary investigations, nouns not in Smith’s smaller dictionary (1933), 272–275: noun-distributions for four separate works of, 275–278; characteristics  $K$ , 277, 284; summary of evidence, 278: conclusions based on vocabulary consistent with conclusions based on sentence-length, 278: quotations from, excluding *De Imitatione Christi*, 75, 76, 227, 241, 245: other mention, 1, 2, 3, 7, 18, 21, 22, 23, 29–31, 35, 38, 40, 51, 58, 74, 83, 120, 121, 124, 129, 208, 221–280 *passim*: refs., Pohl’s edition of the Works, 56, 82, 250: Sommalius’s edition, 249: biographies, authorship controversy, etc., 249, 250, 280
- Alphabetical distributions; mention of the problem, 6: general discussion, 183–209: contrast of Bunyan with Macaulay, 183–186: consistence of samples from the same author, 186–193: distribution is a function of size  $o$ , sample, 193; but the effect is not important within the range of sizes used, 193–197:  $\chi^2$  method not applicable to testing goodness of fit, 197–198: the divergence of Macaulay from Bunyan possibly due to a larger proportion of Latin-Romance nouns, 198–199: basis of etymological analysis of vocabularies into I. Old English-Teutonic and II. Latin-Romance, 199–200: resulting component alphabetical distributions, 200–203: comparison of Macaulay distribution with expected distribution based on Bunyan, 203–205: the ten most frequent initial letters for nouns of Class I and Class II, 205–207: distribution for nouns in St John’s Gospel, A.V., 206–207: the Class II distribution for English nouns compared with a distribution for Mediaeval Latin, 207–208: summary comments, etymological distribution within a given initial, 208–209
- Amort, Eusebius; *Scutum Kempense*, 222; critical discussion of, 247–249: refs., 249
- Anagram; on Thomas Babington Macaulay, 31
- Association; the table, 164: notation and theory, 164–166: coefficient of association  $Q$ , 165: between presence of a word in two samples from the same author, 166–173; table of values of  $Q$ , 167: between presence of a noun in the *Imitatio*, the samples from Thomas à Kempis and the samples from Gerson, 268–271
- Attributes; notation for statistics of, 65–66, 148 *et seq.*: refs., 82, 182
- Babington Smith, B.; *see* Smith, B. Babington
- Bacon, Lord; essay on, *see* Macaulay, T. B.
- Badman, *The Life and Death of Mr*, by John Bunyan (q.v.); 124, 125, 128–131: list of nouns in sample from, 292–294: refs., 147
- Barlow, P.; *Tables of Squares etc.*, 14: refs., 33
- Basic English; noun-distribution for St John’s Gospel in, 25–26: refs., 34

