**SUPPLEMENTARY MATERIAL ACCOMPANYING:**

# Quantification of Short-Range and Long-Range Correlations in Symbolic Sequence Data Generated by Living Systems

*Marc L. Mansfield and Jack F. Douglas*

## TABLE OF CONTENTS

List of texts studied
- English 1
- French 9
- German 13
- Latin 16
- Other 17
- Mixed 17
- Random Digits 17
- Correlated Digits 18
- Computer Lists 18
- Genetic Codes 19
- Letter-sorted texts 20

Sources of the texts 21
Derivations 22
Fitting Parameters 29
Plots of $\langle r_n^2 \rangle$
- Introduction 37
- English 38
- French 72
- German 85
- Latin 94
- Other 97
- Mixed 98
- Random Digits 99
- Correlated Digits 101
- Computer Lists 103
- Genetic Codes 104
- Letter-sorted texts 110

Plots of $g_n$
- Introduction 111
- English 112
- French 141
- German 152
- Latin 157
- Other 159
- Mixed 159
- Genetic Codes 160

**LIST OF TEXTS STUDIED**

**English texts, *N* < 15,000.** (25)

| Text | N | Short-Name | Source |
|---|---|---|---|
| *To Ailsa Rock,* John Keats, 1816 (poetry). | 512 | ailsa_rock | T |
| *For There's Bishop's Teign,* John Keats, (poetry) 1816. | 872 | teign | T |
| *Gettysburg Address*, Abraham Lincoln, 1863. | 1149 | gettysburg | A |
| *Annabel Lee,* Edgar Allan Poe, 1849 (poetry). | 1159 | annabel_lee | T |
| *Charge of the Light Brigade,* Alfred, Lord Tennyson, 1854 (poetry). | 1174 | light_brigade | T |
| *An Epitaph on the Marchioness of Winchester,* John Milton, *1645.* (poetry) | 1935 | winchester | T |
| *Second Inaugural Address*, Abraham Lincoln, 1865. | 3157 | lincoln | A |
| *The Tale of Peter Rabbit,* Beatrix Potter, 1916. | 3989 | peter | G |
| *Paul Revere's Ride,* Henry Wadsworth Longfellow, 1860 (poetry). | 4098 | revere | T |
| *Slavery's Passed Away and Other Songs,* Various authors, ca. 1890. | 4717 | slavery | G |
| *Il Penseroso,* John Milton, 1632 (poetry). | 4737 | penseroso | T |
| *The Monroe Doctrine, Message to Congress, 1823,* James Monroe. | 4810 | monroe | T |
| *The Raven*, Edgar Allan Poe, 1845 (poetry). | 4892 | raven | G |
| *Inaugural Address*, John F. Kennedy, 1961. | 6040 | jfk_inaug | T |
| United Nations Article 102: *Agreement on the Rescue of Astronauts*, 1968. | 6687 | un102 | A |
| *"I Have a Dream,"* Martin Luther King, Jr., 1963. | 6949 | dream_king | T |
| *The Declaration of Independence of the United States of America,* 1776. | 7388 | declaration | A |
| *First Inaugural Address,* Thomas Jefferson, 1801. | 8227 | tj_inaug | T |
| *Life and Adventures of Calamity Jane, by Herself*, 1896. | 8629 | calamity | G |
| *The Gift of the Magi,* O. Henry, 1905 (short story). | 8680 | gift_magi | T |
| *A Ghost Story,* Mark Twain, 1903 (short story). | 10.3 k | ghost_story | T |
| *The Masque of the Red Death,* Edgar Allan Poe, 1842 (short story). | 11.0 k | red_death | T |
| *The Antarctic Treaty,* 1959. (An international agreement that Antarctica "shall be used exclusively for peaceful purposes.") | 12.3 k | antarctica | A |
| *Constitution of New Jersey*, 1776. | 12.8 k | jersey | A |
| *A Modest Proposal for Preventing the Children of Poor People From Being a Burden to Their Parents or Country, and for Making Them Beneficial to the Public,* Jonathan Swift, 1729 (political and social satire). | 13.2 k | poor_children | T |

**English texts, 15,000 < *N* < 40,000.** (24)

| Text | N | Short-Name | Source |
|---|---|---|---|
| *Rime of the Ancient Mariner,* Samuel Taylor Coleridge, 1798 (poetry). | 15.4 k | rime | G |
| *The Life and Public Service of General Zachary Taylor: An Address*, Eulogy by Abraham Lincoln, 1850. | 15.5 k | taylor | G |
| *Isabella,* John Keats, 1816 (poetry). | 16.8 k | isabella | T |
| *On Dreams,* Aristotle, 350 BCE, translated by J.I. Beare. | 18.9 k | dreams | T |
| *Fishing with a Worm,* Bliss Perry, 1916. | 19.8 k | fishing | G |
| *Ion,* Plato, 380 BCE, translated by Benjamin Jowett. | 20.9 k | ion | T |
| *The Hunting of the Snark*, Lewis Carroll, 1876 (poetry). | 21.4 k | snark | G |
| *The U.S. Constitution,* seven original articles, 1787. | 21.5 k | constitution | G |
| *The Adventure of the Sussex Vampire,* Arthur Conan Doyle, 1924. A Sherlock Holmes mystery. | 24.8 k | sussexvamp | T |
| *Treaty of Guadalupe Hidalgo*, 1848. (Peace treaty ending the Mexican-American War.) | 29.1 k | hidalgo | A |
| *Hyperion,* John Keats, 1816 (poetry). | 29.5 k | hyperion | T |
| *The Transcendentalist,* Ralph Waldo Emerson, 1842. | 30.0 k | transcend | T |
| *A Descent into the Maelstrom,* Edgar Allan Poe, 1841. | 30.1 k | maelstrom | T |
| *Piper in the Woods,* Philip K. Dick, 1953. | 30.4 k | piper | G |
| *Literary Ethics,* Ralph Waldo Emerson, 1838. | 31.7 k | literary_ethics | T |
| *The Ballad of Reading Gaol,* Oscar Wilde, 1897 (poetry), two versions appended together. | 32.6 k | gaol | G |
| *The Disappearance of Lady Frances Carfax,* Arthur Conan Doyle, 1917. (A Sherlock Holmes mystery.) | 32.6 k | carfax | T |
| *Comus,* John Milton, 1637 (poetry). | 34.2 k | comus | T |
| *Traditional Nursery Songs of England,* Felix Summerly, ed., 1843. | 34.6 k | nursery | G |
| *On the Gait of Animals,* Aristotle, 350 BCE, translated by A.S.L. Farquharson. | 34.8 k | animal_gait | T |
| *My Escape from Slavery,* Frederick Douglass, 1881. | 36.1 k | escape | T |
| *The Vortex Blaster,* E.E. Smith, 1960. | 36.9 k | vortex | G |
| *The Adductor Muscles of the Jaw in Some Primitive Reptiles*, Richard C. Fox, 1964. | 38.7 k | adductor | G |
| *Civil Disobedience,* Henry David Thoreau, 1849. | 40.0 k | civildis | T |

**English texts, 40,000 < *N* < 100,000.** (26)

| Text | N | Short-Name | Source |
|---|---|---|---|
| *The Boll Weevil: How to Prosper in Boll Weevil Territory*, G. H. Alford, Agricultural Extension Department, International Harvester Company, *ca*. 1915. | 40.6 k | weevil | G |
| *About Sugar Buying: How You Can Lessen Business Risks by Trading in Refined Sugar Futures*, B.W. Dyer, Lamborn & Co., 1921. | 41.9 k | sugar | G |
| *State of the Union Address*, Martin Van Buren, 1840. | 45.1 k | su1840 | G |
| *A Learned Dissertation on Dumpling*, Anonymous, Samuel L. Macey, ed., 1726 (political satire). | 47.5 k | dumpling | G |
| *The Adventure of Wisteria Lodge*, Arthur Conan Doyle, 1908. (A Sherlock Holmes mystery.) | 48.2 k | wisteria | G |
| *Songs of a Sourdough,* Robert W. Service, 1907 (poetry). | 52.5 k | sourdough | G |
| *The Eye of Argon,* Jim Theis, 1970. | 52.8 k | argon | T |
| *The Legend of Sleepy Hollow,* Washington Irving, 1820. | 53.6 k | sleepy_hollow | T |
| *The Thing in the Attic*, James Benjamin Blish, 1954. | 55.7 k | attic | G |
| *State of the Union Address*, Martin Van Buren, 1837. | 56.1 k | su1837 | G |
| *State of the Union Address,* Martin Van Buren, 1838. | 57.0 k | su1838 | G |
| *Catalogue of British Columbia Birds*, Francis Kermode, 1904. | 59.9 k | bc.birds | G |
| *The Stamp Act,* enacted by British Parliament, 1765. | 61.1 k | stamp | A |
| *Ecological Observations on the Woodrat, Neotoma floridana*, Henry S. Fitch and Dennis G. Rainey, University of Kansas Publications, 1956. | 62.8 k | woodrat | G |
| *State of the Union Address*, Martin Van Buren, 1839. | 65.2 k | su1839 | G |
| *Paradise Regained,* John Milton, 1671 (poetry). | 69.6 k | regained | T |
| *Lectures on Stellar Statistics,* C.V.L. Charlier, 1921. | 74.6 k | stellar | G |
| *Macbeth,* William Shakespeare, ca. 1605 (drama). | 75.3 k | macbeth | G |
| *The Creature from Cleveland Depths,* Fritz Leiber, 1962. | 76.2 k | cleveland | G |
| *Parmenides,* Plato, 370 BCE, translated by Benjamin Jowett. | 76.6 k | parmenides | T |
| *The Basis of Early Christian Theism,* Lawrence Thomas Cole, Doctoral Thesis, Columbia University, 1898. | 77.5 k | theism | G |
| *Areopagitica,* John Milton, 1644. | 82.4 k | areopagitica | T |
| *The Scarlet Plague,* Jack London, 1915. | 86.9 k | plague | G |
| *Sketch of the Mythology of the North American Indians,* John Wesley Powell, 1881. | 88.1 k | mythology | G |
| *Rambles in the Mammoth Cave,* Alexander Clark Bullitt, 1845. | 91.1 k | mammoth | G |
| *Comets and Meteors,* Daniel Kirkwood, 1873. | 92.9 k | comet | G |

**English texts, 100,000 < *N* < 175,000.** (24)

| Text | N | Short-Name | Source |
|---|---|---|---|
| *Cratylus,* Plato, 360 BCE, translated by Benjamin Jowett. | 100.0 k | cratylus | T |
| *Beowulf*, Francis Gummere's 1910 translation into Modern English. | 101.2 k | beowulf | G |
| *Alice's Adventures in Wonderland,* Lewis Carroll, 1865. | 107.5 k | alice | T |
| *On Snake-Poison: Its Action and Its Antidote,* A. Mueller, 1893. | 114.5 k | snake | G |
| *Othello,* William Shakespeare, 1603 (drama). | 116.0 k | othello | T |
| *Through the Looking Glass*, Lewis Carroll, 1871. | 119.6 k | glass | G |
| *A Christmas Carol,* Charles Dickens, 1843. | 120.1 k | christmas | T |
| *Oroonoko,* Aphra Behn, 1668. | 122.9 k | oroonoko | T |
| *Marine Protozoa from Woods Hole*, Bulletin of the United States Fish Commission 21: 415-468, 1901, by Gary N. Grealkins. | 127.2 k | woods.hole | G |
| *The Chimes,* Charles Dickens, 1844. | 128.0 k | chimes | T |
| *Hamlet*, William Shakespeare, ca. 1600 (drama). | 129.8 k | hamlet | G |
| *Endymion,* John Keats, 1816 (poetry). | 134.6 k | endymion | T |
| *The Call of the Wild,* Jack London, 1903. | 138.0 k | wild | G |
| *Chats to 'Cello Students,* Arthur Broadley, 1899. | 140.7 k | cello | G |
| *The Time Machine,* H.G. Wells, 1895. | 140.8 k | time_machine | T |
| *Billy Budd,* Herman Melville, 1889. | 143.0 k | billy_budd | T |
| *The Dawn of Amateur Radio in the U.K. and Greece,* Norman F. Joly, 1990. | 143.1 k | radioUKG | T |
| *The Song of Hiawatha,* Henry W. Longfellow, 1855 (poetry). | 144.9 k | hiawatha | G |
| *The Genus Pinus,* George Russell Shaw, 1914. | 147.6 k | pinus | G |
| *Flatland,* 5th ed., Edwin A. Abbott, 1884. | 152.6 k | flatland | T |
| *The Wonderful Wizard of Oz,* L. Frank Baum, 1900. | 158.6 k | oz | G |
| *The Telegraph Boy,* Horatio Alger, Jr., 1879. | 168.3 k | telegraph | G |
| *The Utah Batteries: A History,* Charles R. Mabey, 1900. (Military history, Spanish-American War.) | 171.4 k | utah | G |
| *The Chemical History of a Candle*, (Royal Institution, Christmas Lectures for Young People) Michael Faraday, 1861. | 172.6 k | candle | G |

**English texts, 175,000 < *N* < 300,000.** (30)

| Text | N | Short-Name | Source |
|---|---|---|---|
| *Narrative of the Life of Frederick Douglass, an American Slave,* autobiography, 1845. | 175.3 k | fd_auto | T |
| *The Avalanche,* Gertrude Atherton, 1919. | 176.9 k | avalanche | G |
| *Home Taxidermy for Pleasure and Profit,* Albert B. Farnham, 1944. | 178.5 k | taxidermy | G |
| *With a Highland Regiment in Mesopotamia, 1917-1918,* anonymous. | 185.0 k | highland | G |
| *Autobiography,* Thomas Jefferson, 1821. | 186.8 k | tj_auto | T |
| *Dr. Moreau,* H.G. Wells, 1896. | 187.5 k | DrMoreau | T |
| *Peter Pan,* James M. Barrie, 1904. | 194.1 k | peter_pan | T |
| *The Problems of Philosophy,* Bertrand Russell, 1912. | 195.7 k | problems | G |
| *The Red Badge of Courage,* Stephen Crane, 1895. | 202.0 k | red_badge | T |
| *Thuvia, Maid of Mars,* Edgar Rice Burroughs, 1920. | 209.6 k | maid_mars | T |
| *Shock and Awe: Achieving Rapid Dominance*, Harlan K. Ullman and James P. Wade, 1996. | 213.8 k | shock | D |
| *The Girl at Central*, Geraldine Bonner, 1915. | 216.6 k | ice | G |
| *The Prisoner of Zenda,* Anthony Hope Hawkins, 1894. | 218.3 k | zenda | T |
| *Pudd'nhead Wilson,* Mark Twain, 1894. | 218.4 k | puddnhead | T |
| *Cast Upon the Breakers,* Horatio Alger, Jr., 1893. | 225.9 k | breakers | T |
| *Pilgrim's Progress,* John Bunyan, 1678. | 227.6 k | pilgrim | T |
| *Armour in England,* J. Starkie Gardner and V.A. Farquharson, 1897. | 230.3 k | armour | G |
| *Reflections on the Decline of Science in England,* Charles Babbage, 1830. | 234.0 k | decline | G |
| *Pollyanna,* Eleanor H. Porter, 1913. | 236.3 k | pollyanna | G |
| *The Science of Right*, Immanual Kant, 1790, translated by W. Hastie. | 239.8 k | kantright | T |
| *Other Worlds: Their Nature and Possibilities in Light of the Latest Discoveries,* G. P. Serviss, 1901. | 240.7 k | other | G |
| *The Wind in the Willows,* Kenneth Grahame, 1908. | 250.5 k | willows | G |
| *Outlines of Dairy Bacteriology, 8th ed., H.L. Russell, 1907.* | 258.3 k | dairy | G |
| *Famous Violinists of Today and Yesterday,* Henry C. Lahee, 1899. | 264.6 k | violinists | G |
| *The War of the Worlds,* H. G. Wells, 1898. | 266.0 k | war | G |
| *Treasure Island,* Robert Lewis Stevenson, 1883. | 275.7 k | treasure | T |
| *The Autobiography of Benjamin Franklin,* 1771. | 293.1 k | bf_auto | T |
| *Notes on the State of Virginia,* Thomas Jefferson, 1785. | 293.8 k | virginia | T |
| *The Conduct of Life,* Ralph Waldo Emerson, 1860. | 296.7 k | conduct_life | T |
| *The Adventures of Tom Sawyer,* Mark Twain, 1876. | 296.8 k | tom_sawyer | T |

**English texts, 300,000 < *N* < 500,000.** (29)

| Text | N | Short-Name | Source |
|---|---|---|---|
| *The Prince and the Pauper,* Mark Twain, 1881. | 301.7 k | prince_pauper | T |
| *The Insidious Dr. Fu Manchu,* Sax Rohmer, 1913. | 319.6 k | fu_manchu | T |
| *Kidnapped*, Robert Lewis Stevenson, 1886. | 320.2 k | kidnapped | G |
| *The Lost World,* Arthur Conan Doyle, 1912. | 328.2 k | lost_world | T |
| *Frankenstein,* Mary Wollstonecraft Shelley, 1818. | 332.2 k | frankenstein | T |
| *Northanger Abbey,* Jane Austen, 1803. | 339.8 k | northanger | T |
| Catalog of zip codes (U.S. postal codes)*. | 341.8 k | zc_lett | T |
| *Paradise Lost*, John Milton, 1667 (poetry). | 360.1 k | paradise | G |
| *Canyons of the Colorado,* John Wesley Powell, 1895. | 365.0 k | colorado | G |
| *Physics,* Aristotle, 350 BCE, translated by R.P. Hardie and R.K. Gaye. | 365.8 k | ari_physics | T |
| *Nichomachean Ethics,* Aristotle, 350 BCE, translated by W.D. Ross. | 368.3 k | nicomachean | T |
| *The Scarlet Letter,* Nathaniel Hawthorne, 1850. | 368.6 k | scarlet_letter | T |
| *Tarzan of the Apes,* Edgar Rice Burroughs, 1912. | 376.6 k | tarzan | T |
| *Steep Trails,* John Muir, 1918. | 378.0 k | steep | G |
| *London to Ladysmtih via Pretoria,* Winston S. Churchill, 1900. | 393.3 k | pretoria | G |
| *The Sceptical Chymist*, Robert Boyle, 1661. | 397.1 k | chymist | G |
| *Huckleberry Finn*, Mark Twain, 1884. | 422.3 k | huckleberry | G |
| *Anne of Green Gables,* Lucy Maud Montgomery, 1908. | 432.8 k | green_gables | T |
| *Belgium From the Roman Invasion to the Present Day*, Emile Cammaerts, 1921. | 440.1 k | belgium | G |
| *Riders of the Purple Sage,* Zane Grey, 1912. | 446.0 k | purple | G |
| *Gulliver's Travels*, Johnathan Swift, 1726. | 454.9 k | gulliver | G |
| *Odyssey of Homer*, William Cowper's 1791 English translation. | 458.6 k | odyssey | G |
| *Virgil's Aeneid,* translated by John Dryden, 1909. | 475.7 k | aeneid_english | T |
| *Robinson Crusoe,* Daniel Defoe, 1719. | 480.4 k | crusoe | G |
| *Metaphysics,* Aristotle, 350 BCE, translated by W.D. Ross. | 483.7 k | ari_meta | T |
| *The House of the Seven Gables,* Nathaniel Hawthorne, 1815. | 484.1 k | seven_gables | T |
| *A Connecticut Yankee in King Arthur's Court,* Mark Twain, 1889. | 493.7 k | connyank | T |
| *Walden,* Henry David Thoreau , 1854. | 496.6 k | walden | G |
| *The Book of the Bush, Containing Many Truthful Sketches Of The Early Colonial Life of Squatters, Whalers, Convicts, Diggers, And Others Who Left Their Native Land And Never Returned*, George Dunderdale, 1893. (Anecdotal history of Australia.) | 498.6 k | bush | G |

**English texts, 500,000 < *N* < 1,000,000.** (27)

| Text | N | Short-Name | Source |
|---|---|---|---|
| *The Republic,* Plato, 360 BCE, translated by Benjamin Jowett. | 529.4 k | republic | T |
| *Pride and Prejudice,* Jane Austen, 1813. | 537.1 k | pnp | T |
| *Leaves of Grass,* Walt Whitman, 1855. | 542.3 k | leaves | G |
| *The Alhambra,* Washington Irving, 1832. | 547.9 k | alhambra | T |
| *The American,* Henry James, 1877. | 568.6 k | american | T |
| *Knickerbocker's History of New York*, Washington Irving, 1809. (Social satire.) | 569.7 k | knickerbocker | G |
| *A Tale of Two Cities*, Charles Dickens, 1859. | 584.6 k | 2cities | G |
| *Far From the Madding Crowd,* Thomas Hardy, 1874. | 590.2 k | madcrowd | T |
| *Thrilling Narratives of Mutiny, Murder and Piracy,* Anon. *ca.* 1882. | 620.7 k | narratives | G |
| *The Song of the Lark,* Willa Cather, 1915. | 644.4 k | song_lark | T |
| *Tess of the d'Urbervilles*, Thomas Hardy, 1891. | 648.9 k | tess | G |
| *The Last of the Mohicans*, James Fenimore Cooper, 1826. | 668.2 k | mohicans | G |
| *The Last Days of Pompeii,* Edward George Bulwer-Lytton, 1834. | 669.6 k | pompeii | T |
| *Mansfield Park,* Jane Austen, 1814. | 691.5 k | mansfield | G |
| *Two Years Before the Mast,* Richard Henry Dana, 1840. | 692.2 k | 2years | T |
| *The Fourteen Gilbert and Sullivan Plays*, W.S. Gilbert and A. Sullivan, 1871-1896. | 694.8 k | GandS | G |
| *A Child's History of England,* Charles Dickens, 1853. | 714.5 k | childs_history | T |
| *Uncle Tom's Cabin,* Harriett Beecher Stowe, 1852. | 763.2 k | cabin | G |
| *Applied Eugenics,* P. Popenoe and R. H. Johnson, 1918. | 766.2 k | eugenics | G |
| *Little Women,* Louisa May Alcott, 1869. | 771.7 k | little_women | T |
| *The Expedition to Borneo of H.M.S. Dido For the Suppression of Piracy*, H. Keppel, J. Brooke, 1846. | 787.0 k | dido | G |
| *Jane Eyre,* Charlotte Brontë, 1847. | 788.4 k | eyre | G |
| *The Remarkable History of the Hudson's Bay Company,* George Bryce, 3rd ed., 1910. | 839.0 k | hudson | G |
| *Ivanhoe,* Sir Walter Scott, 1820. | 865.0 k | ivanhoe | T |
| *Argentine Ornithology*, P. L. Sclater and W. H. Hudson, 1888. Two volumes, concatenated. | 916.5 k | argentine | G |
| *Leviathan*, Thomas Hobbes, 1651. | 949.3 k | leviathan | G |
| *Moby Dick*, Herman Melville, 1851. | 952.0 k | mobydick | G |

**English texts, *N* > 1,000,000.** (16)

| Text | N | Short-Name | Source |
|---|---|---|---|
| *The Origin of Species,* 6th ed., Charles Darwin, 1872 | 1.01 M | origin | G |
| *A Treatise of Human Nature*, David Hume, 1739. | 1.06 M | human.nature | G |
| *The History of the Conquest of Mexico,* William Hickling Prescott, 1843. | 1.07 M | mexcon | T |
| *Ulysses,* James Joyce, 1922 | 1.18 M | ulysses | G |
| *Special Report on Diseases of Cattle*, U.S. Department of Agriculture, Bureau of Animal Industry, 1923. | 1.20 M | cattle | G |
| *Memoirs of Extraordinary Popular Delusions and the Madness of Crowds,* Charles Mackay, 1852. | 1.28 M | delusions | G |
| *The Pickwick Papers,* Charles Dickens, 1837. | 1.34 M | pickwick | T |
| *David Copperfield,* Charles Dickens, 1850. | 1.48 M | copperfield | G |
| *The Rámáyan of Válmíki* (Sanskrit poet of 4th century BCE), English translation of 1870-74 by Ralph T. H. Griffith. | 1.73 M | ramayan | G |
| *Wealth of Nations,* Adam Smith, 1776. | 1.77 M | wealth | G |
| *Introduction to the Science of Sociology,* Robert E. Park and Ernest W. Burgess, 1921. | 1.97 M | sociology | G |
| *Les Misérables,* Victor Hugo, translated by Isabel F. Hapgood. | 2.51 M | lesmis_english | T |
| *The Bible,* King James Version, Old and New Testaments. | 3.22 M | bible | G |
| *History of Friedrich II of Prussia*, Thomas Carlyle, 1858. 21 volumes concatenated into a single text. | 5.86 M | fritz | G |
| *The History of the Decline and Fall of the Roman Empire,* Edward Gibbon (1845 revised edition). Six volumes concatenated into a single text. | 7.43 M | empire | G |
| *CIA World Factbook 2010** | 7.44 M | CIA2010let | G |

* Streaming list of letters

**Texts in French, *N* < 100,000. (21)**

| Text | N | Short-Name | Source |
|---|---|---|---|
| *Déclaration des droits de l'Homme et du citoyen de 1789* (*Declaration of Human and Civil Rights of 1789*), French National Assembly. | 3927 | droits | F |
| *J'accuse (I accuse),* Émile Zola, 1898. | 22.7 k | dreyfus | AB |
| *Tableau analytique du cocuage (Analytical Table of Cuckoldry),* Charles Fourier. | 24.3 k | cocuage | AB |
| *Les Revoltés de la Bounty (The Mutineers of the Bounty),* Jules Verne, 1879, short story. | 32.6 k | bounty_verne | AB |
| *Adelaïde,* Arthur de Gobineau, 1869. | 36.5 k | adelaide | AB |
| *L'Île des esclaves (Slave Island*), Pierre de Marivaux, 1725. | 38.2 k | esclaves | AB |
| *Recueil des Expressions Vicieuses et des Anglicismes les Plus Fréquents (Collection of Incorrect Expressions and Frequently Used Anglicisms*) Jules Fabien Gingras, 1861. | 41.9 k | anglicismes | G |
| *Esther,* Jean Racine, 1689 (drama). | 50.3 k | esther_racine | G |
| *Reflexions ou sentences et maximes morales (Moral Maxims),* François de La Rochefoucauld, 1680. | 59.7 k | maximes | AB |
| *La Chasse au Caribou (The Caribou Hunt),* Arthur de Gobineau, 1871. | 59.9 k | caribou | AB |
| *Le droit à la paresse (The Right to Be Lazy),* Paul Lafargue, 1883. | 60.8 k | paresse | AB |
| *Andromaque,* Jean Racine, 1667 (drama). | 62.9 k | andromaque | AB |
| *Les fourberies de Scapin (Scapin's Deceits),* Molière, 1671 (drama). | 64.4 k | fourberies | AB |
| *Le jeu de l'amour et du hazard (The game of love and chance),* Pierre de Marivaux, 1730 (drama). | 65.4 k | jeu | AB |
| *Supplément au Voyage de Bougainville (Addendum to the Voyage of Bougainville)*, Denis Diderot, 1796. | 70.3 k | bougainville | AB |
| *La visite à Buffon, ou voyage à Montbard (A Visit with Buffon, or Voyage to Montbard),* Marie-Jean Hérault de Séchelles, 1785. | 70.3 k | montbard | AB |
| *Le Cid,* Pierre Corneille, 1682 (drama). | 70.5 k | LeCid | AB |
| *Le Misanthrope (The Misanthrope),* Molière, 1666 (drama). | 71.6 k | misanthrope | AB |
| *Smarra,* Charles Nodier, 1821. | 76.5 k | smarra | AB |
| *Tartuffe,* Molière, 1664 (drama). | 78.4 k | tartuffe | AB |
| *Harmonies poétiques et religieuses (Poetic and Religious Harmonies),* Alphonse Lamartine, 1830. | 97.5 k | harmonies | AB |

**Texts in French, 100,000 < *N* < 300,000. (22)**

| Text | N | Short-Name | Source |
|---|---|---|---|
| *L'âne (The Donkey)* Victor Hugo, 1880 (poetry). | 100.7 k | ane | AB |
| *Les Fleurs du Mal* (*The Flowers of Evil*), Charles Baudelaire, 1857 (poetry). | 115.3 k | fleurs | G |
| *Claire de Lune (Moonlight)*, Guy de Maupassant, 1884 (collection of short stories). | 116.3 k | lune | G |
| *La Fille aux yeux d'or (The Girl with the Golden Eyes),* Honoré de Balzac, 1834-35 (novel). | 124.4 k | yeux_d_or | AB |
| *Lettres à un gentilhomme russe sur l'Inquisition espagnole (Letters to a Russian Gentleman on the Spanish Inquisition),* Joseph de Maistre, 1815. | 130.8 k | inquisition | AB |
| *Henri Poincaré, 2nd ed.,* Ernest Lebon, 1912 (biography). | 141.0 k | hp_bio | G |
| *Candide*, Voltaire, 1759 (novel). | 155.7 k | candide | G |
| *Contes de la Bécasse* (*Tales of the Woodcock*), Guy de Maupassant, 1894 (collection of short stories). | 157.8 k | becasse | G |
| *Entretiens sur la pluralité des mondes (Conversations on the Plurality of Worlds),* Bernard Le Bovier de Fontenelle, 1686. | 167.3 k | mondes | AB |
| *Du Contrat Social (The Social Contract),* Jean-Jacques Rousseau, 1762. | 187.1 k | contrat1762 | AB |
| *François le Champi,* George Sand*,* 1850 (novel). | 207.0 k | champi | AB |
| *L'Étui de nacre (Mother of Pearl),* Anatole France, 1892 (collection of short stories). | 211.6 k | nacre | AB |
| *Considérations sur la France (Considerations on France),* Joseph de Maistre*.1796.* | 212.3 k | considerations | AB |
| *Égarement de la Raison (Loss of Reason),* Charles Fourier, 1806. | 215.1 k | egarement | AB |
| *Correspondance de Voltaire avec le roi de Prusse (Correspondence of Voltaire with the King of Prussia),* Voltaire and Friedrich II, 1889*.* | 224.8 k | correspondance | G |
| *Mémoires d'un touriste: Voyage en Bretagne et en Normandie (Memoirs of a tourist: Voyage in Brittany and Normandy),* Stendhal, 1837 (travelogue). | 235.9 k | bn1837 | AB |
| *Aziyadé,* Pierre Loti, 1879 (novel). | 236.2 k | aziyade | AB |
| *Pensées (Thoughts),* Blaise Pascal, 1669 (collection of vignettes). | 244.6 k | pensees | AB |
| *Manon Lescaut,* Antoine-François Prévost 1728, (novel). | 266.0 k | manon | AB |
| *De la terre à la lune (From the Earth to the Moon),* Jules Verne, 1865 (science fiction)*.* | 266.6 k | terre_lune | AB |
| *La vie électrique (Electric Life),* Albert Robida, 1890 (novel) | 277.6 k | robida | G |
| *Le Tour de L'Espagne en Automobile (Tour of Spain in Automobile),* Pierre Marge, 1909, (travelogue). | 285.9 k | Espagne | G |

**Texts in French, 300,000 < *N* < 1,000,000. (21)**

| Text | N | Short-Name | Source |
|---|---|---|---|
| *Hokousaï.* Edmond de Goncourt, *ca*. 1892. (Biography of Hokusai.) | 328.6 k | hokousai | G |
| *Le tour du monde en quatre-vingts jours (Around the World in Eighty Days),* Jules Verne, 1872 (novel). | 330.3 k | 80jours | AB |
| *Dominique,* Eugène Fromentin, 1863 (novel). | 339.9 k | dominique | AB |
| *Une vie (A Woman's Life),* Guy de Maupassant, 1883 (novel). | 343.4 k | une_vie | AB |
| *Le Téléphone, le Microphone et le Phonographe,* Théodore du Moncel, 1878. | 370.3 k | telephone | G |
| *L'Île des Pingouins (Penguin Island),* Anatole France, 1908 (novel). | 393.7 k | pingouins | G |
| *Le père Goriot (Old Father Goriot),* Honoré de Balzac, 1835 (novel). | 416.3 k | goriot | AB |
| *Les preuves (Proofs),* Jean Jaurès, 1898. | 444.3 k | preuves | AB |
| *Salammbô,* Gustave Flaubert, 1862. | 490.2 k | salammbo | AB |
| *La curée (The Kill),* Émile Zola, 1871. | 502.6 k | curee | AB |
| *Voyages en France pendant les années 1787, 1788, 1789,* Arthur Young, translated from the English original, *Travels in France*, by M.H.J. Lesage in 1882, travelogue. | 515.4 k | young | AB |
| *Madame Bovary*, Gustave Flaubert, 1856. | 534.4 k | bovary | G |
| *Discours de la Méthode* (*Discourse of the Method*), René Descartes, 1637. | 538.6 k | methode | G |
| *De l'Esprit des lois (The Spirit of Laws),* Charles de Secondat Montesquieu, 2nd ed. With explicative notes by Paul Janet, 1892. | 554.7 k | loi | G |
| *Les liaisons dangereuses (Dangerous Liaisons),* Pierre Laclos, 1796. | 621.9 k | liaisons | AB |
| *20000 Lieues Sous les Mers* (*20,000 Leagues Under the Sea)*, Jules Verne, 1869. | 688.1 k | 20000lieues | G |
| *La Vie de Jeanne d'Arc (The Life of Joan of Arc),* Volume 1, Anatole France, 1908. | 707.3 k | arc | G |
| *Le rouge et le noir (The Red and the Black*), Stendhal, 1830. | 788.5 k | rouge_noir | AB |
| *Notre-Dame de Paris (The Hunchback of Notre-Dame),* Victor Hugo, 1831. | 799.2 k | notre_dame | AB |
| *Les Soirées de Saint-Pétersbourg (The Saint Petersburg Dialogues),* Joseph de Maistre, 1821. | 828.7 k | petersbourg | AB |
| *Douze ans dans la Haute-Ethiopie* (*Twelve Years in Upper Ethiopia*), Volume 1, Arnauld d'Abbadie, 1868. | 831.2 k | ethiopie | G |

**Texts in French, *N* > 1,000,000. (9)**

| Text | N | Short-Name | Source |
|---|---|---|---|
| *Les Trois Mousquetaires* (*The Three Musketeers*), Alexendre Dumas, 1844. | 1.02 M | 3mousquetaires | G |
| *Mémoires de Hector Berlioz*, Louis Hector Berlioz, 1870. | 1.04 M | berlioz | G |
| *Histoire de Paris depuis le temps des Gaulois jusqu'a nos jours (History of Paris from the Time of the Gauls to our Day), 2nd, ed.,* Théophile Lavallée, 1857. Two volumes, concatenated. | 1.09 M | gaulois | G |
| *Les Confessions (Confessions),* Jean-Jacques Rousseau, 1762. | 1.17 M | confessions | AB |
| *Gil Blas de Santillane (Gil Blas),* Alain René Lesage, 1715-35. | 1.25 M | santillane | AB |
| *Mémoires de Constant, premier valet de chambre de l'empereur, sur la vie privée de Napoléon, sa famille et sa cour. (Memoirs of Constant, First Valet of the Emperor, on the Private Life of Napoleon, His Family, and His Court)* Louis Constant Wairy, 1830. | 2.30 M | constant | G |
| *Histoire du Canada Depuis sa Découverte jusqu'à nos Jours* (*History of Canada from its Discovery to Modern Times*), François-Xavier Garneau, 1845. (All four tomes concatenated into a single document.) | 2.38 M | canada | G |
| *Les Misérables*, Victor Hugo, 1862. | 2.43 M | lesmis | G |
| *Mémoires d'Outre-tombe (Memories from Beyond the Grave),* François-René de Chateaubriand, 1848-50. | 3.43 M | outre_tombe | AB |

**Texts in German, N < 125,000.** In every case, the source is G. (20)

| Text | N | Short-Name |
|---|---|---|
| *Lolo*, Kasimir Edschmid, 1915. | 5658 | lolo |
| *Gedenkrede auf Wolfgang Amade Mozart,* (Tribute to Wolfgang Amadeus Mozart) Richard Beer-Hofmann, 1906. | 8902 | mozart |
| *Der Abend: Ein kleines Gespräch (The Evening: A Little Conversation),* Ferdinand Hardekopf, 1913. | 9722 | abend |
| *Der Struwwelpeter,* Heinrich Hoffmann, 1871. | 14.6 k | struwwelpeter |
| *Gesänge gegen den Tod (Chants against Death),* Gottfried Kölwel, 1914. | 16.0 k | tod |
| *Das Himmlische Licht (The Heavenly Light),* Ludwig Rubiner, 1916. | 26.8 k | licht |
| *Viola Tricolor,* Theodor Storm, 1873. | 47.1 k | viola |
| *Über die Geometrie der alten Aegypter, (On the Geometry of the Ancient Egyptians),* Emil Weyr, 1884. | 47.9 k | aegypter |
| *Die Jungfrau von Treiden (The Virgin of Treiden),* Adelbert Cammerer, 1848. | 55.6 k | jungfrau |
| *Acht Briefe an eine Freundin über Clavier-Unterricht (Letters to a Friend about Piano Lessons),* Johanna Kinkel, 1852. | 56.4 k | freundin |
| *Shakespeare und die Bacon-Mythen (Shakespeare and the Bacon Myths) Kuno Fischer, 1895.* | 70.1 k | mythen |
| *Der Schuß von der Kanzel (The Shot from the Pulpit), Conrad Ferdinand Meyer, 1878.* | 76.8 k | schuss |
| *Das Liebeskonzil (The Council of Love),* Oskar Panizza, 1895. | 86.7 k | liebeskonzil |
| *Die Brüder Wright (The Wright Brothers),* Alfred Hildebrandt, 1909. | 96.3 k | wright |
| *Darwinismus und Sozialismus (Darwinism and Socialism) Ludwig Büchner, 1894.* | 100.6 k | darwinismus |
| *Platons Gastmahl (Plato's Symposium),* German translation by Rodolf Kassner. 380 BCE. | 101.2 k | gastmahl |
| *Der Alpenkönig und der Menschenfeind, (The King of the Alps and the Misanthrope),* Ferdinand Raimund, play, 1828. | 105.0 k | alpenkonig |
| *Tonio Kröger,* Thomas Mann, 1903. | 111.1 k | tonio |
| *Der rote Komet (The Red Comet),* Robert Heymann, 1909. | 115.3 k | komet |
| *Das Abendmahl im Zusammenhang mit dem Leben Jesu und der Geschichte des Urchristentums (The Last Supper in Connection with the Life of Jesus and the History of Early Christianity),* Albert Schweitzer, 1901. | 122.9 k | abendmahl |

**Texts in German, 125,000 < N < 400,000. (18)** In every case, the source is G.

| Text | N | Short-Name |
|---|---|---|
| *Geschichten vom lieben Gott (Stories of the Good Lord),* Rainer Maria Rilke, 1921. | 137.1 k | geschichten |
| *Der Unterkiefer des Homo Heidelbergensis Aus den Sanden von Mauer bei Heidelberg (The Mandible of Homo Heidelbergensis from the Sands of Mauer near Heidelberg),* Otto Schoetensack, 1908. | 139.9 k | homo |
| *Der Goldene Topf (The Golden Pot),* E.T.A. Hoffmann, 1913. | 153.9 k | topf |
| *Ecce Homo,* Friedrich Wilhelm Nietzsche, 1888. | 158.5 k | EcceHomo |
| *Der Rote Kampfflieger (The Red Fighter Pilot),* Manfred von Richthofen, 1917. Autobiography. | 167.3 | redbaron |
| *Rembrandt,* Hermann Knackfuß, 1897. (Biography) | 176.9 k | rembrandt |
| *Volks-Kochbuch (Cookbook for the People*), Hedwig Heyl, 1905. | 180.6 k | kochbuch |
| *Alaeddin und die Wunderlampe Aus Tausend und eine Nacht (Alladin and the Magic Lamp),* F. Staeger. 1919. | 182.4 k | wunderlampe |
| *Schelmuffskys wahrhaftige, kuriöse und sehr gefährliche Reisebeschreibung zu Wasser und zu Lande (Schelmuffsky's True, Curious, and Very Dangerous Journey over Water and over Land),* Christian Reuter, 1912. (Satire) | 184.0 k | Schelmuffsky |
| *Zur Psychopathologie des Alltagslebens* (The Psychopathology of Everyday Life), Sigmund Freud, 1901. | 189.2 k | freud |
| *Mein Weg, als Deutscher und Jude (My Way as a German and a Jew),* Jakob Wassermann, 1921. (Autobiography) | 189.9 k | weg |
| *Rousseau,* Paul Hensel, 1912, biography. | 191.5 k | rousseau |
| *Himmelsvolk: Ein Buch von Blumen, Tieren, und Gott (Sky People: A Book about Flowers, Animals, and God),* Waldemar Bonsels, 1920. Children's literature. | 234.2 k | himmelsvolk |
| *Kleine Dichtungen (Little Poems),* Robert Walser, 1914, collection of poetry and other vignettes. | 238.2 k | dichtungen |
| *Märchen für Kinder* (*Fairy Tales for Children*), Hans Christian Andersen, German translation of Paul Arndt. | 274.6 k | marchen |
| *Professor Unrat oder Das Ende eines Tyrannen (Professor Unrat or the End of a Tyrant),* Heinrich Mann, 1905. | 306.6 k | unrat |
| *Celsissimus: Salzburger Roman,* Arthur Achleitner, 1900. (Historical novel.) | 375.2 k | celsissimus |
| *Faust: Der Tragödie* (*The Tragedy of Faust*), Johann Wolfgang von Goethe, 1829, 1832. (Two parts, concatenated.) | 377.2 k | faust |

**Texts in German, N > 400,000.** In every case, the source is G. (11)

| Text | N | Short-Name |
|---|---|---|
| *Also Sprach Zarathustra* (*Thus Spoke Zarathustra*), Friedrich Nietzsche, 1883-85. | 410.7 k | zarathustra |
| *Der Moloch (The Juggernaut*), Jakob Wassermann, 1908. | 421.3 k | moloch |
| *Charles Fourier: Sein Leben und seine Theorien (Charles Fourier: His Life and His Theories),* August Bebel, 1890. | 458.5 k | fourier |
| *Singapore, Malacca, Java,* Fedor Jagor, 1866, travelogue. | 542.9 k | jagor |
| *Lichtenstein,* Wilhelm Hauff, 1826, historical novel. | 549.1 k | lichtenstein |
| *Königliche Hoheit (Royal Highness),* Thomas Mann, 1909 (novel). | 596.1 k | hoheit |
| *Aus meinem Leben (From my Life),* Paul von Hindenburg, 1920. | 632.2 k | hindenburg |
| *Italienische Reise (Italian Journey),* Johann Wolfgang Goethe, 1817. Two volumes concatenated (travelogue). | 916.7 k | italienische |
| *Kritik der reinen Vernunft* (*Critique of Pure Reason*), Immanuel Kant, First edition, 1781. | 1.01 M | kritik |
| *Reise in die Aequinoctial-Gegenden des neuen Continents* (Voyage to the Equinoctial Regions of the New Continents), Alexander von Humboldt, 1865. Translated from French to German by Hermann Hauff. (Three volumes, concatenated). | 1.70 M | continent |
| *Nach Amerika! (To America!)* Friedrich Gerstäcker, 1855. Six volumes concatenated into a single document. | 1.85 M | amerika_bang |

**Texts in Latin.** In every case, the source is G. (17)

| Text | N | Short-Name |
|---|---|---|
| *Disputatio pro Declaratione Virtutis Indulgentiarum (Disputation on the Power and Efficacy of Indulgences)* Martin Luther, 1517. | 9673 | 95 |
| *Magna Carta*, 1215. | 14.0 k | magna |
| *Sexti Properti Elegiarum (Elegies of Sextus Propertius),* Sextus Propertius, ca. 15 BCE. | 23.6 k | sexti |
| *Encomium artis medicae (Praise of the Art of Medicine),* Desiderius Erasmus, 1518. | 28.2 k | encomium |
| *De Pronuntiatione Graecae et Latinae Linguae (Pronunciation of the Greek and Latin Languages),* John Caius, 1574. | 29.6 k | caius |
| *Liber Esther* (*Book of Esther*) Latin Vulgate Bible. | 35.0 k | esther |
| *Confessio Augustana (Augsberg Confession),* Philip Melanchthon, 1530. | 56.7 k | augsberg |
| *Prophetia Danielis (Book of Daniel)* Latin Vulgate Bible. | 59.2 k | danielis |
| *Georgicon (The Georgics),* Virgil | 81.8 k | georgics |
| *De francicae linguae recta pronuntiatione (Correct pronunciation of the French language*), Théodore de Bèze, 1519-1605. | 82.3 k | beze |
| *Ars Grammaticae Iaponicae Linguae, (Grammar of the Japanese Language)* Diego Collado, 1632. | 117.4 k | iaponicae |
| *Fasti (The Festivals),* Ovid, AD 8. | 170.1 k | fasti |
| *Commentarii de Bello Galico* (*Commentaries on the Gallic Wars*), Julius Ceasar, ca 50 BCE. | 194.1 k | bello |
| *Aeneidos (The Aeneid),* Virgil, 29-19 BCE. | 367.6 k | aeneidos |
| *Confessiones,* St. Augustine, 397-398. | 723.1 k | augustine |
| *Colloquium heptaplomeres de rerum sublimium arcanis abditis (Colloquium of the Seven Regarding Hidden Secrets of Sublime Things),* Jean Bodin, 1588. | 799.9 k | bodin |
| *Procès des Templiers* (*Trial of the Templars*), M. Michelet, ed., Paris, 1841, 1851. (Two original tomes combined into a single document. This book was published with a French title, but its text is Latin.) | 2.10 M | templars |

**Texts in other languages. (4)**

| | | | |
|---|---|---|---|
| *Li congiés Adan,* Adam de la Halle, ca. 1300, Picard (French dialect). | 3828 | conge | AB |
| *La Chanson de Roland*, ca. 1100, Old French, Anon. | 118.0 k | roland | AB |
| *Per luchtschip "de Argonaut" naar Mars (The dirigible "The Argonaut" to Mars),* Albert Daiber, ca. 1910. Dutch translation from the original German by Cath. A. Visser. | 190.4 k | luchtschip | G |
| *Kalevala*, Elias Lönnrot, 1849. Finnish epic poetry. | 451.6 k | kalevala | G |

**Mixed-Language Texts.** In every case, the source is G. (3)

| Text | N | Short-Name |
|---|---|---|
| *Le Catalogue des Livres Censurez par la Faculté de Théologie de Paris* (*Catalogue of Books Censured by the Faculty of Theology of Paris*), 1544. (Latin/French) | 30.4 k | blacklist |
| *Dictionary of the Chinook Jargon, or Indian Trade Language, of the North Pacific Coast,* Canadian Ministry of Agriculture, 1889. (English/Chinook Jargon) | 34.5 k | chinook |
| *A Collection of Gospel Hymns in Ojibway and English,* Anon., 1899. (Ojibway/English) | 42.3 k | ojibway |

**Random Numerical Sequences. (10) In every case, the source is [G]**

| Sequence | N | Short-Name |
|---|---|---|
| 1000th Fibonacci number | 209 | Fib1000 |
| $(9000!)/10^{2248}$ (†) | 29.4 k | fac9000div |
| $(10000!)/10^{2499}$ (†) | 33.2 k | fac10000div |
| List of the first 1000 Fibonacci numbers | 104.8 k | fiblist |
| The 32nd Mersenne prime in base 10. This is an interesting sequence, because in base 2 it is a string of 1's, and therefore highly correlated. What happens to the correlations when we convert to base 10? | 227.8 k | 32prime |
| Catalan's constant, $G$ or $\beta(2)$. See [AS], Chapter 23. $G = \beta(2) = \sum_{k=0}^{\infty} (-1)^k (2k+1)^{-2}$ | 300.0 k | catalan |
| $\pi$, 1,000,000 digits | 1.00 M | pi |
| Golden ratio = $(1+\sqrt{5})/2$, 1,000,025 digits | 1.00 M | gold |
| $e$, 2,000,068 digits | 2.00 M | e |
| $\sqrt{2}$, 5,000,005 digits | 5.00 M | root2 |

† Large factorials in base 10 contain many trailing zeroes. Dividing by the appropriate power of 10 removes all these.