- Baxter, J. H.; Baxter and Johnson, *Medieval Latin Word-List*, 238, 245: refs., 250
- Beta coefficients; ratios of powers of moments, 54
- Bible; Hebrew text of O.T., statistics, 7–8; refs., 8: *Dictionary of the Bible*, 7; refs., 8: the word ‘ensample’ in, 35: the words ‘ink’, ‘string’ and ‘variance’ in, 141
- Biblical nouns; in Bunyan and Macaulay, generally, 140–146: meaning and limitations of the term, 140–142: distribution in Bunyan, 142–143: sample investigation for Macaulay and comparison with Bunyan, 143–146
- Bunyan, John; nouns in four samples from his works, generally, 124–133: difficulties as to what is a noun, 125–126: distributions, 128: characteristics  $K$ , 129, 284: comment on the characteristics, 130–132: distribution of pool of samples, 132; and reduced value of characteristic, 133: certain classes of nouns in, 133–146: verbal nouns in *-ing*, 134–137: distribution, 134: proportions, 135; obsolete idioms, 135–137: monosyllabic nouns, 137–140; distribution, 138; proportions, 139; comparison of distribution with that for all nouns, 140: biblical nouns, 140–146; meaning and limitation of the term, 140–142; distribution, 143; comparison with sample from Macaulay, 144–146: Macaulay on Bunyan’s style, 146: actual nouns, classified as in the noun-distribution, in samples from *Pilgrim’s Progress*, Pt I, 286; *Pilgrim’s Progress*, Pt II, 289; *Mr Badman*, 292; *Holy War*, 295: quotations from, 126, 130, 131, 135–136; other mention, 6, 124–220 *passim*, 272, 282: refs., editions of the tales used, 147
- Busch, J.; 221 footnote: refs., 280
- Campbell, Lewis; work on Plato, 113
- Chambers, E. G.; 47: refs., *Theory and Observation in the Investigation of Accident Causation*, 56: *A Study of Accident Proneness among Motor Drivers*, 56
- Characteristic  $K$ ; the problem, 5, 41: deduction of, elementary, 44–48; more detailed, 48–54: equation defining, 47, 53, 57: limiting form for large samples, 54: values of  $K$  for tables in Chapter 2, 57: experiment on sampling to control constancy of  $K$  with increasing size of sample, 58–69; values of  $K$  for successive sizes of sample, 64: empirical relation between  $K$  and mean, 77; not a theoretical necessity, 78–82: the characteristic as implying concentration in a certain sense, 78, 82, 283: effect on of altering all frequencies in the same proportion, 82: variation for different but similar works of the same author, generally, 117–133; four of Macaulay’s essays, 117–124; values of  $K$ , 119–120; their standard deviation, 120; four of Bunyan’s tales, 124–133; values of  $K$  and their standard deviation, 129: reduction of  $K$  below the mean on pooling, suggested, 123, 231; tested by examples, Macaulay, 123–124, Bunyan, 132–133: values of  $K$  for *Imitatio*, and samples from Thomas à Kempis and Gerson, 237: for four separate works of Thomas à Kempis, 277; and their standard deviation, 278: theorem on the characteristic, the stable point, 283: summary table of all the characteristics for noun-distributions, 284; chart, 285: standard deviation of, in groups of works from the same author, 120–121, 129, 278
- Chaucer, Geoffrey; Mersand’s work on examined, 104–112: verbal nouns in *-ing* in, 137: quotations from, 137: other mention, 18, 32, 97, 104–112 *passim*: refs., Mersand, J., *Chaucer’s Romance Vocabulary*, 34: *The Complete Works*, ed. W. W. Skeat, 116; *The Student’s Chaucer*, ed. W. W. Skeat, 147
- Clark or Clarke, Samuel; quotation from, 131: refs., *A Mirror or Looking-Glasse both for Saints and Sinners*, 147
- Concise Oxford Dictionary of Current English*, *The*, estimate of nouns, adjectives and verbs in 1911 edition, 70–73: refs., 82
- Concordance; to the Bible, A.V., *see* Cruden: to the Vulgate, 22–23, 238; refs., 34, 250: to the *De Imitatione Christi*, 9, 10, 18–19, 20, 223, 224, 241; refs., 34, 249
- Concordance, coefficient of; refs., 220
- Contingency, coefficient of; mentioned, 163: for occurrences of nouns in the data for the *Imitatio*, and samples from à Kempis and Gerson, 254: refs., 280
- Correlation, coefficient of; used for measuring resemblance of alphabetical distributions, 192, 203, 204; and of two columns in a different form of table, 197: rank correlation, 191–193: correlations between numbers of occurrences of a noun in data for *Imitatio*, and samples from à Kempis and Gerson, 254; stability of these correlations, 255: refs., 220, 280
- Cruden, Alexander; *A Complete Concordance to the Old and New Testament*, comments on, 22 and footnote, 141–142: refs., 147
- Cruise, F. R.; refs., *Thomas à Kempis*, 249: *Who was the author of the Imitation of Christ?*, 249
- De Backer, A.; refs., *Essai bibliographique sur le livre De Imitatione Christi*, 249
- De Imitatione Christi*; noun-distribution, 9–14, 235–238: adjective-distribution, 18–20: verb-distribution, 20: characteristics  $K$  for the three preceding distributions, 57: attempt to estimate total vocabulary of nouns of author of, 71–73; and possible value of  $v_\lambda$ ,

## INDEX

301

*De Imitatione Christi* (cont.)

- 76–77: nouns not in Lewis and Short, 101–102, 241–242: authorship controversy in relation to, Thomas à Kempis and Gerson, 221–250, 251–280: *editio princeps*, 221: some of those to whom the work has been attributed, 221: outline of method of work on, 223–225: difficulties as to what is a noun, 224: brief comments on results, 225–227: statistical comparison of the three noun-distributions, 235–238: exceptional character of Lib. iv, 235: concordance and dictionaries used for obtaining special classes of nouns, 238: non-Vulgate nouns, 239–241: nouns not in Lewis and Short, 241–242: nouns not in Vulgate nor in dictionaries used, 242–244: non-religious, non-Vulgate, non-classical nouns borrowed from the Greek, 244–245: summary of evidence to this point, 245–247: the *Scutum Kempense* of Amort, critical discussion, 247–249: correlation and contingency tables, 251–253; and values of the coefficients, 254–255; stability of these correlations, 255; linearity of regression, 255–258: ‘frequent’ nouns in, general discussion, 258–268; table of, 260; summary of evidence, 267–268: associations, 268–270: alphabetical distribution of nouns, 271–272: supplementary investigations, nouns not in Smith’s *Smaller Latin Dictionary* (1933), 272–275; noun distributions for four works of à Kempis and comparison with the *Imitatio*, 275–278: evidence from vocabulary is consistent with that from sentence-length, 278: criticism of the argument that the *Imitatio* cannot be by Thomas à Kempis since it is so much better than any admitted work, 278–280: quotations from, 1, 226, 240, 249: other mention, 1, 2, 3, 4, 9, 15, 16, 18, 21, 24, 29–31, 32, 40, 46, 49–51, 58, 120–122, 124, 129, 208, 221–280 *passim*: refs., including authorship controversy, 249–250, 280
- De Montmorency, J. E. G.; refs., *Thomas à Kempis: his age and book*, 249
- De Morgan, Augustus; *A Budget of Paradoxes*, anagram on Thomas Babington Macaulay, 31
- Deviation, standard; defined, 12; expressed in terms of moments about zero as origin, 13: arithmetical calculation, 14
- Dickens, Charles; *David Copperfield*, 116
- Distribution of vocabulary over the alphabet; *see* Alphabetical distributions
- Distribution of vocabulary over several samples from the same author, generally, 148–173: notation, 148–150: the uniformities observed, 150–153: sampling scheme, random partitioning, 153–155; comparison with observation, 155–162: attempted measures of heterogeneity of the samples, 162–164: distribution over two samples, associations, 164–173; associations observed and expected on random partitioning, 167–168
- Distribution of words (nouns, etc.) occurring once, twice, thrice, and so on; *see* Word-distributions
- Ducange, C. D.; 243: refs., *Glossarium mediae et infimae Latinitatis*, 250
- Eldridge, R. C.; 33: refs., *Six Thousand Common English Words*, 34
- Ephesians, Epistle to the; 1
- Error, standard; of the mean, 37; ordinary formula not applicable to word-distributions, 55: of a percentage or proportion, used, 145, refs., 147, 220
- Etienne, le R. P.; refs., *Concordance to the Vulgate*, 34, 250
- Etymological analysis of English vocabularies; generally, 198–220: the two classes used, I. Old English-Teutonic, II. Latin-Romance, 199: words not classed, 199: *Shorter Oxford Dictionary* as authority, 199: analysis of alphabetical distributions for Bunyan and Macaulay, 200–203: alphabetical distribution for Macaulay and distribution expected on basis of Bunyan, 203–204: differentiating characters of Class I and Class II alphabetical distributions, 204–207: etymological analysis of word-distributions for Bunyan and Macaulay, 210–220; table for Bunyan, all nouns, 211; for Macaulay, all nouns, 212; table for Bunyan, monosyllabic nouns, 216; for Macaulay, monosyllabic nouns, 217; proportion of monosyllabic nouns in Class I and Class II, 218: Class II on the whole the older group, 218
- Experiment in sampling, to test the constancy of the characteristic *K* against increasing size of sample, 5–6, 58–69; details, 58–60: difficulties with certain classes of nouns, 59: tables of results, 61–64: discussion of results, 61, 65–69: values of *K* for increasing sizes of sample, 64
- Exterius; used in sense of ‘by heart’ in *Imitatio*, 2, 221 footnote
- Fallacies; generally, 83–116: vocabulary ratios, 83–97: proportions or percentages, 97–116. *See also* Proportions, Vocabulary ratios
- Farmer, E.; refs., *A Study of Accident Proneness among Motor Drivers*, 56
- Fit, goodness of;  $\chi^2$  method of testing, 61, 108: not applicable to alphabetical distributions of vocabulary, 197–198: refs., 82
- Florentius (Radewijn); 221 footnote, 228
- Fowler, H. W. & F. G.; refs., *The Concise Oxford Dictionary of Current English*, 82
- Frederic the Great, Macaulay’s essay on; 117–123. *See also* Macaulay, T. B.