**Non-Random Numerical Sequences. (10)**

| Sequence | N | Short-Name | Source |
|---|---|---|---|
| Actinic Flux Tables. Tables 3.7 and 3.11 of [FP], concatenated | 4479 | actinic | FP |
| Pascal's Triangle, first 50 rows. | 9102 | pascal | M |
| Table 4.1, Common Logarithms, from [AS]. | 16.9 k | log | MX, AS |
| Trigonometric Table. Sines, cosines, tangents, cotangents, cosecants, secants for angles in degrees and radians. | 17.3 k | trig | M |
| 9000! | 31.7 k | 9000fac | G |
| 10000! | 35.7 k | 10000fac | G |
| The First 10,000 Primes | 49.0 k | primes | G |
| Meteorological data, Salt Lake City, Utah. (approximately 20 years of daily temperature, daily precipitation, etc.) | 82.3 k | RedButte | UCC |
| Catalog of zip codes (US postal codes).* | 166.0 k | zc_dig | T |
| CIA World Factbook 2010* | 1.06 M | CIA2010dig | G |

* Streaming list of digits extracted from the text.

**Fortran Computer Codes, other computer lists. (3)**

| | N | Short-Name |
|---|---|---|
| Full suite of Fortran codes from [NR], concatenated in alphabetical order. | 60.9 k | recipes |
| Zeno V3.3, the authors' program for calculating transport properties of macromolecules; see, for example Ref. [MDG, MD] | 132.9 k | zeno |
| "ls –R" listing of the author's MacBook Pro home directory (a listing of all folders and files in the directory). | 2.7 M | lsRMacBook |

**Genetic Codes.** *(33)* Seven organisms are represented. Approximate length in basepairs (bp) is also displayed.

Eukaryotes:

*Caenorhabditis elegans* (a nematode), six nuclear chromosomes (I, II, III, IV, V, and X), and one mitochondrial chromosome (M).

*Saccharomyces cerevisiae* (brewer's yeast), 16 nuclear chromosomes (I through XVI), and one mitochondrial chromosome (M).

*Pan troglodytes* (chimpanzee), 23 non-sex chromosomes (1, 2A, 2B, 3, 4, …, 22), two sex chromosomes (X, Y), and one mitochondrial chromosome (MT). (We believe one of the chimpanzees was persuaded to leave his work at the keyboard long enough to provide a cheek swab. But not before he almost succeeded in generating *Hamlet*. Alas, he typed the entire text perfectly except for two typos. He typed "to pee or not to pee.")

Prokaryote:

*Escherichia coli*.

Viruses:

Ovine herpesvirus 2, fowlpox virus, koi herpesvirus.

| **Chromosome, length in basepairs is also shown.** | **Short-Name** | **Source** |
|---|---|---|
| *C. elegans*, M, 13.8 kbp | CelegansM | Ge |
| *P. troglodytes*, MT, 16.6 kbp | PanMT | Pan |
| *S. cerevisiae*, M, 85.8 kbp | yeastM | Ge |
| ovine herpesvirus 2, strain BJ1035, top strand, 135.1 kbp | herpes2BJ | V |
| *S. cerevisiae*, I, 230.2 kbp | yeastI | Ge |
| *S. cerevisiae*, VI, 270.1 kbp | yeastVI | Ge |
| fowlpox virus, strain Iowa, top strand, 288.5 kbp | Iowa | V |
| Koi herpesvirus, strain TUMST1, top strand, 295.1 kbp | koi | V |
| *S. cerevisiae*, III, 316.6 kbp | yeastIII | Ge |
| *S. cerevisiae*, IX, 439.9 kbp | yeastIX | Ge |
| *S. cerevisiae*, VIII, 562.6 kbp | yeastVIII | Ge |
| *S. cerevisiae*, V, 576.9 kbp | yeastV | Ge |
| *S. cerevisiae*, XI, 666.5 kbp | yeastXI | Ge |
| *S. cerevisiae*, X, 745.7 kbp | yeastX | Ge |
| *S. cerevisiae*, XIV, 784.3 kbp | yeastXIV | Ge |
| *S. cerevisiae*, II, 813.2 kbp | yeastII | Ge |
| *S. cerevisiae*, XIII, 924.4 kbp | yeastXIII | Ge |

| | | |
|---|---|---|
| *S. cerevisiae*, XVI, 948.1 kbp | yeastXVI | Ge |
| *S. cerevisiae*, XII, 1.08 Mbp | yeastXII | Ge |
| *S. cerevisiae*, VII, 1.09 Mbp | yeastVII | Ge |
| *S. cerevisiae*, XV, 1.09 Mbp | yeastXV | Ge |
| *S. cerevisiae*, IV, 1.53 Mbp | yeastIV | Ge |
| *E. coli*, strain K12-MG1655, 4.09 Mbp | Ecoli | C |
| *C. elegans*, III, 13.8 Mbp | CelegansIII | Ge |
| *C. elegans*, I, 15.1 Mbp | CelegansI | Ge |
| *C. elegans*, II, 15.3 Mbp | CelegansII | Ge |
| *C. elegans*, IV, 17.5 Mbp | CelegansIV | Ge |
| *C. elegans,* X, 17.7 Mbp | CelegansX | Ge |
| *C. elegans*, V, 20.9 Mbp | CelegansV | Ge |
| *P. troglodytes*, Y, 25.3 Mbp | Pan_Y | Pan |
| *P. troglodytes*, 21, 32.7 Mbp | Pan21 | Pan |
| *P. troglodytes*, 22, 33.3 Mbp | Pan22 | Pan |
| *P. troglodytes*, 19, 53.9 Mbp | Pan19 | Pan |

**Table S23. Letter-sorted texts (3)**

| | |
|---|---|
| songlark_sort (sort of song_lark) | $N = 6.44 \times 10^5$ |
| wealth_sort (sort of wealth) | $N = 1.77 \times 10^6$ |
| bible_sort (sort of bible) | $N = 3.22 \times 10^6$ |

**Sources of the texts.**

| Code | Description |
|---|---|
| A | Avalon Project: Documents in Law, History, and Diplomacy. avalon.law.yale.edu |
| AB | ABU: La Bibliothèque Universelle. abu.cnam.fr |
| AS | M. Abramowitz & I.A. Stegen, *Handbook of Mathematical Functions*, Dover, New York (1965). |
| C | Comprehensive Microbial Resource, J. Craig Venter Institute. cmr.jcvi.org |
| D | U.S. Department of Defense, Command and Control Research Program, dodccrp.org, via Project Gutenberg. |
| E | 1 Million Digits of Pi. eveandersson.com/pi/digits/1000000 |
| F | Ministère de la Justice et des Libertés, République Française, textes.justice.gouv.fr |
| FP | B.J. Findlayson-Pitts and J.N. Pitts, *Chemistry of the Upper and Lower Atmosphere*, Academic Press, San Diego, California (2000). |
| G | Free eBooks by Project Gutenberg. gutenberg.org |
| Ge | University of California Santa Cruz, Genome Bioinformatics Site. genome.ucsc.edu |
| M | Machine-generated by the authors. |
| MD | M.L. Mansfield & J.F. Douglas, Physical Review E, 78, 046712 (2008). |
| MDG | M.L. Mansfield, J.F. Douglas, E. J. Garbozci, Physical Review E, 64, 061401 (2001). |
| MX | Machine-generated by the authors to be an exact replica of the indicated table. |
| NR | W.H. Press, S.A. Teukolsky, W.T. Vettering, B.P. Flannery, *Numerical Recipes in Fortran 77*, 2nd ed., Cambridge University Press (1992). |
| T | www.textfiles.com/etext/ |
| UCC | Utah Climate Center, https://climate.usu.edu/ |
| V | Viral Bioinformatics Resource Center, virology.ca |
| Pan | ftp.ncbi.nih.gov/genomes/Pan_troglodytes |

**DERIVATIONS**

This section contains mathematical demonstrations and arguments alluded to in the main text.

**A. R-walks are embedded in ($d$ – 1) dimensions.**

The mapping from the *S*-walk to the *R*-walk projects the walk onto a ($d$ – 1)-dimensional subspace. This becomes obvious if we project step $k$ of the *R*-walk onto **D**:

$$(\mathbf{R}_k - \mathbf{R}_{k-1}) \cdot \mathbf{D} = (\mathbf{e}_k - \mathbf{s}) \cdot \mathbf{D} = 1 - 1 = 0$$

Because this is true for arbitrary $k$, it follows that the *R*-walks are embedded in the subspace orthogonal to **D**.

**B. Initial value of $\langle r_n^2 \rangle$.**

Prove that $\langle r_1^2 \rangle = 1 - s^2$

$$\langle r_1^2 \rangle = \frac{1}{N}\sum_{j=0}^{N-1} \left|\mathbf{R}_{j+1} - \mathbf{R}_j\right|^2 = \frac{1}{N}\sum_{j=0}^{N-1} \left|\mathbf{e}_{j+1} - \mathbf{s}\right|^2 = \frac{1}{N}\sum_{j=0}^{N-1} \left(e_{j+1}^2 - 2\mathbf{e}_{j+1} \cdot \mathbf{s} + s^2\right)$$

$$= 1 - \frac{2}{N}\sum_{j=0}^{N-1} \mathbf{e}_{j+1} \cdot s + s^2 \;=\; 1 - \frac{2}{N}\mathbf{S}_{\boldsymbol{N}} \cdot \mathbf{s} + s^2 = 1 - s^2$$

**C. $g_n$ sums to zero.**

Show that

$$\sum_{n=0}^{N-1} g_n = 0$$

$$\sum_{n=0}^{N-1} g_n = \frac{1}{N}\sum_{j=1}^{N} \mathbf{e}_j \cdot \sum_{n=0}^{N-1} \mathbf{e}_{j+n} - Ns^2 = \frac{1}{N}\sum_{j=1}^{N} \mathbf{e}_j \cdot \mathbf{S}_N - Ns^2 = \mathbf{s} \cdot \mathbf{S}_N - Ns^2 = 0$$

**D. Value of $g_0$.**

Show that $g_0 = 1 - s^2$

$$g_0 = \frac{1}{N}\sum_{j=1}^{N} e_j^2 - s^2 = 1 - s^2$$

**E. Symmetry properties of $g_n$.**

Prove that $g_{-n} = g_n$ and $g_{N-n} = g_n$.

$$g_{-n} = \frac{1}{N}\sum_{j=1}^{N} \mathbf{e}_j \cdot \mathbf{e}_{j-n} - s^2$$

$$\text{Let } q = j - n$$

$$g_{-n} = \frac{1}{N}\sum_{q=1-n}^{N-n} \mathbf{e}_q \cdot \mathbf{e}_{q+n} - s^2 = g_n$$

Because all indices are defined modulo *N*, we also have $g_{N-n} = g_n$.

**F. Limits on $s^2$.**

Prove that

$$\frac{1}{d} \le s^2 \le 1$$

The first inequality follows from an application of Lagrange multipliers – extremize $\sum f_j^2$ subject to the constraint $\sum f_j = 1$ – which also indicates that $1/d = s^2$ only when all the $f_j$'s are equal.

The second inequality is obtained as follows: $f_j \le 1$; hence $f_j^2 \le f_j$; summing over *j* yields $s^2 \le 1$.

**G. Relationship between $\langle r_n^2 \rangle$ and $g_n$.**

Prove:

$$\langle r_n^2 \rangle = n(1 - s^2) + 2\sum_{k=1}^{n-1} (n - k) g_k$$

$$\mathbf{R}_{j+n} - \mathbf{R}_j = \mathbf{S}_{j+n} - \mathbf{S}_j - n\mathbf{s} = \sum_{k=j+1}^{j+n} \mathbf{e}_k - n\mathbf{s}$$

$$\left(\mathbf{R}_{j+n} - \mathbf{R}_j\right)^2 = \sum_{k=j+1}^{j+n} \sum_{l=j+1}^{j+n} \mathbf{e}_k \cdot \mathbf{e}_l - 2n\mathbf{s} \cdot \sum_{k=j+1}^{j+n} \mathbf{e}_k + n^2 s^2$$

$$\langle r_n^2 \rangle = \frac{1}{N} \sum_{j=0}^{N-1} \sum_{k=j+1}^{j+n} \sum_{l=j+1}^{j+n} \mathbf{e}_k \cdot \mathbf{e}_l - 2n\mathbf{s} \cdot \frac{1}{N} \sum_{j=0}^{N-1} \sum_{k=j+1}^{j+n} \mathbf{e}_k + n^2 s^2$$

The double sum

$$\sum_{k=j+1}^{j+n} \sum_{l=j+1}^{j+n} \mathbf{e}_k \cdot \mathbf{e}_l$$

is the sum over all elements of a square matrix of order *n*, each of whose terms is the dot product of two **e**-vectors carrying indexes of the form *j* + *u* and *j* + *v*, for *u* and *v* integers lying somewhere between *j* + 1 and *j* + n. Each of these dot products is then acted upon by the operator

$$\frac{1}{N} \sum_{j=0}^{N-1}$$

Recall the definition

$$p_n = \frac{1}{N} \sum_{j=1}^{N} \mathbf{e}_j \cdot \mathbf{e}_{j+n}$$

When the *j*-sum operator operates on the dot product $\mathbf{e}_k \cdot \mathbf{e}_l$ it produces the term $p_q$ where *q* is the absolute value of the difference between *k* and *l*. But this means that all terms along the main diagonal of the matrix, of which there are *n*, become $p_0 = 1$. All the terms in the two first off-diagonals, of which there are $2(n-1)$, become $p_1$. All terms in the second off-diagonals, of which there are $2(n-2)$, become $p_2$. And so forth, until we reach the two terms in the upper-right and lower-left corners of the matrix, which become $p_{n-1}$. Therefore, we can write

$$\frac{1}{N}\sum_{j=0}^{N-1}\sum_{k=j+1}^{j+n}\sum_{l=j+1}^{j+n}\mathbf{e}_k\cdot\mathbf{e}_l = n + 2\sum_{q=1}^{n-1}(n-q)p_q$$

When the same operator operates on a term $\mathbf{e}_{j+q}$ it produces **s**. Therefore,

$$\frac{1}{N}\sum_{j=0}^{N-1}\sum_{k=j+1}^{j+n}\mathbf{e}_k = n\mathbf{s}$$

Recombining yields the desired result.

**H. Properties of Brownian trajectories constrained to return to the origin.**

The concept of an entropic spring is useful in understanding the properties of Brownian trajectories. The mean-square displacement of segments *j* and *j* + *n* can be derived by assuming that they are connected by a spring of strength $d/n$. The spring has Hamiltonian

$$\frac{d}{2n}r_n^2$$

and $\langle r_n^2\rangle$ is seen to scale as *n* when we evaluate the appropriate Gaussian integrals. But if the trajectory is constrained to return to the origin after *N* steps, then it is as if segment *n* is coupled to the origin with two springs, having respective strengths $d/n$ and $d/(N - n)$. The effective spring is equivalent to the sum of these, or

$$\frac{Nd}{n(N-n)}$$

with Hamiltonian

$$\frac{Nd\,r_n^2}{2n(N-n)}$$

Then the appropriate Gaussian integrals yield

$$\langle r_n^2\rangle \sim n(N-n)$$

Of course, this simplifies to $\langle r_n^2\rangle \sim n$ when $n \ll N$. It is consistent with the expectation that the final result must be invariant to the exchange $n \leftrightarrow N - n$, and could have been anticipated by realizing that $\langle r_n^2\rangle \sim N - n$ when *n* is near *N*.

## I. Results for texts without correlations.

When a text has no correlations, we expect $g_0 = 1 - s^2$, while $g_n$ for $n \neq 0$ is, aside from fluctuations, independent of $n$. Since all $g_n$ sum to 0, we obtain

$$g_{n\neq 0} \cong \frac{-(1-s^2)}{N-1}$$

When these values of $g_n$ are used to estimate the mean-square displacement, we obtain

$$\langle r_n^2 \rangle = (1-s^2)\frac{(n)(N-n)}{(N-1)}$$

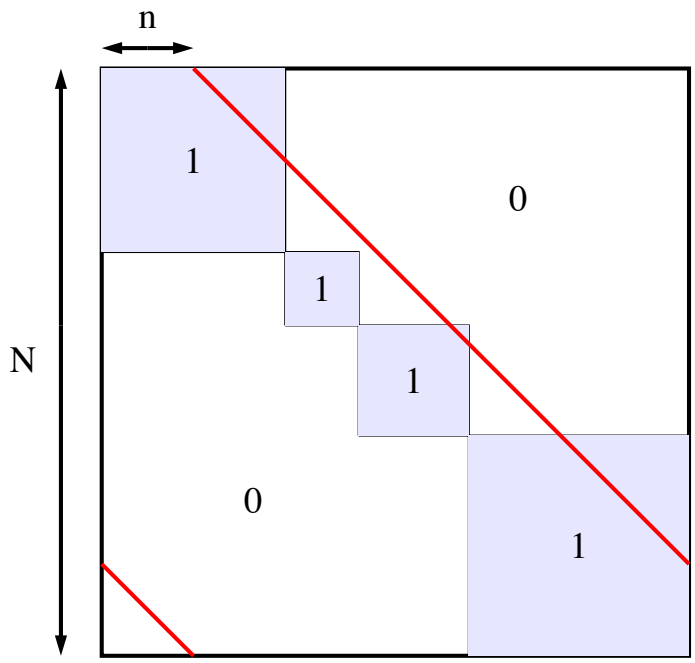
consistent with the scaling relation $\langle r_n^2 \rangle \sim n(N-n)$ for Brownian trajectories given above.

## J. Results for letter-sorted texts.

Our starting point is this formula:

$$g_n = \frac{1}{N}\sum_{j=1}^{N} \mathbf{e}_j \cdot \mathbf{e}_{j+n} - s^2$$

Envision each dot product to be an element of an order-$N$ matrix: $E_{jk} = \mathbf{e}_j \cdot \mathbf{e}_k$. Because letters in the text are sorted, the matrix is block-diagonal with a total of d blocks. All elements of each block are 1:



The sum over $E_{j,j+n} = \mathbf{e}_j \cdot \mathbf{e}_{j+n}$ includes all matrix elements along the red trace, offset by $n$ from the main diagonal. In other words, contributions to the sum occur when the red trace intersects the light-blue blocks. Let $L_1$, $L_2$, $L_3$, ..., $L_d$ represent the order of each block, and equivalently, the number of times letter 1, letter 2, etc., appear in the complete text. If the red trace intersects block q then the contribution from the block is $L_q - n$. If it does not intersect the block, then the contribution is 0. Therefore, we can write

$$p_n = \frac{1}{N}\sum_{q=1}^{d} V_{qn}; \qquad g_n = p_n - s^2$$

where $V_{qn} = \max(L_q - n, 0)$

Then

$$\langle r_n^2 \rangle = n(1 - s^2) + 2\sum_{k=1}^{n-1}(n-k)(p_k - s^2)$$

First consider $n$ so small that it is no larger than the smallest $L_q$. Then

$$p_k = \frac{1}{N}\sum_{q=1}^{d}\left(L_q - k\right) = 1 - \frac{kd}{N}$$

Combining all the above, we obtain

$$\langle r_n^2 \rangle = (1 - s^2)n^2 - O\left(\frac{d}{N}\right)$$

when is smaller than the smallest $L_q$. Generally, the $O(d/N)$ has a complex expression. It results from the $kd/N$ term in $p_k$, and is negligible since $d << N$.

## K. Statistical tests of the equivalence of $\alpha_1$ language distributions

Kolmogorov-Smirnoff tests (see the references listed below) were applied to the $\alpha_1$ samples for each of the four languages. The tests indicate that we can reject with confidence $(1 - p)$ the null hypotheses that any two samples were drawn from the same distribution, where $p$ is given in this table:

| | French | German | Latin |
|---|---|---|---|
| English | $< 10^{-6}$ | $< 10^{-6}$ | 0.40 |
| French | | $2 \times 10^{-4}$ | $4 \times 10^{-4}$ |
| German | | | $7 \times 10^{-6}$ |

Except for the English-Latin combination, we can safely reject the null hypotheses. A larger Latin sample would probably also yield a small $p$ value.

References:

Wikipedia, Kolmogorov-Smirnov test,
https://en.wikipedia.org/wiki/Kolmogorov%E2%80%93Smirnov_test

NIST Engineering Statistics Handbook, 1.3.5.16. Kolmogorov-Smirnov Goodness-of-Fit Test,
https://www.itl.nist.gov/div898/handbook/eda/section3/eda35g.htm

**Fitting Parameters**

Each text walk was fit to the single exponential form when only $\alpha$ is listed below.
Otherwise, each text walk was fit to the crossover form with $\alpha_1$ and $\alpha_2$ both listed.

$\beta$ appears whenever positive long-range correlations extend over a decade or more. For this reason, $\beta$ does not always appear below, especially for shorter texts.

Class Codes

| | |
|---|---|
| En | English language texts |
| Fr | French language texts |
| Ge | German language texts |
| La | Latin language texts |
| Ot | Other language texts (Picard, Old French, Dutch, Finnish) |
| Mx | Mixed-language texts (English + Chinook, English + Ojibway) |
| UD | Uncorrelated strings of digits |
| CD | Correlated strings of digits |
| CC | Computer lists |
| GC | Genetic codes |
| SL | Letter-sorted texts |

| NAME | A | alpha | n* | A1 | alpha1 | A2 | alpha2 | N | n1 | n2 | beta | Class |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ailsa_rock | - | - | 22.0 | 1.88E-03 | 0.877 | 1.08E-03 | 1.089 | 512 | -- | -- | -- | En |
| teign | 1.03E-03 | 1.003 | - | - | - | - | - | 872 | -- | -- | -- | En |
| gettysburg | 7.32E-04 | 1.009 | - | - | - | - | - | 1149 | -- | -- | -- | En |
| annabel_lee | 6.50E-04 | 1.030 | - | - | - | - | - | 1159 | -- | -- | -- | En |
| light_brigade | - | - | 6.0 | 7.87E-04 | 0.923 | 6.50E-04 | 1.043 | 1174 | -- | -- | -- | En |
| winchester | - | - | 28.3 | 4.77E-04 | 0.937 | 3.45E-04 | 1.045 | 1935 | -- | -- | -- | En |
| lincoln | - | - | 44.6 | 2.95E-04 | 0.935 | 1.46E-04 | 1.158 | 3157 | -- | -- | -- | En |
| peter | - | - | 734.1 | 2.30E-04 | 0.964 | 2.61E-02 | 0.569 | 3989 | -- | -- | -- | En |
| revere | - | - | 35.7 | 2.26E-04 | 0.929 | 1.39E-04 | 1.083 | 4098 | -- | -- | -- | En |
| slavery | - | - | 46.8 | 1.96E-04 | 0.976 | 8.08E-05 | 1.281 | 4717 | -- | -- | -- | En |
| penseroso | - | - | 27.8 | 1.96E-04 | 0.948 | 1.31E-04 | 1.083 | 4737 | -- | -- | -- | En |
| monroe | - | - | 43.4 | 1.89E-04 | 0.943 | 1.43E-04 | 1.023 | 4810 | -- | -- | -- | En |
| raven | - | - | 13.1 | 1.87E-04 | 0.968 | 1.41E-04 | 1.092 | 4892 | -- | -- | -- | En |
| jfk_inaug | - | - | 30.9 | 1.53E-04 | 0.948 | 8.56E-05 | 1.149 | 6040 | 7.9 | 446.7 | -0.532 | En |
| un102 | - | - | 102.9 | 1.39E-04 | 0.917 | 3.45E-05 | 1.326 | 6687 | -- | -- | -- | En |
| dream_king | - | - | 30.8 | 1.32E-04 | 0.960 | 7.23E-05 | 1.173 | 6949 | -- | -- | -- | En |
| declaration | - | - | 68.5 | 1.24E-04 | 0.958 | 4.85E-05 | 1.242 | 7388 | -- | -- | -- | En |
| tj_inaug | - | - | 44.3 | 1.13E-04 | 0.946 | 7.15E-05 | 1.079 | 8227 | -- | -- | -- | En |
| calamity | - | - | 51.0 | 1.06E-04 | 0.961 | 5.46E-05 | 1.165 | 8629 | -- | -- | -- | En |
| gift_magi | - | - | 19.1 | 1.07E-04 | 0.956 | 7.94E-05 | 1.069 | 8680 | -- | -- | -- | En |
| ghost_story | - | - | 70.6 | 9.01E-05 | 0.954 | 3.58E-05 | 1.227 | 10256 | -- | -- | -- | En |
| red_death | - | - | 36.2 | 8.34E-05 | 0.952 | 6.39E-05 | 1.032 | 11000 | -- | -- | -- | En |
| antarctica | - | - | 77.8 | 7.38E-05 | 0.940 | 2.30E-05 | 1.298 | 12338 | 14.1 | 1413 | -0.231 | En |
| jersey | - | - | 162.7 | 7.23E-05 | 0.961 | 2.48E-05 | 1.226 | 12753 | 22.4 | 1413 | -0.678 | En |
| poor_children | - | - | 52.5 | 7.00E-05 | 0.939 | 4.65E-05 | 1.051 | 13219 | -- | -- | -- | En |
| rime | - | - | 32.7 | 5.96E-05 | 0.957 | 3.34E-05 | 1.155 | 15423 | 5.6 | 1585 | -0.884 | En |
| taylor | - | - | 50.5 | 5.90E-05 | 0.950 | 3.08E-05 | 1.146 | 15546 | 7.1 | 1585 | -0.400 | En |
| isabella | - | - | 16.1 | 5.52E-05 | 0.954 | 4.17E-05 | 1.065 | 16848 | -- | -- | -- | En |
| dreams | - | - | 65.4 | 4.83E-05 | 0.953 | 2.37E-05 | 1.154 | 18880 | 8.9 | 1122 | -0.839 | En |
| fishing | - | - | 150.6 | 4.66E-05 | 0.957 | 2.11E-05 | 1.143 | 19800 | -- | -- | -- | En |
| ion | - | - | 51.3 | 4.39E-05 | 0.947 | 1.73E-05 | 1.250 | 20949 | 5.6 | 2239 | -0.261 | En |
| snark | - | - | 40.2 | 4.30E-05 | 0.957 | 2.20E-05 | 1.175 | 21389 | 12.6 | 2239 | -1.222 | En |
| constitution | - | - | 36.4 | 4.29E-05 | 0.943 | 2.29E-05 | 1.149 | 21488 | 8.9 | 3758 | -0.437 | En |
| sussexvamp | - | - | 25.1 | 3.72E-05 | 0.954 | 2.74E-05 | 1.058 | 24793 | 11.2 | 891.3 | -0.807 | En |
| hidalgo | - | - | 60.7 | 3.19E-05 | 0.932 | 1.31E-05 | 1.195 | 29067 | 7.1 | 5623 | -0.392 | En |
| hyperion | - | - | 60.8 | 3.13E-05 | 0.950 | 1.55E-05 | 1.151 | 29474 | 12.6 | 1585 | -0.513 | En |
| transcend | - | - | 34.2 | 3.05E-05 | 0.952 | 1.91E-05 | 1.102 | 30014 | 5.0 | 1413 | -0.206 | En |
| maelstrom | - | - | 58.7 | 3.06E-05 | 0.955 | 1.85E-05 | 1.093 | 30051 | 17.8 | 3758 | -1.085 | En |
| piper | - | - | 49.3 | 3.05E-05 | 0.954 | 1.47E-05 | 1.180 | 30379 | 8.9 | 2818 | -0.697 | En |
| literary_ethics | - | - | 10.6 | 2.89E-05 | 0.952 | 2.41E-05 | 1.036 | 31694 | 7.1 | 707.9 | -0.949 | En |
| gaol | - | - | 3.6 | 2.83E-05 | 0.948 | 2.63E-05 | 1.012 | 32557 | -- | -- | -- | En |
| carfax | - | - | 72.2 | 2.83E-05 | 0.954 | 1.34E-05 | 1.162 | 32593 | 17.8 | 2818 | -0.802 | En |
| comus | - | - | 85.7 | 2.72E-05 | 0.946 | 1.51E-05 | 1.094 | 34159 | -- | -- | -- | En |
| nursery | - | - | 43.1 | 2.66E-05 | 0.964 | 1.02E-05 | 1.305 | 34616 | 5.0 | 4467 | -0.579 | En |
| animal_gait | - | - | 57.1 | 2.64E-05 | 0.943 | 1.01E-05 | 1.245 | 34771 | 7.1 | 4467 | -0.649 | En |
| escape | - | - | 99.7 | 2.54E-05 | 0.953 | 8.12E-06 | 1.276 | 36130 | 14.1 | 5623 | -0.467 | En |
| vortex | - | - | 57.9 | 2.49E-05 | 0.961 | 1.17E-05 | 1.185 | 36889 | 10.0 | 2818 | -0.748 | En |
| adductor | - | - | 66.6 | 2.39E-05 | 0.945 | 8.80E-06 | 1.248 | 38670 | 10.0 | 8414 | -0.585 | En |
| civildis | - | - | 53.0 | 2.28E-05 | 0.953 | 1.20E-05 | 1.141 | 40225 | 14.1 | 1778 | -0.617 | En |
| weevil | - | - | 74.3 | 2.28E-05 | 0.956 | 9.00E-06 | 1.225 | 40621 | 12.6 | 7079 | -0.796 | En |

Fitting Parameters, continued.

| NAME | A | alpha | n* | A1 | alpha1 | A2 | alpha2 | N | n1 | n2 | beta | Class |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| sugar | - | - | 70.7 | 2.19E-05 | 0.961 | 6.61E-06 | 1.348 | 41906 | 6.3 | 8913 | -0.519 | En |
| su1840 | - | - | 174.1 | 2.04E-05 | 0.947 | 7.89E-06 | 1.167 | 45067 | -- | -- | -- | En |
| dumpling | - | - | 85.7 | 1.94E-05 | 0.963 | 7.25E-06 | 1.244 | 47503 | 6.3 | 3981 | -0.383 | En |
| wisteria | - | - | 81.7 | 1.91E-05 | 0.953 | 8.82E-06 | 1.162 | 48219 | 15.9 | 3981 | -0.647 | En |
| sourdough | - | - | 38.2 | 1.75E-05 | 0.952 | 7.68E-06 | 1.240 | 52542 | 5.6 | 3981 | -0.461 | En |
| argon | - | - | 108.0 | 1.76E-05 | 0.945 | 5.71E-06 | 1.250 | 52831 | 20.0 | 6310 | -0.536 | En |
| sleepy_hollow | - | - | 84.2 | 1.73E-05 | 0.954 | 8.49E-06 | 1.140 | 53572 | 17.8 | 2818 | -0.465 | En |
| attic | - | - | 77.1 | 1.66E-05 | 0.950 | 7.71E-06 | 1.159 | 55659 | 7.9 | 8913 | -0.577 | En |
| su1837 | - | - | 297.0 | 1.64E-05 | 0.945 | 4.60E-06 | 1.222 | 56166 | 28.2 | 5012 | -0.396 | En |
| su1838 | - | - | 117.3 | 1.61E-05 | 0.947 | 6.58E-06 | 1.171 | 57022 | 28.2 | 2239 | -0.363 | En |
| bc.birds | - | - | 60.2 | 1.53E-05 | 0.978 | 4.70E-06 | 1.387 | 59853 | 2.8 | 1.12E+04 | -0.395 | En |
| stamp | - | - | 79.4 | 1.51E-05 | 0.946 | 3.96E-06 | 1.377 | 61095 | 7.1 | 1.12E+04 | -0.355 | En |
| woodrat | - | - | 82.0 | 1.47E-05 | 0.951 | 4.79E-06 | 1.284 | 62807 | 12.6 | 8913 | -0.637 | En |
| su1839 | - | - | 24.0 | 1.41E-05 | 0.949 | 9.74E-06 | 1.078 | 65222 | 56.2 | 6683 | -0.543 | En |
| regained | - | - | 33.5 | 1.32E-05 | 0.947 | 8.25E-06 | 1.100 | 69561 | 12.6 | 5623 | -0.870 | En |
| stellar | - | - | 66.6 | 1.23E-05 | 0.948 | 4.11E-06 | 1.293 | 74577 | 11.2 | 1.78E+04 | -0.634 | En |
| macbeth | - | - | 28.7 | 1.22E-05 | 0.963 | 8.11E-06 | 1.101 | 75297 | 7.1 | 3981 | -0.597 | En |
| cleveland | - | - | 109.1 | 1.21E-05 | 0.957 | 5.02E-06 | 1.185 | 76229 | 8.9 | 5012 | -0.397 | En |
| parmenides | - | - | 43.1 | 1.18E-05 | 0.940 | 3.81E-06 | 1.361 | 76632 | 5.0 | 1.41E+04 | -0.576 | En |
| theism | - | - | 36.8 | 1.18E-05 | 0.951 | 6.84E-06 | 1.125 | 77486 | 5.0 | 8913 | -0.515 | En |
| areopagitica | - | - | 98.0 | 1.12E-05 | 0.951 | 6.17E-06 | 1.096 | 82382 | 11.2 | 2512 | -0.511 | En |
| plague | - | - | 63.2 | 1.06E-05 | 0.958 | 5.00E-06 | 1.175 | 86896 | 5.6 | 7079 | -0.690 | En |
| mythology | - | - | 78.5 | 1.04E-05 | 0.946 | 2.65E-06 | 1.392 | 88055 | 7.1 | 1.41E+04 | -0.308 | En |
| mammoth | - | - | 104.4 | 1.01E-05 | 0.949 | 3.26E-06 | 1.263 | 91143 | 14.1 | 2.37E+04 | -0.564 | En |
| comet | - | - | 95.4 | 9.89E-06 | 0.941 | 3.49E-06 | 1.227 | 92893 | 10.0 | 2.11E+04 | -0.529 | En |
| cratylus | - | - | 75.1 | 9.17E-06 | 0.951 | 3.10E-06 | 1.280 | 99987 | 7.1 | 1.41E+04 | -0.616 | En |
| beowulf | - | - | 44.3 | 9.07E-06 | 0.947 | 5.84E-06 | 1.075 | 101194 | 20.0 | 3162 | -0.897 | En |
| alice | - | - | 77.2 | 8.58E-06 | 0.952 | 3.13E-06 | 1.248 | 107462 | 11.2 | 8913 | -0.620 | En |
| snake | - | - | 228.8 | 8.03E-06 | 0.950 | 1.64E-06 | 1.351 | 114479 | 14.1 | 2.24E+04 | -0.420 | En |
| othello | - | - | 73.4 | 7.92E-06 | 0.970 | 2.91E-06 | 1.273 | 115991 | 4.0 | 2.24E+04 | -0.650 | En |
| glass | - | - | 67.4 | 7.70E-06 | 0.957 | 3.23E-06 | 1.213 | 119620 | 10.0 | 5623 | -0.222 | En |
| christmas | - | - | 58.4 | 7.68E-06 | 0.957 | 3.79E-06 | 1.163 | 120081 | 7.9 | 5012 | -0.562 | En |
| oroonoko | - | - | 110.8 | 7.45E-06 | 0.955 | 2.91E-06 | 1.199 | 122924 | 8.9 | 1.00E+04 | -0.638 | En |
| woods.hole | - | - | 77.1 | 7.25E-06 | 0.949 | 2.13E-06 | 1.332 | 127169 | 10.0 | 1.12E+04 | -0.463 | En |
| chimes | - | - | 62.4 | 7.16E-06 | 0.963 | 3.18E-06 | 1.206 | 128002 | 4.5 | 1.59E+04 | -0.559 | En |
| hamlet | - | - | 57.3 | 7.11E-06 | 0.960 | 3.09E-06 | 1.215 | 129785 | 5.6 | 7079 | -0.478 | En |
| endymion | - | - | 38.8 | 6.90E-06 | 0.952 | 4.33E-06 | 1.095 | 134597 | 10.0 | 1.19E+04 | -0.610 | En |
| wild | - | - | 98.2 | 6.68E-06 | 0.951 | 2.47E-06 | 1.222 | 138045 | 8.9 | 1.78E+04 | -0.600 | En |
| cello | - | - | 114.5 | 6.55E-06 | 0.946 | 1.80E-06 | 1.311 | 140747 | 10.0 | 2.51E+04 | -0.376 | En |
| time_machine | - | - | 123.9 | 6.56E-06 | 0.950 | 2.50E-06 | 1.193 | 140769 | 17.8 | 1.26E+04 | -0.699 | En |
| billy_budd | - | - | 163.2 | 6.47E-06 | 0.951 | 2.31E-06 | 1.198 | 142980 | 15.9 | 1.59E+04 | -0.562 | En |
| radioUKG | - | - | 112.8 | 6.42E-06 | 0.955 | 2.17E-06 | 1.246 | 143120 | 10.0 | 2.82E+04 | -0.592 | En |
| hiawatha | - | - | 43.7 | 6.37E-06 | 0.952 | 2.87E-06 | 1.213 | 144913 | 7.9 | 1.78E+04 | -0.707 | En |
| pinus | - | - | 61.4 | 6.23E-06 | 0.952 | 1.90E-06 | 1.351 | 147572 | 7.1 | 2.24E+04 | -0.556 | En |
| flatland | - | - | 127.0 | 6.02E-06 | 0.951 | 1.69E-06 | 1.299 | 152607 | 11.2 | 2.82E+04 | -0.456 | En |
| oz | - | - | 57.1 | 5.85E-06 | 0.944 | 2.78E-06 | 1.161 | 158578 | 14.1 | 2.66E+04 | -0.415 | En |
| telegraph | - | - | 128.4 | 5.51E-06 | 0.957 | 1.75E-06 | 1.260 | 168254 | 11.2 | 2.24E+04 | -0.410 | En |
| utah | - | - | 122.6 | 5.38E-06 | 0.949 | 1.19E-06 | 1.393 | 171403 | 10.0 | 1.41E+04 | -0.406 | En |
| candle | - | - | 98.1 | 5.33E-06 | 0.952 | 1.63E-06 | 1.292 | 172605 | 6.3 | 1.78E+04 | -0.438 | En |
| fd_auto | - | - | 89.4 | 5.25E-06 | 0.954 | 1.78E-06 | 1.264 | 175261 | 7.1 | 1.41E+04 | -0.458 | En |
| avalanche | - | - | 139.4 | 5.19E-06 | 0.957 | 1.85E-06 | 1.216 | 176916 | 7.9 | 1.12E+04 | -0.220 | En |

Fitting parameters, continued.