- Frequency distribution; meaning of the term, 3, 9; mean, standard deviation, 12; coefficient of variation, 13; variance (square of standard deviation), 45. *See also* Accident distributions, Alphabetical distributions, Word distributions
- Frequent nouns; extent to which they colour the text, illustrations, 12, 16, 17, 18; summary table, 18; in *De Imitatione Christi* and samples from Thomas à Kempis and Gerson (nouns used 20 times or more), 258–268; tables, 260, 262, 265; summary of discussion, 267–268
- Gantois, le R. P.; refs., *Concordance to the Vulgate*, 34, 250
- Gersen, John; of Vercelli, *Imitatio* attributed to, 221, 247
- Gerson, Jean Charlier de; noun-distribution, 16; characteristic  $K$  of, 57, 120, 129, 237, 284; non-Vulgate nouns in, 19, 32, 239–241; characteristic  $K$ , 57; attempt to estimate total nouns in vocabulary, 70–73, thence possible value of  $v_\lambda$ , 76–77; nouns not in Lewis and Short, 72, 101–102, 241–242; authorship controversy in relation to *De Imitatione Christi* and Thomas à Kempis, generally, 221–250, 251–280; method of sampling the works, 231–232; list of samples, 232; some comments on the methods of sampling, 232–233; characteristics  $K$  at intervals during the work, 233–235; statistical comparison of the three noun-distributions, 235–238; concordance and dictionaries used for obtaining special classes of nouns, 238; non-Vulgate nouns, 239–241; nouns not in Lewis and Short, 241–242; nouns not in Vulgate nor in dictionaries used, 242–244; non-religious, non-Vulgate, non-classical nouns borrowed from the Greek, 244–245; summary of evidence to this point, 245–246; correlation and contingency tables, 251–253, and values of the coefficients, 254–255; linearity of regression, 255–258; ‘frequent’ nouns in, general discussion, 258–268; table, 265; overdriving metaphors and its effect, 264–266; ecclesiastical nouns, 266–267; summary of discussion, 267–268; associations, 268–270; alphabetical distribution of nouns, 271–272; supplementary investigations, nouns not in Smith’s *Smaller Latin Dictionary* (1933), 272–275; complete divergence of all the data from those for *De Imitatione Christi* and Thomas à Kempis, 275; other mention, 3, 4, 7, 17, 18, 21, 22, 23, 29–31, 35, 38, 40, 58, 74, 83, 120, 121, 124, 208, 221–280 *passim*; refs., *Opera... in partes quatuor distributa*, 250
- Gerund; risk of confusion with nouns, 59
- Goodness of fit;  $\chi^2$  method of testing, 61, 108; not applicable to alphabetical distributions of vocabulary, 197–198; refs., 82
- Greenwood, M.; 48, 55; refs., ‘Frequency distributions representative of multiple happenings’, 56; *The Incidence of Industrial Accidents*, 56
- Groote, Gerard; founder of the Brothers of the Common Life; *Imitatio* attributed to, 221; mention, 228; refs., 249
- Haemmerlein or Hemerken, surname of Thomas à Kempis, 221. *See also* à Kempis
- Hall, T. D.; William Smith and, refs., *Smaller Latin-English Dictionary*, 250
- Hampden, John; Macaulay’s essay on, 117–123. *See also* Macaulay, T. B.
- Hastings, J.; 7; refs., *Dictionary of the Bible*, 8
- Heterogeneity of samples; effect of, on distribution of vocabulary over several samples, 159–164; attempted measures of heterogeneity, 162–164; heterogeneity of samples from Bunyan as compared with those from Macaulay, 133, 163
- Hirsche, Carolus; 223, 226; refs., *Thomae Kempensis De Imitatione Christi Libri Quatuor*, 250
- Holy War*, by John Bunyan (q.v.): 124–127, 130–132; list of nouns in, 295–297; refs., 147
- Hyphened nouns; treatment of, in noun-distributions, 59, 125
- Imitatio*; *see De Imitatione Christi*
- Impressionist judgment of richness of vocabulary; obscure foundation of, 83, 90–97
- Indifferent words; *see* Neutral words
- ing; *see* Verbal nouns in -ing
- Innocent III, Pope; *De Imitatione Christi* attributed to, 221
- Irwin, J. O.; refs., Discussion on accidents, 56
- Johnson, C.; Baxter and, refs., *Medieval Latin Word-List*, 250
- $K$ ; symbol for the characteristic of the word-distribution; *see* Characteristic  $K$
- Kempis, Thomas à; *see* à Kempis, Thomas
- Kendall, M. G.; 36; refs., *An Introduction to the Theory of Statistics*, 33 and end of each chapter; *Tables of random sampling numbers*, 56; ‘The problem of  $m$  rankings’, 220
- Learning vocabulary of a new language; light thrown on the difficulty by the word-distribution, 29–30
- Lewis, C. T. and Short, C.; *Latin Dictionary*; estimate of numbers of nouns, adjectives and verbs in, 70–73; table, 72; nouns not in, proportion to vocabularies of *Imitatio* etc., 101–102; distributions of, 241–242; nouns not in this dictionary, nor the Vulgate, nor certain other dictionaries, 242–244; summary, not a very satisfactory dictionary for