| NAME | A | alpha | n* | A1 | alpha1 | A2 | alpha2 | N | n1 | n2 | beta | Class |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| taxidermy | - | - | 104.0 | 5.19E-06 | 0.951 | 1.21E-06 | 1.397 | 178508 | 8.9 | 3.55E+04 | -0.419 | En |
| highland | - | - | 83.3 | 5.00E-06 | 0.953 | 9.52E-07 | 1.542 | 184997 | 7.1 | 2.24E+04 | -0.336 | En |
| tj_auto | - | - | 92.4 | 4.91E-06 | 0.953 | 1.63E-06 | 1.268 | 186751 | 12.6 | 2.24E+04 | -0.603 | En |
| DrMoreau | - | - | 58.9 | 4.91E-06 | 0.950 | 2.44E-06 | 1.153 | 187507 | 7.1 | 1.26E+04 | -0.358 | En |
| peter_pan | - | - | 111.9 | 4.74E-06 | 0.956 | 1.49E-06 | 1.273 | 194175 | 10.0 | 2.24E+04 | -0.572 | En |
| problems | - | - | 73.7 | 4.67E-06 | 0.959 | 1.48E-06 | 1.319 | 195674 | 6.3 | 1.78E+04 | -0.519 | En |
| red_badge | - | - | 106.8 | 4.57E-06 | 0.949 | 2.09E-06 | 1.144 | 201956 | 12.6 | 7079 | -0.551 | En |
| maid_mars | - | - | 142.3 | 4.40E-06 | 0.945 | 1.22E-06 | 1.282 | 209642 | 15.9 | 2.82E+04 | -0.575 | En |
| shock | - | - | 173.4 | 4.30E-06 | 0.951 | 1.15E-06 | 1.285 | 213824 | 15.9 | 1.41E+04 | -0.446 | En |
| ice | - | - | 94.0 | 4.26E-06 | 0.956 | 1.76E-06 | 1.191 | 216578 | 11.2 | 2.37E+04 | -0.463 | En |
| zenda | - | - | 99.7 | 4.22E-06 | 0.954 | 1.52E-06 | 1.232 | 218326 | 10.0 | 2.82E+04 | -0.652 | En |
| puddnhead | - | - | 78.1 | 4.21E-06 | 0.958 | 1.43E-06 | 1.283 | 218393 | 7.1 | 2.24E+04 | -0.422 | En |
| breakers | - | - | 94.1 | 4.10E-06 | 0.958 | 1.64E-06 | 1.206 | 225918 | 7.1 | 4.73E+04 | -0.503 | En |
| pilgrim | - | - | 68.2 | 4.02E-06 | 0.954 | 1.37E-06 | 1.290 | 227561 | 4.5 | 1.41E+04 | -0.378 | En |
| armour | - | - | 74.7 | 4.00E-06 | 0.950 | 1.52E-06 | 1.231 | 230262 | 10.0 | 3.35E+04 | -0.439 | En |
| decline | - | - | 94.4 | 3.92E-06 | 0.947 | 1.31E-06 | 1.258 | 234047 | 8.9 | 2.51E+04 | -0.351 | En |
| pollyanna | - | - | 102.0 | 3.94E-06 | 0.965 | 1.48E-06 | 1.230 | 236343 | 8.9 | 2.51E+04 | -0.520 | En |
| kantright | - | - | 128.9 | 3.82E-06 | 0.953 | 8.44E-07 | 1.395 | 239823 | 7.1 | 3.55E+04 | -0.286 | En |
| other | - | - | 119.5 | 3.82E-06 | 0.944 | 1.01E-06 | 1.317 | 240666 | 8.9 | 5.96E+04 | -0.314 | En |
| willows | - | - | 91.8 | 3.68E-06 | 0.957 | 1.55E-06 | 1.189 | 250518 | 5.6 | 1.78E+04 | -0.483 | En |
| dairy | - | - | 189.7 | 3.57E-06 | 0.948 | 7.00E-07 | 1.385 | 258299 | 10.0 | 3.55E+04 | -0.303 | En |
| violinists | - | - | 137.4 | 3.46E-06 | 0.960 | 7.72E-07 | 1.392 | 264642 | 8.9 | 4.47E+04 | -0.536 | En |
| war | - | - | 127.7 | 3.46E-06 | 0.950 | 1.12E-06 | 1.245 | 266023 | 7.1 | 4.47E+04 | -0.348 | En |
| treasure | - | - | 81.5 | 3.35E-06 | 0.953 | 1.55E-06 | 1.159 | 275699 | 10.0 | 1.41E+04 | -0.536 | En |
| bf_auto | - | - | 221.8 | 3.14E-06 | 0.954 | 7.90E-07 | 1.289 | 293104 | 17.8 | 4.47E+04 | -0.483 | En |
| virginia | - | - | 83.8 | 3.11E-06 | 0.958 | 8.25E-07 | 1.380 | 293806 | 5.6 | 3.55E+04 | -0.384 | En |
| conduct_life | - | - | 42.1 | 3.09E-06 | 0.953 | 1.82E-06 | 1.114 | 296731 | 7.1 | 1.78E+04 | -0.755 | En |
| tom_sawyer | - | - | 44.6 | 3.11E-06 | 0.957 | 1.59E-06 | 1.166 | 296809 | 5.6 | 3.35E+04 | -0.511 | En |
| prince_pauper | - | - | 74.6 | 3.06E-06 | 0.950 | 1.30E-06 | 1.191 | 301660 | 7.1 | 2.82E+04 | -0.488 | En |
| fu_manchu | - | - | 66.4 | 2.89E-06 | 0.955 | 1.60E-06 | 1.114 | 319629 | 10.0 | 3.16E+04 | -0.502 | En |
| kidnapped | - | - | 235.8 | 2.88E-06 | 0.956 | 7.71E-07 | 1.266 | 320163 | 10.0 | 3.98E+04 | -0.474 | En |
| lost_world | - | - | 221.6 | 2.81E-06 | 0.957 | 7.05E-07 | 1.293 | 328203 | 14.1 | 4.47E+04 | -0.488 | En |
| frankenstein | - | - | 168.5 | 2.77E-06 | 0.950 | 8.89E-07 | 1.226 | 332235 | 22.4 | 2.82E+04 | -0.604 | En |
| northanger | - | - | 135.3 | 2.71E-06 | 0.951 | 8.83E-07 | 1.240 | 339821 | 12.6 | 4.47E+04 | -0.596 | En |
| zc_lett | - | - | 9.7 | 2.71E-06 | 0.973 | 1.05E-06 | 1.706 | 341835 | 5.6 | 4.47E+04 | -0.464 | En |
| paradise | - | - | 114.8 | 2.57E-06 | 0.945 | 8.63E-07 | 1.234 | 360069 | 8.9 | 5.62E+04 | -0.461 | En |
| colorado | - | - | 128.2 | 2.52E-06 | 0.946 | 6.60E-07 | 1.317 | 365033 | 7.9 | 5.62E+04 | -0.473 | En |
| ari_physics | - | - | 62.8 | 2.47E-06 | 0.954 | 6.63E-07 | 1.415 | 365836 | 4.5 | 5.62E+04 | -0.311 | En |
| nicomachean | - | - | 72.4 | 2.48E-06 | 0.951 | 6.95E-07 | 1.365 | 368285 | 7.1 | 5.62E+04 | -0.468 | En |
| scarlet_letter | - | - | 499.8 | 2.50E-06 | 0.951 | 3.68E-07 | 1.383 | 368619 | 28.2 | 5.62E+04 | -0.381 | En |
| tarzan | - | - | 170.6 | 2.45E-06 | 0.947 | 5.90E-07 | 1.317 | 376628 | 11.2 | 5.62E+04 | -0.477 | En |
| steep | - | - | 103.1 | 2.44E-06 | 0.948 | 8.78E-07 | 1.222 | 378017 | 12.6 | 3.55E+04 | -0.569 | En |
| pretoria | - | - | 246.6 | 2.35E-06 | 0.951 | 6.72E-07 | 1.236 | 393293 | 17.8 | 4.47E+04 | -0.479 | En |
| chymist | - | - | 112.3 | 2.31E-06 | 0.954 | 8.30E-07 | 1.224 | 397149 | 7.1 | 3.55E+04 | -0.590 | En |
| huckleberry | - | - | 86.0 | 2.18E-06 | 0.959 | 7.34E-07 | 1.277 | 422275 | 7.1 | 7.08E+04 | -0.542 | En |
| green_gables | - | - | 86.7 | 2.14E-06 | 0.963 | 8.74E-07 | 1.210 | 432840 | 7.9 | 8.91E+04 | -0.721 | En |
| belgium | - | - | 151.6 | 2.09E-06 | 0.945 | 5.51E-07 | 1.295 | 440091 | 8.9 | 7.08E+04 | -0.434 | En |
| purple | - | - | 104.3 | 2.07E-06 | 0.954 | 6.64E-07 | 1.270 | 446033 | 7.1 | 5.62E+04 | -0.392 | En |
| gulliver | - | - | 259.2 | 2.02E-06 | 0.947 | 4.72E-07 | 1.290 | 454867 | 28.2 | 8.41E+04 | -0.464 | En |
| odyssey | - | - | 274.5 | 2.02E-06 | 0.948 | 5.73E-07 | 1.228 | 458573 | 22.4 | 5.01E+04 | -0.530 | En |
| aeneid_english | - | - | 83.7 | 1.94E-06 | 0.937 | 7.71E-07 | 1.190 | 475729 | 14.1 | 5.62E+04 | -0.653 | En |

Fitting parameters, continued.

| NAME | A | alpha | n* | A1 | alpha1 | A2 | alpha2 | N | n1 | n2 | beta | Class |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| crusoe | - | - | 137.4 | 1.92E-06 | 0.952 | 5.06E-07 | 1.313 | 480411 | 14.1 | 8.91E+04 | -0.440 | En |
| ari_meta | - | - | 50.0 | 1.88E-06 | 0.947 | 6.54E-07 | 1.309 | 483717 | 5.0 | 3.55E+04 | -0.505 | En |
| seven_gables | - | - | 302.7 | 1.91E-06 | 0.951 | 5.51E-07 | 1.220 | 484120 | 28.2 | 2.82E+04 | -0.506 | En |
| connyank | - | - | 63.4 | 1.86E-06 | 0.956 | 8.42E-07 | 1.188 | 493657 | 7.1 | 4.47E+04 | -0.694 | En |
| walden | - | - | 111.3 | 1.85E-06 | 0.955 | 6.24E-07 | 1.248 | 496622 | 7.1 | 5.62E+04 | -0.516 | En |
| bush | - | - | 96.9 | 1.85E-06 | 0.953 | 6.26E-07 | 1.256 | 498560 | 10.0 | 5.62E+04 | -0.584 | En |
| republic | - | - | 92.4 | 1.73E-06 | 0.950 | 5.54E-07 | 1.278 | 529425 | 5.6 | 1.00E+05 | -0.479 | En |
| pnp | - | - | 265.7 | 1.72E-06 | 0.954 | 4.12E-07 | 1.288 | 537054 | 10.0 | 7.94E+04 | -0.417 | En |
| leaves | - | - | 33.0 | 1.69E-06 | 0.957 | 8.42E-07 | 1.204 | 542291 | 5.6 | 2.24E+04 | -0.767 | En |
| alhambra | - | - | 187.9 | 1.68E-06 | 0.949 | 4.37E-07 | 1.286 | 547853 | 17.8 | 6.31E+04 | -0.431 | En |
| american | - | - | 93.6 | 1.61E-06 | 0.959 | 5.81E-07 | 1.245 | 568555 | 6.3 | 1.68E+05 | -0.621 | En |
| knickerbocker | - | - | 161.5 | 1.62E-06 | 0.951 | 5.83E-07 | 1.194 | 569679 | 10.0 | 5.62E+04 | -0.613 | En |
| 2cities | - | - | 77.2 | 1.57E-06 | 0.958 | 6.04E-07 | 1.235 | 584594 | 7.1 | 5.01E+04 | -0.500 | En |
| madcrowd | - | - | 168.2 | 1.56E-06 | 0.961 | 4.62E-07 | 1.267 | 590211 | 8.9 | 4.47E+04 | -0.470 | En |
| narratives | - | - | 230.0 | 1.48E-06 | 0.949 | 2.83E-07 | 1.374 | 620685 | 11.2 | 1.12E+05 | -0.505 | En |
| song_lark | - | - | 100.8 | 1.43E-06 | 0.954 | 5.01E-07 | 1.242 | 644387 | 8.9 | 7.08E+04 | -0.545 | En |
| tess | - | - | 102.4 | 1.42E-06 | 0.956 | 6.02E-07 | 1.178 | 648879 | 7.1 | 3.55E+04 | -0.514 | En |
| mohicans | - | - | 238.9 | 1.38E-06 | 0.948 | 4.57E-07 | 1.192 | 668189 | 35.5 | 7.08E+04 | -0.556 | En |
| pompeii | - | - | 109.3 | 1.38E-06 | 0.947 | 4.65E-07 | 1.239 | 669604 | 17.8 | 6.31E+04 | -0.577 | En |
| mansfield | - | - | 166.9 | 1.33E-06 | 0.958 | 4.24E-07 | 1.241 | 691470 | 7.9 | 8.91E+04 | -0.580 | En |
| 2years | - | - | 123.8 | 1.33E-06 | 0.954 | 3.88E-07 | 1.290 | 692184 | 10.0 | 7.08E+04 | -0.515 | En |
| GandS | - | - | 39.7 | 1.32E-06 | 0.974 | 5.29E-07 | 1.306 | 694841 | 3.5 | 7.94E+04 | -0.516 | En |
| childs_history | - | - | 164.3 | 1.29E-06 | 0.947 | 3.58E-07 | 1.271 | 714472 | 8.9 | 1.12E+05 | -0.529 | En |
| cabin | - | - | 103.6 | 1.21E-06 | 0.959 | 3.95E-07 | 1.271 | 763206 | 5.6 | 7.08E+04 | -0.468 | En |
| eugenics | - | - | 111.2 | 1.20E-06 | 0.954 | 3.34E-07 | 1.317 | 766187 | 11.2 | 1.50E+05 | -0.432 | En |
| little_women | - | - | 107.5 | 1.20E-06 | 0.956 | 4.76E-07 | 1.196 | 771627 | 8.9 | 3.55E+04 | -0.524 | En |
| dido | - | - | 270.9 | 1.17E-06 | 0.957 | 2.13E-07 | 1.384 | 787012 | 11.2 | 1.78E+05 | -0.480 | En |
| eyre | - | - | 94.1 | 1.17E-06 | 0.958 | 4.55E-07 | 1.215 | 788378 | 7.1 | 8.91E+04 | -0.628 | En |
| hudson | - | - | 130.5 | 1.09E-06 | 0.950 | 2.57E-07 | 1.363 | 839050 | 7.1 | 7.94E+04 | -0.323 | En |
| ivanhoe | - | - | 193.6 | 1.07E-06 | 0.948 | 2.67E-07 | 1.293 | 865021 | 14.1 | 1.00E+05 | -0.424 | En |
| argentine | - | - | 54.2 | 1.00E-06 | 0.966 | 2.66E-07 | 1.463 | 916430 | 5.6 | 1.33E+05 | -0.266 | En |
| leviathan | - | - | 112.5 | 9.65E-07 | 0.951 | 2.15E-07 | 1.407 | 949336 | 5.0 | 2.24E+05 | -0.453 | En |
| mobydick | - | - | 98.6 | 9.67E-07 | 0.959 | 3.47E-07 | 1.241 | 951981 | 5.6 | 5.01E+04 | -0.374 | En |
| origin | - | - | 160.6 | 9.10E-07 | 0.953 | 1.90E-07 | 1.388 | 1009400 | 14.1 | 7.94E+04 | -0.319 | En |
| human.nature | - | - | 215.3 | 8.60E-07 | 0.949 | 1.64E-07 | 1.383 | 1064260 | 22.4 | 1.41E+05 | -0.430 | En |
| mexcon | - | - | 596.3 | 8.59E-07 | 0.941 | 1.27E-07 | 1.350 | 1073300 | 35.5 | 1.26E+05 | -0.398 | En |
| ulysses | - | - | 119.0 | 7.78E-07 | 0.967 | 1.68E-07 | 1.436 | 1182412 | 5.6 | 1.12E+05 | -0.391 | En |
| cattle | - | - | 126.3 | 7.69E-07 | 0.948 | 1.47E-07 | 1.453 | 1197303 | 12.6 | 1.78E+05 | -0.437 | En |
| delusions | - | - | 178.1 | 7.19E-07 | 0.948 | 1.42E-07 | 1.391 | 1281301 | 17.8 | 1.78E+05 | -0.592 | En |
| pickwick | - | - | 99.6 | 6.87E-07 | 0.963 | 2.41E-07 | 1.254 | 1339981 | 5.6 | 5.62E+04 | -0.383 | En |
| copperfield | - | - | 114.7 | 6.24E-07 | 0.961 | 2.02E-07 | 1.267 | 1479411 | 7.1 | 1.78E+05 | -0.573 | En |
| ramayan | - | - | 114.8 | 5.32E-07 | 0.952 | 9.66E-08 | 1.501 | 1733928 | 7.9 | 2.82E+05 | -0.327 | En |
| wealth | - | - | 113.5 | 5.20E-07 | 0.944 | 1.34E-07 | 1.333 | 1767096 | 14.1 | 2.24E+05 | -0.545 | En |
| sociology | - | - | 65.4 | 4.64E-07 | 0.958 | 1.34E-07 | 1.377 | 1968230 | 6.3 | 1.59E+05 | -0.376 | En |
| lesmis_english | - | - | 76.6 | 3.65E-07 | 0.954 | 1.19E-07 | 1.295 | 2514927 | -- | -- | -- | En |
| bible | - | - | 81.5 | 2.85E-07 | 0.941 | 5.57E-08 | 1.514 | 3224197 | -- | -- | -- | En |
| fritz | - | - | 263.4 | 1.57E-07 | 0.957 | 3.34E-08 | 1.333 | 5856280 | -- | -- | -- | En |
| empire | - | - | 486.9 | 1.24E-07 | 0.944 | 1.41E-08 | 1.465 | 7433938 | -- | -- | -- | En |
| CIA2010let | - | - | 41.4 | 1.22E-07 | 0.968 | 2.89E-08 | 1.604 | 7443719 | -- | -- | -- | En |
| droits | 2.10E-04 | 0.970 | - | - | - | - | - | 3927 | -- | -- | -- | Fr |
| dreyfus | - | - | 54.9 | 3.98E-05 | 0.947 | 2.04E-05 | 1.143 | 22701 | 39.8 | 2239 | -0.601 | Fr |

Fitting parameters, continued.

| NAME | A | alpha | n* | A1 | alpha1 | A2 | alpha2 | N | n1 | n2 | beta | Class |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| cocuage | - | - | 18.0 | 3.75E-05 | 0.945 | 3.17E-05 | 1.005 | 24189 | -- | -- | -- | Fr |
| bounty_verne | - | - | 87.8 | 2.77E-05 | 0.946 | 1.07E-05 | 1.208 | 32604 | 11.2 | 4467 | -0.459 | Fr |
| adelaide | - | - | 26.9 | 2.47E-05 | 0.951 | 1.71E-05 | 1.076 | 36539 | 7.1 | 1884 | -0.274 | Fr |
| esclaves | - | - | 31.0 | 2.37E-05 | 0.937 | 1.26E-05 | 1.158 | 38218 | 8.9 | 3548 | -0.474 | Fr |
| anglicismes | - | - | 17.5 | 2.15E-05 | 0.950 | 1.46E-05 | 1.105 | 41936 | 7.1 | 1995 | -0.662 | Fr |
| esther_racine | - | - | 41.5 | 1.79E-05 | 0.942 | 9.48E-06 | 1.143 | 50337 | 20.0 | 2818 | -1.019 | Fr |
| maximes | - | - | 29.3 | 1.51E-05 | 0.935 | 9.32E-06 | 1.096 | 59732 | 7.1 | 2512 | -0.437 | Fr |
| caribou | - | - | 57.5 | 1.51E-05 | 0.949 | 7.68E-06 | 1.145 | 59920 | 5.6 | 1.33E+04 | -0.596 | Fr |
| paresse | - | - | 33.8 | 1.48E-05 | 0.938 | 8.48E-06 | 1.123 | 60841 | 7.1 | 3162 | -0.739 | Fr |
| andromaque | - | - | 42.8 | 1.44E-05 | 0.942 | 7.74E-06 | 1.134 | 62945 | 6.3 | 1122 | -0.635 | Fr |
| fourberies | - | - | 50.3 | 1.40E-05 | 0.944 | 6.14E-06 | 1.204 | 64353 | 7.9 | 1.00E+04 | -0.670 | Fr |
| jeu | - | - | 31.0 | 1.38E-05 | 0.943 | 8.31E-06 | 1.113 | 65408 | 7.1 | 8414 | -0.574 | Fr |
| bougainville | - | - | 26.8 | 1.28E-05 | 0.953 | 7.45E-06 | 1.148 | 70269 | 4.0 | 7079 | -0.570 | Fr |
| montbard | - | - | 51.8 | 1.28E-05 | 0.948 | 7.08E-06 | 1.122 | 70296 | 5.6 | 6310 | -0.388 | Fr |
| LeCid | - | - | 24.5 | 1.29E-05 | 0.941 | 8.05E-06 | 1.109 | 70540 | 10.0 | 1.06E+04 | -0.620 | Fr |
| misanthrope | - | - | 56.3 | 1.26E-05 | 0.942 | 6.49E-06 | 1.133 | 71635 | -- | -- | -- | Fr |
| smarra | - | - | 74.1 | 1.18E-05 | 0.941 | 4.22E-06 | 1.247 | 76482 | 5.6 | 1.78E+04 | -0.551 | Fr |
| tartuffe | - | - | 52.8 | 1.15E-05 | 0.947 | 5.37E-06 | 1.181 | 78409 | 7.9 | 1.00E+04 | -0.620 | Fr |
| harmonies | - | - | 44.1 | 9.21E-06 | 0.950 | 4.38E-06 | 1.191 | 97473 | 4.0 | 1.59E+04 | -0.744 | Fr |
| ane | - | - | 27.0 | 8.97E-06 | 0.949 | 5.13E-06 | 1.150 | 100649 | 5.6 | 7079 | -0.765 | Fr |
| fleurs | - | - | 33.3 | 7.84E-06 | 0.941 | 4.23E-06 | 1.150 | 115293 | 4.0 | 1.26E+04 | -0.728 | Fr |
| lune | - | - | 73.4 | 7.75E-06 | 0.954 | 3.21E-06 | 1.208 | 116330 | 4.0 | 1.26E+04 | -0.524 | Fr |
| yeux d or | - | - | 72.4 | 7.23E-06 | 0.948 | 2.53E-06 | 1.268 | 124444 | 5.6 | 1.78E+04 | -0.485 | Fr |
| inquisition | - | - | 64.2 | 6.89E-06 | 0.945 | 3.19E-06 | 1.166 | 130756 | 5.6 | 1.26E+04 | -0.671 | Fr |
| hp_bio | - | - | 59.5 | 6.40E-06 | 0.954 | 1.64E-06 | 1.452 | 140849 | 4.5 | 2.24E+04 | -0.428 | Fr |
| candide | - | - | 52.5 | 5.81E-06 | 0.949 | 2.64E-06 | 1.193 | 155742 | 5.6 | 2.82E+04 | -0.722 | Fr |
| becasse | - | - | 63.6 | 5.72E-06 | 0.952 | 2.40E-06 | 1.212 | 157841 | 4.0 | 2.66E+04 | -0.365 | Fr |
| mondes | - | - | 72.2 | 5.36E-06 | 0.943 | 1.99E-06 | 1.236 | 167266 | 10.0 | 3.55E+04 | -0.709 | Fr |
| contrat1762 | - | - | 81.1 | 4.82E-06 | 0.940 | 1.75E-06 | 1.230 | 187179 | 11.2 | 2.24E+04 | -0.530 | Fr |
| champi | - | - | 72.5 | 4.37E-06 | 0.951 | 1.62E-06 | 1.246 | 207000 | 5.6 | 1.41E+04 | -0.310 | Fr |
| nacre | - | - | 78.6 | 4.27E-06 | 0.945 | 1.57E-06 | 1.235 | 211556 | 10.0 | 1.78E+04 | -0.496 | Fr |
| considerations | - | - | 67.9 | 4.26E-06 | 0.939 | 1.77E-06 | 1.194 | 212264 | 8.9 | 2.24E+04 | -0.509 | Fr |
| egarement | - | - | 108.7 | 4.19E-06 | 0.942 | 1.36E-06 | 1.250 | 215032 | 11.2 | 1.78E+04 | -0.536 | Fr |
| correspondance | - | - | 61.5 | 4.00E-06 | 0.942 | 1.87E-06 | 1.163 | 224816 | 5.6 | 3.98E+04 | -0.714 | Fr |
| bn1837 | - | - | 50.8 | 3.82E-06 | 0.943 | 2.00E-06 | 1.136 | 235893 | 8.9 | 3.35E+04 | -0.527 | Fr |
| aziyade | - | - | 32.4 | 3.80E-06 | 0.953 | 2.19E-06 | 1.140 | 236174 | 4.0 | 2.11E+04 | -0.541 | Fr |
| pensees | - | - | 48.9 | 3.67E-06 | 0.944 | 1.45E-06 | 1.251 | 244638 | 5.6 | 2.24E+04 | -0.509 | Fr |
| manon | - | - | 101.9 | 3.40E-06 | 0.944 | 1.48E-06 | 1.156 | 266035 | 28.2 | 1.41E+04 | -0.407 | Fr |
| terre_lune | - | - | 102.5 | 3.39E-06 | 0.939 | 1.05E-06 | 1.268 | 266562 | 8.9 | 2.24E+04 | -0.384 | Fr |
| robida | - | - | 72.5 | 3.25E-06 | 0.948 | 1.21E-06 | 1.242 | 277621 | 4.0 | 4.47E+04 | -0.544 | Fr |
| Espagne | - | - | 45.4 | 3.14E-06 | 0.945 | 1.57E-06 | 1.163 | 285748 | 5.6 | 3.55E+04 | -0.668 | Fr |
| hokousai | - | - | 87.5 | 2.73E-06 | 0.965 | 6.32E-07 | 1.451 | 328602 | 4.0 | 5.62E+04 | -0.409 | Fr |
| 80jours | - | - | 83.5 | 2.74E-06 | 0.950 | 1.08E-06 | 1.210 | 330319 | 7.1 | 6.68E+04 | -0.515 | Fr |
| dominique | - | - | 77.3 | 2.66E-06 | 0.945 | 1.21E-06 | 1.159 | 339925 | 5.0 | 1.26E+04 | -0.293 | Fr |
| une_vie | - | - | 117.3 | 2.63E-06 | 0.953 | 9.36E-07 | 1.225 | 343400 | 5.6 | 2.82E+04 | -0.509 | Fr |
| telephone | - | - | 85.4 | 2.44E-06 | 0.939 | 8.38E-07 | 1.246 | 370290 | 5.6 | 4.47E+04 | -0.452 | Fr |
| pingouins | - | - | 80.7 | 2.29E-06 | 0.943 | 7.70E-07 | 1.265 | 393713 | 5.6 | 5.62E+04 | -0.490 | Fr |
| goriot | - | - | 90.6 | 2.16E-06 | 0.952 | 7.43E-07 | 1.257 | 416278 | 5.6 | 9.44E+04 | -0.520 | Fr |
| preuves | - | - | 111.8 | 2.03E-06 | 0.946 | 5.85E-07 | 1.296 | 444296 | 7.1 | 4.47E+04 | -0.402 | Fr |
| salammbo | - | - | 129.1 | 1.85E-06 | 0.936 | 7.20E-07 | 1.167 | 490228 | 8.9 | 1.59E+04 | -0.440 | Fr |
| curee | - | - | 141.0 | 1.80E-06 | 0.949 | 5.37E-07 | 1.265 | 502591 | 6.3 | 7.50E+04 | -0.372 | Fr |

Fitting parameters, continued.

| NAME | A | alpha | n* | A1 | alpha1 | A2 | alpha2 | N | n1 | n2 | beta | Class |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| young | - | - | 113.7 | 1.75E-06 | 0.943 | 6.61E-07 | 1.194 | 515393 | 5.6 | 5.62E+04 | -0.463 | Fr |
| bovary | - | - | 204.7 | 1.70E-06 | 0.946 | 5.18E-07 | 1.225 | 534365 | 6.3 | 5.62E+04 | -0.441 | Fr |
| methode | - | - | 121.0 | 1.66E-06 | 0.947 | 2.95E-07 | 1.499 | 538635 | 4.0 | 1.12E+05 | -0.242 | Fr |
| loi | - | - | 70.8 | 1.63E-06 | 0.938 | 6.02E-07 | 1.233 | 554660 | 7.1 | 5.01E+04 | -0.524 | Fr |
| liaisons | - | - | 59.3 | 1.45E-06 | 0.946 | 6.68E-07 | 1.175 | 621410 | 8.9 | 3.16E+04 | -0.550 | Fr |
| 20000lieues | - | - | 78.2 | 1.31E-06 | 0.940 | 4.60E-07 | 1.247 | 688086 | 5.6 | 3.98E+04 | -0.427 | Fr |
| arc | - | - | 80.0 | 1.28E-06 | 0.945 | 4.62E-07 | 1.240 | 707290 | 5.6 | 1.50E+05 | -0.552 | Fr |
| rouge_noir | - | - | 171.8 | 1.15E-06 | 0.948 | 3.35E-07 | 1.254 | 788460 | 11.2 | 1.12E+05 | -0.461 | Fr |
| notre_dame | - | - | 76.3 | 1.13E-06 | 0.951 | 4.09E-07 | 1.251 | 799206 | 4.0 | 1.00E+05 | -0.608 | Fr |
| petersbourg | - | - | 53.9 | 1.09E-06 | 0.950 | 4.81E-07 | 1.202 | 828683 | 5.6 | 6.31E+04 | -0.508 | Fr |
| ethiopie | - | - | 114.6 | 1.08E-06 | 0.949 | 2.76E-07 | 1.347 | 831189 | 5.6 | 1.12E+05 | -0.425 | Fr |
| 3mousquetaires | - | - | 82.0 | 8.83E-07 | 0.956 | 2.83E-07 | 1.300 | 1021742 | 4.0 | 2.00E+05 | -0.462 | Fr |
| berlioz | - | - | 112.5 | 8.67E-07 | 0.945 | 2.88E-07 | 1.241 | 1038977 | 5.6 | 7.08E+04 | -0.488 | Fr |
| gaulois | - | - | 127.1 | 8.26E-07 | 0.940 | 1.98E-07 | 1.346 | 1090002 | 5.6 | 2.24E+05 | -0.412 | Fr |
| confessions | - | - | 119.5 | 7.69E-07 | 0.952 | 2.74E-07 | 1.221 | 1173112 | 6.3 | 1.41E+05 | -0.631 | Fr |
| santillane | - | - | 116.7 | 7.25E-07 | 0.944 | 2.51E-07 | 1.222 | 1247846 | 22.4 | 1.78E+05 | -0.604 | Fr |
| constant | - | - | 135.4 | 3.92E-07 | 0.948 | 8.67E-08 | 1.382 | 2295004 | -- | -- | -- | Fr |
| canada | - | - | 147.7 | 3.78E-07 | 0.947 | 9.54E-08 | 1.319 | 2377083 | -- | -- | -- | Fr |
| lesmis | - | - | 66.7 | 3.71E-07 | 0.951 | 1.26E-07 | 1.293 | 2425910 | -- | -- | -- | Fr |
| outre_tombe | - | - | 117.0 | 2.63E-07 | 0.947 | 6.61E-08 | 1.346 | 3426335 | -- | -- | -- | Fr |
| lolo | 1.57E-04 | 0.956 | - | - | - | - | - | 5658 | -- | -- | -- | Ge |
| mozart | - | - | 72.3 | 1.01E-04 | 0.949 | 5.17E-05 | 1.135 | 8902 | -- | -- | -- | Ge |
| abend | - | - | 2.5 | 9.47E-05 | 0.941 | 8.96E-05 | 1.007 | 9722 | -- | -- | -- | Ge |
| struwwelpeter | - | - | 40.5 | 6.24E-05 | 0.944 | 2.88E-05 | 1.202 | 14647 | 22.4 | 1778 | -1.183 | Ge |
| tod | - | - | 47.1 | 5.67E-05 | 0.941 | 2.30E-05 | 1.241 | 15988 | 7.9 | 2239 | -0.681 | Ge |
| licht | - | - | 65.7 | 3.37E-05 | 0.937 | 1.41E-05 | 1.194 | 26824 | 8.9 | 2818 | -0.194 | Ge |
| esther | - | - | 122.1 | 2.58E-05 | 0.957 | 9.38E-06 | 1.219 | 35000 | 12.6 | 4467 | -0.310 | Ge |
| viola | - | - | 113.4 | 1.93E-05 | 0.936 | 7.67E-06 | 1.170 | 47115 | 10.0 | 3162 | -0.199 | Ge |
| aegypter | - | - | 45.1 | 1.88E-05 | 0.941 | 9.83E-06 | 1.141 | 47867 | 7.9 | 5957 | -0.239 | Ge |
| jungfrau | - | - | 47.1 | 1.66E-05 | 0.910 | 8.31E-06 | 1.115 | 55550 | -- | -- | -- | Ge |
| freundin | - | - | 134.1 | 1.60E-05 | 0.935 | 5.31E-06 | 1.215 | 56440 | 15.9 | 1.50E+04 | -0.642 | Ge |
| mythen | - | - | 73.2 | 1.29E-05 | 0.946 | 4.67E-06 | 1.249 | 70098 | 5.0 | 7079 | -0.229 | Ge |
| schuss | - | - | 51.3 | 1.18E-05 | 0.937 | 6.75E-06 | 1.097 | 76768 | 7.9 | 1.06E+04 | -0.538 | Ge |
| liebeskonzil | - | - | 58.0 | 1.04E-05 | 0.955 | 3.86E-06 | 1.275 | 86724 | 5.0 | 7943 | -0.484 | Ge |
| wright | - | - | 78.8 | 9.41E-06 | 0.946 | 2.09E-06 | 1.461 | 96268 | 4.0 | 8913 | -0.168 | Ge |
| darwinismus | - | - | 101.4 | 9.01E-06 | 0.937 | 3.26E-06 | 1.209 | 100620 | 6.3 | 1.12E+04 | -0.402 | Ge |
| gastmahl | - | - | 59.0 | 8.97E-06 | 0.947 | 2.72E-06 | 1.353 | 101163 | 5.6 | 2.37E+04 | -0.328 | Ge |
| alpenkonig | - | - | 73.2 | 8.69E-06 | 0.959 | 3.05E-06 | 1.278 | 104969 | 5.6 | 2.00E+04 | -0.418 | Ge |
| tonio | - | - | 84.1 | 8.17E-06 | 0.943 | 3.16E-06 | 1.208 | 111113 | 12.6 | 1.41E+04 | -0.604 | Ge |
| komet | - | - | 127.4 | 7.83E-06 | 0.957 | 2.98E-06 | 1.202 | 115268 | 10.0 | 1.59E+04 | -0.382 | Ge |
| abendmahl | - | - | 75.1 | 7.39E-06 | 0.939 | 2.70E-06 | 1.234 | 122881 | 5.0 | 1.12E+04 | -0.360 | Ge |
| geschichten | - | - | 76.4 | 6.58E-06 | 0.943 | 2.69E-06 | 1.196 | 137120 | 5.0 | 1.41E+04 | -0.433 | Ge |
| homo | - | - | 83.4 | 6.47E-06 | 0.952 | 1.90E-06 | 1.329 | 139878 | 5.0 | 2.24E+04 | -0.457 | Ge |
| topf | - | - | 57.1 | 5.89E-06 | 0.942 | 2.91E-06 | 1.147 | 153870 | 4.0 | 2.00E+04 | -0.307 | Ge |
| EcceHomo | - | - | 42.1 | 5.71E-06 | 0.942 | 3.13E-06 | 1.128 | 158502 | 5.0 | 1.00E+04 | -0.644 | Ge |
| redbaron | - | - | 131.7 | 5.45E-06 | 0.946 | 1.91E-06 | 1.211 | 167321 | 8.9 | 1.41E+04 | -0.748 | Ge |
| rembrandt | - | - | 45.5 | 5.12E-06 | 0.935 | 2.99E-06 | 1.093 | 176910 | 8.9 | 1.00E+04 | -0.781 | Ge |
| kochbuch | - | - | 61.9 | 5.12E-06 | 0.943 | 1.99E-06 | 1.229 | 180591 | 5.0 | 3.76E+04 | -0.434 | Ge |
| wunderlampe | - | - | 265.6 | 5.00E-06 | 0.940 | 1.01E-06 | 1.326 | 182417 | 11.2 | 3.55E+04 | -0.669 | Ge |
| Schelmuffsky | - | - | 180.5 | 4.97E-06 | 0.940 | 1.34E-06 | 1.264 | 183959 | 11.2 | 2.24E+04 | -0.616 | Ge |
| freud | - | - | 99.0 | 4.78E-06 | 0.943 | 1.69E-06 | 1.228 | 189232 | 6.3 | 2.51E+04 | -0.554 | Ge |

Fitting parameters, continued.

| NAME | A | alpha | n* | A1 | alpha1 | A2 | alpha2 | N | n1 | n2 | beta | Class |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| weg | - | - | 57.5 | 4.78E-06 | 0.939 | 2.51E-06 | 1.121 | 189873 | 5.0 | 5623 | -0.547 | Ge |
| rousseau | - | - | 79.0 | 4.72E-06 | 0.939 | 1.86E-06 | 1.201 | 191467 | 7.9 | 3.76E+04 | -0.612 | Ge |
| himmelsvolk | - | - | 121.3 | 3.87E-06 | 0.938 | 1.07E-06 | 1.290 | 234184 | 7.1 | 2.82E+04 | -0.386 | Ge |
| dichtungen | - | - | 63.3 | 3.79E-06 | 0.956 | 7.43E-07 | 1.601 | 238261 | 4.0 | 2.24E+04 | -0.169 | Ge |
| marchen | - | - | 84.9 | 3.31E-06 | 0.939 | 1.05E-06 | 1.277 | 274617 | 5.6 | 2.24E+04 | -0.510 | Ge |
| unrat | - | - | 117.9 | 2.97E-06 | 0.942 | 1.04E-06 | 1.215 | 306614 | 8.9 | 2.24E+04 | -0.566 | Ge |
| celsissimus | - | - | 230.4 | 2.44E-06 | 0.937 | 5.22E-07 | 1.316 | 375184 | 7.9 | 7.08E+04 | -0.456 | Ge |
| faust | - | - | 82.2 | 2.42E-06 | 0.941 | 7.78E-07 | 1.277 | 377202 | 7.9 | 5.62E+04 | -0.566 | Ge |
| zarathustra | - | - | 56.0 | 2.20E-06 | 0.949 | 7.75E-07 | 1.291 | 410702 | 4.0 | 5.62E+04 | -0.644 | Ge |
| moloch | - | - | 121.5 | 2.16E-06 | 0.939 | 7.94E-07 | 1.193 | 421277 | 7.9 | 5.62E+04 | -0.353 | Ge |
| fourier | - | - | 85.7 | 1.97E-06 | 0.939 | 6.62E-07 | 1.253 | 458521 | 4.0 | 8.91E+04 | -0.486 | Ge |
| jagor | - | - | 83.6 | 1.66E-06 | 0.948 | 5.22E-07 | 1.295 | 542914 | 4.0 | 7.08E+04 | -0.569 | Ge |
| lichtenstein | - | - | 96.7 | 1.65E-06 | 0.941 | 5.94E-07 | 1.220 | 549135 | 7.1 | 5.62E+04 | -0.598 | Ge |
| hoheit | - | - | 95.4 | 1.53E-06 | 0.940 | 5.29E-07 | 1.233 | 596107 | 5.0 | 5.62E+04 | -0.428 | Ge |
| hindenburg | - | - | 127.4 | 1.44E-06 | 0.935 | 4.73E-07 | 1.221 | 632233 | 7.9 | 5.01E+04 | -0.487 | Ge |
| italienische | - | - | 128.4 | 9.89E-07 | 0.938 | 3.15E-07 | 1.236 | 916655 | 7.1 | 8.91E+04 | -0.467 | Ge |
| kritik | - | - | 222.4 | 8.96E-07 | 0.942 | 1.66E-07 | 1.382 | 1010711 | 7.9 | 1.78E+05 | -0.523 | Ge |
| continent | - | - | 115.6 | 5.31E-07 | 0.947 | 1.53E-07 | 1.293 | 1702629 | 4.0 | 1.78E+05 | -0.482 | Ge |
| amerika_bang | - | - | 203.9 | 4.93E-07 | 0.941 | 1.17E-07 | 1.298 | 1845558 | 7.1 | 4.47E+05 | -0.519 | Ge |
| 95 | - | - | 12.9 | 9.39E-05 | 0.943 | 7.13E-05 | 1.063 | 9673 | -- | -- | -- | La |
| magna | - | - | 23.0 | 6.46E-05 | 0.959 | 3.25E-05 | 1.241 | 14038 | 5.6 | 707.9 | -0.432 | La |
| sexti | - | - | 13.5 | 3.86E-05 | 0.951 | 3.14E-05 | 1.037 | 23624 | -- | -- | -- | La |
| encomium | - | - | 45.7 | 3.21E-05 | 0.958 | 1.92E-05 | 1.112 | 28179 | -- | -- | -- | La |
| caius | - | - | 2.8 | 3.07E-05 | 0.959 | 2.87E-05 | 1.035 | 29294 | 4.5 | 2113 | -0.494 | La |
| augsberg | - | - | 39.5 | 1.60E-05 | 0.953 | 8.94E-06 | 1.137 | 56706 | 5.6 | 7499 | -0.540 | La |
| danielis | - | - | 45.8 | 1.52E-05 | 0.950 | 5.99E-06 | 1.268 | 59191 | 5.6 | 6310 | -0.311 | La |
| beze | - | - | 23.8 | 1.09E-05 | 0.995 | 5.76E-06 | 1.259 | 81445 | 3.5 | 4467 | -0.586 | La |
| georgics | - | - | 139.6 | 1.12E-05 | 0.941 | 4.52E-06 | 1.157 | 81806 | 22.4 | 6310 | -0.705 | La |
| iaponicae | - | - | 43.8 | 7.70E-06 | 0.982 | 2.82E-06 | 1.352 | 116164 | 2.5 | 1.12E+04 | -0.459 | La |
| fasti | - | - | 94.3 | 5.34E-06 | 0.957 | 2.71E-06 | 1.128 | 170104 | 7.1 | 4467 | -0.436 | La |
| bello | - | - | 75.0 | 4.68E-06 | 0.952 | 2.22E-06 | 1.157 | 194106 | 10.0 | 1.78E+04 | -0.552 | La |
| aeneidos | - | - | 223.2 | 2.48E-06 | 0.944 | 7.67E-07 | 1.211 | 367582 | 25.1 | 2.51E+04 | -0.472 | La |
| augustine | - | - | 79.3 | 1.25E-06 | 0.961 | 3.06E-07 | 1.432 | 723039 | 4.5 | 1.41E+05 | -0.402 | La |
| bodin | - | - | 88.6 | 1.14E-06 | 0.950 | 3.76E-07 | 1.271 | 799796 | 5.6 | 1.41E+05 | -0.566 | La |
| templars | - | - | 93.1 | 4.30E-07 | 0.963 | 7.97E-08 | 1.554 | 2099300 | -- | -- | -- | La |
| conge | - | - | 20.9 | 2.34E-04 | 0.923 | 1.47E-04 | 1.100 | 3828 | -- | -- | -- | Ot |
| roland | - | - | 49.9 | 7.60E-06 | 0.932 | 3.00E-06 | 1.234 | 117979 | 7.9 | 8913 | -0.392 | Ot |
| luchtschip | - | - | 172.6 | 4.80E-06 | 0.963 | 1.56E-06 | 1.238 | 190365 | 5.0 | 1.59E+04 | -0.451 | Ot |
| kalevala | - | - | 24.2 | 1.95E-06 | 1.020 | 1.18E-06 | 1.222 | 451592 | 2.8 | 4.73E+04 | -0.734 | Ot |
| blacklist | - | - | 56.8 | 2.97E-05 | 0.961 | 6.27E-06 | 1.589 | 30357 | 4.5 | 5623 | -0.139 | Mx |
| chinook | - | - | 14.8 | 2.61E-05 | 1.028 | 1.47E-05 | 1.319 | 34512 | 2.8 | 5012 | -0.607 | Mx |
| ojibway | - | - | 50.1 | 2.16E-05 | 0.996 | 2.83E-05 | 1.050 | 42258 | 3.5 | 2239 | -0.227 | Mx |
| Fib1000 | 3.39E-03 | 1.143 | - | - | - | - | - | 209 | -- | -- | -- | UD |
| fac9000div | 2.65E-05 | 1.035 | - | - | - | - | - | 29434 | -- | -- | -- | UD |
| fac10000div | 2.30E-05 | 1.049 | - | - | - | - | - | 33161 | -- | -- | -- | UD |
| fiblist | 8.38E-06 | 1.007 | - | - | - | - | - | 104750 | -- | -- | -- | UD |
| 32prime | 4.47E-06 | 0.970 | - | - | - | - | - | 227832 | -- | -- | -- | UD |
| catalan | 3.03E-06 | 1.000 | - | - | - | - | - | 300000 | -- | -- | -- | UD |
| pi | 7.71E-07 | 1.028 | - | - | - | - | - | 1000001 | -- | -- | -- | UD |
| gold | 1.05E-06 | 0.969 | - | - | - | - | - | 1000025 | -- | -- | -- | UD |
| e | 4.04E-07 | 1.019 | - | - | - | - | - | 2000068 | -- | -- | -- | UD |

Fitting parameters, continued.

| NAME | A | alpha | n* | A1 | alpha1 | A2 | alpha2 | N | n1 | n2 | beta | Class |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| root2 | 1.63E-07 | 1.020 | - | - | - | - | - | 5000005 | -- | -- | -- | UD |
| actinic | - | - | 5.4 | 1.94E-04 | 1.103 | 1.27E-04 | 1.521 | 4479 | -- | -- | -- | CD |
| pascal | - | - | 37.7 | 9.69E-05 | 1.031 | 4.05E-05 | 1.375 | 9102 | -- | -- | -- | CD |
| log | - | - | 35.3 | 5.26E-05 | 1.025 | 1.38E-05 | 1.676 | 16933 | -- | -- | -- | CD |
| trig | - | - | 65.0 | 5.18E-05 | 0.997 | 1.40E-05 | 1.472 | 17252 | -- | -- | -- | CD |
| 9000fac | - | - | 13.9 | 2.63E-05 | 1.258 | 1.30E-05 | 1.765 | 31682 | -- | -- | -- | CD |
| 10000fac | - | - | 14.6 | 2.34E-05 | 1.256 | 1.12E-05 | 1.780 | 35660 | -- | -- | -- | CD |
| primes | - | - | 4.2 | 1.77E-05 | 1.051 | 1.09E-05 | 1.683 | 48982 | -- | -- | -- | CD |
| RedButte | - | - | 10.1 | 9.78E-06 | 0.957 | 3.60E-06 | 1.828 | 82305 | -- | -- | -- | CD |
| zc_dig | - | - | 4.9 | 5.27E-06 | 1.041 | 2.98E-06 | 1.706 | 166028 | -- | -- | -- | CD |
| CIA2010dig | - | - | 19.2 | 7.71E-07 | 1.071 | 2.68E-07 | 1.739 | 1059524 | -- | -- | -- | CD |
| recipes | - | - | 3.6 | 1.53E-05 | 1.047 | 1.18E-05 | 1.336 | 60926 | -- | -- | -- | CC |
| zeno | - | - | 17.2 | 6.67E-06 | 1.050 | 3.16E-06 | 1.447 | 132883 | -- | -- | -- | CC |
| lsRMacBook | - | - | 11.8 | 3.34E-07 | 0.952 | 1.02E-07 | 1.912 | 2714710 | -- | -- | -- | CC |
| CelegansM | - | - | 25.8 | 5.00E-05 | 1.072 | 3.15E-05 | 1.278 | 13794 | 2.4 | 1259 | -0.432 | GC |
| PanMT | - | - | 53.3 | 4.39E-05 | 1.060 | 1.79E-05 | 1.406 | 16554 | 7.5 | 2239 | -0.473 | GC |
| yeastM | - | - | 6.3 | 7.47E-06 | 0.995 | 8.26E-06 | 1.258 | 85779 | 1.8 | 1.19E+04 | -0.745 | GC |
| herpes2BJ | - | - | 11.2 | 5.52E-06 | 1.105 | 3.62E-06 | 1.377 | 135135 | 0.9 | 1.59E+04 | -0.414 | GC |
| yeastI | - | - | 15.5 | 3.20E-06 | 1.088 | 1.89E-06 | 1.388 | 230208 | 0.9 | 2.00E+04 | -0.383 | GC |
| yeastVI | - | - | 14.6 | 2.74E-06 | 1.084 | 1.72E-06 | 1.345 | 270148 | 0.9 | 2.82E+04 | -0.483 | GC |
| Iowa | - | - | 50.8 | 2.43E-06 | 1.044 | 7.34E-07 | 1.569 | 288539 | 3.2 | 5.62E+04 | -0.452 | GC |
| koi | - | - | 31.9 | 2.35E-06 | 1.093 | 8.29E-07 | 1.647 | 295052 | 4.0 | 7.08E+04 | -0.426 | GC |
| yeastIII | - | - | 20.6 | 2.32E-06 | 1.092 | 1.27E-06 | 1.404 | 316617 | 0.9 | 5.31E+04 | -0.506 | GC |
| yeastIX | - | - | 8.5 | 1.68E-06 | 1.084 | 1.23E-06 | 1.305 | 439885 | 0.9 | 2.24E+04 | -0.411 | GC |
| yeastVIII | - | - | 8.5 | 1.32E-06 | 1.080 | 9.70E-07 | 1.294 | 562643 | 0.9 | 3.55E+04 | -0.469 | GC |
| yeastV | - | - | 12.1 | 1.28E-06 | 1.084 | 8.71E-07 | 1.316 | 576869 | 0.9 | 5.31E+04 | -0.502 | GC |
| yeastXI | - | - | 4.0 | 1.11E-06 | 1.079 | 9.55E-07 | 1.254 | 666454 | 0.9 | 4.22E+04 | -0.520 | GC |
| yeastX | - | - | 11.8 | 9.93E-07 | 1.073 | 6.78E-07 | 1.302 | 745742 | 0.9 | 4.47E+04 | -0.439 | GC |
| yeastXIV | - | - | 4.6 | 9.43E-07 | 1.080 | 7.88E-07 | 1.265 | 784333 | 0.9 | 5.01E+04 | -0.595 | GC |
| yeastII | - | - | 5.9 | 9.12E-07 | 1.077 | 7.40E-07 | 1.255 | 813178 | 0.9 | 3.16E+04 | -0.421 | GC |
| yeastXIII | - | - | 3.2 | 8.00E-07 | 1.079 | 7.15E-07 | 1.243 | 924429 | 0.9 | 5.31E+04 | -0.486 | GC |
| yeastXVI | - | - | 100.0 | 7.80E-07 | 1.081 | 8.31E-07 | 1.201 | 948062 | 0.9 | 4.73E+04 | -0.501 | GC |
| yeastXII | - | - | 8.4 | 6.86E-07 | 1.080 | 5.03E-07 | 1.302 | 1078175 | 0.9 | 9.44E+04 | -0.494 | GC |
| yeastVII | - | - | 3.0 | 6.78E-07 | 1.079 | 6.11E-07 | 1.240 | 1090947 | 0.9 | 4.47E+04 | -0.479 | GC |
| yeastXV | - | - | 100.0 | 6.78E-07 | 1.081 | 7.22E-07 | 1.194 | 1091289 | 0.9 | 4.47E+04 | -0.618 | GC |
| yeastIV | - | - | 7.9 | 4.82E-07 | 1.081 | 3.64E-07 | 1.286 | 1531919 | 0.9 | 4.47E+04 | -0.510 | GC |
| Ecoli | - | - | 35.9 | 1.83E-07 | 1.021 | 7.72E-08 | 1.383 | 4089838 | -- | -- | -- | GC |
| CelegansIII | - | - | 1.7 | 5.51E-08 | 1.126 | 5.35E-08 | 1.271 | 13783681 | -- | -- | -- | GC |
| CelegansI | - | - | 27.8 | 5.02E-08 | 1.127 | 2.70E-08 | 1.416 | 15072421 | -- | -- | -- | GC |
| CelegansII | - | - | 28.2 | 4.93E-08 | 1.120 | 2.63E-08 | 1.410 | 15279323 | -- | -- | -- | GC |
| CelegansIV | - | - | 60.2 | 4.31E-08 | 1.120 | 1.59E-08 | 1.516 | 17493785 | -- | -- | -- | GC |
| CelegansX | - | - | 12.1 | 4.21E-08 | 1.124 | 3.07E-08 | 1.313 | 17718854 | -- | -- | -- | GC |
| CelegansV | - | - | 22.2 | 3.60E-08 | 1.110 | 2.13E-08 | 1.364 | 20919568 | -- | -- | -- | GC |
| Pan_Y | - | - | 10.0 | 2.91E-08 | 1.151 | 1.85E-08 | 1.488 | 25272412 | -- | -- | -- | GC |
| Pan21 | - | - | 57.7 | 2.24E-08 | 1.174 | 6.28E-09 | 1.779 | 32706069 | -- | -- | -- | GC |
| Pan22 | - | - | 32.0 | 2.22E-08 | 1.185 | 9.37E-09 | 1.629 | 33251066 | -- | -- | -- | GC |
| Pan19 | - | - | 43.8 | 1.37E-08 | 1.197 | 4.96E-09 | 1.692 | 53859967 | -- | -- | -- | GC |
| songlark_sort | 1.88E-06 | 1.913 | - | - | - | - | - | 644387 | -- | -- | -- | SL |
| wealth_sort | 6.96E-07 | 1.913 | - | - | - | - | - | 1767096 | -- | -- | -- | SL |
| bible_sort | 3.65E-07 | 1.930 | - | - | - | - | - | 3224197 | -- | -- | -- | SL |

**Plots of $\langle r_n^2 \rangle$**

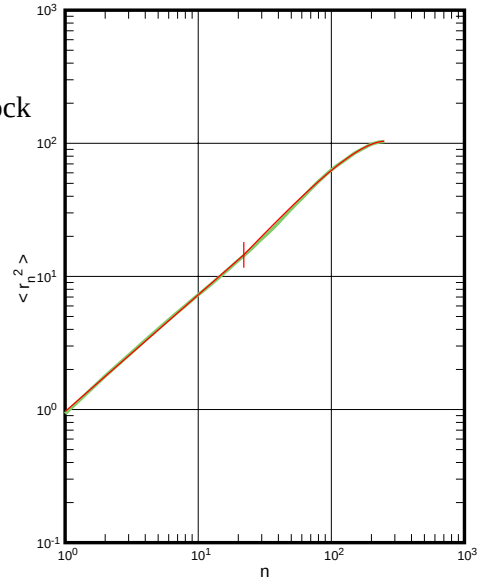
Introduction: Log-log plots of $\langle r_n^2 \rangle$ as a function of *n* appear on the following pages.

Light green traces are actual values $\langle r_n^2 \rangle$ as calculated from each text.

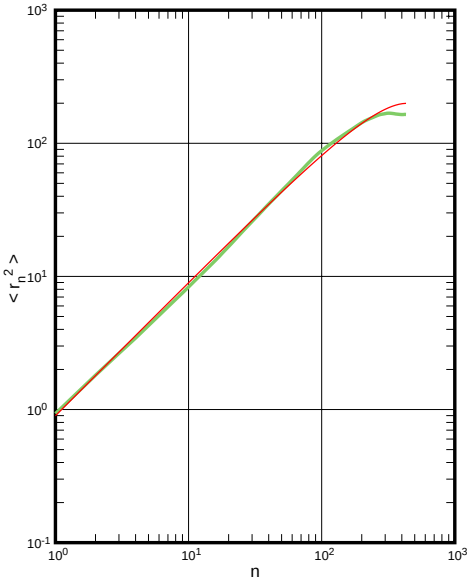
Red traces are fits according to either the single-exponent form, Eq. (12), or to the crossover form, Eq. (13). Whenever a text walk is fit to the crossover form, the value of n* is displayed with a short, vertical, red trace.