## INDEX

303

- Lewis, C. T. and Short, C. (*cont.*)  
 classificatory purposes, 246: other mention, 97, 238, 246: refs., 82, 250
- Liabilities  $\lambda$ , of persons to accident or of words to use, 45, 47, 51, 52; thereafter *passim*
- Linearity of regression; *see* Regression
- Lockwood, J. F.; William Smith and; 272: refs., *Smaller Latin-English Dictionary* (1933), 280
- Lutoslawski, W.; statistics of peculiarities in Plato, recognition of influence of size of sample, 113–115: refs., *The Origin and Growth of Plato's Logic etc.*, 116
- Macaulay, T. B., Lord; anagram on, 31: *Times* newspaper on (1839), 31: noun distribution for essay on Bacon, 17, 63; its characteristic  $K$ , 57, 64, 284: comparison with Thomas à Kempis, 30–31: experiment in sampling with essay on Bacon to test constancy of  $K$ , 58–69; details, 58; difficulties as to what is a noun, 59; general procedure, 59–60; tables, 61–64; discussion of results, 61, 65–69; characteristics  $K$ , 64: attempt to estimate total nouns in vocabulary, 70–74: simile, 76: possible value of  $v_k$ , 76–77: once-nouns in the sampling experiment, 103–104: three additional essays, Milton, Hampden, Frederic the Great, generally, 117–124; noun-distributions, 118; characteristics  $K$ , 119–120, 129, 284, and standard deviation of  $K$ , 120–121; most frequent nouns, 121–122: distribution for the four essays pooled, 123–124: reduction of  $K$ , 124: certain classes of nouns in, 133–146; verbal nouns in *-ing*, 134–135; monosyllabic nouns, 137–140; distribution, 138; comparison with Bunyan, 139; comparison of distribution with that for all nouns, 139–141: biblical nouns, 140–146; meaning and limitation of term, 140–141; estimate of proportion by way of sample, and comparison with Bunyan, 143–144; the sample possibly giving  $p_w$  in excess, 145–146: his tribute to Bunyan's style, 146: alphabetical distributions (q.v.), 183–209; contrast with Bunyan, 183–186; but consistence of samples from different essays, 186–193; contrast may be due to larger proportion of Latin-Romance nouns, 198–199; basis of etymological analysis (q.v.) of vocabularies, 199–200; resulting alphabetical distributions, 200–203; and discussion, 203–207: etymological analysis of word-distributions, 210–220; table, all nouns, 212; monosyllabic nouns, 217: proportions of monosyllabic nouns in all nouns of Class I and Class II, 218: quotations from, 40, 76, 146: other mention, 6, 18, 29, 30, 33, 78, 117–220 *passim*, 282, 283: refs., *Critical and Historical Essays*, 82, 147
- Machinists; accidents to, 42
- Maigne d'Arnis; 221 footnote, 238, 243, 245: refs., *Lexicon Manuale ad Scriptores mediae et infimae Latinitatis*, 250
- Masora, Masorettes; numerical work on Hebrew text of the O.T., 7, 8: refs., 8
- Mersand, Joseph; discussion of his terminology for vocabulary and occurrences, 31–32: discussion of his data for Romance words in Chaucer,  $p_v$  and  $p_w$ , 104–109; and the cross-percentage  $p_c$ , 109–112: mention, 97, 100: refs., *Chaucer's Romance Vocabulary*, 34, 116
- Metaphors; *see* Similes and Metaphors
- Milton, John, Macaulay's essay on; 117–123. *See also* Macaulay, T. B.
- Moffatt, James; 137: refs., *A new Translation of the Bible*, 147
- Moments, of a frequency distribution; first and second, 13: higher, 54: refs., 56
- Monosyllabic nouns; in Bunyan and Macaulay, generally, 137–141; table of distributions, 138; comparison with those for all nouns, 139–141: etymological analysis (q.v.) of distributions, 215–220; tables, 216, 217: proportions of, in the two etymological classes, and reasons for difference, 218: comparison of the distributions with those for all nouns, 218–220
- Multiple happenings; 41: refs., 56. *See also* Accident distributions
- Names, in Bunyan; not taken as nouns save in two exceptional instances, 126
- Neutral or indifferent words, that may occur in connection with almost any subject, may nevertheless be characteristic of an author, 21: comparison of nouns, adjectives and verbs as regards neutrality, 21; in *Imitatio* or Thomas à Kempis and in Gerson, 261, 264
- Newbold, Ethel M.; refs., 'Applications of the statistics of repeated events, particularly to industrial accidents', 56
- Noun-distributions; *see* Word-distributions
- Nouns; reasons for choice as basis for work, 21: discussion as to correctness of choice, 21–22: difficulties in deciding what is a noun for purposes of classification, 59, 125–126, 223–224: the general problem what words to admit, 281: lists of nouns in the samples from Bunyan, 286–297
- Occurrences; the term introduced, 12, 14: discussion of terminology, 31–32
- Omnibus drivers; accidents to, 42–44
- Once-nouns, once-words; terms introduced for nouns (or words) occurring only once in the given sample, 26: slight variation in proportion of once-nouns in four tables otherwise showing considerable differences, 30: fall in the proportion with increasing size of sample, table, 64; discussion, fall relatively slow, 68

- Onions, C. T.; refs., *The Shorter Oxford English Dictionary*, 56, 82, 220  
*Oxford English Dictionary*, the *Concise* and the *Shorter*; estimates of numbers of nouns, adjectives and verbs in, by sampling, 37–38, 70–72: refs., 56, 82, 220
- Partitioning, random: 153–154, 156, 159, 161–164, 168–171  
 Pearson, Karl; binomial curve, 55: *Tables for Statisticians and Biometricians*, 61; refs., 82  
 Percentages; see Proportions  
 Percentiles; 278: refs. 280  
 Peultier, le R. P.; refs. *Concordance to the Vulgate*, 34, 250  
*Pilgrim's Progress*, by John Bunyan (q.v.); 124–126: noun-distributions, 128: the characteristics  $K$ , 130–132: diversity of the tales illustrated by vocabulary, 133, 162–163: relatively high association between the presence of a noun in the vocabulary of Pt I and Pt II, 167–168: style, Macaulay on, 146: lists of nouns in, 286, 289  
 Plato; 1, 113  
 Pohl, Michael Joseph; 38, 221 footnote, 227, 228, 229, 233, 241: refs., *Thomae Hemerken à Kempis Opera Omnia*, 56, 82, 250  
 Poisson distribution; mean equal to variance, 45: form of the series, 48, 51: mean and variance of the compound Poisson distribution in terms of mean and variance of  $\lambda$  distribution, 52: theory of bi-compound Poisson distribution, 78–80, and examples, 80–82, 89: example of a triply-compound distribution, 100  
 Pooling of several samples: effect on characteristic  $K$ , suggested, 123, 231; illustrations, 123–124, 132–133: effect on association, 170; refs., 182  
 Positive classes; definition and table of frequencies, 150: refs., 182  
 Proportions or percentages; generally, 97–116; the two principal ways of stating, (1) on occurrences,  $p_w$ , (2) on vocabulary,  $p_v$ , terminology, 32, 97;  $p_w$  independent of size of sample,  $p_v$  a function of size of sample, 97–101: examples, nouns not in Lewis and Short, 101–102; once-nouns in the sampling experiment, 103–104; Mersand's data for Romance words in Chaucer, 104–109: the cross-proportion, special vocabulary on occurrences,  $p_c$ , decreases with increasing size of sample, 109–112: application of conclusions to a wider field than statistics of vocabulary, 111–115
- $Q$ ; symbol for the coefficient of association, 165
- Random numbers, tables of, for sampling, 36–38: refs., 56
- Random partitioning; see Partitioning, random  
 Random sampling, simple random sampling, 36–39. See also Sampling  
 Rank, ranking, methods; applied to alphabetical distributions, generally, 184–209: rank correlation, coefficient of, 191; for all possible pairs of  $m$  rankings, 192: refs., 220  
 References; lists at ends of chapters; 8, 33–34, 56, 82, 116, 147, 182, 220, 249–250, 280  
 Regression, linearity of; in word-distribution tables, 255–258; diagrams, 256–257  
 Repeated events; statistics of, 41, refs., 56. See also Accident distributions  
 Repetitive style of St John's Gospel; illustrative quotations, 27  
 Richness of vocabulary; difficulty of the conception, 83–84: obscure foundation of the impressionist judgment, 83, 96–97. See also Vocabulary ratios  
 Robinson, H. Wheeler; 7: refs., *The Bible in its Ancient and English Versions*, 8  
 Rosweyd, H.; 278: refs., *Vindiciae Kempenses*, 280
- St Augustine; *Confessions*, 30: on memory, 69  
 St Bernard; *Imitatio* attributed to, 221  
 St Bonaventure; *Imitatio* attributed to, 221  
 St John's Gospel; noun-distribution, Basic version, 26; A.V., 28; characteristics  $K$ , 57, 120, 129, 284; repetitiveness of style, quotations from A.V., 27: high concentration, 122: alphabetical distribution of nouns, A.V., 207: other mention, 40: refs., Basic version, 34  
 St Paul; 1  
 Sample; meaning of term, 35: conditions necessary for a useful sample, 35: continuous passages give 'clumped' samples, 39–40  
 Sample, size of; effect on word distribution, 4, 41, 44: difference between meanings of the term as applied to (1) accident distributions and (2) word distributions, 5, 43–44, 115: influence on vocabulary ratios, 83–97; on proportion of vocabulary of a special class to total vocabulary, 97–109; on proportion of special vocabulary to total occurrences, 109–112; recognition of influence of size of sample by Lutoslawski, 113–115: influence on alphabetical distribution of vocabulary, 193–197  
 Sampling; generally, 35–41; simple random sampling, 36–39; sampling by random numbers, 36–38; necessity for prior enumeration of the population or universe, 38–39; standard error of the mean, 37: spread sampling, 39–41; warning that if continuous passages are taken samples are 'clumped', 39–40: process used in sampling experiment with essay on Bacon, 58–60; with other essays of Macaulay, 118–119; improved pro-