English texts: 500 < N < 2000

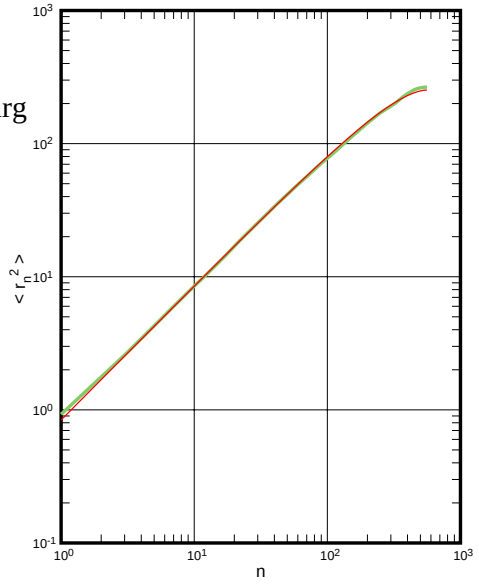
ailsa_rock



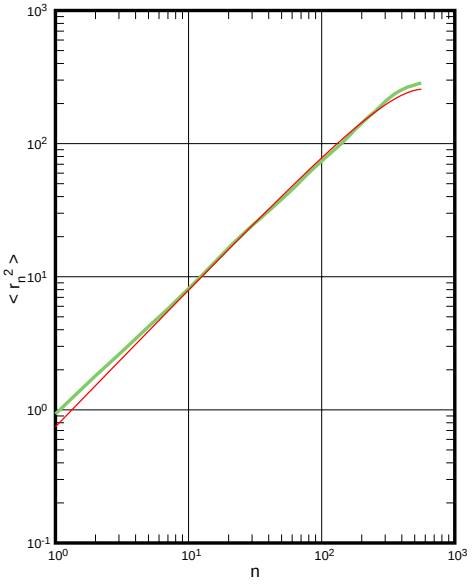

teign



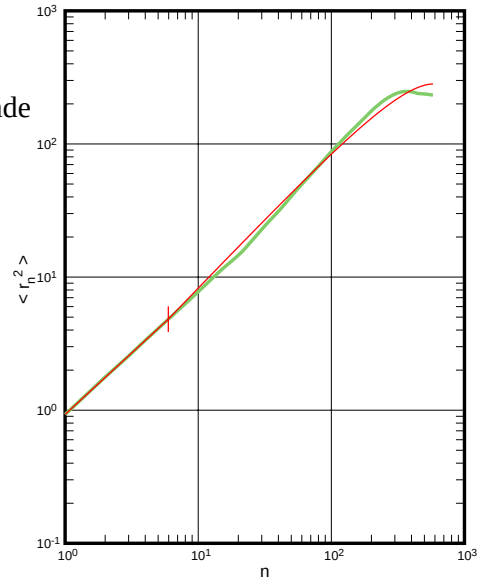

gettysburg



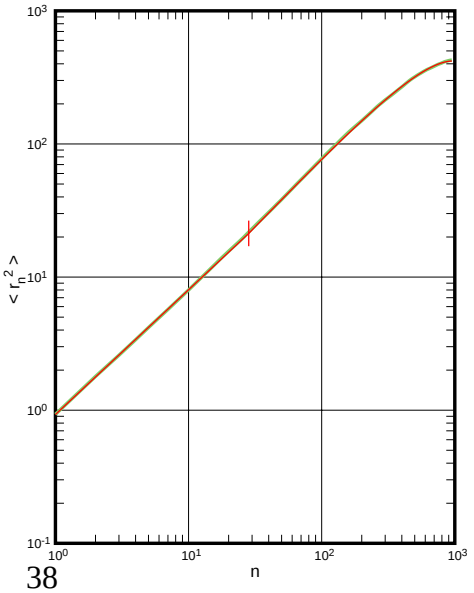

annabel_lee




light_brigade

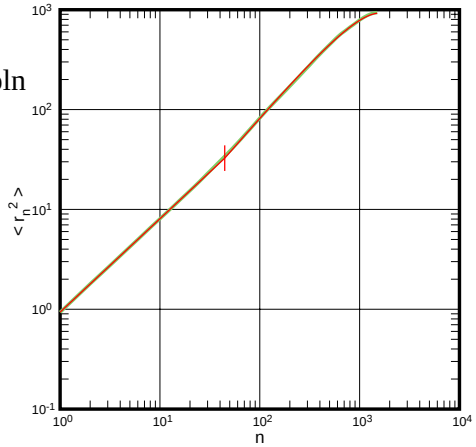



winchester

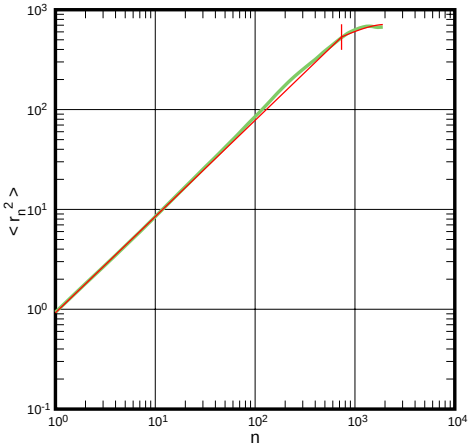


English texts: 3000 < N < 4850

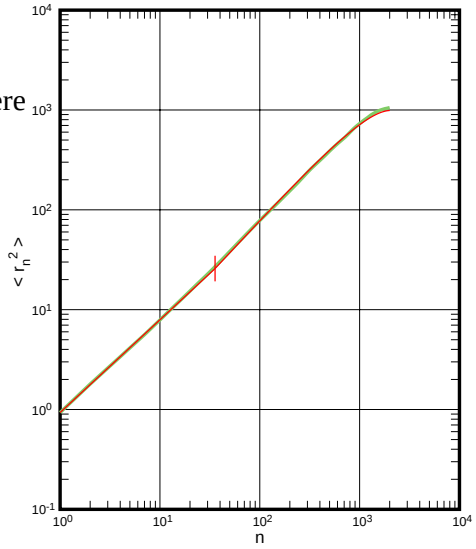
lincoln



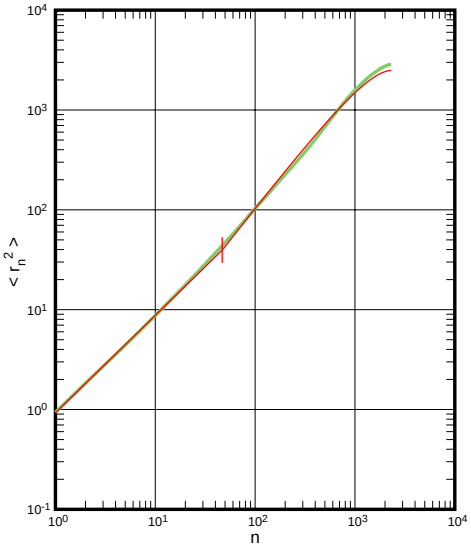

peter



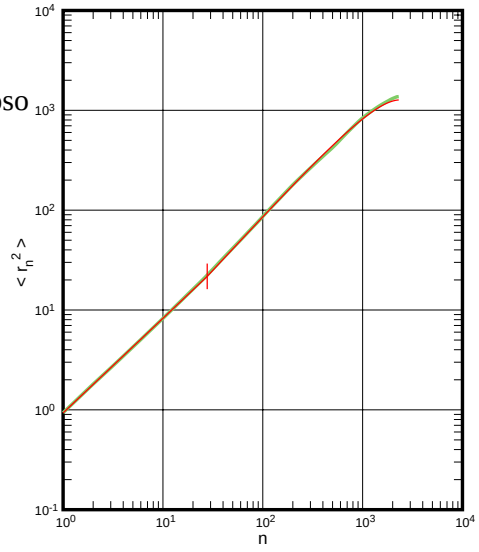

revere



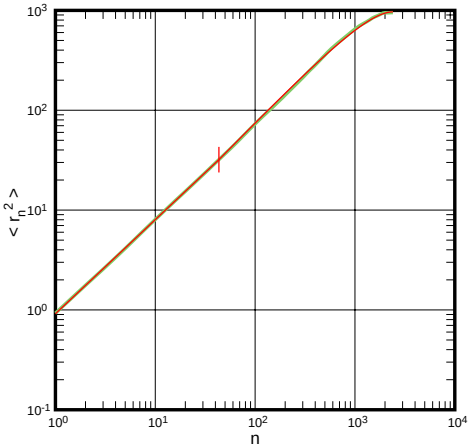

slavery




penseroso

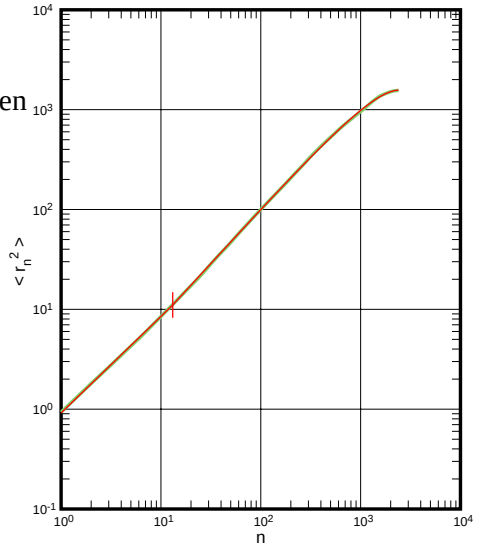



monroe

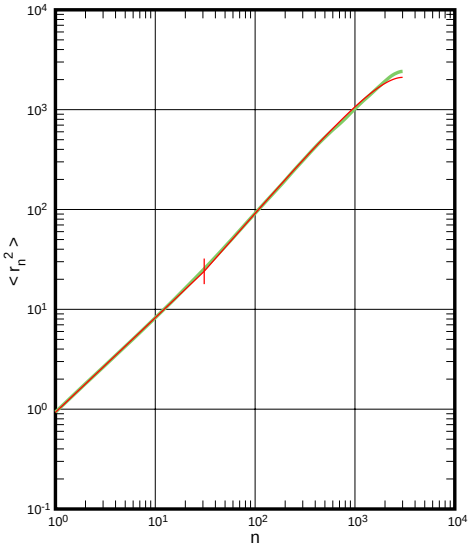


English texts: 4890 < N < 8500

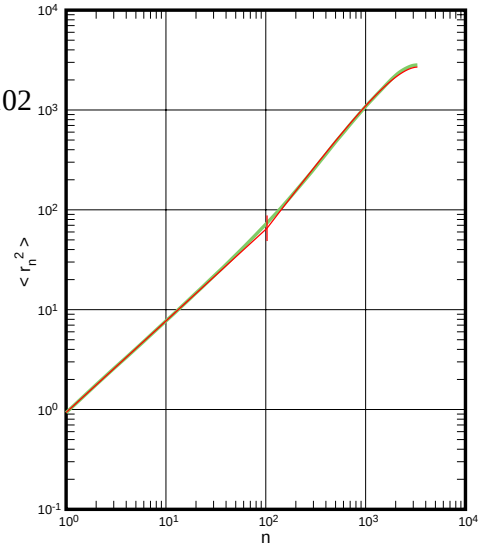
raven



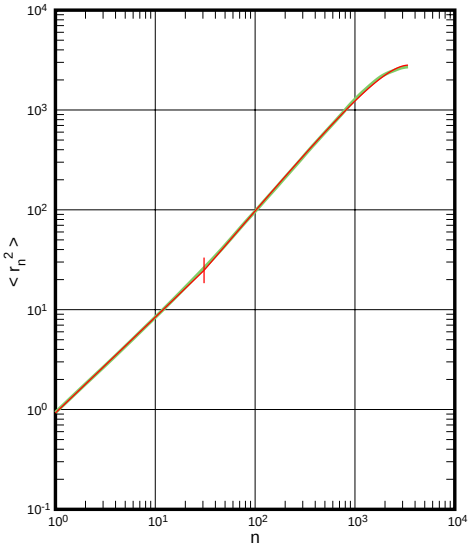

jfk_inaug



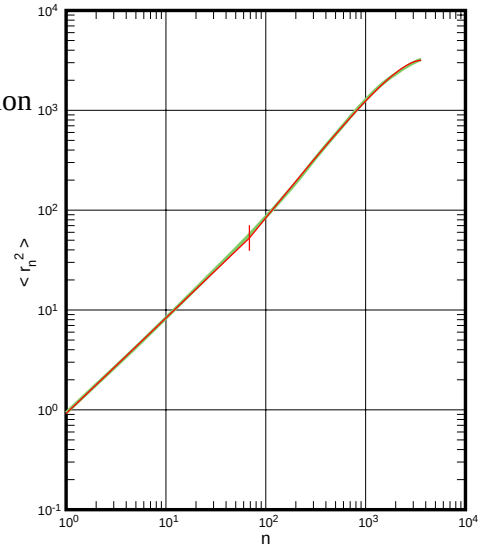

un102



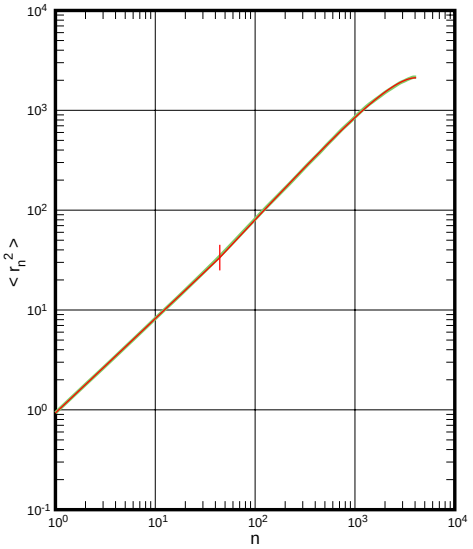

dream_king




declaration

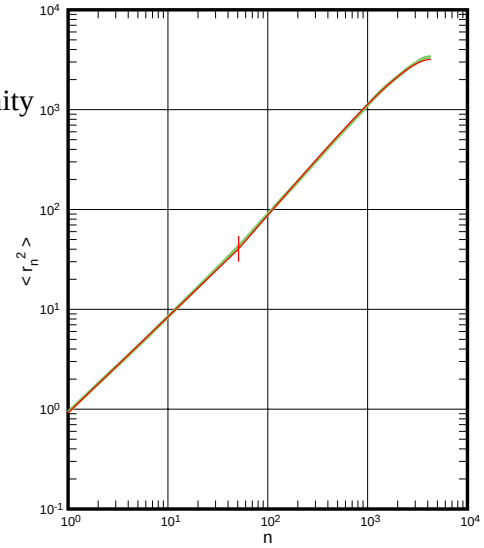


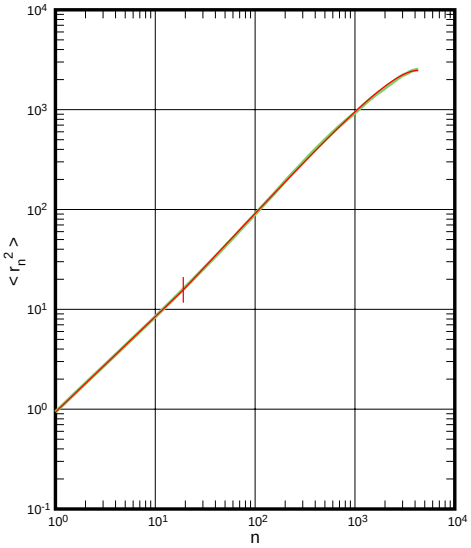

tj_inaug



English texts: 8600 < N < 13,000

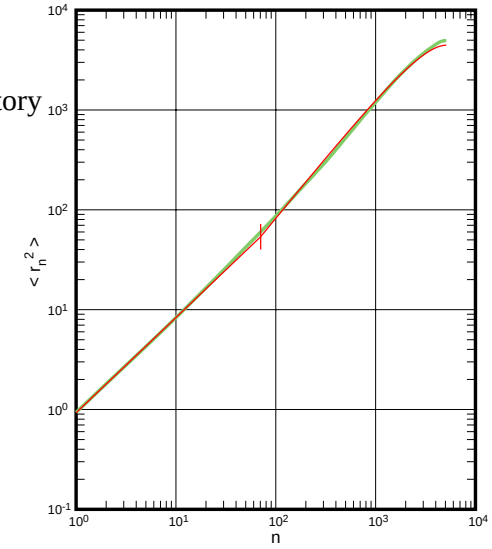
calamity

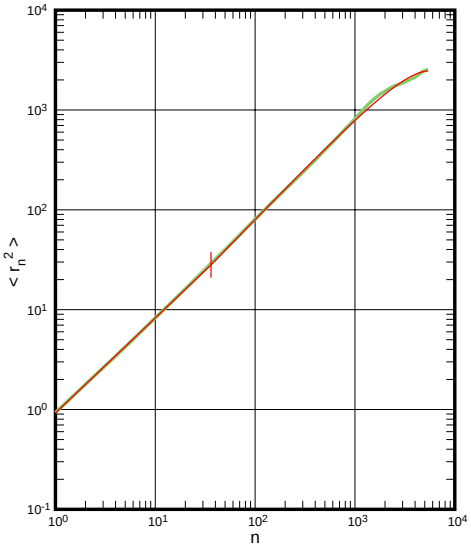



gift_magi

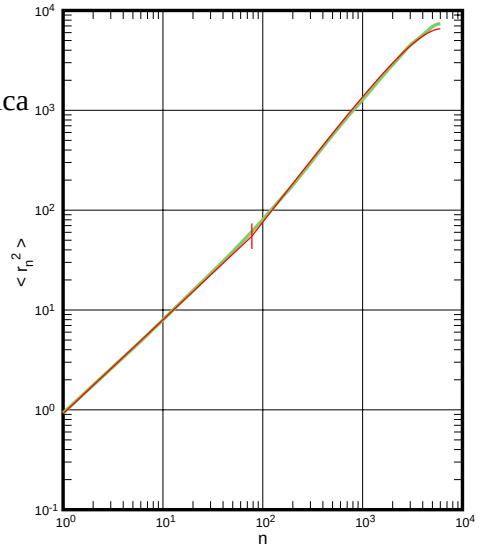
ghost_story

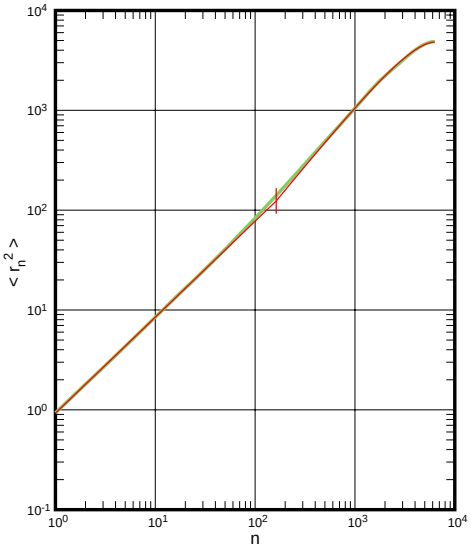



red_death

antarctica

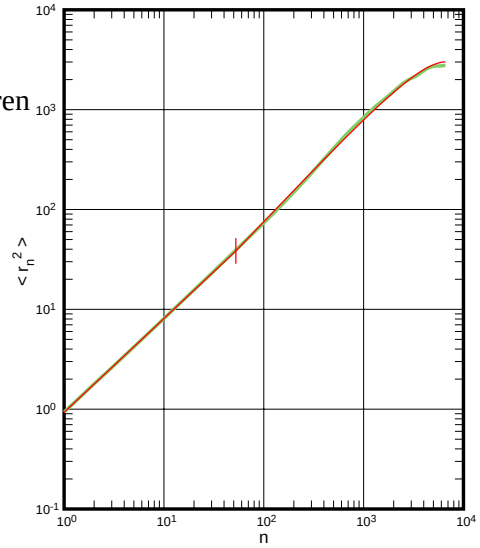


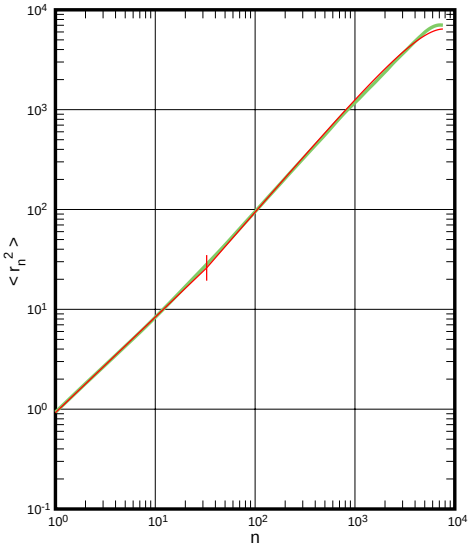

jersey

English texts: 13,000 < N < 20,000

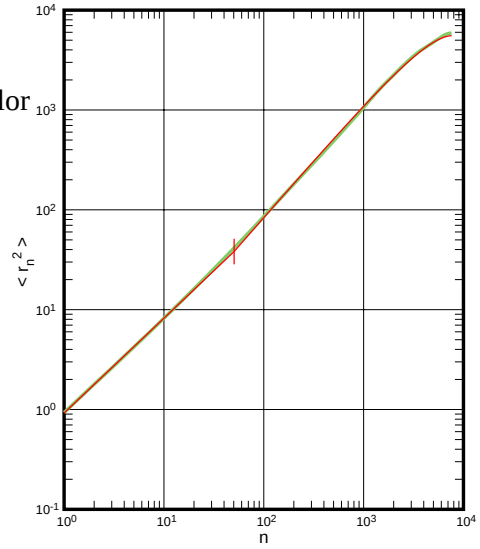
poor_children



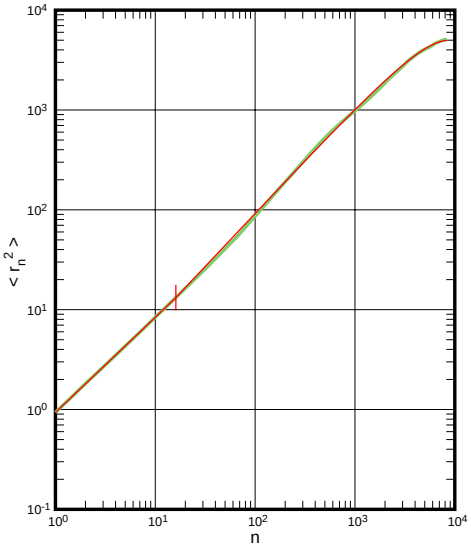

rime



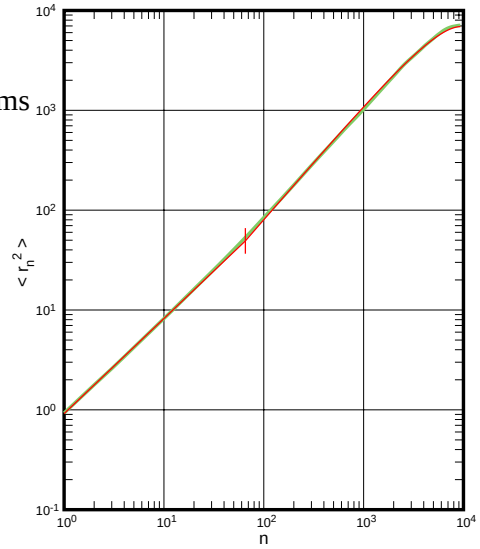

taylor



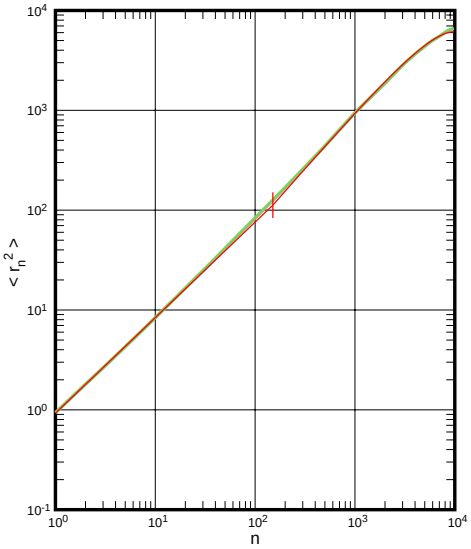

isabella




dreams

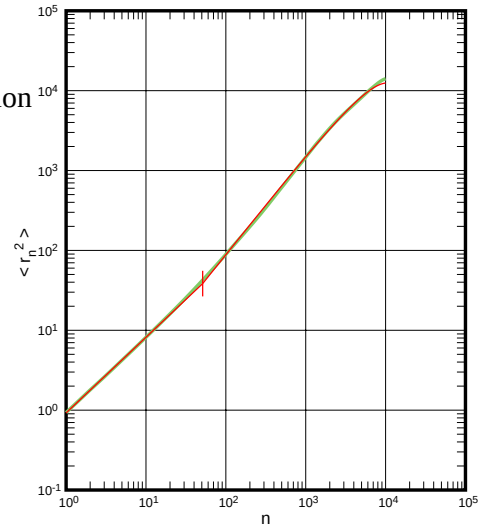



fishing

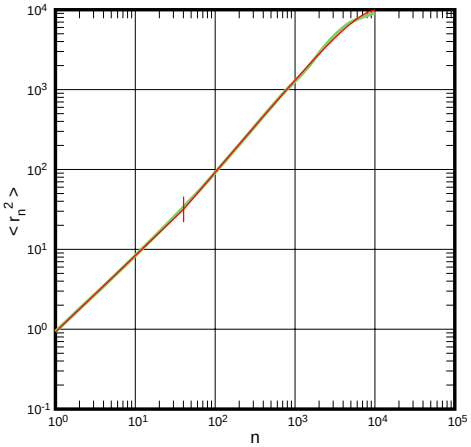


English texts: 20,000 < N < 30,000

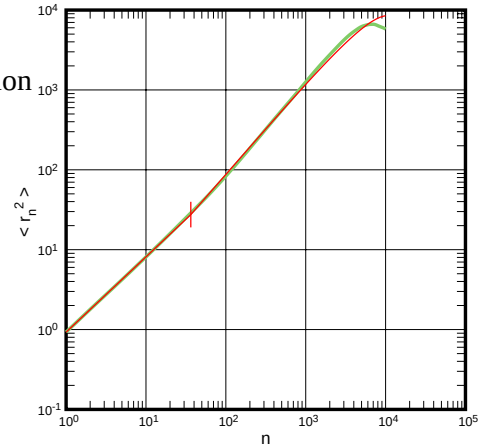
ion



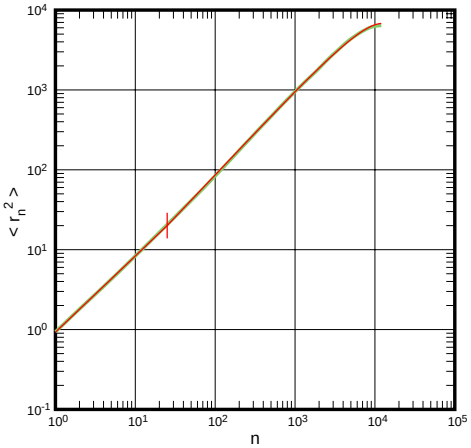

snark



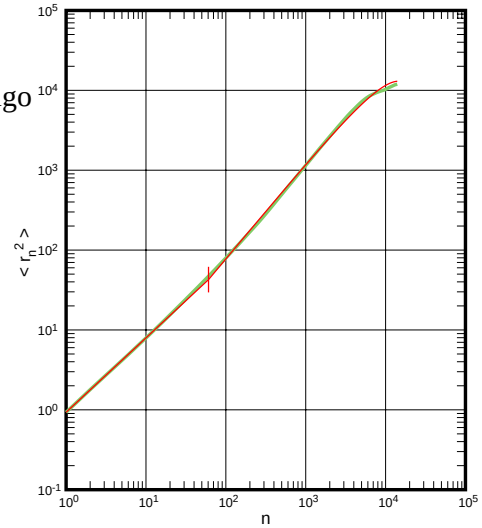

constitution



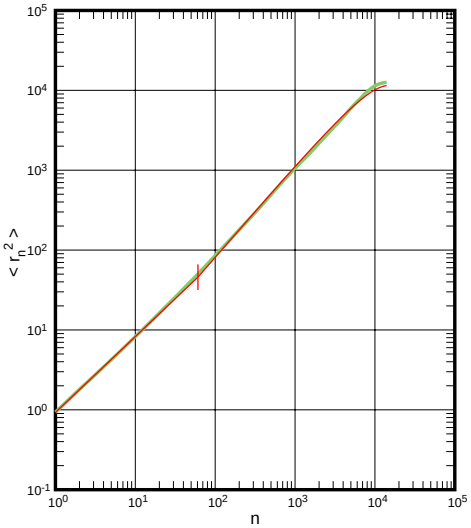

sussexvamp




hidalgo

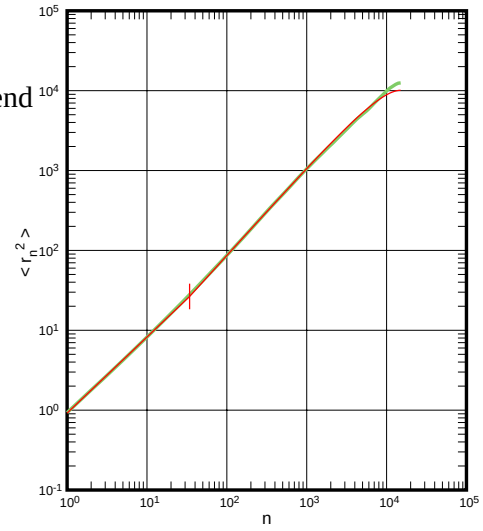


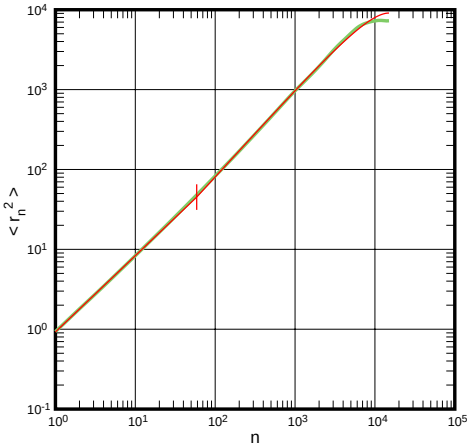

hyperion



English texts: 30,000 < N < 33,000

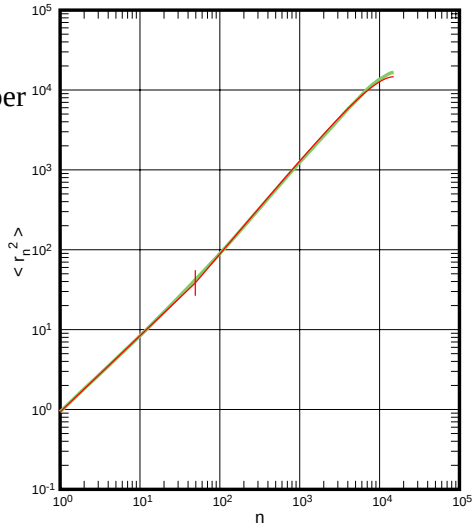
transcend

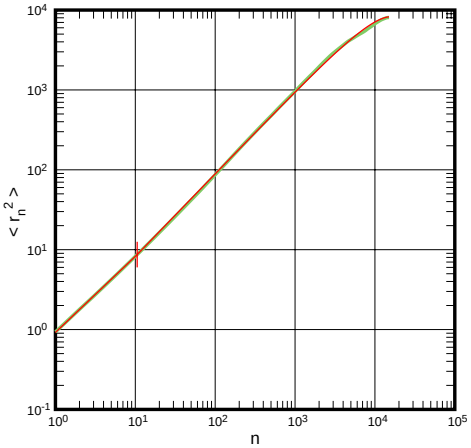



maelstrom

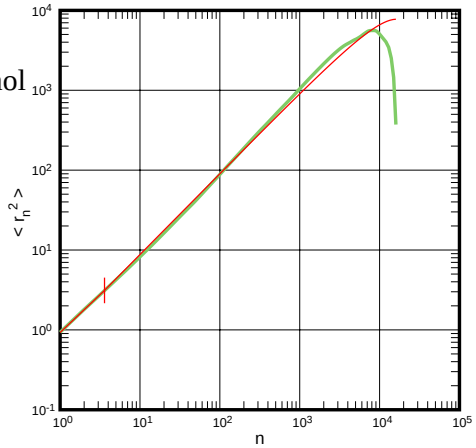
piper

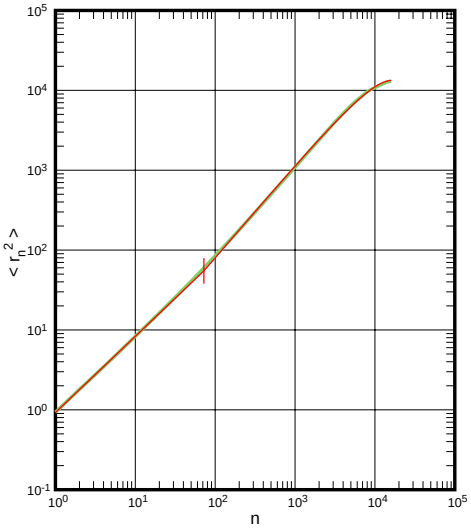



literary_ethics

gaol

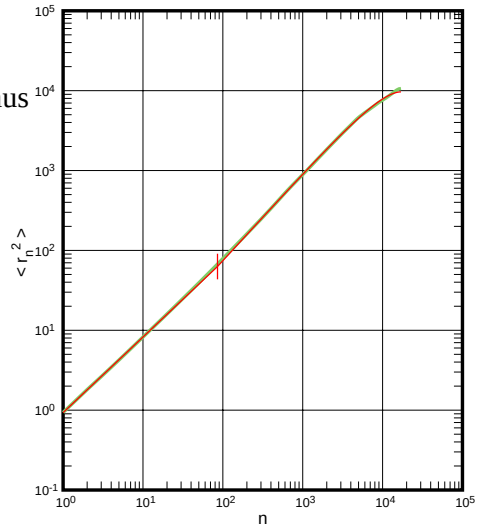


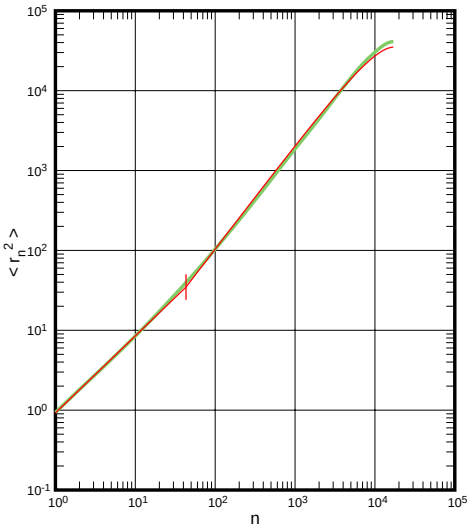

carfax

English texts: 33,000 < N < 40,000

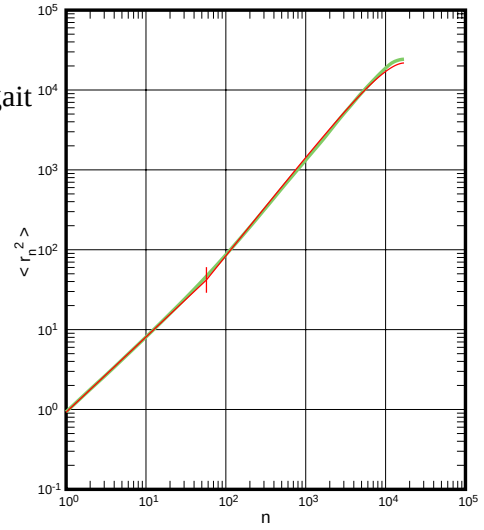
comus



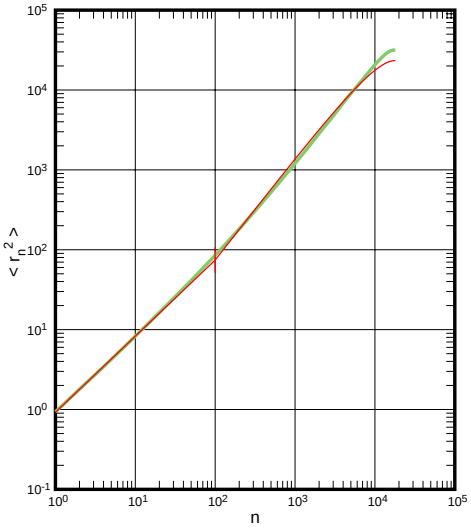

nursery



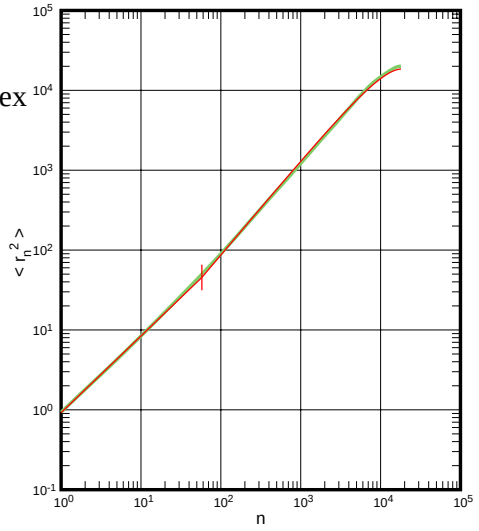

animal_gait



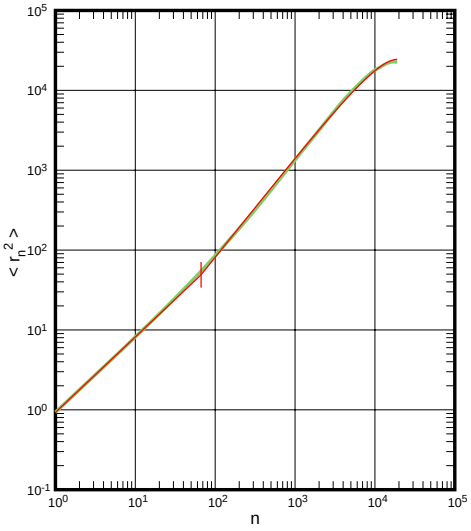

escape




vortex

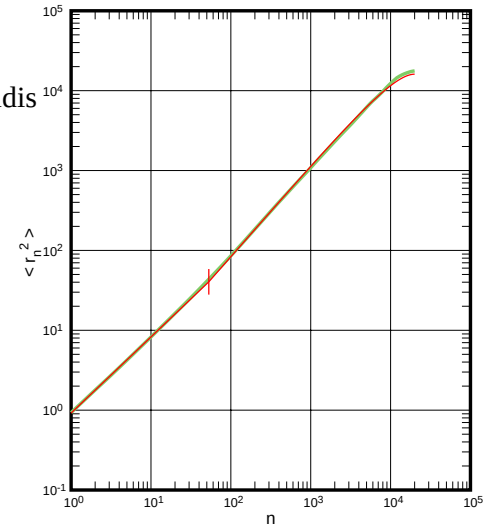



adductor

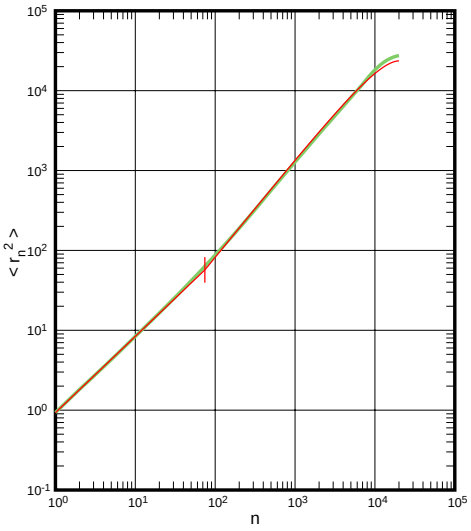


English texts: 40,000 < N < 50,000

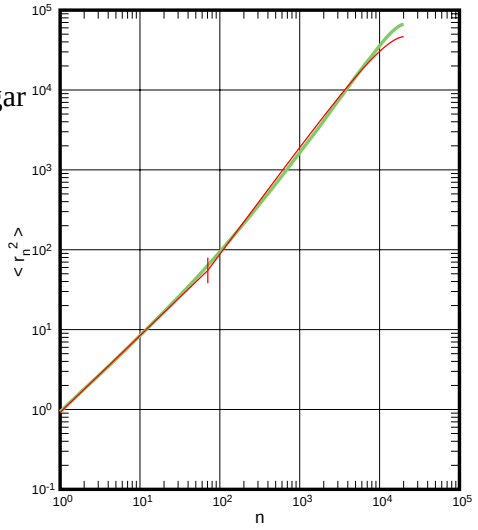
civildis



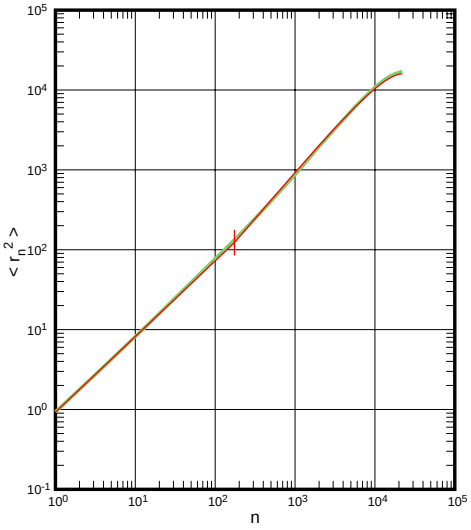

weevil



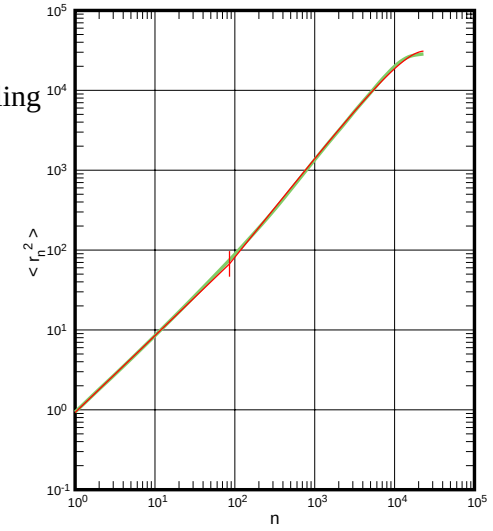

sugar



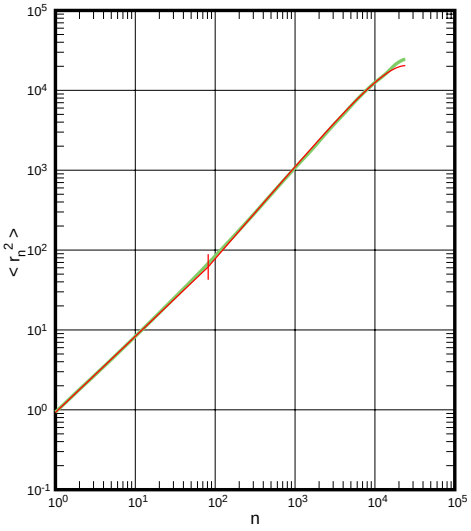

su1840




dumpling

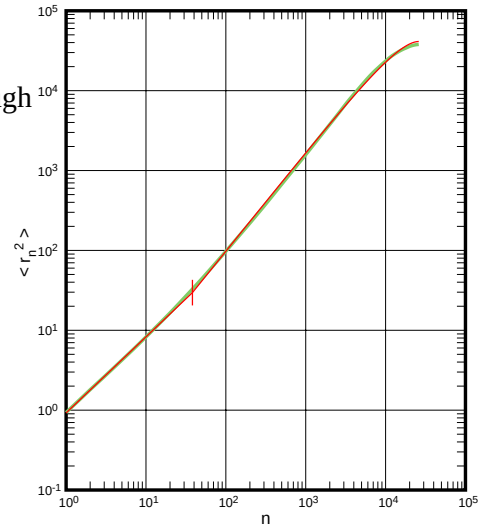



wisteria

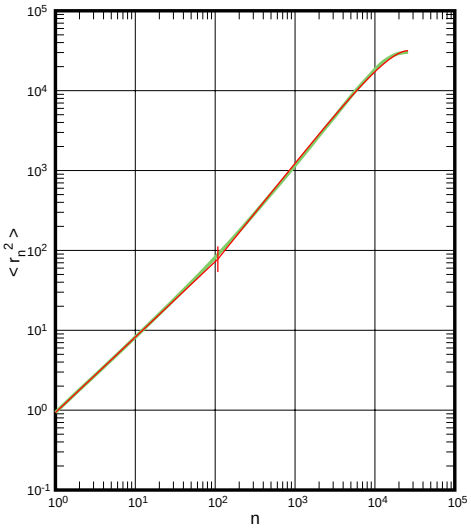


English texts: 50,000 < N < 58,000

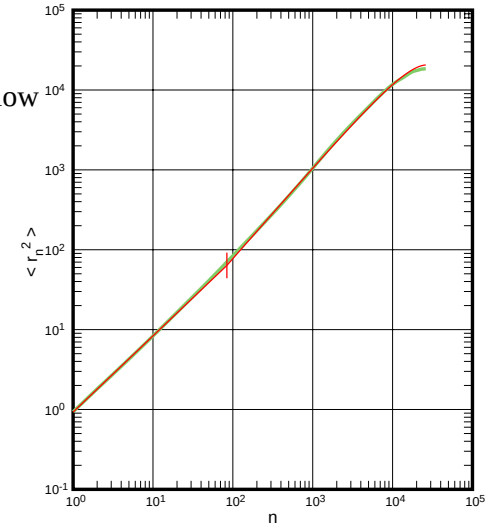
sourdough



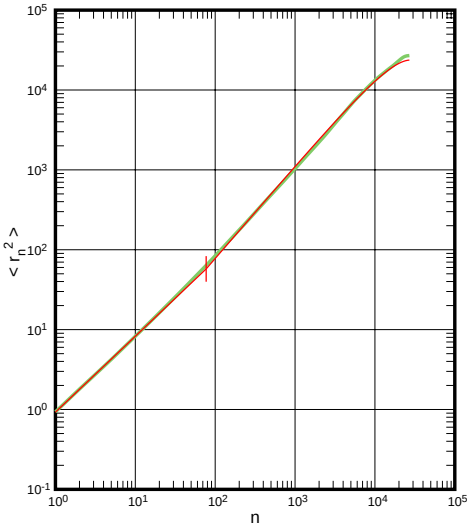

argon



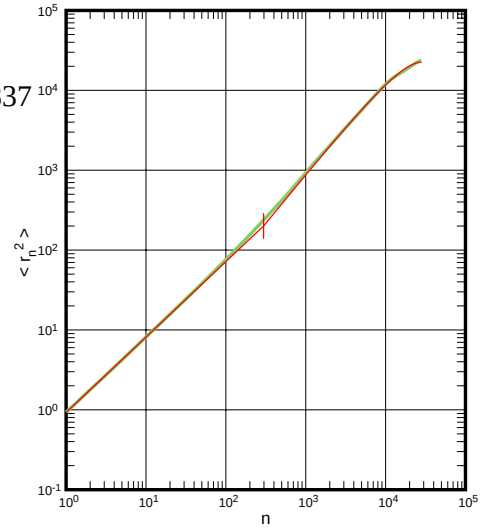

sleepy_hollow



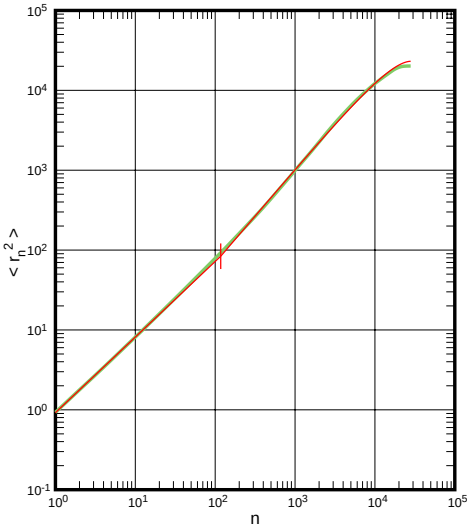

attic




su1837

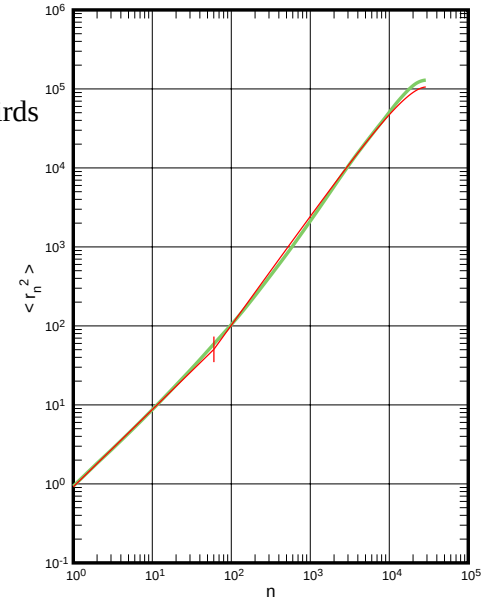



su1838

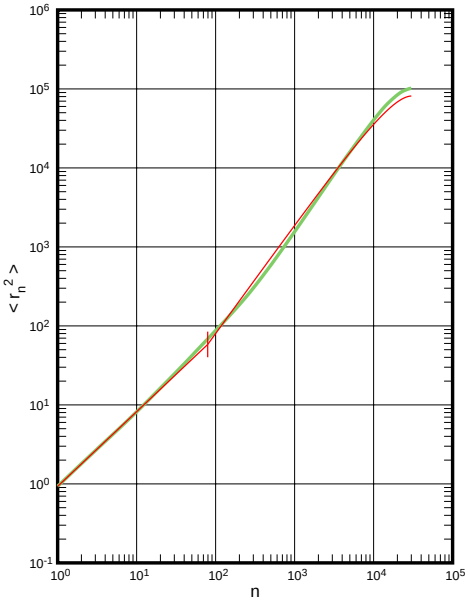


English texts: 60,000 < N < 75,000