## INDEX

305

- Sampling (*cont.*)  
 cedure for Bunyan's tales, 124–125; early procedure for Thomas à Kempis, 228–229; and Gerson, 231–232; for biblical nouns in Macaulay, 142–146: standard error of a percentage, 145; refs., 147: standard error of a small frequency equal to the square root of the frequency, 180; refs., 182: refs., general, 56
- Sampling experiment, to test constancy of characteristic  $K$ ; *see* Experiment in sampling *Scutum Kempense*; *see* Amort, E.
- Sentence-length; mention, 228: evidence from vocabulary as to authorship of *Imitatio* consistent with that from sentence-length, 278: refs., 250, 280
- Shakespeare, W.; 1
- Short, C.; *Latin Dictionary*; *see* Lewis, C. T.
- Shorter Oxford English Dictionary; estimate of numbers of nouns, adjectives and verbs in, 37–38, 70–72: used as authority for etymological classifications, 199: other mention, 127: refs., 56, 82, 220
- Similes and metaphors; effect on vocabulary, 74–76: illustrations from Thomas à Kempis, 75–76; and example from Macaulay, 76: metaphors in Gerson, 264–266
- Singer, I.; refs., *Jewish Encyclopaedia*, 8
- Size of sample; *see* Sample, size of
- Skeat, W. W.; 110, 137: refs., *The Complete Works of Chaucer*, 116; *The Student's Chaucer*, 147
- Smith, B. Babington; 36: refs., *Random Sampling Numbers*, 56: 'The problem of  $m$  rankings', 220
- Smith, Edwin W.; refs., *The Basic St John*, 34
- Smith, Sir William; 245, 272–274: refs., *A Smaller Latin English Dictionary* (1881), 250; with Lockwood, J. F., a new edition (1933), 280
- Sommalius, Henricus; refs., *Thomae Malleoli à Kempis Opera Omnia*, 249
- Southey, Robert; edition of *Pilgrim's Progress*, 146
- Spread sampling; general notion, 39–41. *See also* Sampling
- Standard deviation; *see* Deviation, standard
- Standard, error; *see* Error, standard
- Storr, Rayner; mention, 9, 10, 18–19, 20, 223, 224, 241: refs., *Concordantia ad quatuor libros latine scriptos De Imitatione Christi*, 34, 249
- Swinburne, A. C.; quotation from, 1
- Tables; of squares, cubes, square roots etc., Barlow's, refs., 33: of random numbers, refs., 56: *Tables for Statisticians and Biometricians*, refs., 82
- Theophilus; Acts i. 1, 141
- Thompson, Francis; polysyllabic Latin words in the poems of, 222
- Thomson, Godfrey H. and Thompson, J. Ridley; 116: refs., 'Outlines of a method for the quantitative analysis of writing vocabularies', 116
- Times*; the newspaper, on Macaulay (1839), 31
- Tippett, L. H. C.; 36, 37: refs., *Random Sampling Numbers*, 56
- Titles; such as Lord, Lady, treatment of in compiling noun-distributions, 59, 126
- Tyre; fairs of, 141
- $v_{\lambda}$ ; coefficient of variation of the liabilities  $\lambda$  of persons to accident or of words to use, 45, 52: relation to the characteristic  $K$  and constants of the word-distribution, 52, 54, 77: high magnitude of for *Imitatio*, Thomas à Kempis, Gerson and Macaulay, 76–77
- Van Ginneken, Jac.; refs., *Trois textes pré-Kempistes du premier livre de l'Imitation*, 249
- Variance; term used for the square of the standard deviation, 45: the word occurs in the N.T. (Mat. x. 35, Gal. v. 20), 141
- Variation, coefficient of; ratio of standard deviation to mean, 13: calculation of, 14
- Verbal nouns in *-ing*; danger of confusion with the gerund, 59: in Bunyan and Macaulay, 134–137: idioms in Bunyan, quotations, 135–136; in Chaucer, 137
- Verbs; distribution for *Imitatio*, and summary comparison with nouns, 20: appear to be very indifferent or neutral words, 21: estimated numbers in Latin and English dictionaries, 72
- Vocabulary; in numerical sense of sum  $S_0$ , 13, 14, 32: in sense of the list of words used, 32: vocabulary per thousand occurrences a dangerous measure of 'richness', 14, as it is a function of size of sample, 68, 110–113: discussion of terminology, 31–32: estimates of nouns, adjectives and verbs in certain dictionaries, 72: attempted estimates of nouns at risk for Thomas à Kempis and Gerson, and Macaulay, 71–74: effect of similes on, 74–76: actual vocabularies of samples from Bunyan's tales, 286–298. *See also* Alphabetical distributions, Etymological analysis
- Vocabulary ratios; generally, 83–97: a function of size of sample, 84: form of the function and theoretical examples, 85–97; graphs, 87, 91, 92: of Bunyan to Macaulay, generally, 173–182; fallacious method, 173–178; valid method based on noun-distributions, 179–180; graph, 179; advantages of this method, 181–182
- Vulgate; concordance to, 22, 23, 238; difficulties for statistical work, 22–23; refs., 34, 250: noun distribution estimated by sample, 23–24: non-Vulgate nouns in *Imitatio*, Thomas à Kempis and Gerson, 19, 239–241,