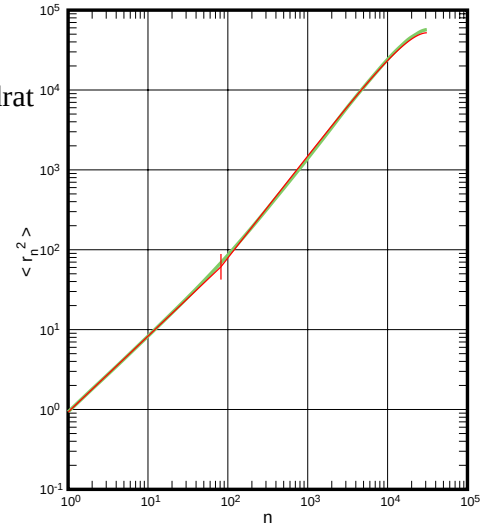
bc.birds



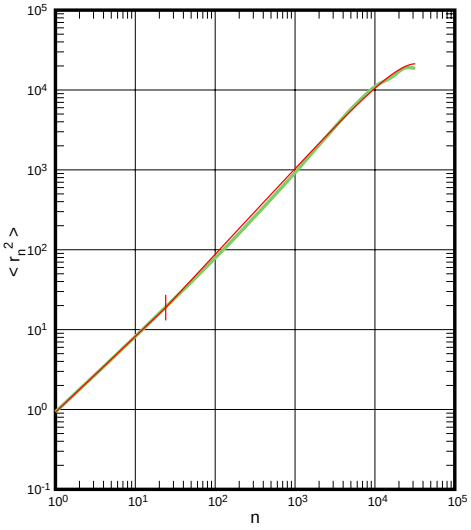

stamp



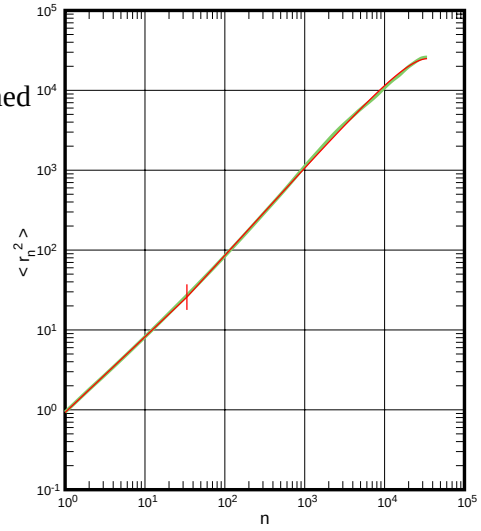

woodrat



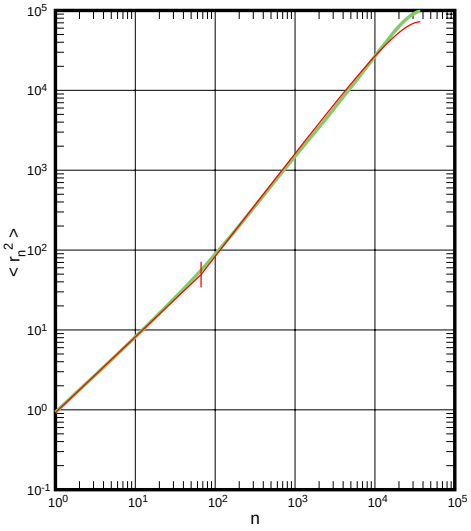

su1839




regained

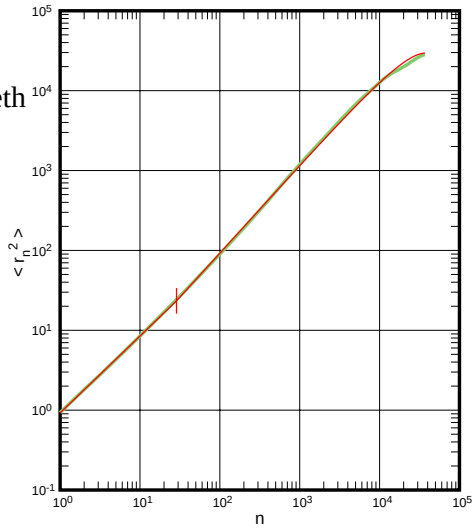



stellar

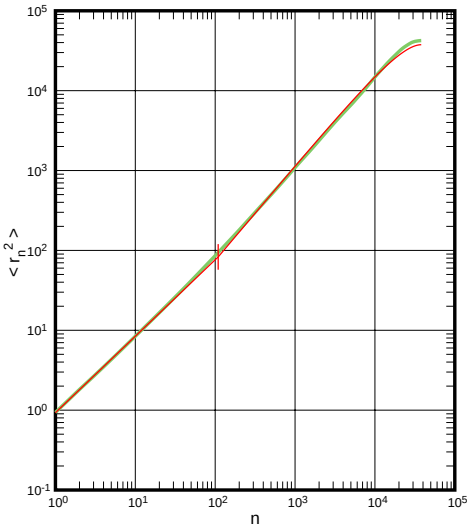


English texts: 75,000 < N < 88,000

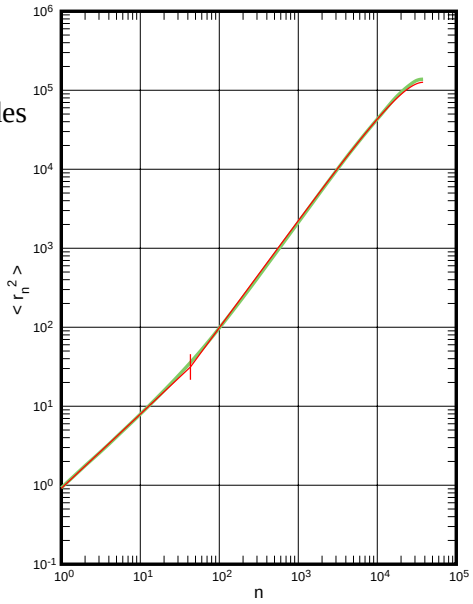
macbeth



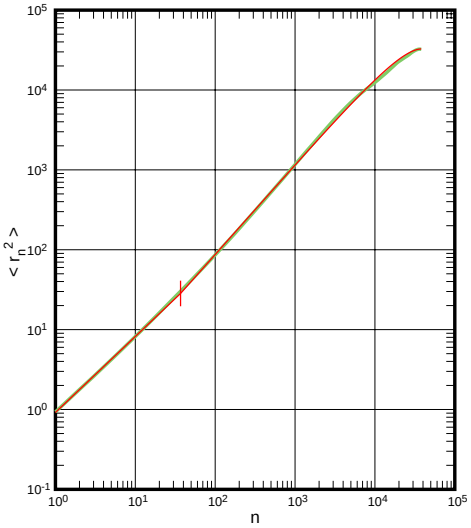

cleveland



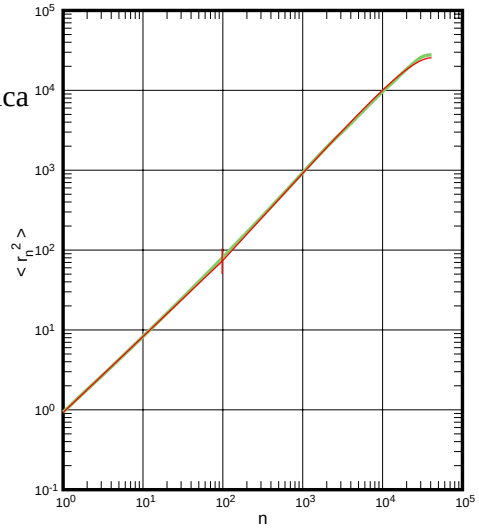

parmenides



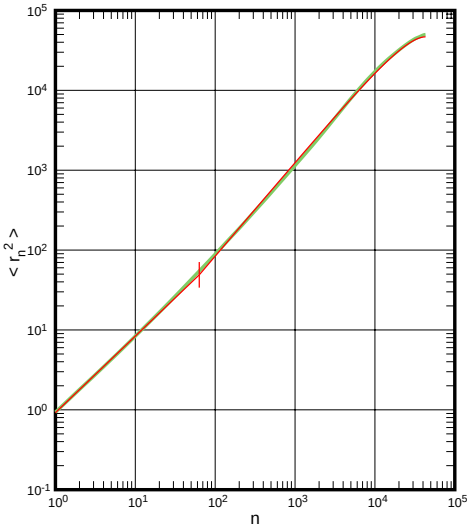

theism




areopagitica

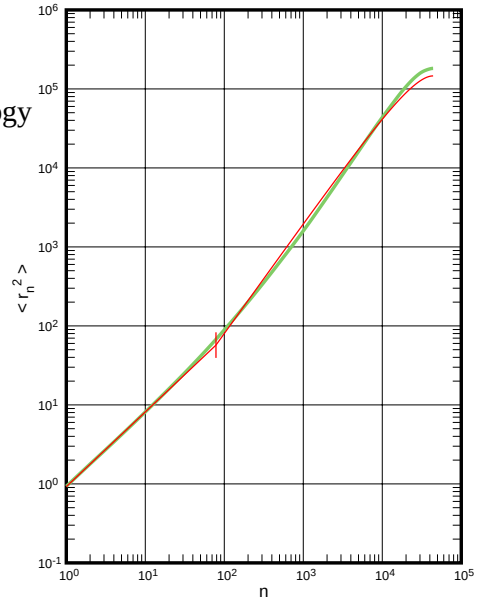


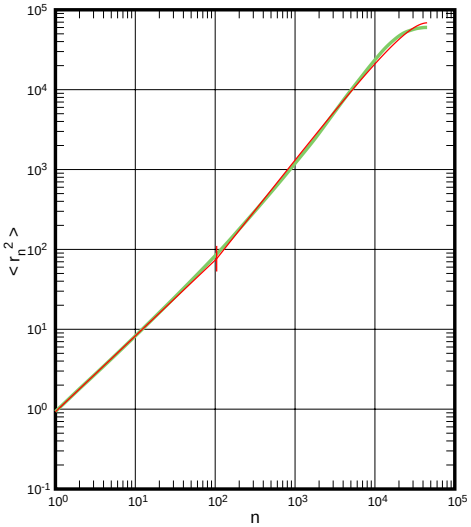

plague



English texts: 88,000 < N < 110,000

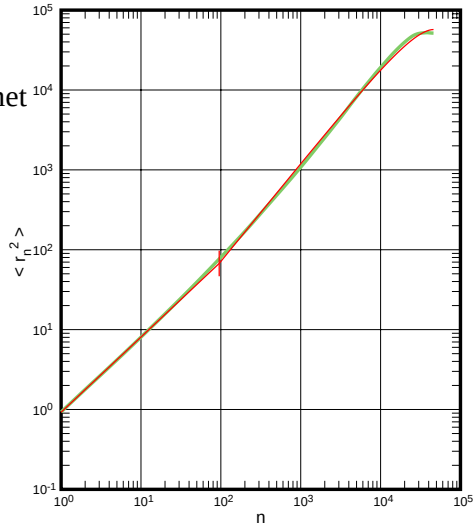
mythology

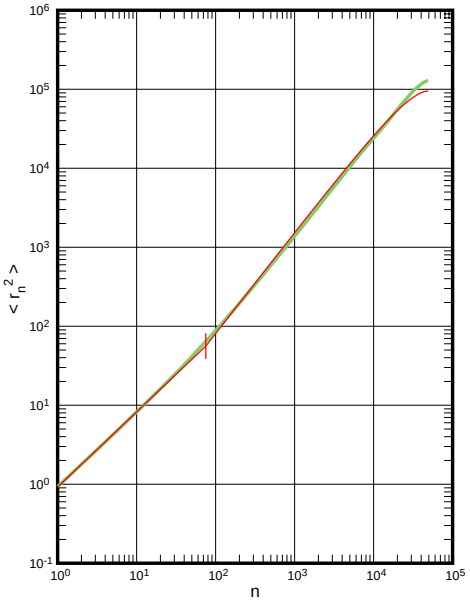



mammoth

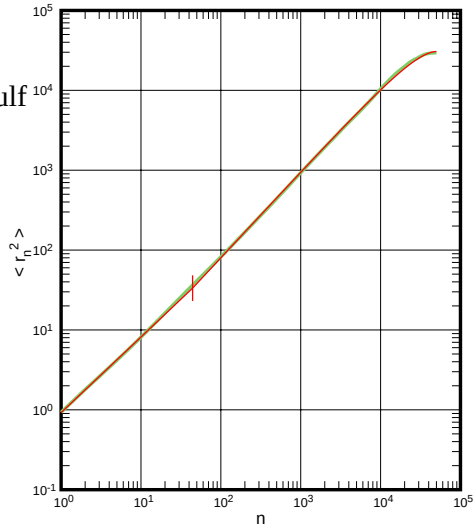
comet

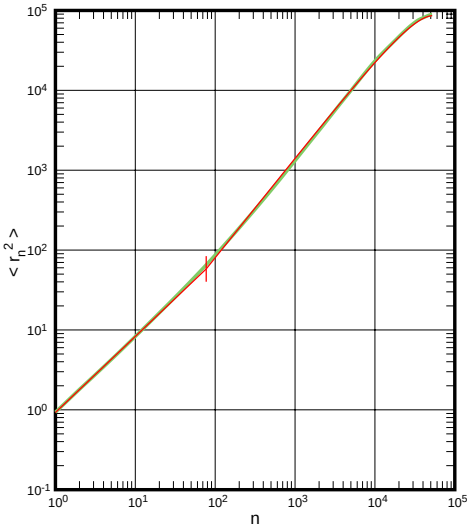



cratylus

beowulf

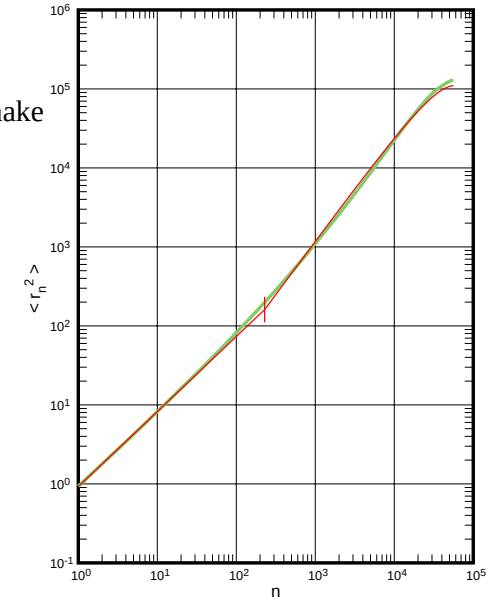


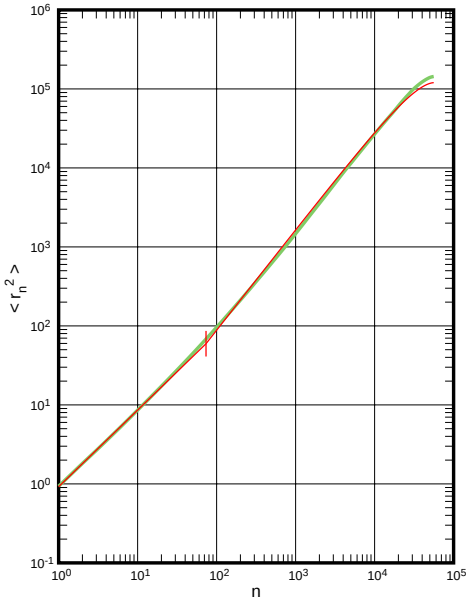

alice

English texts: 110,000 < N < 128,000

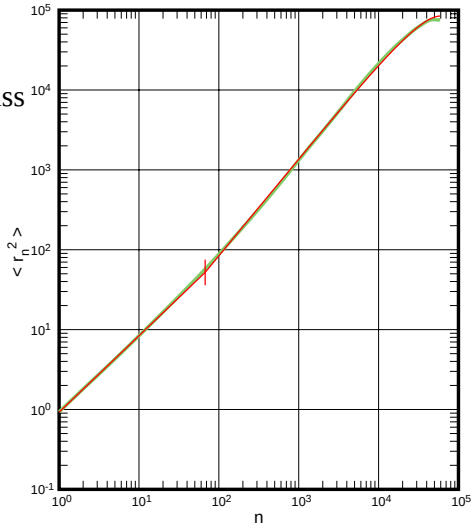
snake



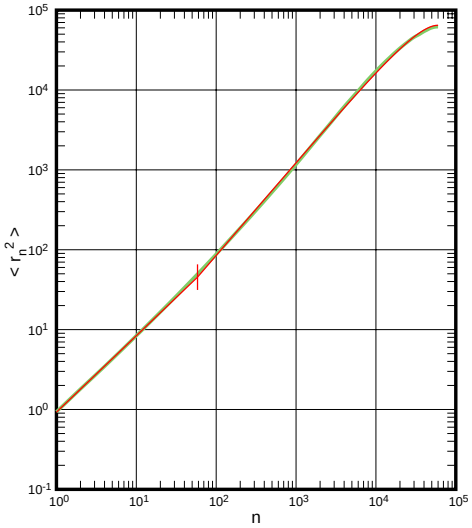

othello



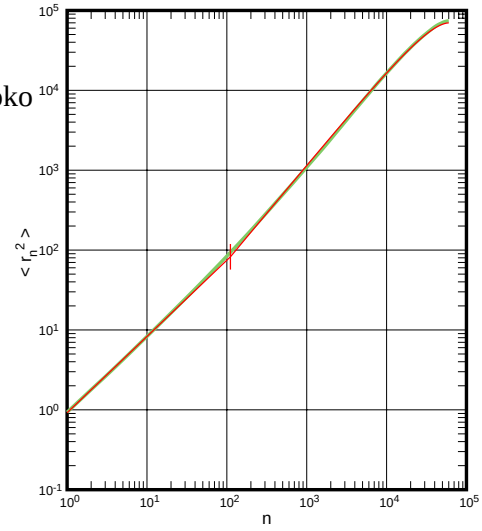

glass



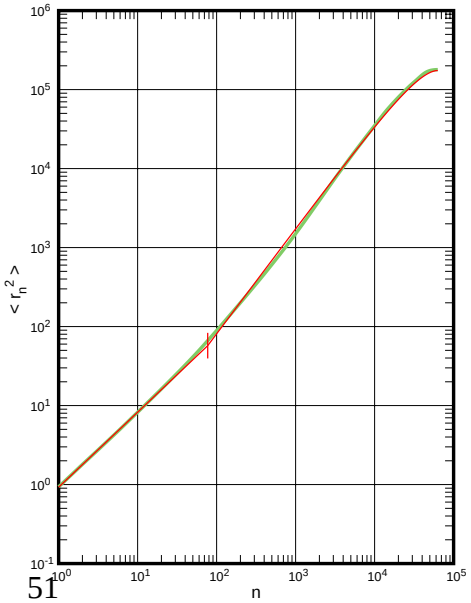

christmas




oroonoko

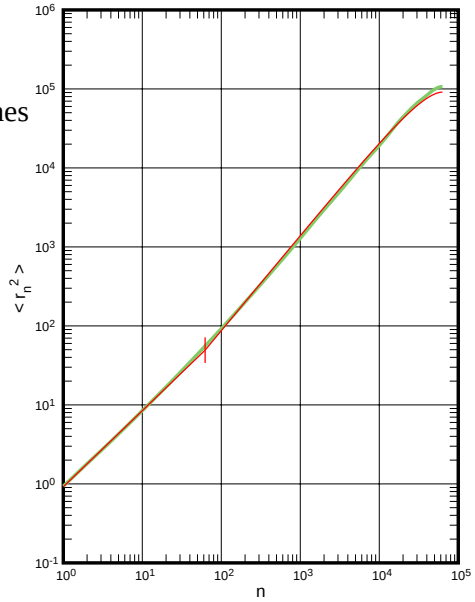



woods.hole

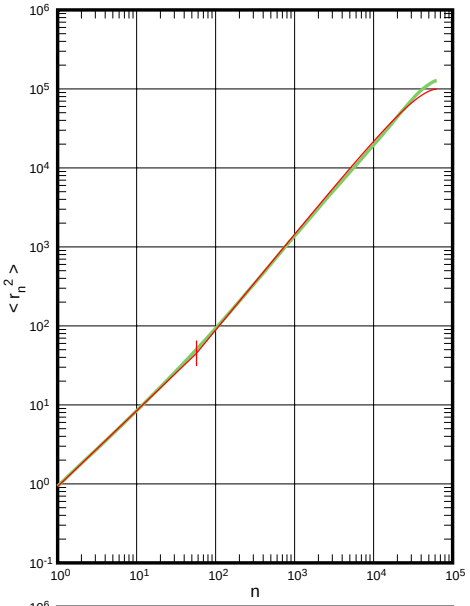


English texts: 128,000 < N < 141,000

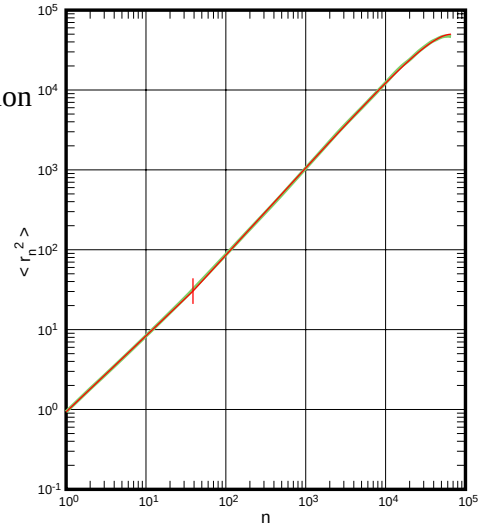
chimes



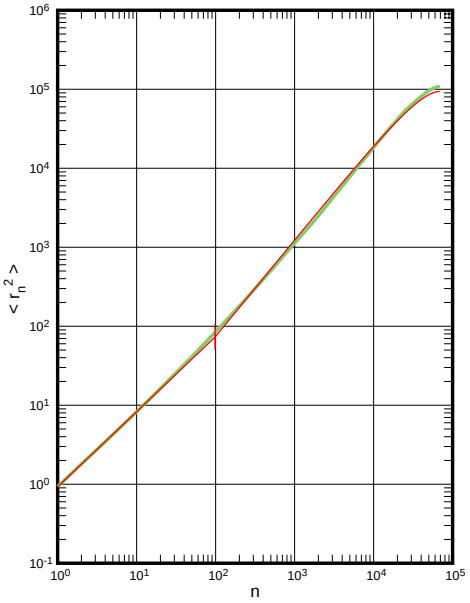

hamlet



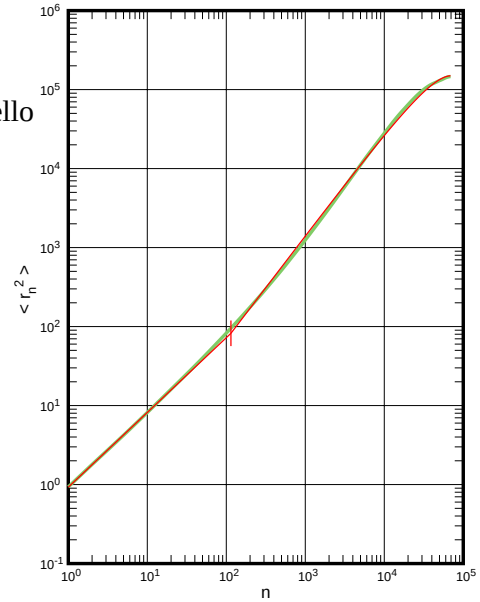

endymion



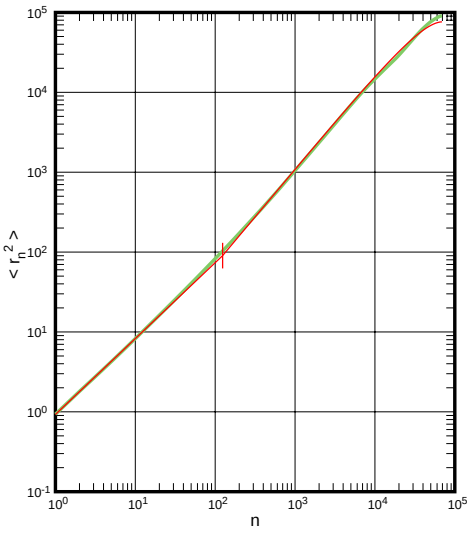

wild




cello

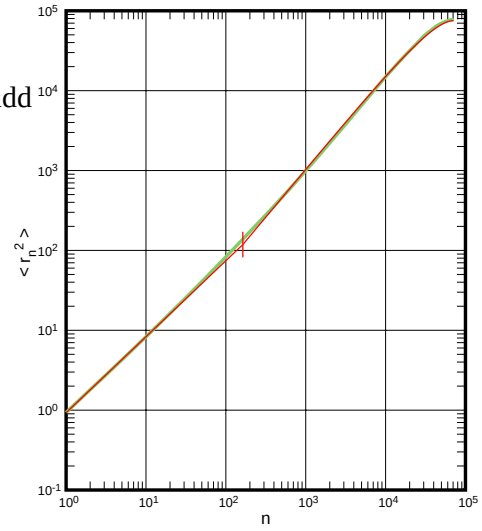



time_machine

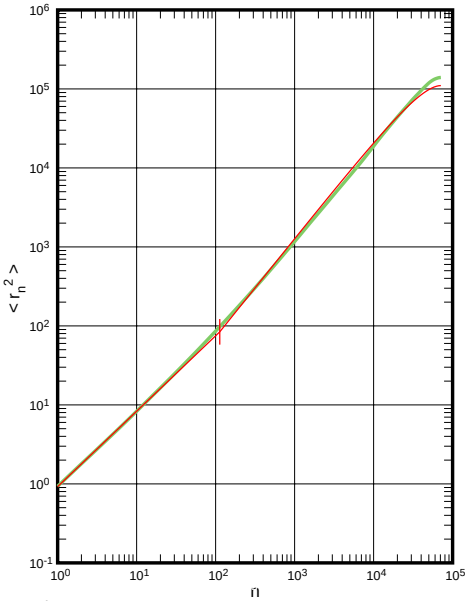


English texts: 143,000 < N < 160,000

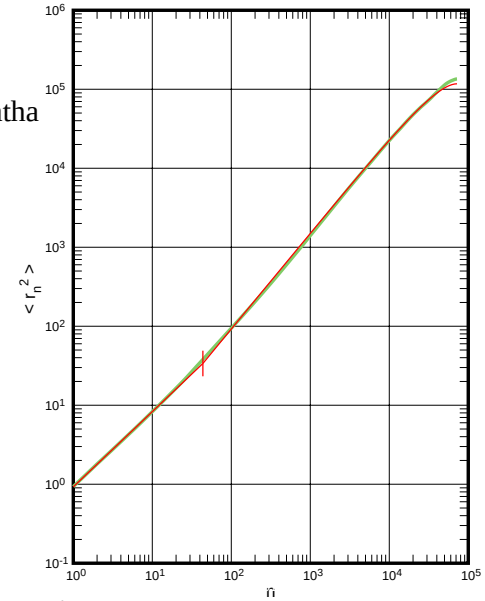
billy_budd



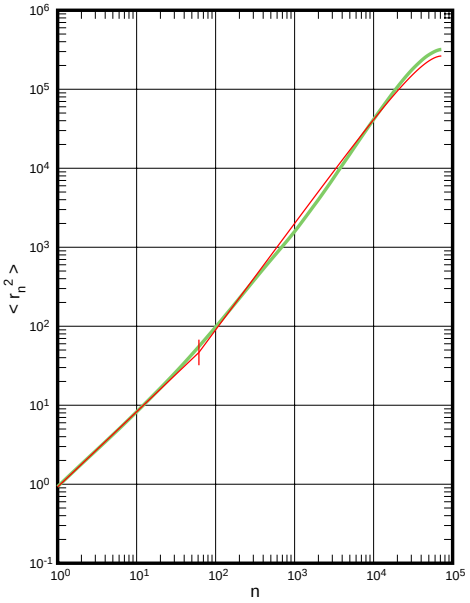

radioUKG



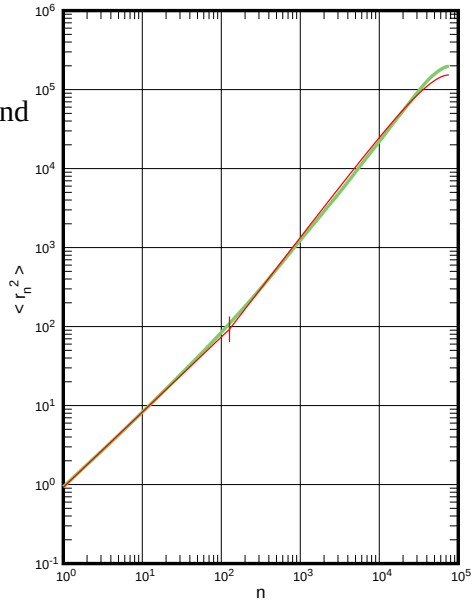

hiawatha



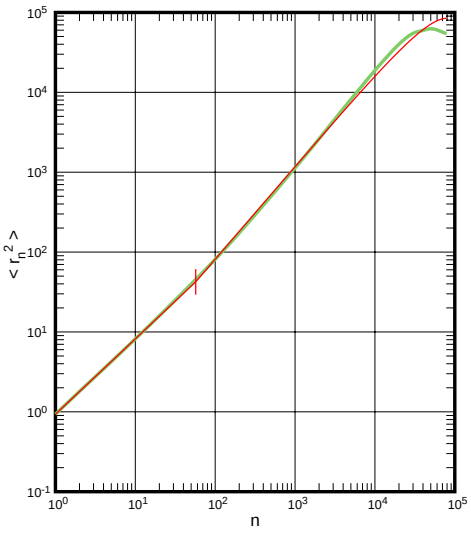

pinus




flatland

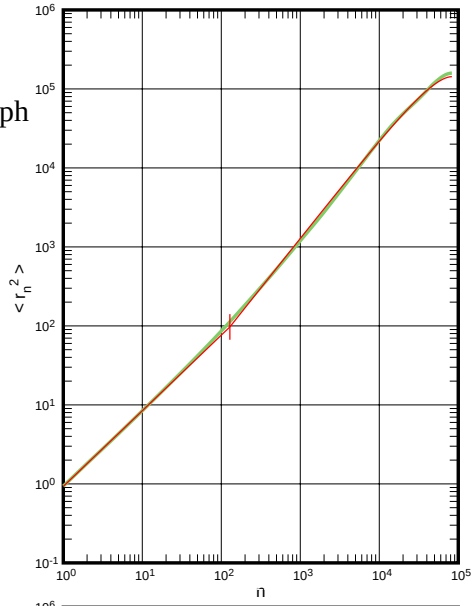



oz

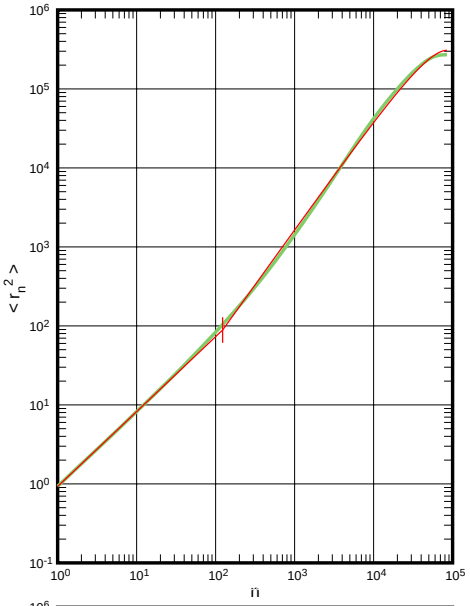


English texts: 165,000 < N < 180,000

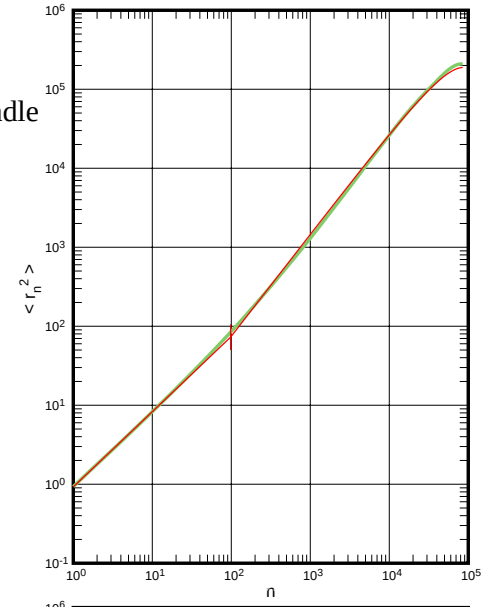
telegraph



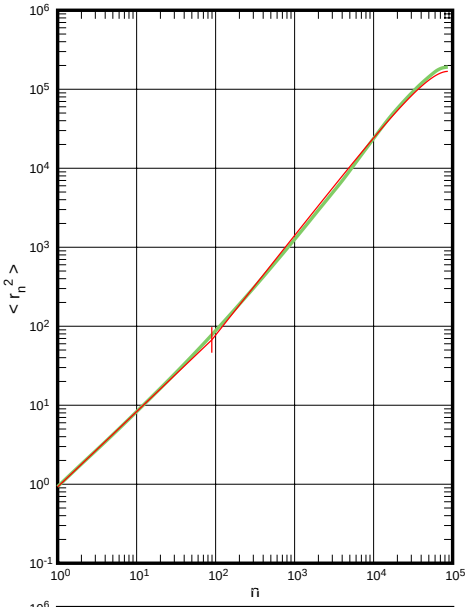

utah



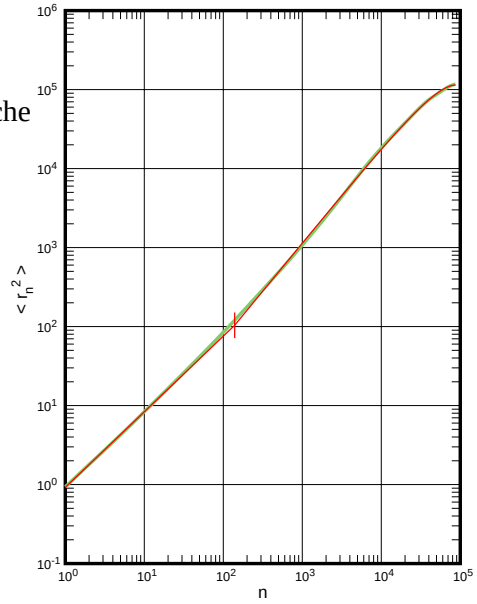

candle



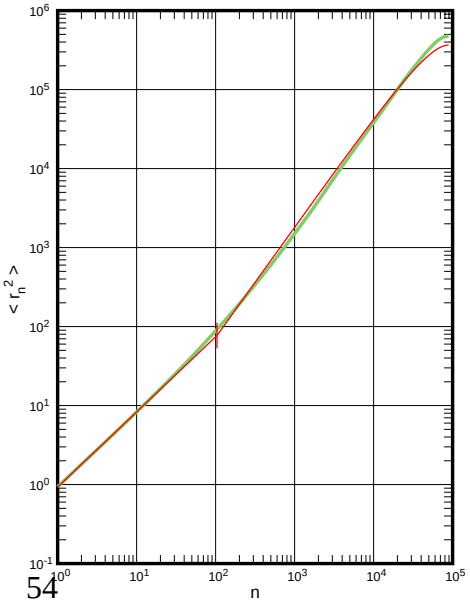

fd_auto




avalanche

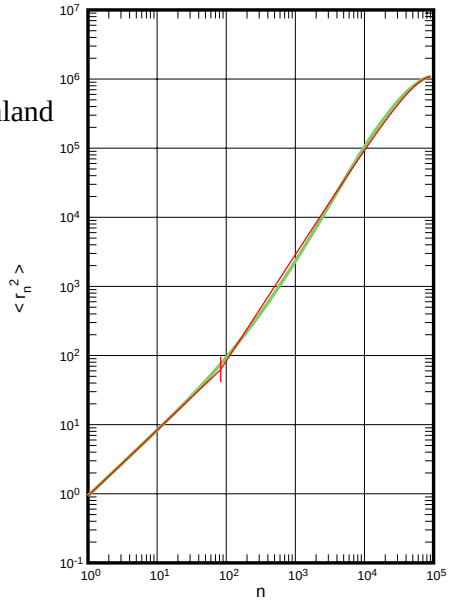



taxidermy

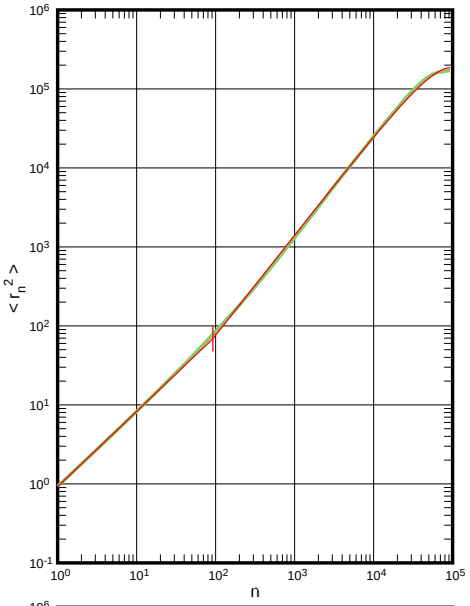


English texts: 185,000 < N < 205,000

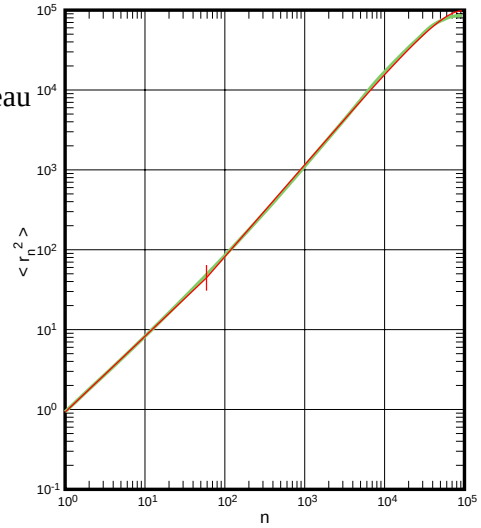
highland



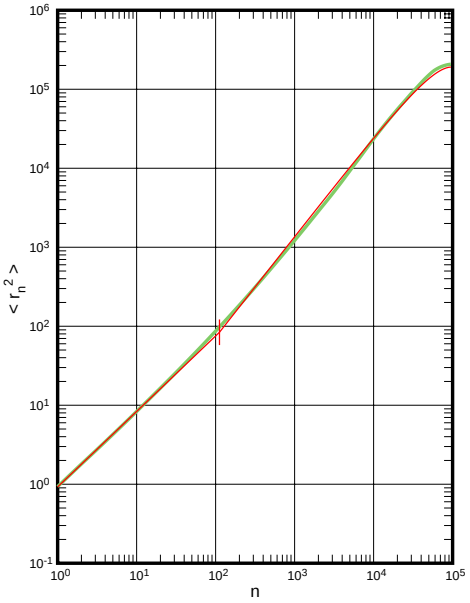

tj_auto



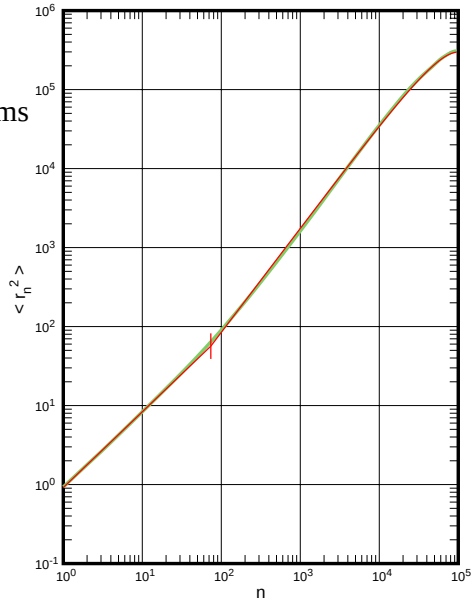

DrMoreau



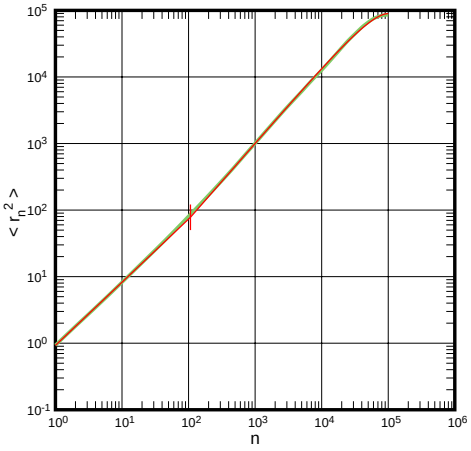

peter_pan




problems

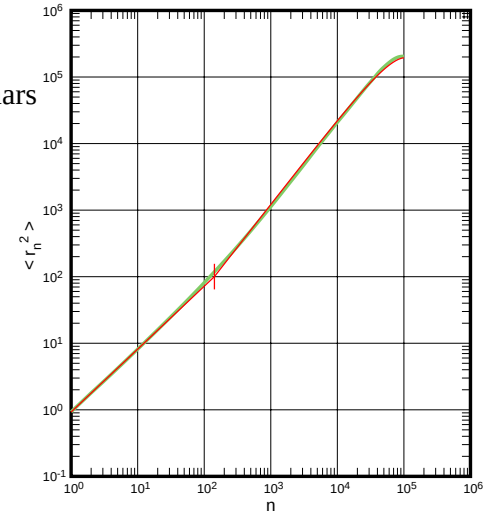



red_badge

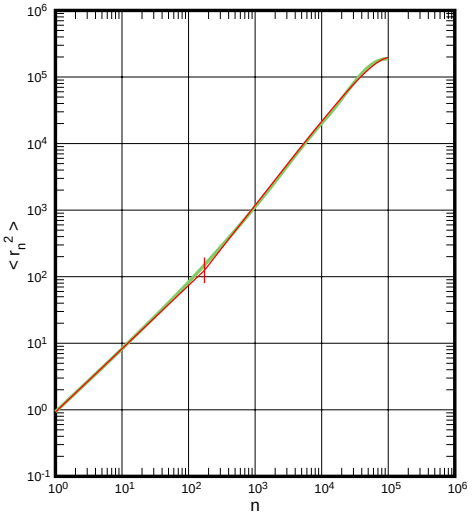


English texts: 209,000 < N < 226,000

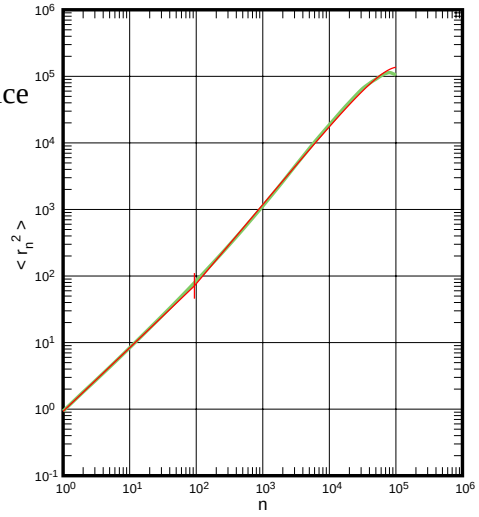
maid_mars



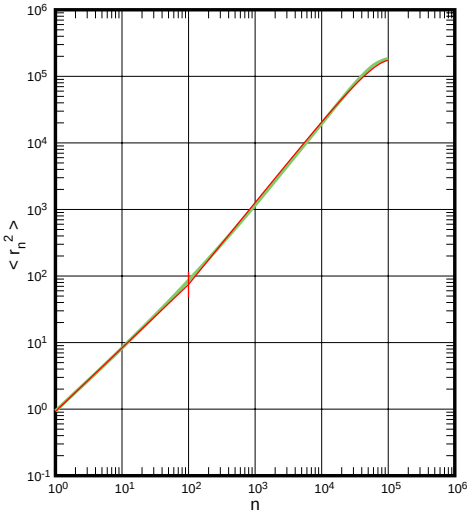

shock



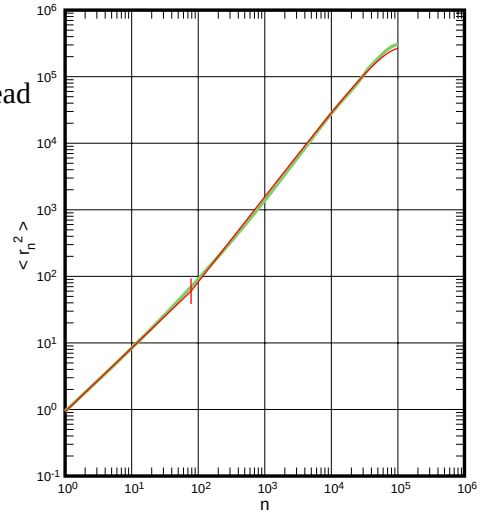

ice



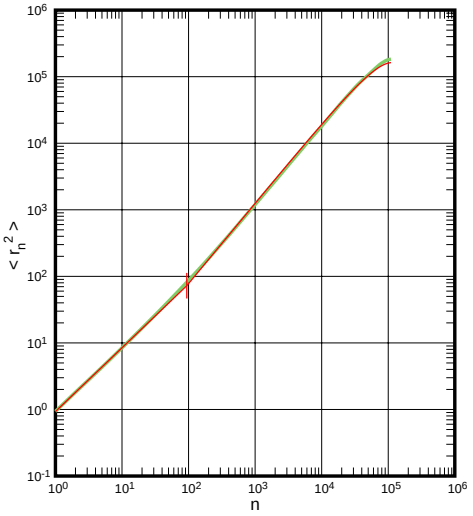

zenda




puddnhead

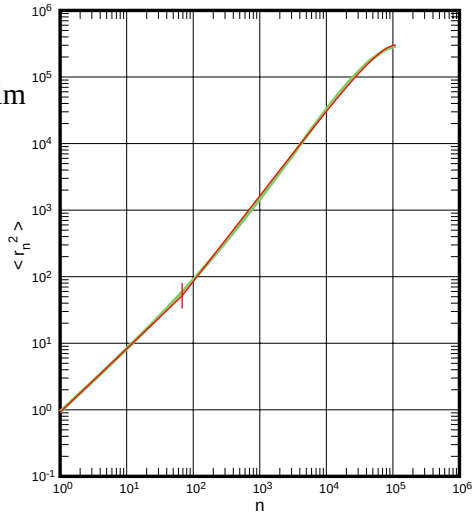



breakers

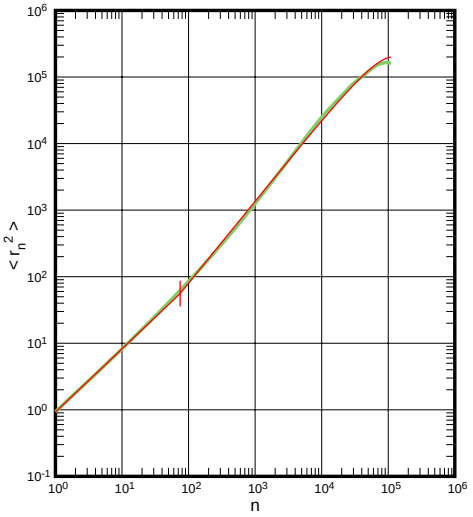


English texts: 227,000 < N < 241,000

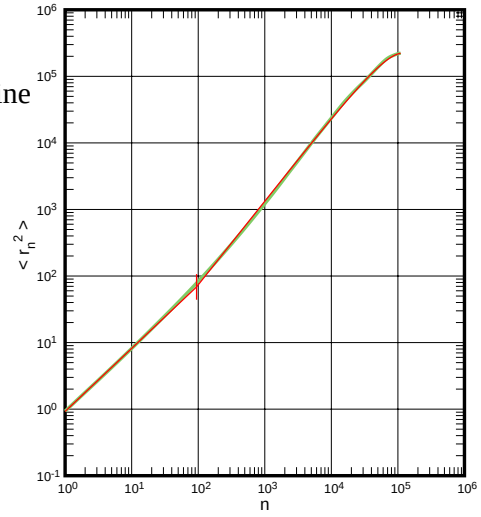
pilgrim



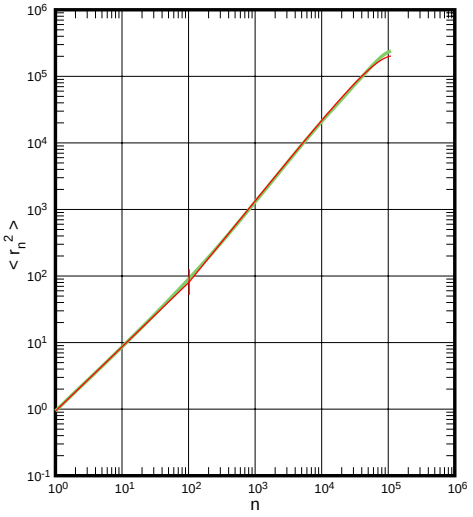

armour



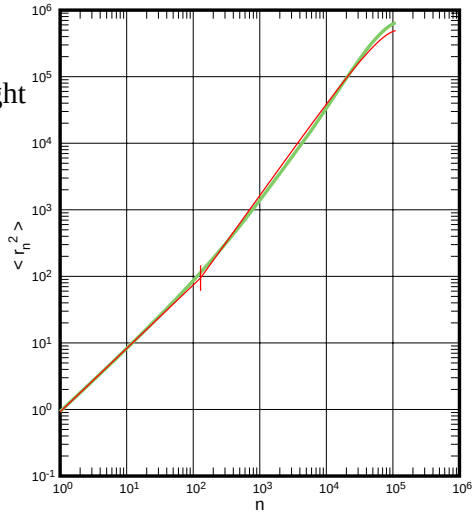

decline



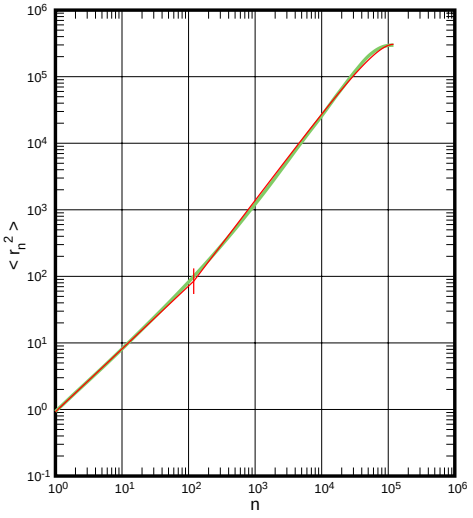

pollyanna




kantright

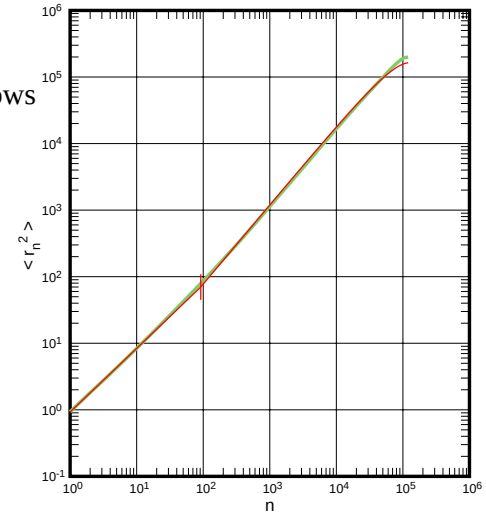



other

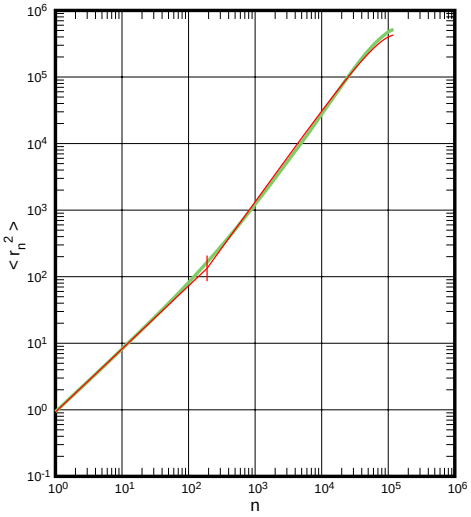


English texts: 250,000 < N < 293,000

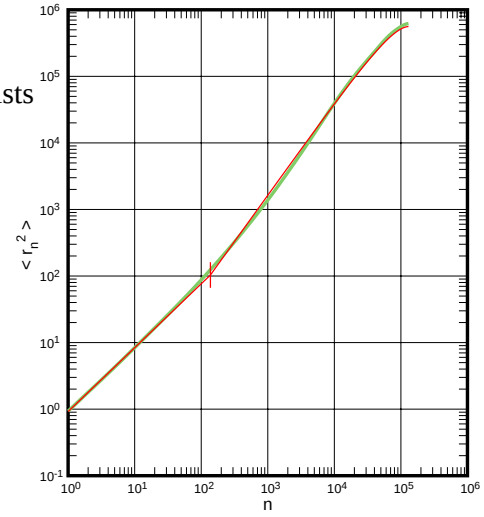
willows



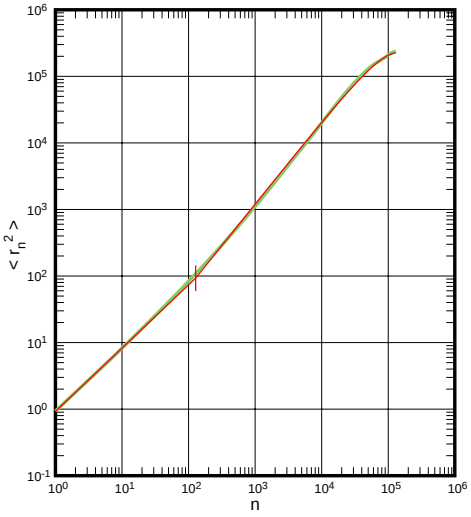

dairy



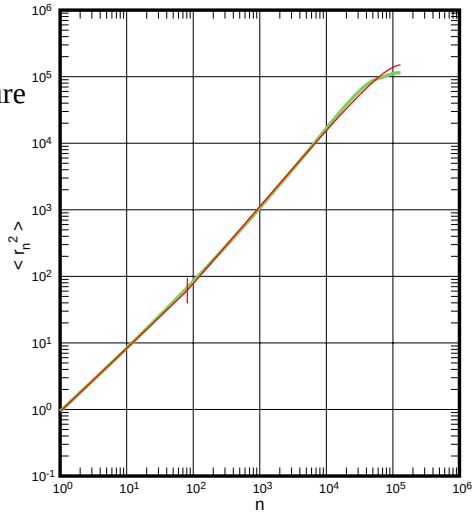

violinists



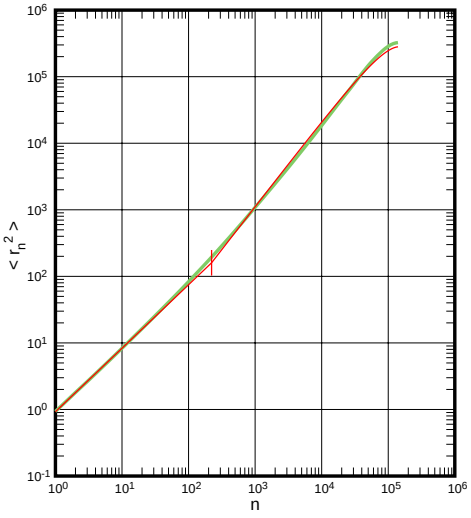

war




treasure

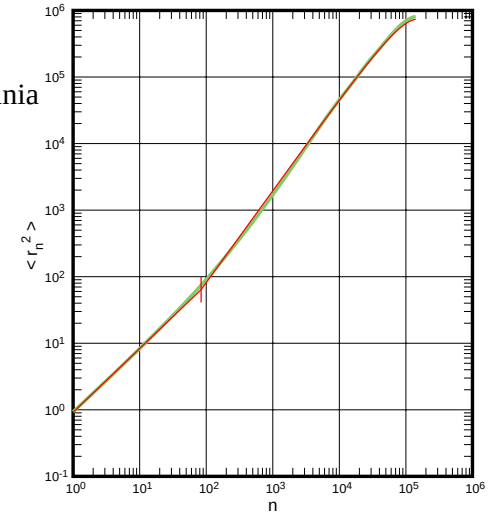



bf_auto

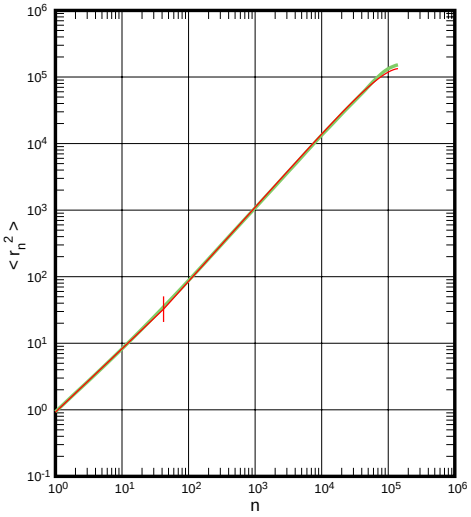


English texts: 293,000 < N < 320,000

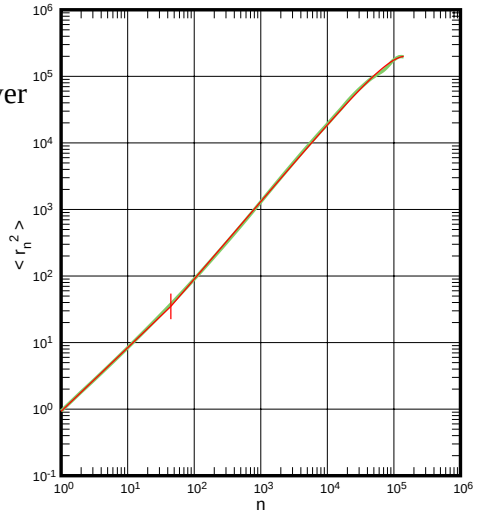
virginia



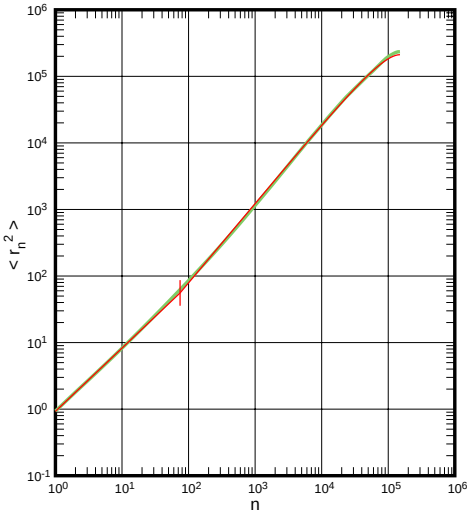

conduct_life



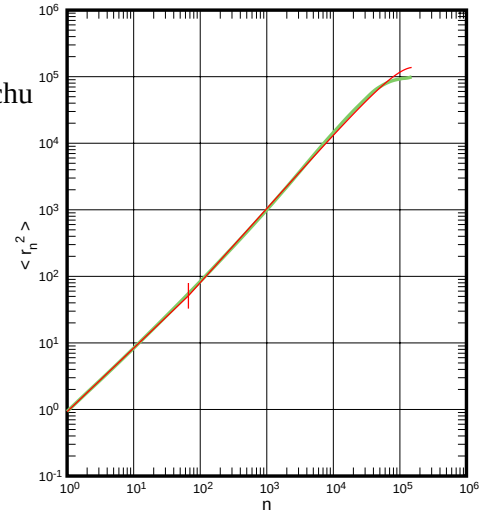

tom_sawyer



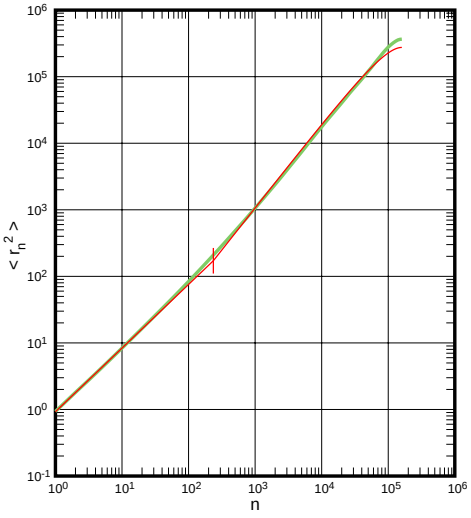

prince_pauper




fu_manchu

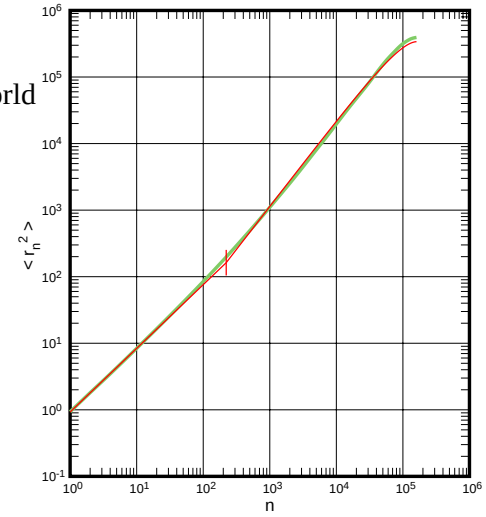



kidnapped

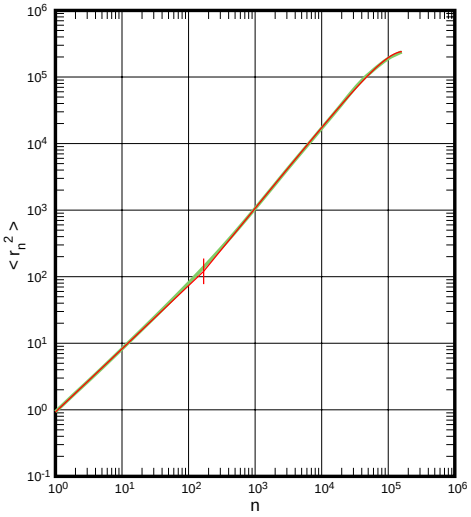


English texts: 328,000 < N < 365,000

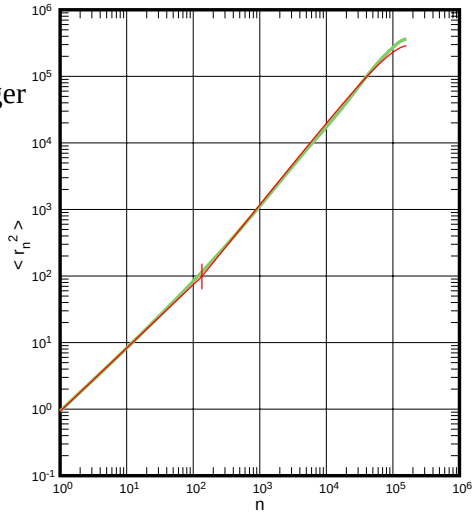
lost_world



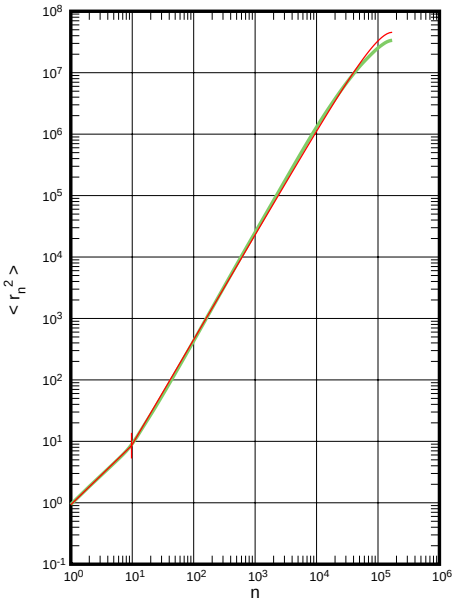

frankenstein



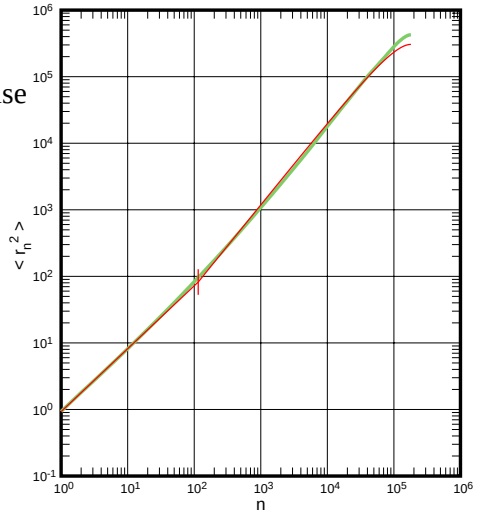

northanger



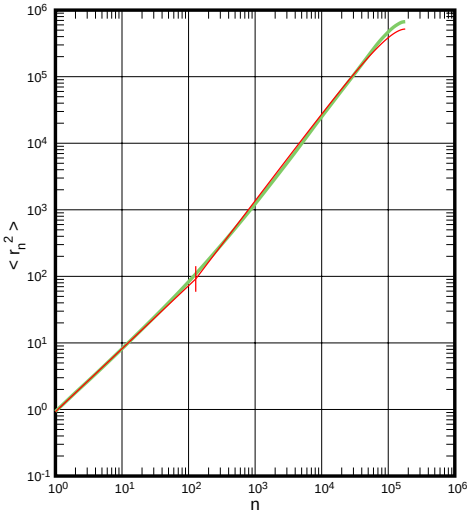

zc_lett




paradise

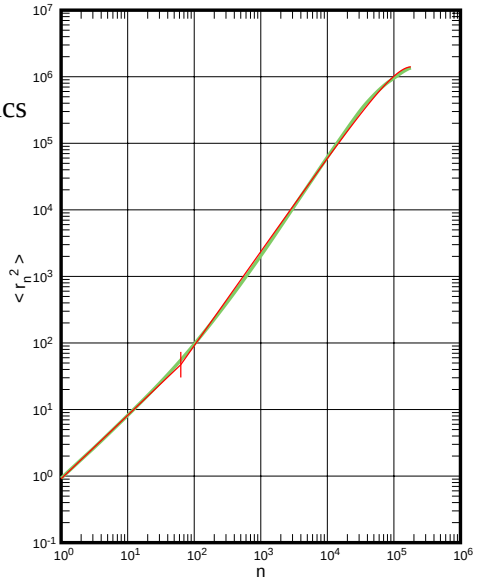



colorado

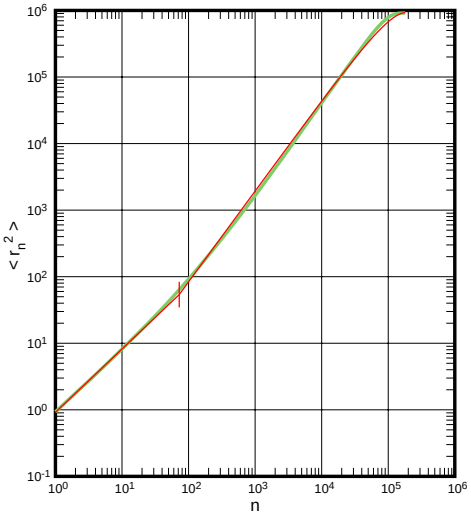


English texts: 365,000 < N < 395,000

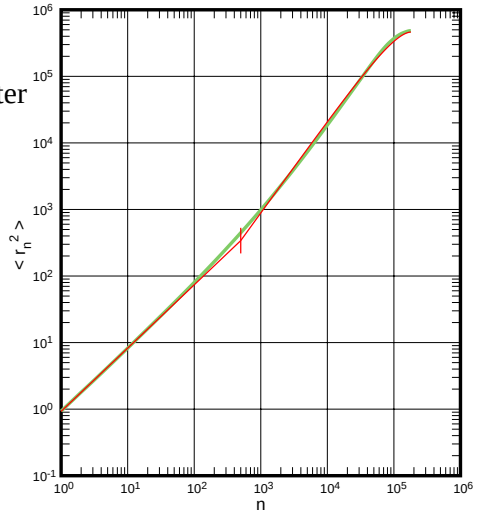
ari_physics



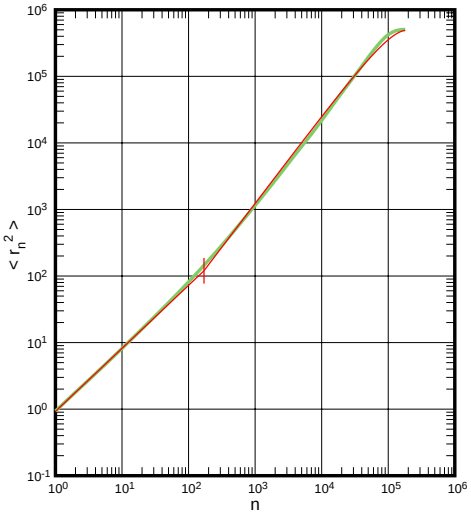

nicomachean



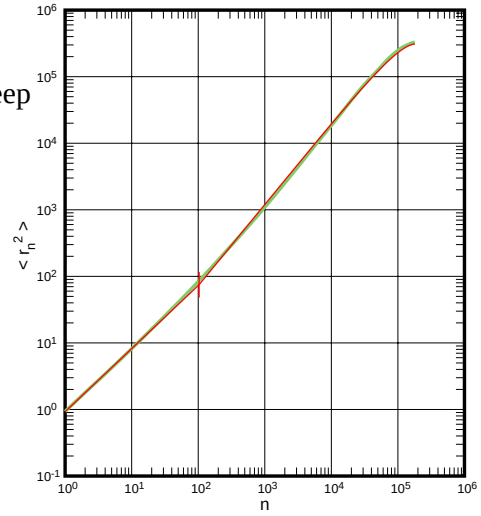

scarlet_letter



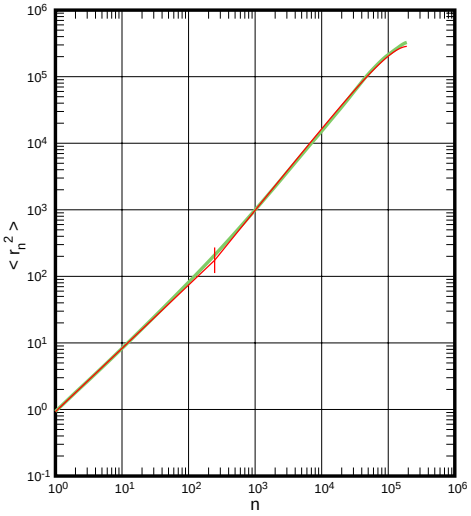

tarzan




steep

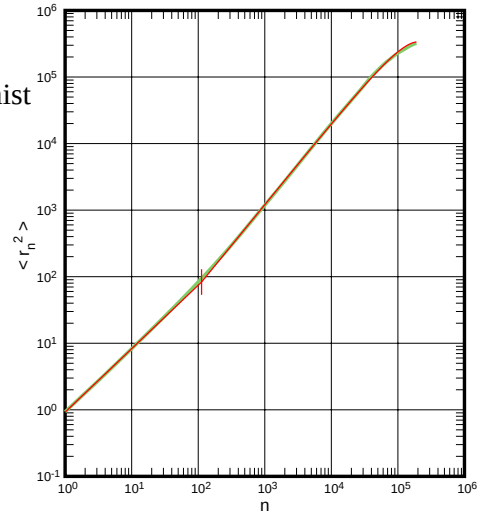



pretoria

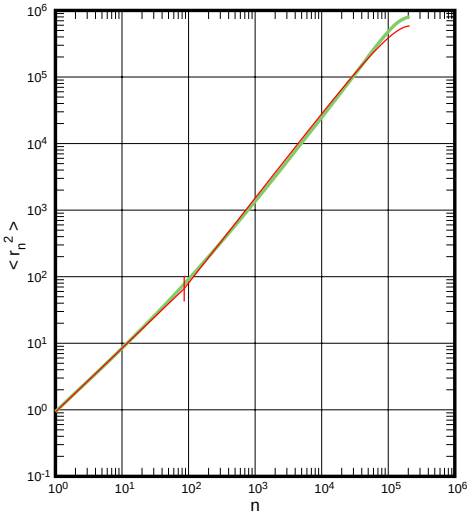


English texts: 397,000 < N < 455,000

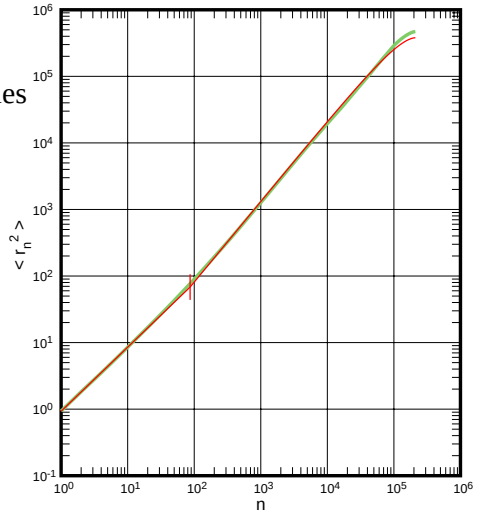
chymist



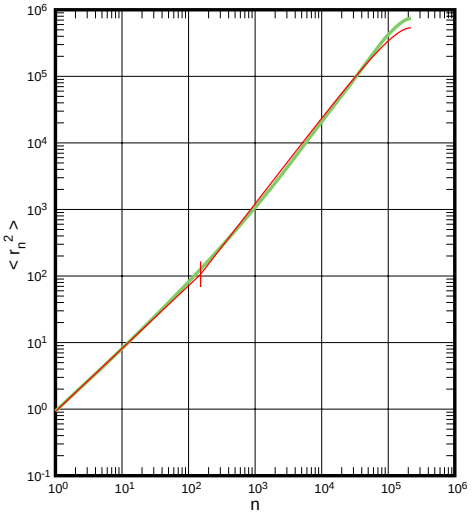

huckleberry



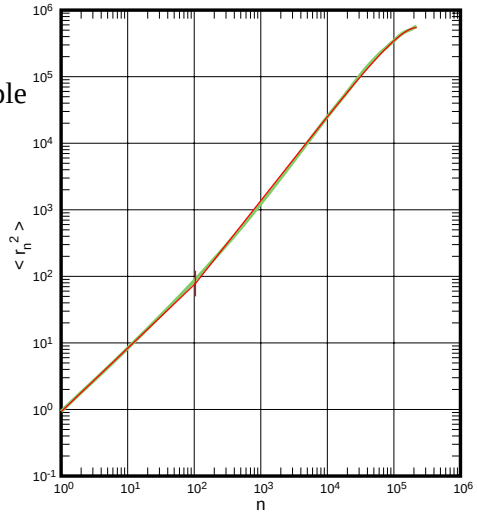

green_gables



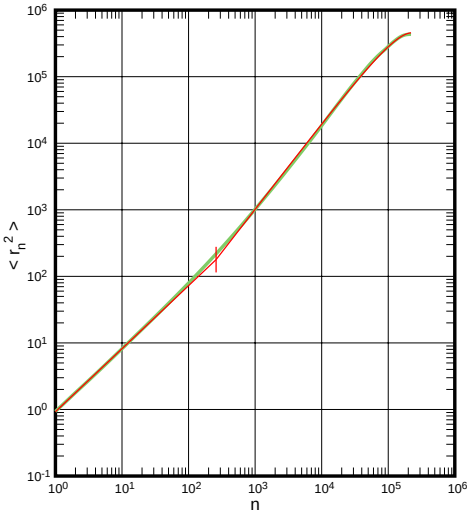

belgium




purple

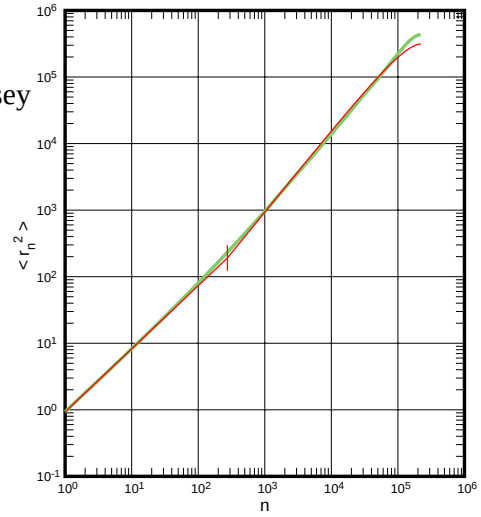



gulliver

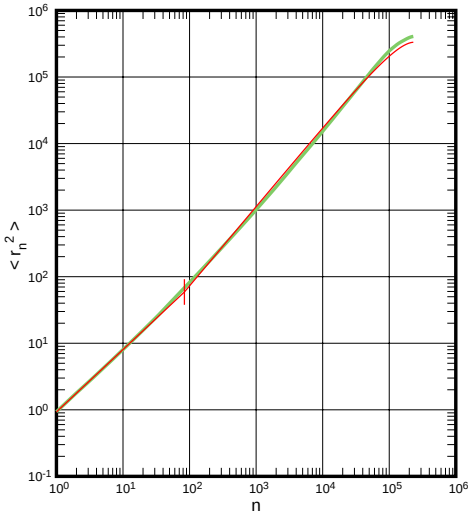


English texts: 458,000 < N < 495,000

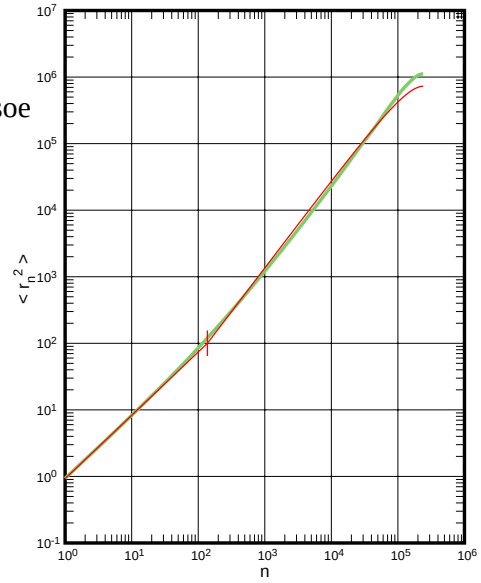
odyssey



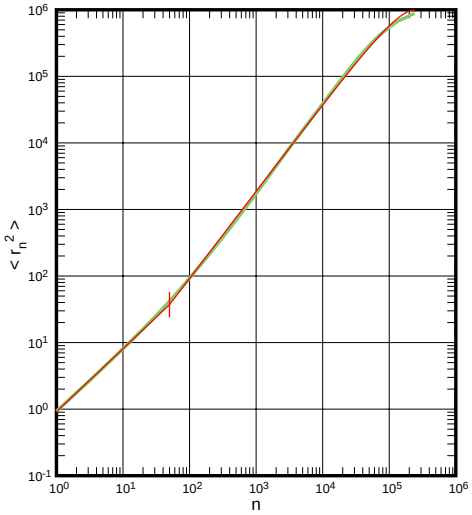

aeneid_english



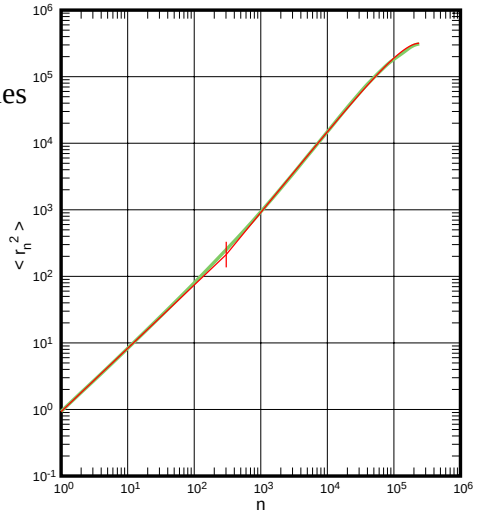

crusoe



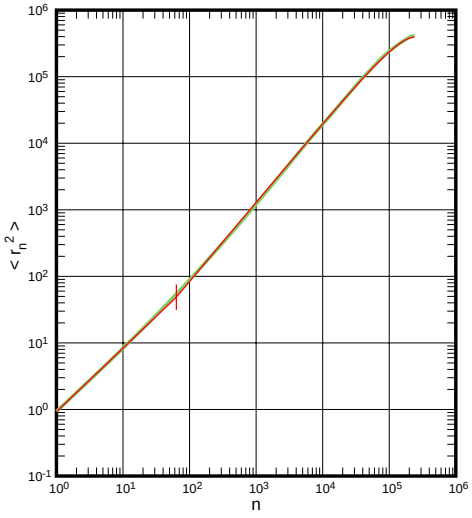

ari_meta




seven_gables

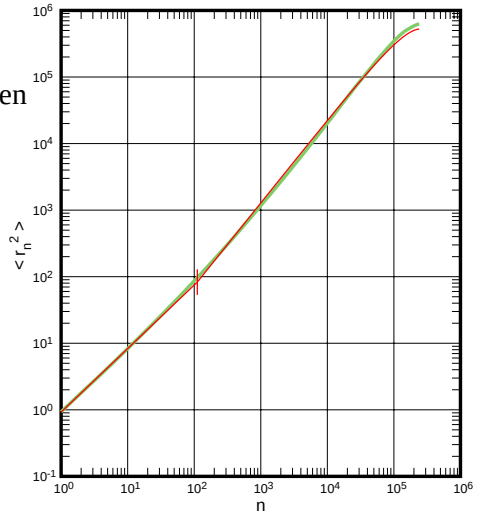



connyank

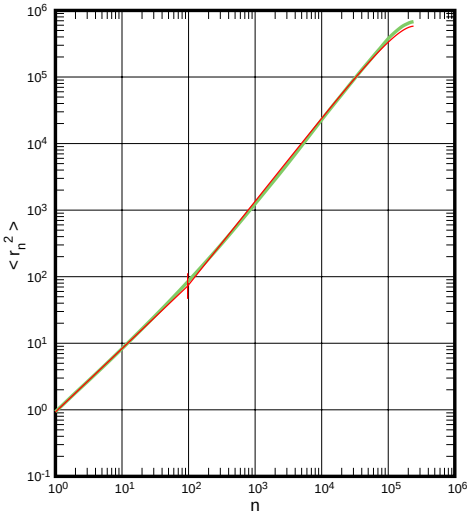


English texts: 496,000 < N < 550,000

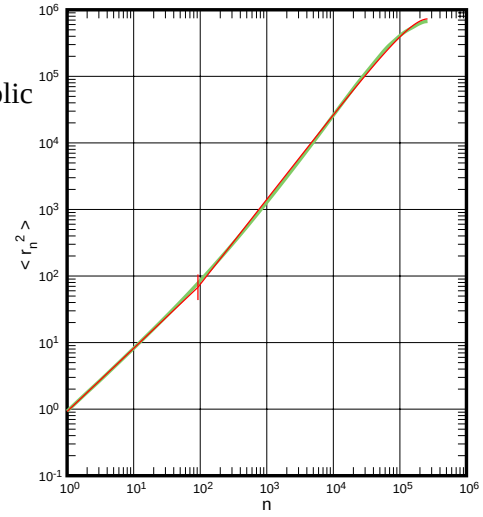
walden



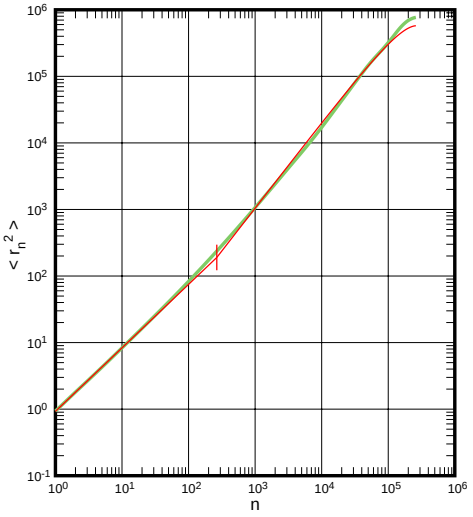

bush



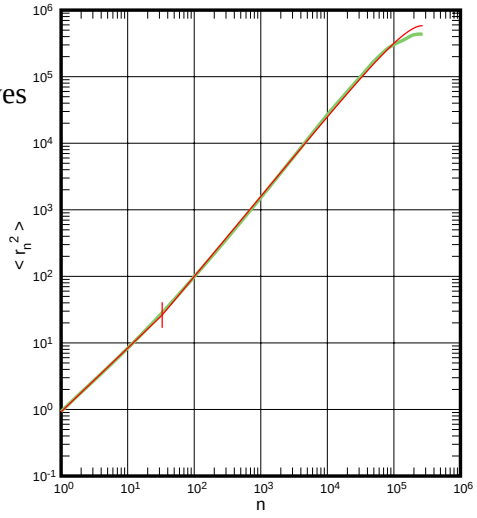

republic



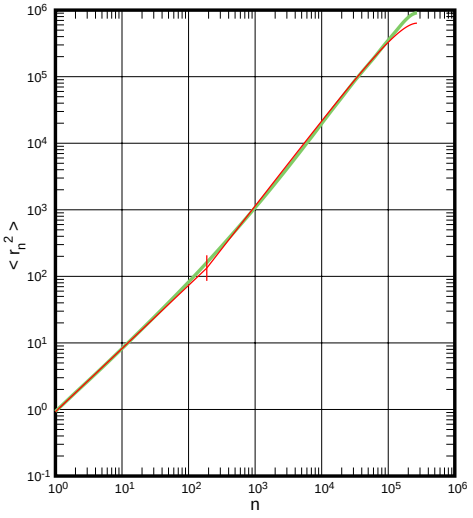

pnp




leaves

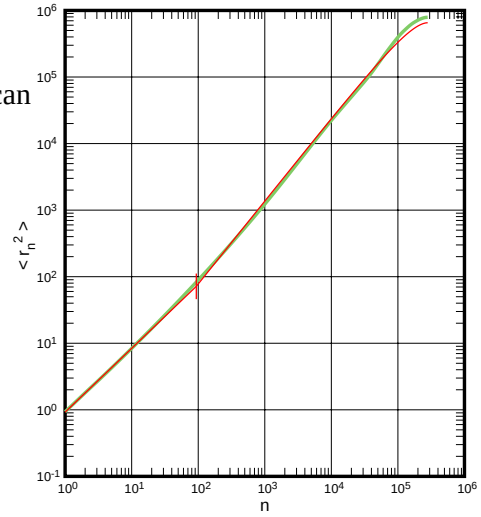



alhambra

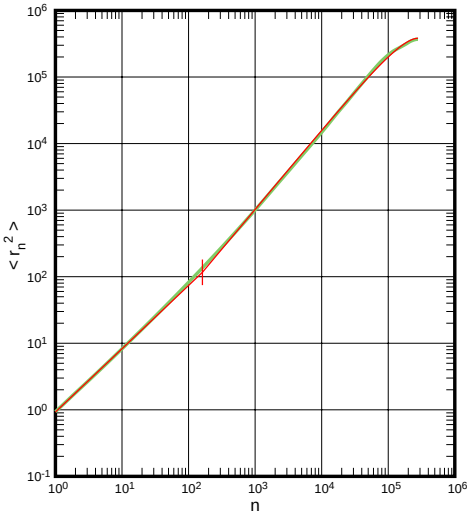


English texts: 568,000 < N < 645,000

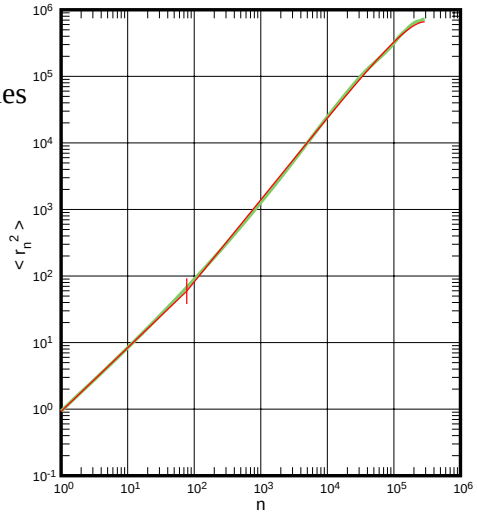
american



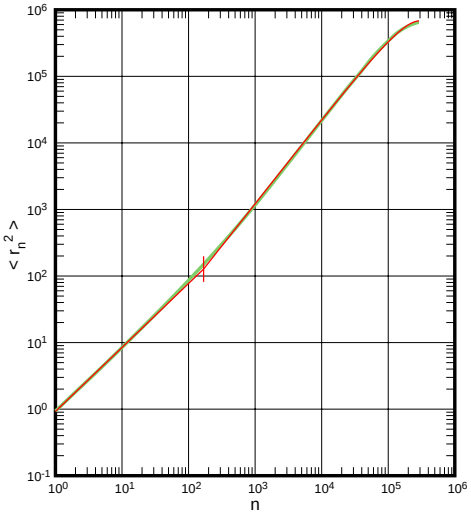

knickerbocker



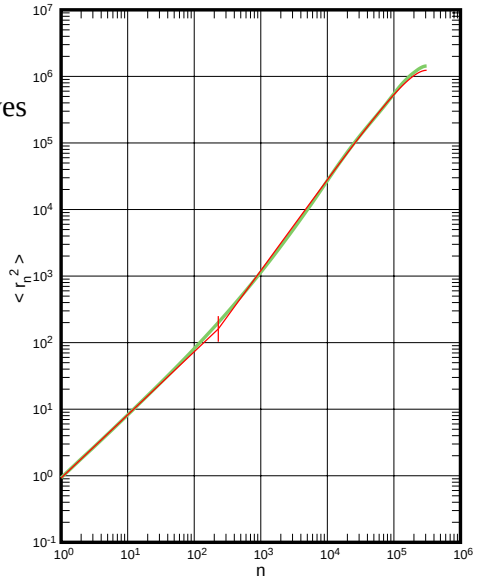

2cities



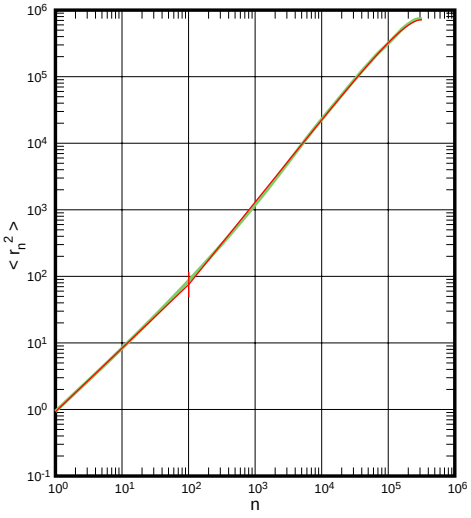

madcrowd




narratives

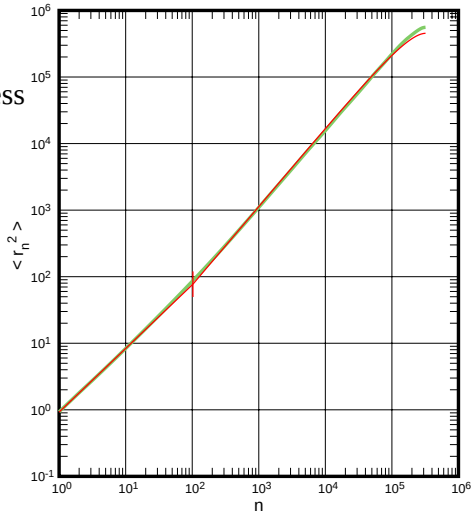



song_lark

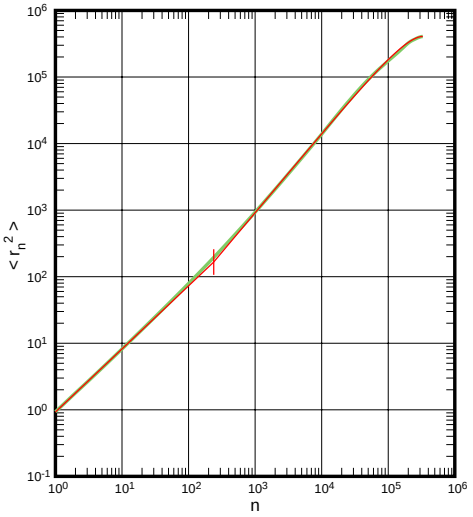


English texts: 648,000 < N < 695,000

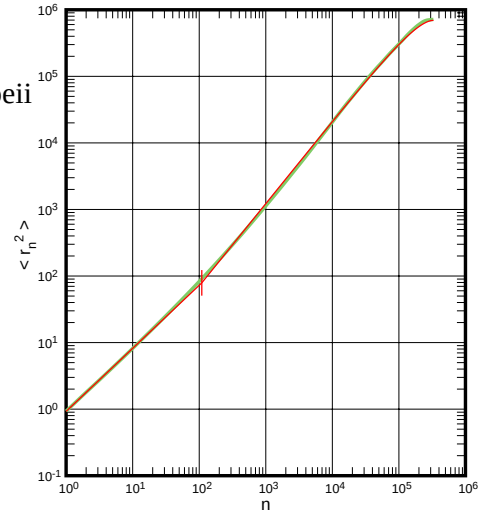
tess



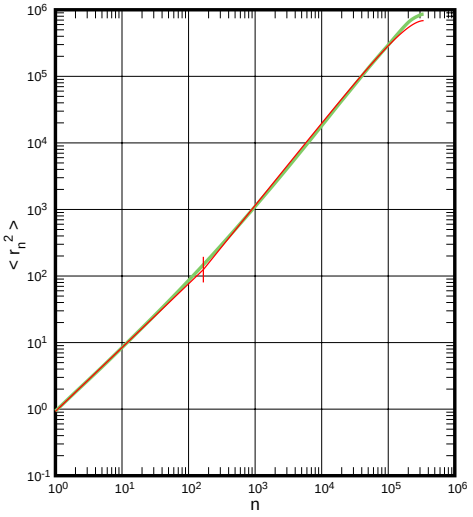

mohicans



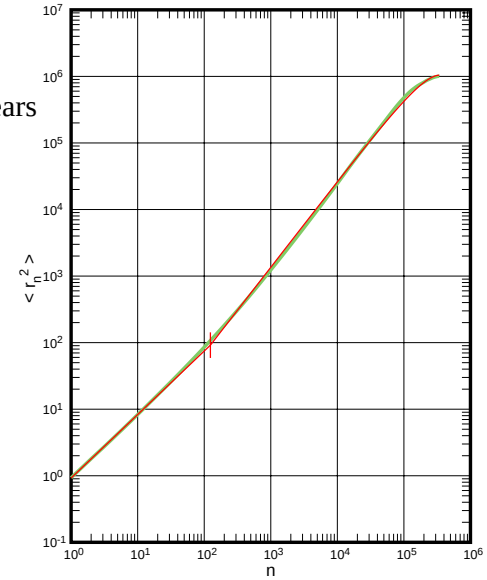

pompeii



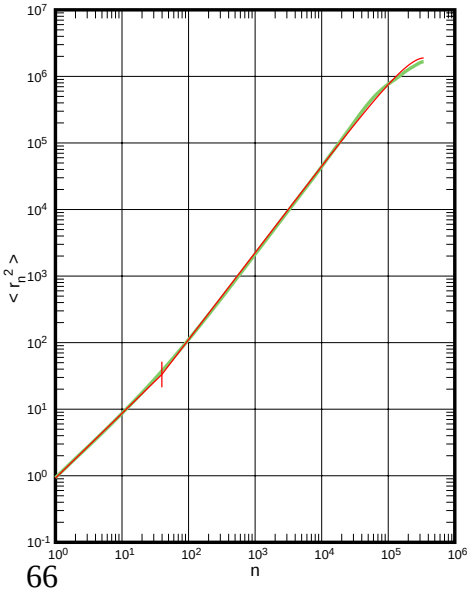

mansfield




2years

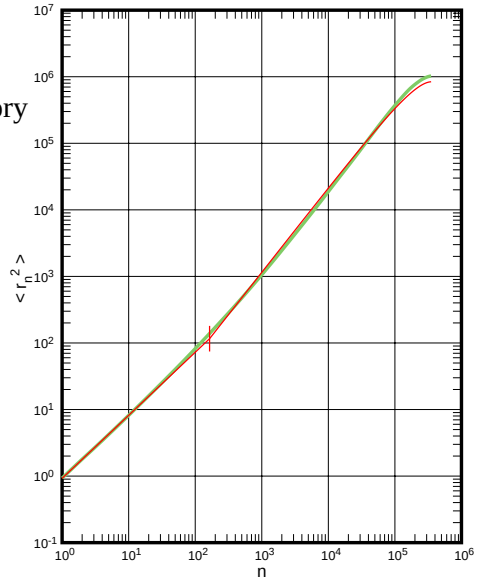



GandS

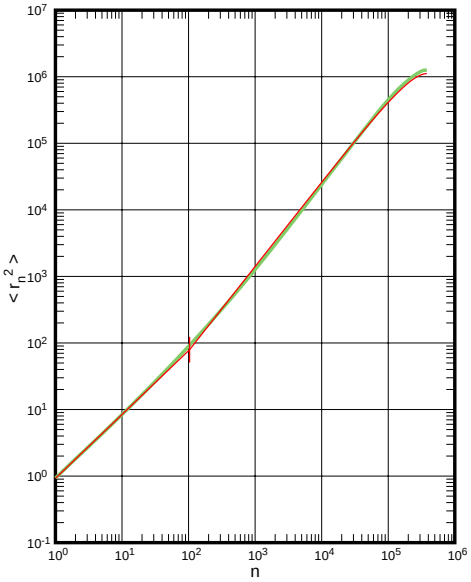


English texts: 714,000 < N < 790,000

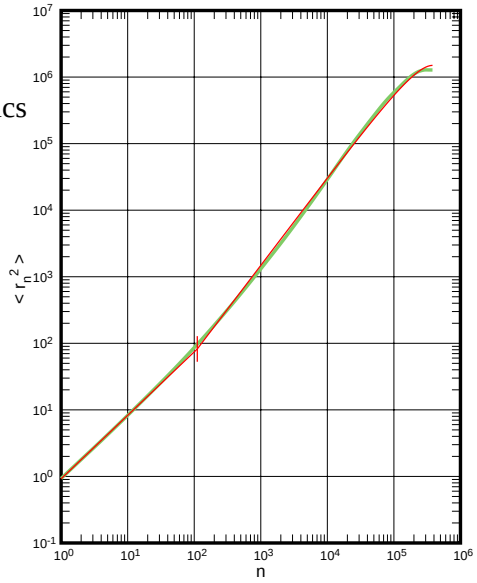
childs_history



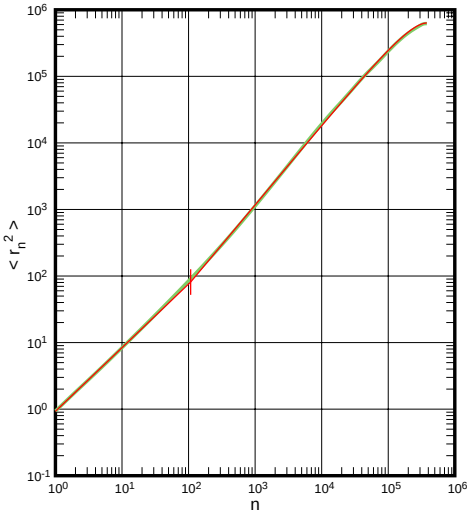

cabin



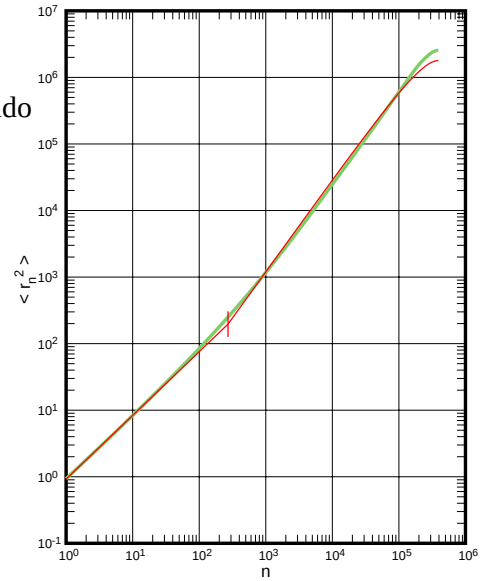

eugenics



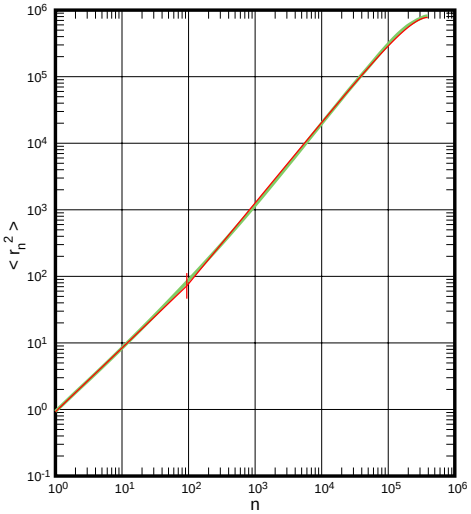

little_women




dido

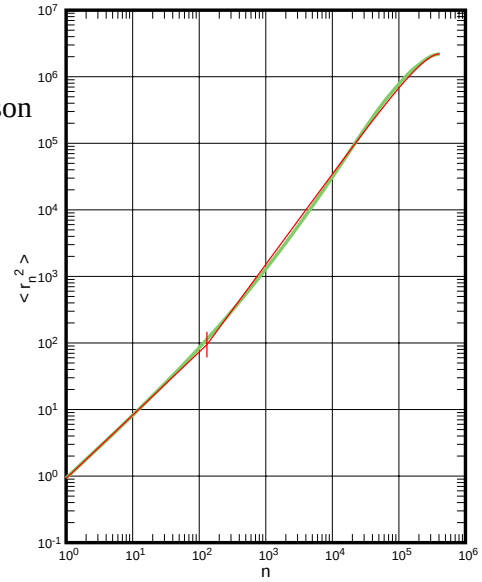



eyre

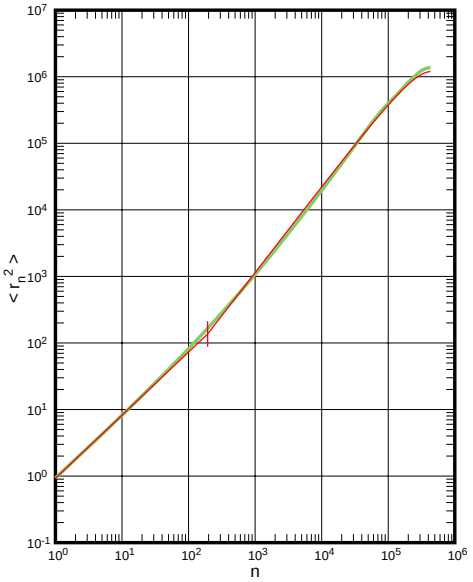


English texts: 835,000 < N < 1,010,000