Vulgate (*cont.*)

242–245: manuscript written by Thomas à Kempis, 51: citations and quotations, 248–249: mention, 18, 97, 101, 246

Wheatley, L. A.; refs., *The Story of the Imitatio Christi*, 249

Wheeler Robinson, H.; see Robinson, H. Wheeler

Woods, Hilda M.; refs., *The incidence of industrial accidents*, 56

Word-distributions; empirical form of, 3, 9–33: weighting of the more frequent words in occurrences, 11–12, 16, 17, 18; light thereby thrown on the difficulty of acquiring a new language, 29–30: analogy with accident distributions (q.v.) and differences, 4–5, 41, 44: theory and deduction of characteristic *K*, elementary exposition, 44–48; more detailed discussion, 48–54; equation defining *K*, 47, 53, 57: note on mathematical formulae suggested for distribution, 55. See also Characteristic *K*

Illustrations; *De Imitatione Christi*, nouns, table 2.1, 10–11: Thomas à Kempis, nouns, table 2.2, 15: Gerson, nouns, table 2.3, 16: Macaulay, essay on Bacon, nouns, table 2.4, 17: Gerson, non-Vulgate nouns, table 2.5, 19: *De Imitatione Christi*, adjectives, table 2.6, 19: *De Imitatione Christi*, verbs, table 2.7, 20: sample of nouns from the Vulgate, table 2.8, 24: St John's Gospel, Basic nouns, table 2.9, 26: St John's Gospel, A.V., nouns, table 2.10, 28: Macaulay's essay on Bacon, nouns, samples of increasing size, tables 4.1–4.4, 61–63: Macaulay's essays on Milton, Hampden and Frederic the Great, nouns, table 6.1, 118: Macaulay, pool of the four essays, nouns, table 6.4, 123: Bunyan, four tales, nouns, table 6.5, 128: Bunyan, pool of the four tales, nouns, table 6.7, 132: Bunyan and Macaulay, nouns in *-ing*, table 6.8, 134: Bunyan and Macaulay, monosyllabic nouns, table 6.9, 138: Bunyan, biblical and non-biblical nouns, table 6.12, 143: Bunyan, nouns with certain initial letters, table 8.8,

195: Bunyan, all nouns, etymological analysis, table 8.17, 211; Macaulay, the same, table 8.18, 212: Bunyan, monosyllabic nouns, etymological analysis, table 8.21, 216; Macaulay, the same, table 8.22, 217: *Imitatio*, Thomas à Kempis and Gerson, nouns, table 9.1, 236; the same, non-Vulgate nouns, table 9.3, 239: the same, nouns not in Lewis and Short, table 9.4, 242: the same, nouns not in Vulgate nor in certain dictionaries, table 9.5, 243: the same, non-religious, non-Vulgate, non-classical nouns from the Greek, table 9.6, 244: the same, nouns not in Smith's *Smaller Latin-English Dictionary* (1933), table 10.13, 274: Thomas à Kempis, nouns, four of his works, table 10.14, 277

Yule, G. Udney; 48, 55: refs., *An Introduction to the Theory of Statistics*, 33 and end of each chapter: 'Frequency distributions representative of multiple happenings', 56: 'Note on the theory of accidents with special reference to the time factor', 56: 'On sentence-length as a statistical characteristic of style in prose', 250, 280

Zainer, Günther (Gintheus); printer of the *editio princeps* of the *De Imitatione Christi*, 221

Zipf, G. K.; word-distributions for Latin, English and Chinese, 32, 33: different conception of 'word' for the purpose of these tables, 33: formula suggested for word-distribution unacceptable, 55: refs., *Studies of the principle of relative frequency in language*, 34

$\lambda$ ; symbol for the 'liability' of a person to accident or of a word to use, 45, 46: conception of the  $\lambda$ -distribution, 45: rough picture of its form, 77

$\rho$ ; symbol used for a vocabulary ratio, 84–97

$\sigma$ ; symbol used for the standard deviation, 12

$\chi^2$  method, for testing goodness of fit, 61, 108: not applicable to alphabetical distributions of vocabulary, 197–198: refs., 82