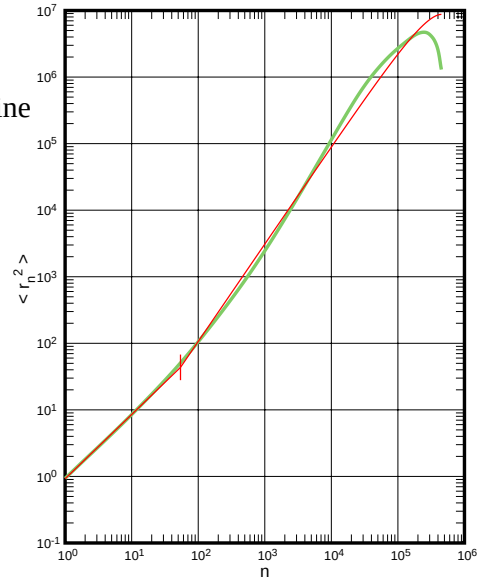
hudson



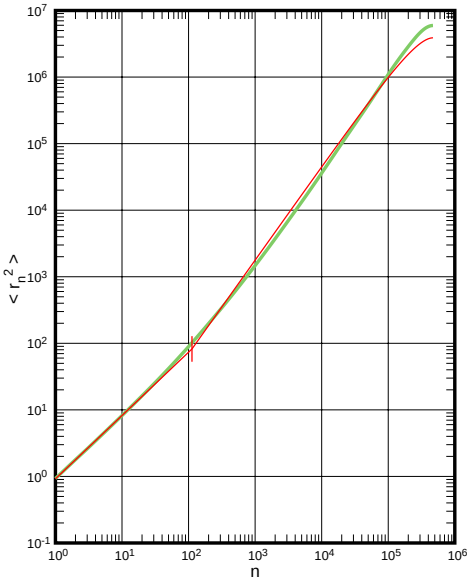

ivanhoe



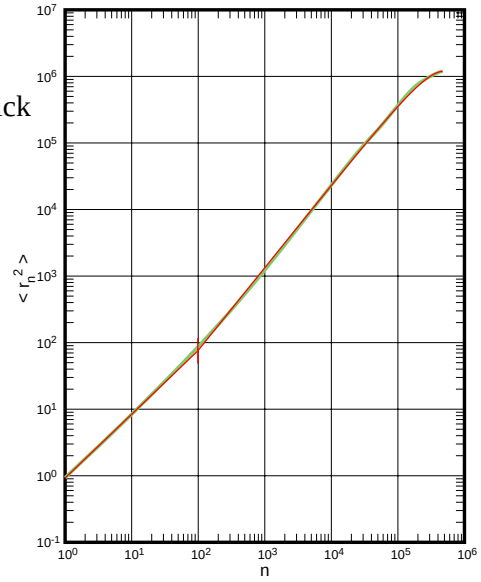

argentine



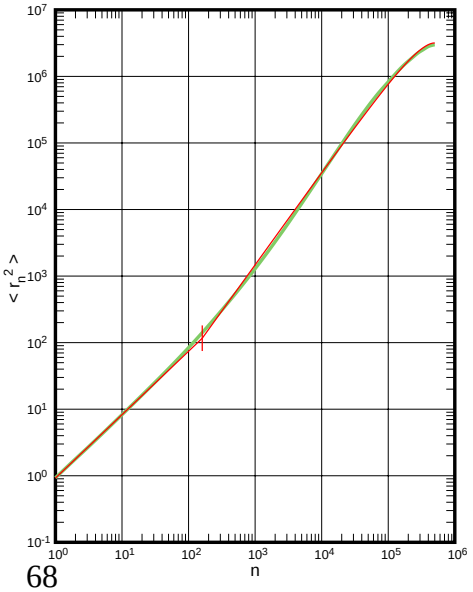

leviathan




mobydick

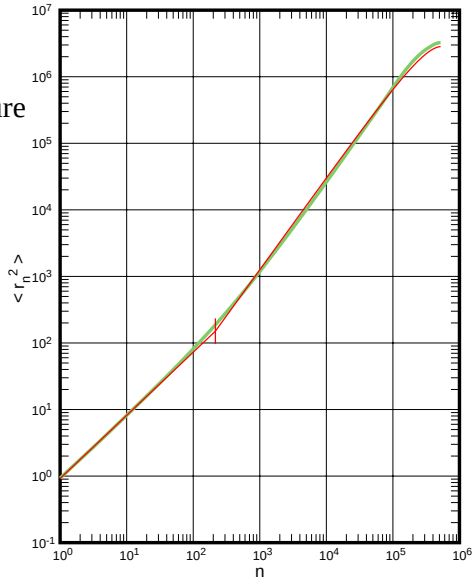



origin

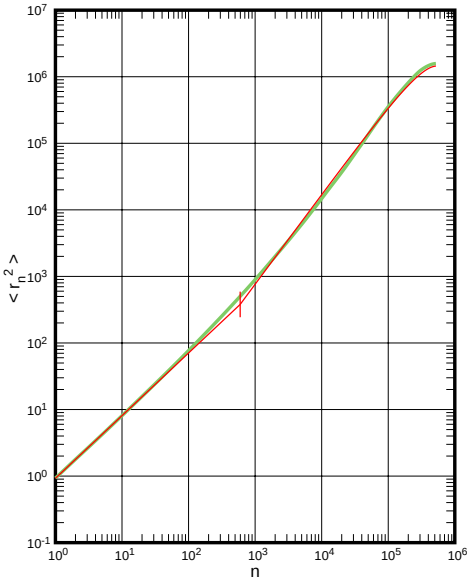


English texts: 1,050,000 < N < 1,350,000

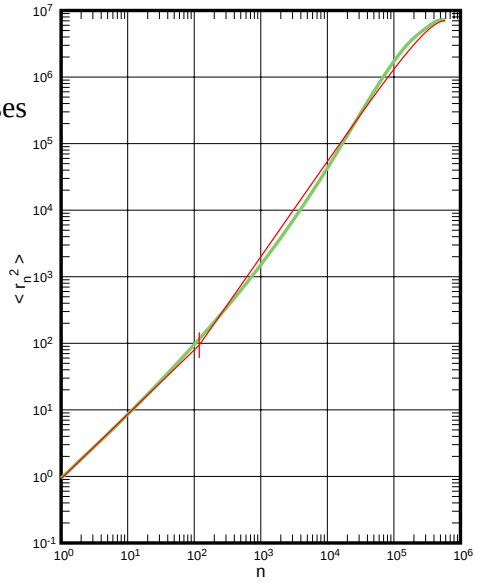
human.nature



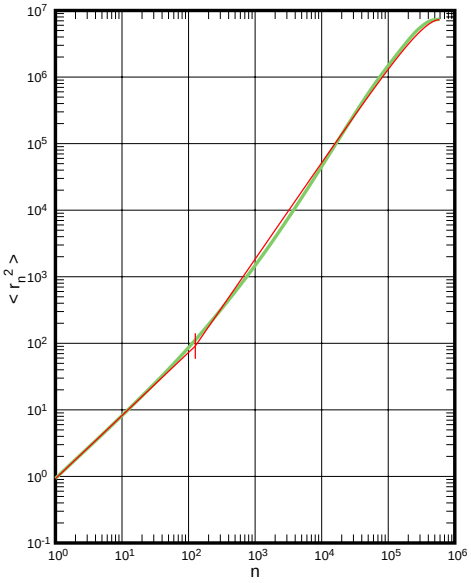

mexcon



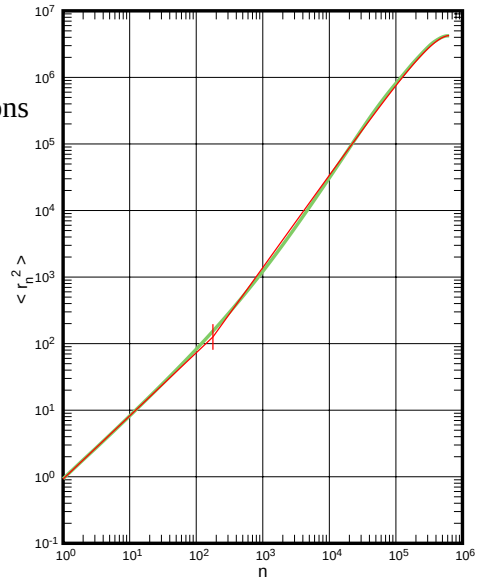

ulysses



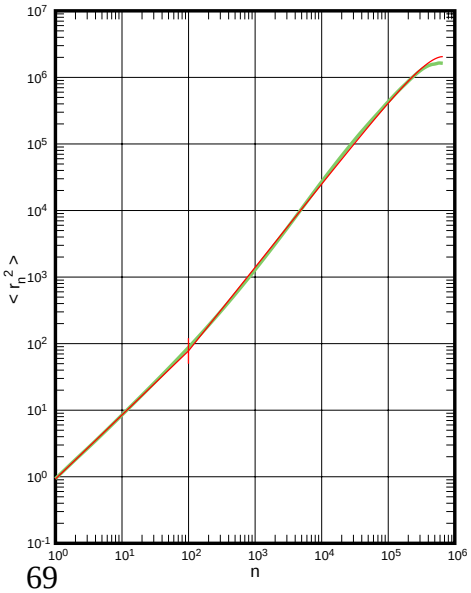

cattle




delusions

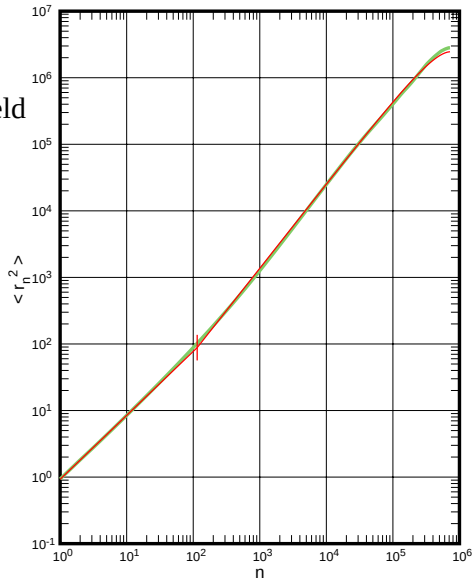



pickwick

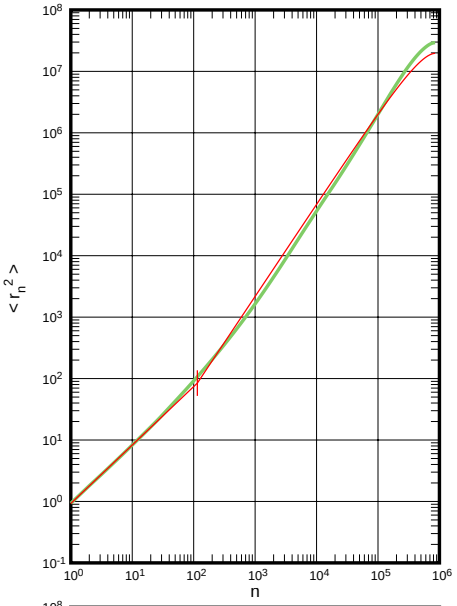


English texts: 1,480,000 < N < 2,520,000

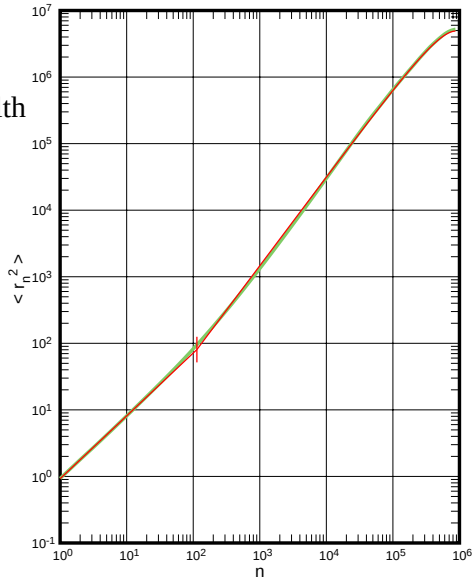
copperfield



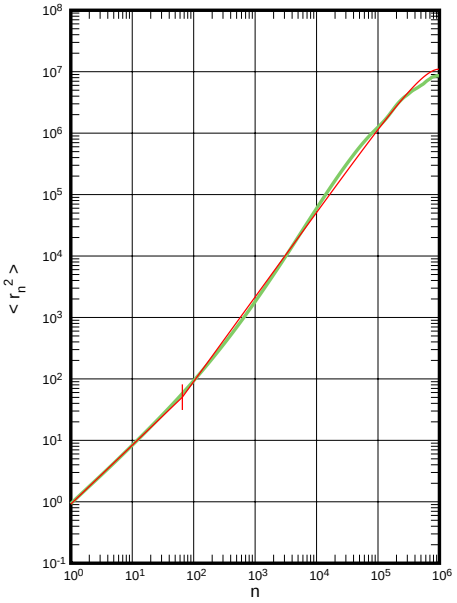

ramayan



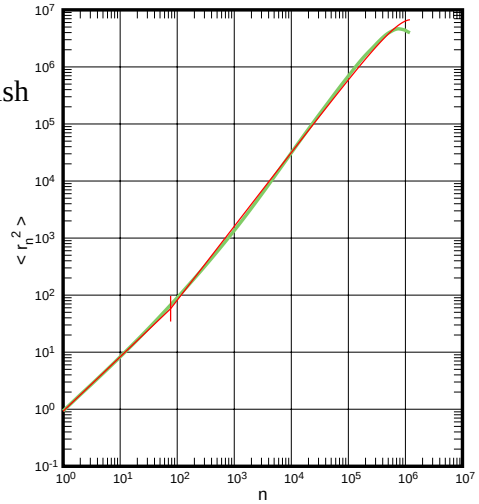

wealth




sociology

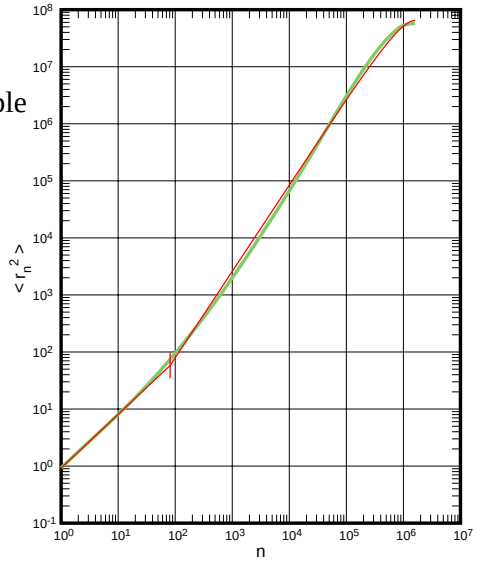



lesmis_english

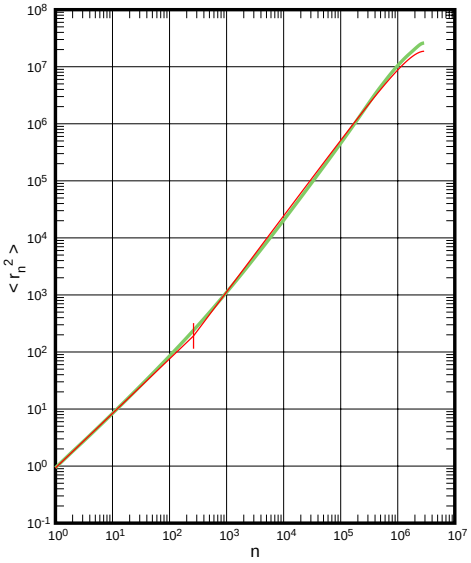


English texts:  3,200,000 < N < 7,500,000

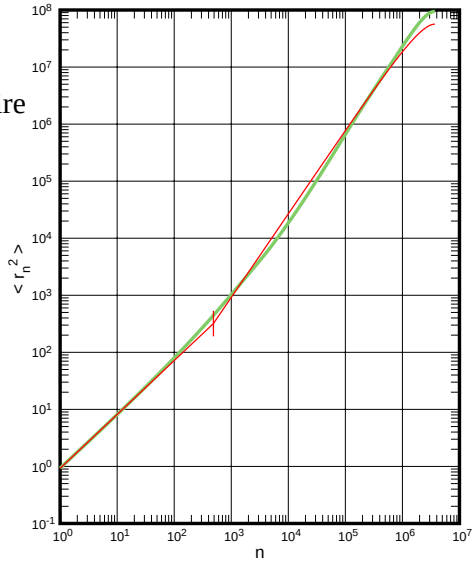
bible

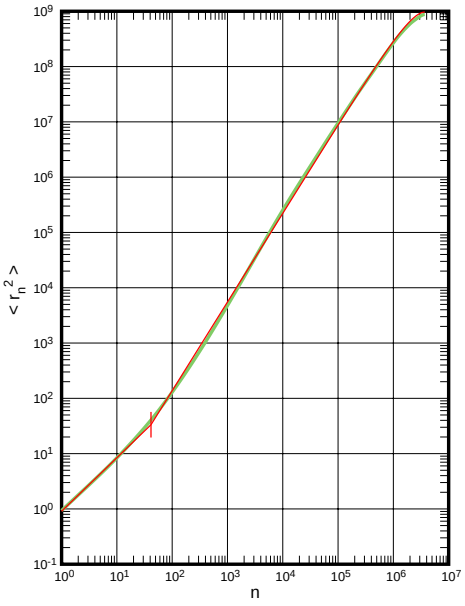



fritz




empire

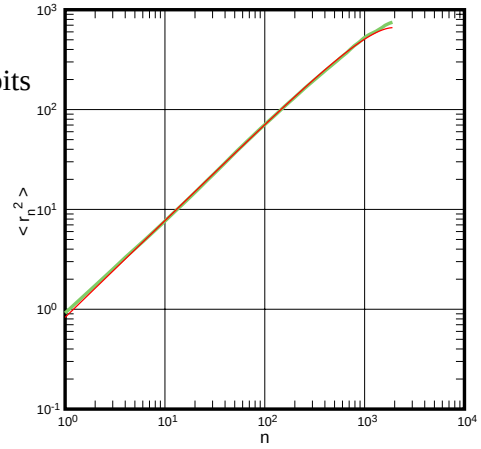


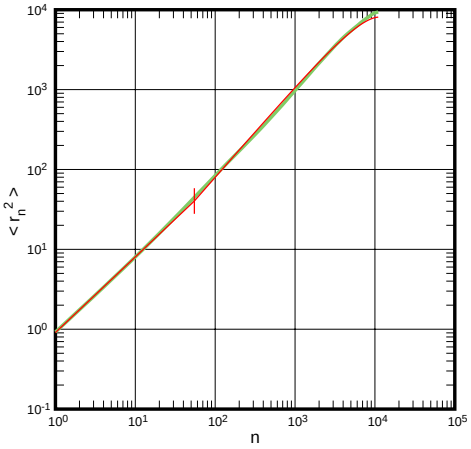

CIA2010let

French texts: 3900 < N < 40,000

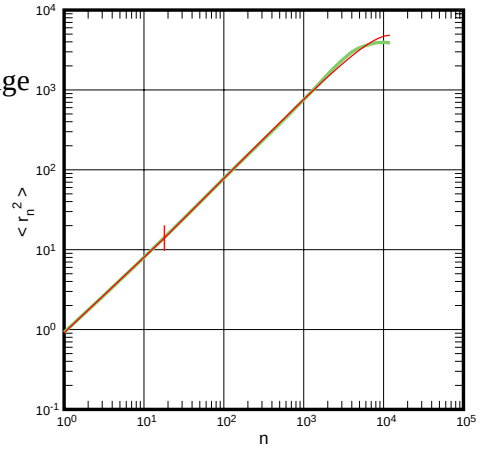
droits



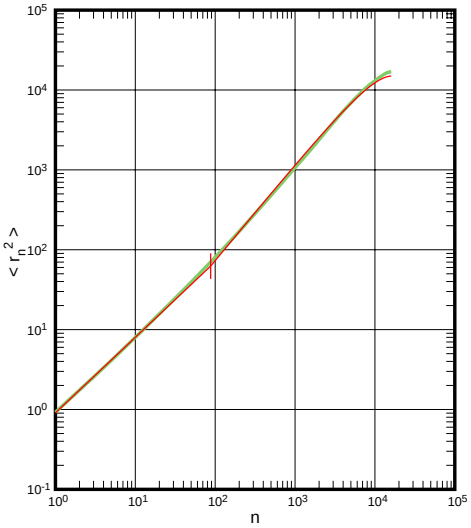

dreyfus



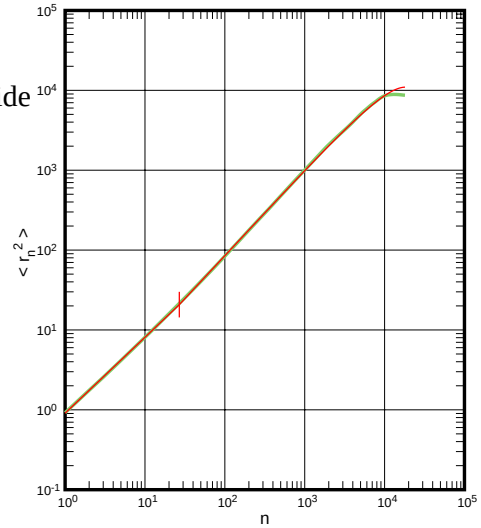

cocuage



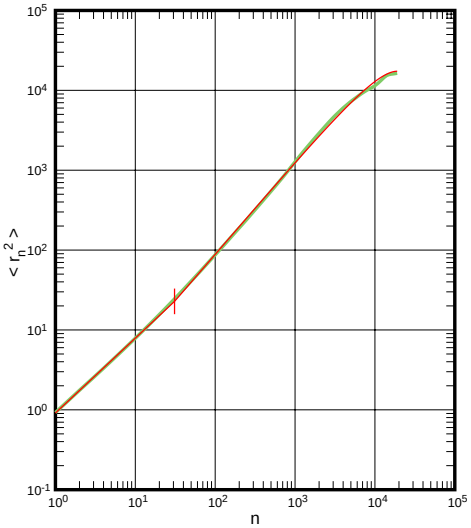

bounty_verne




adelaide

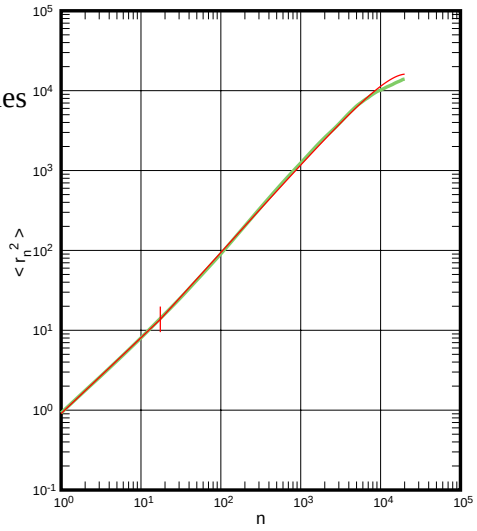



esclaves

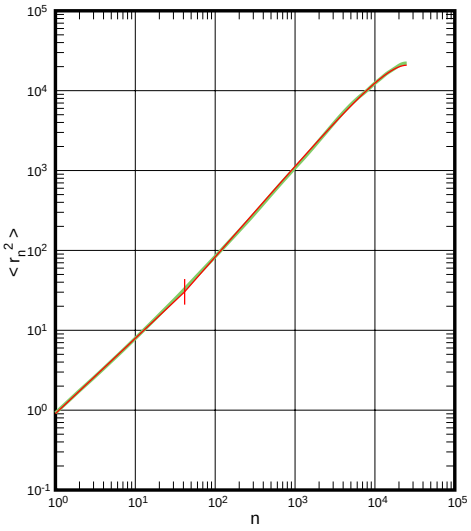


French texts: 40,000 < N < 63,000

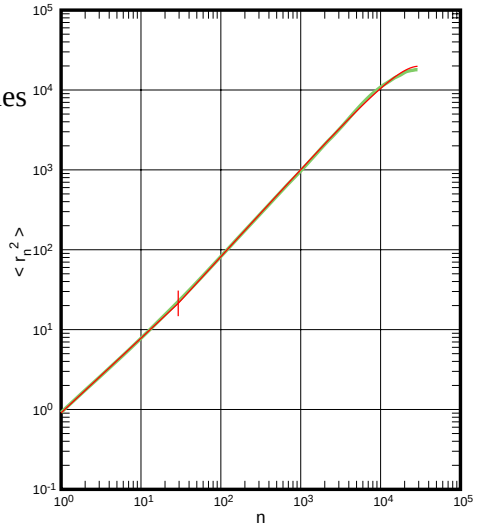
anglicismes



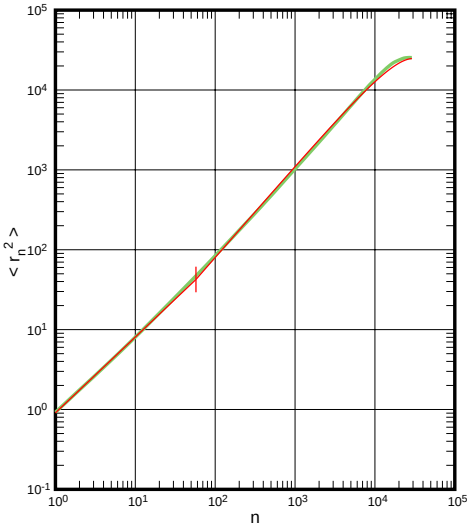

esther_racine



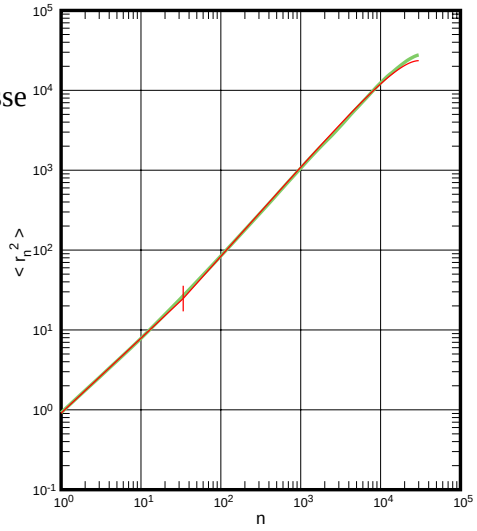

maximes



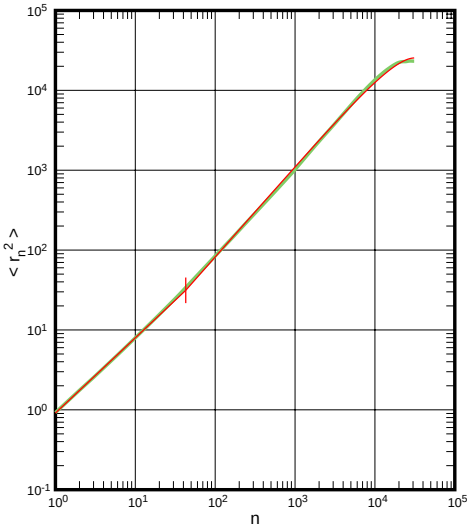

caribou




paresse

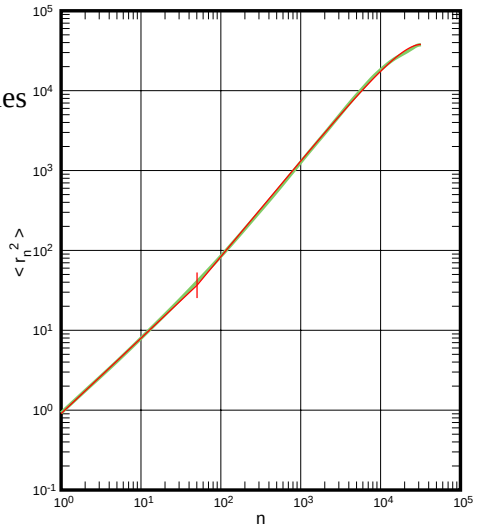



andromaque

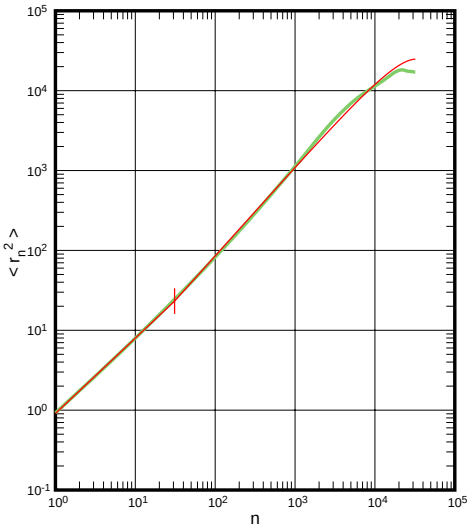


French texts: 64,000 < N < 75,000

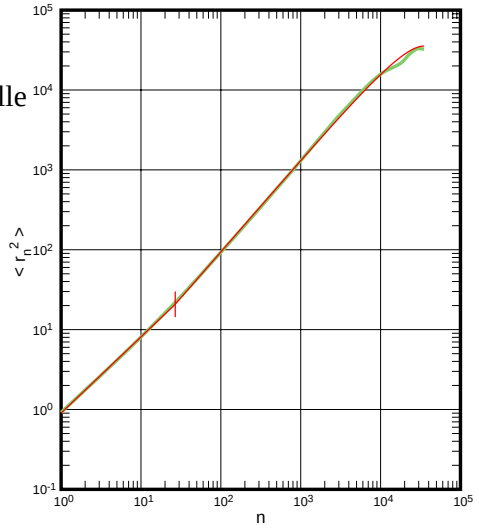
fourberies



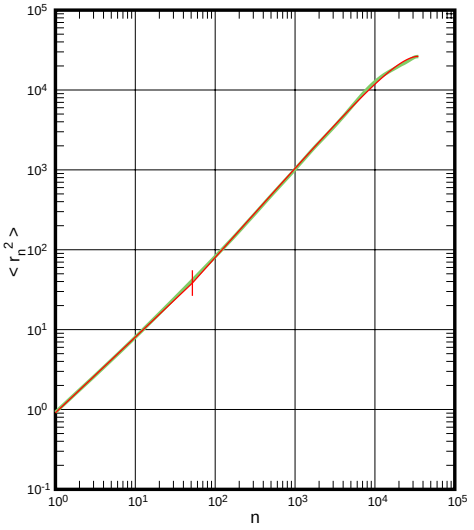

jeu



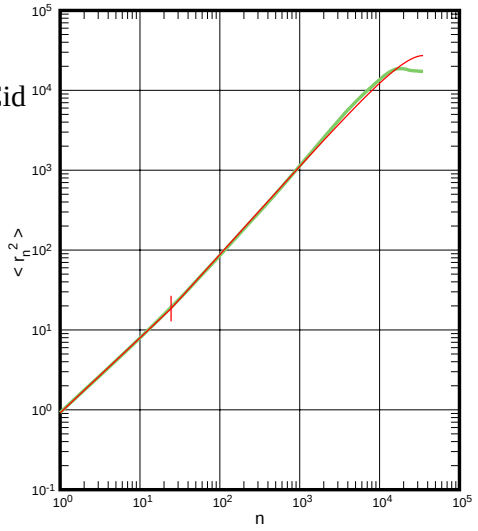

bougainville



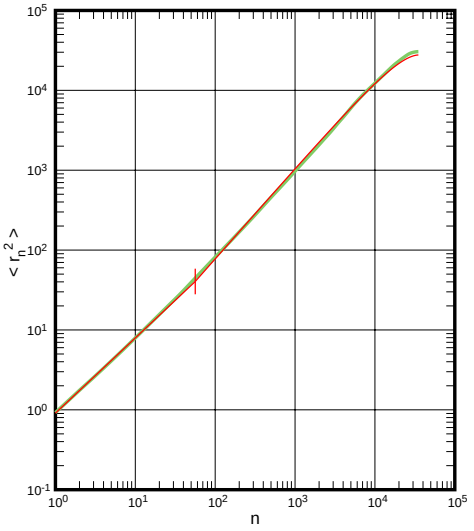

montbard




LeCid

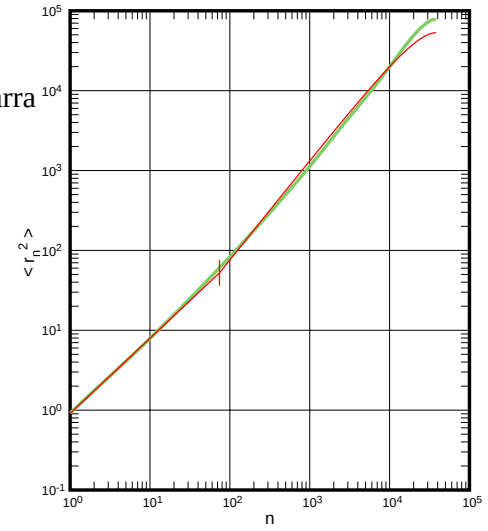



misanthrope

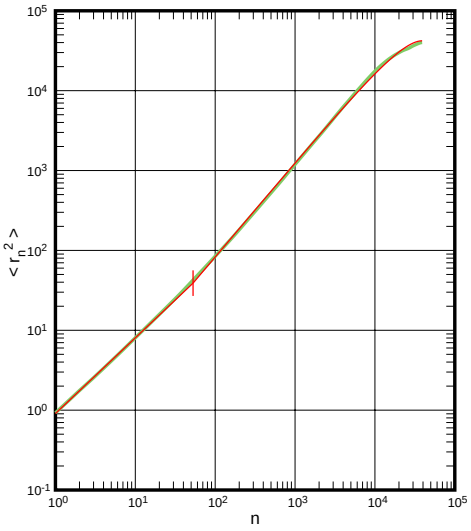


French texts: 76,000 < N < 120,000

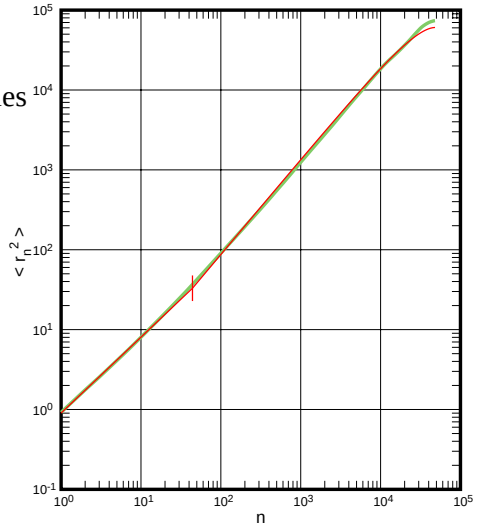
smarra



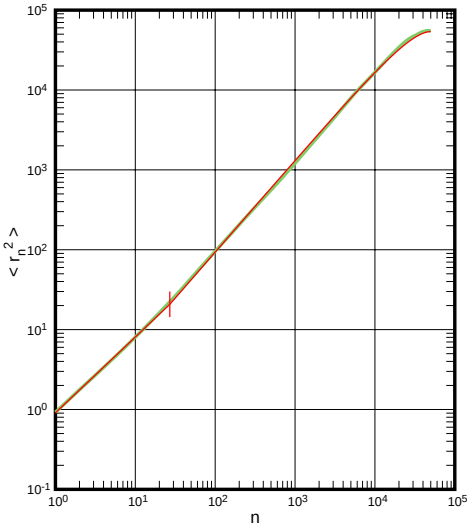

tartuffe



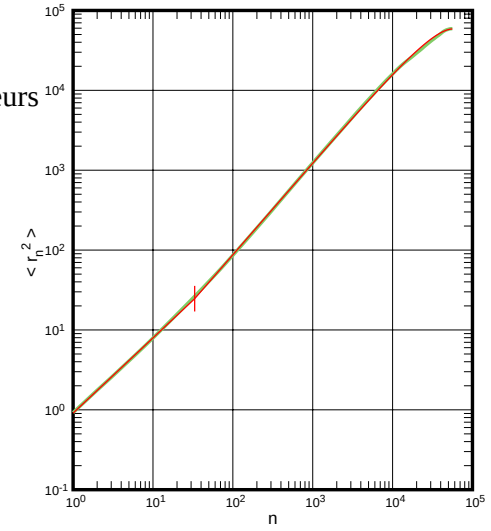

harmonies



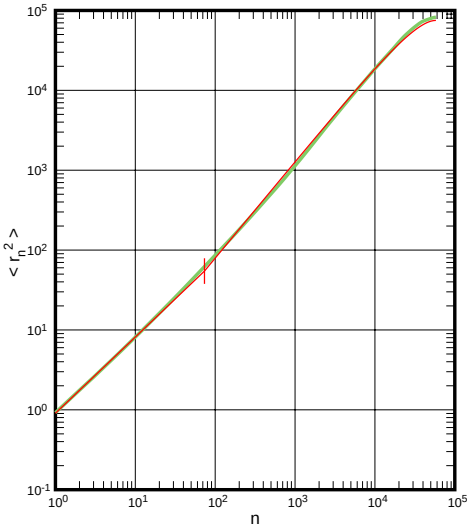

ane




fleurs

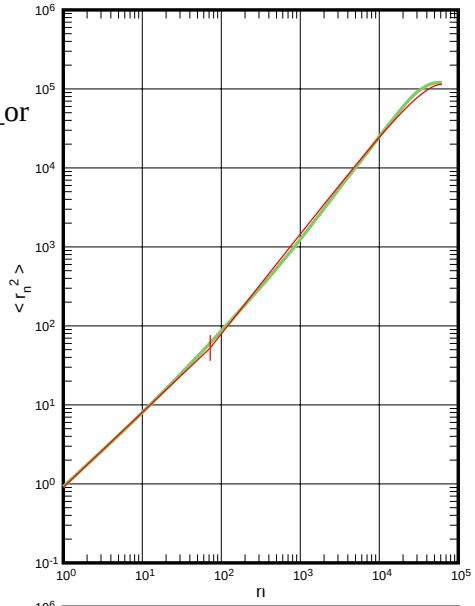


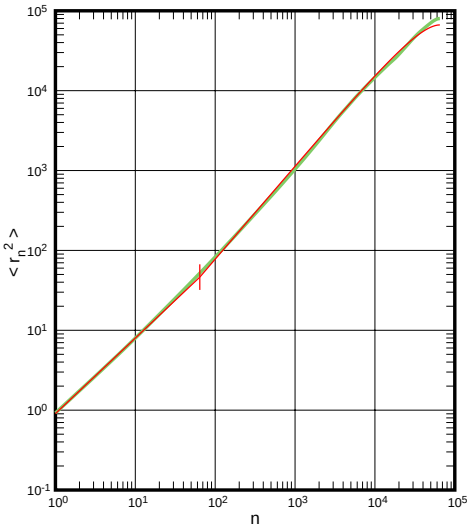

lune



French texts: 124,000 < N < 170,000

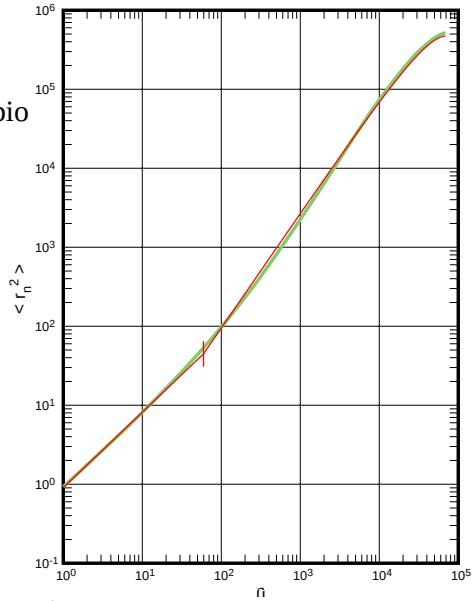
yeux_d_or

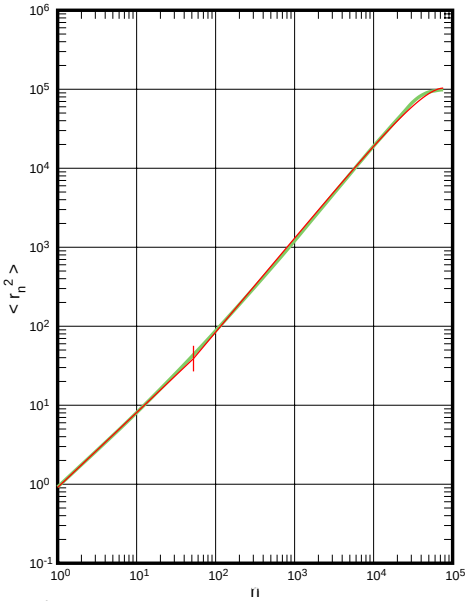



inquisition

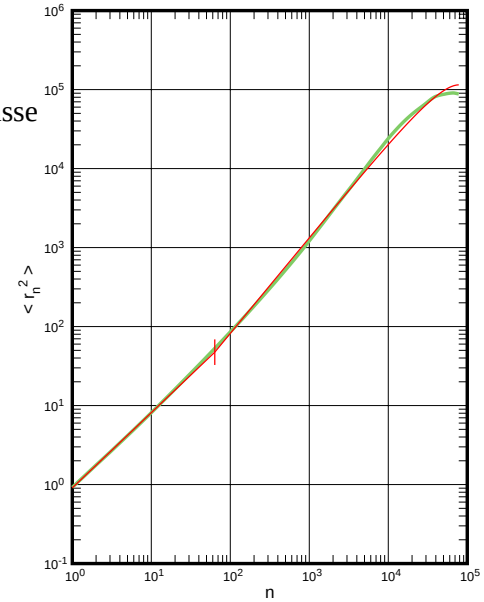
hp_bio

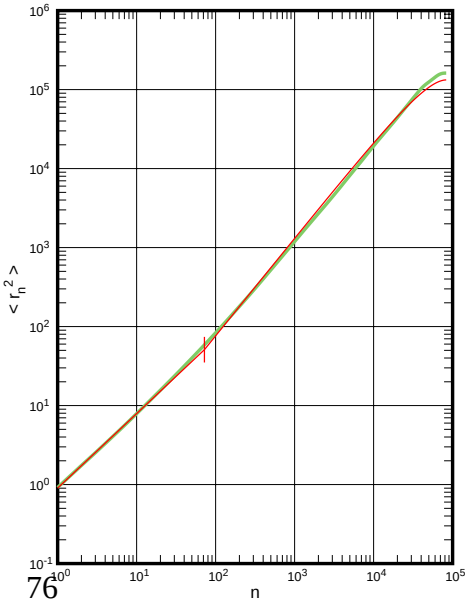



candide

becasse

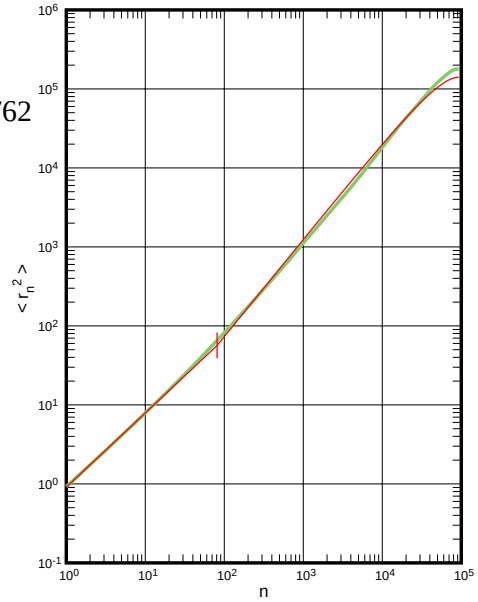


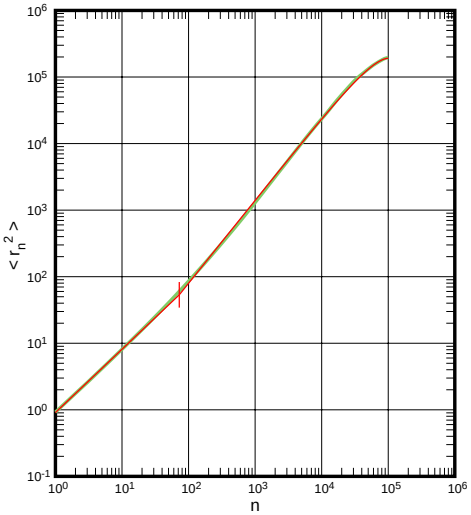

mondes

French texts: 185,000 < N < 225,000

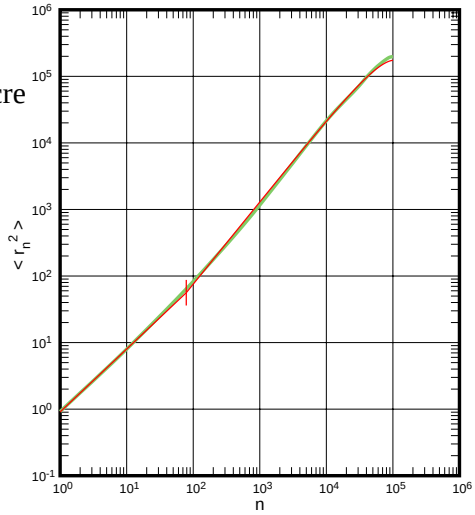
contrat1762



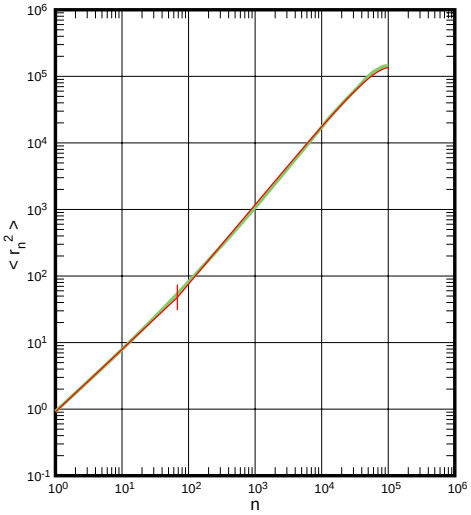

champi



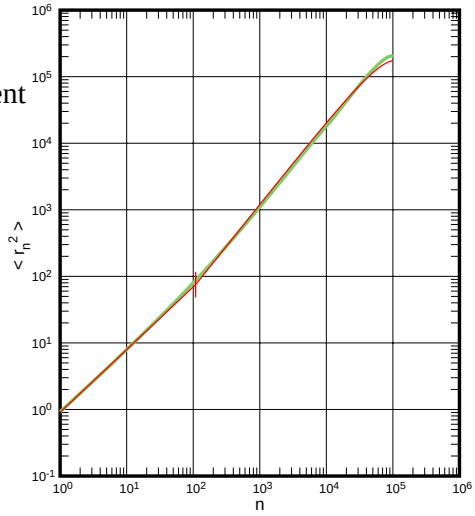

nacre



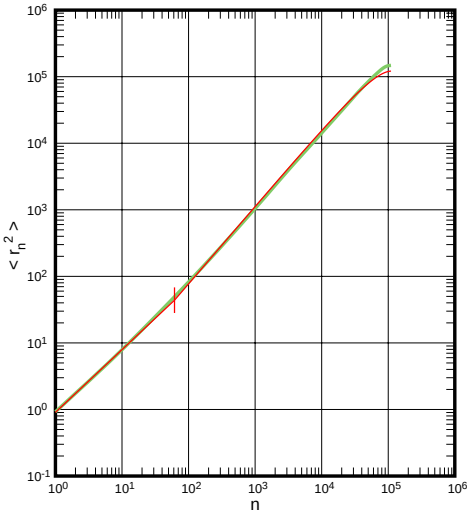

considerations




egarement

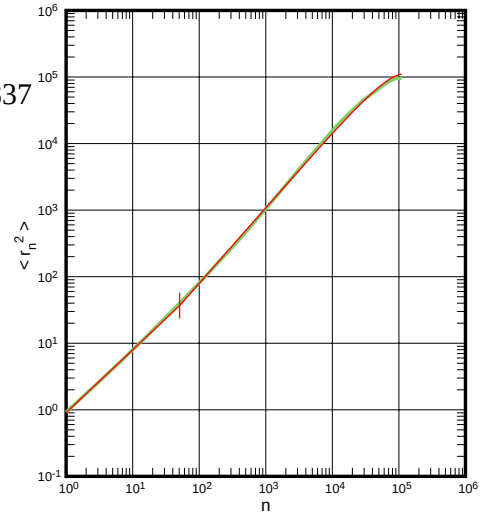



correspondance

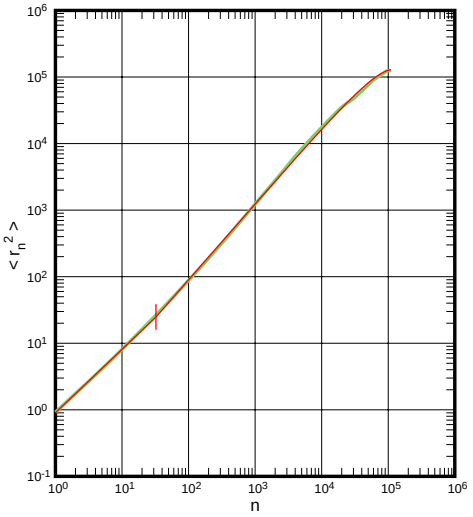


French texts: 235,000 < N < 280,000

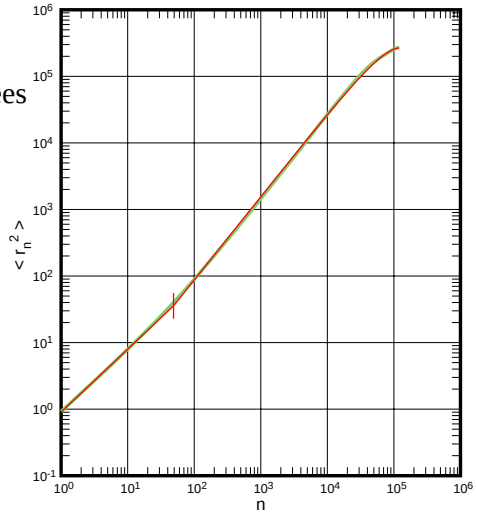
bn1837



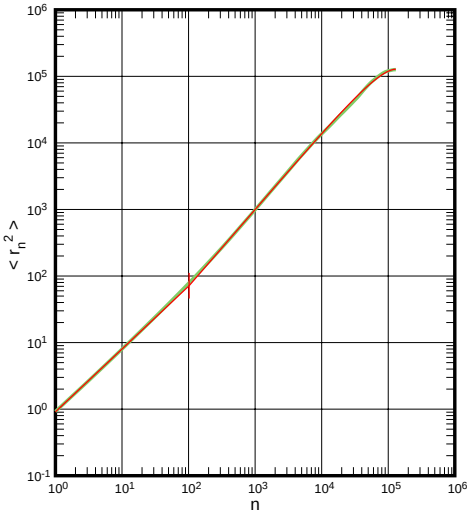

aziyade



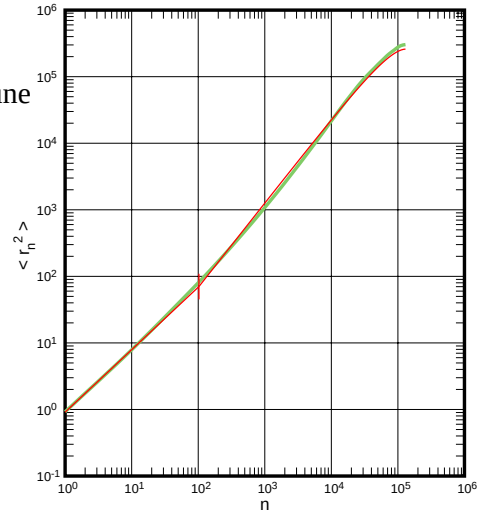

pensees



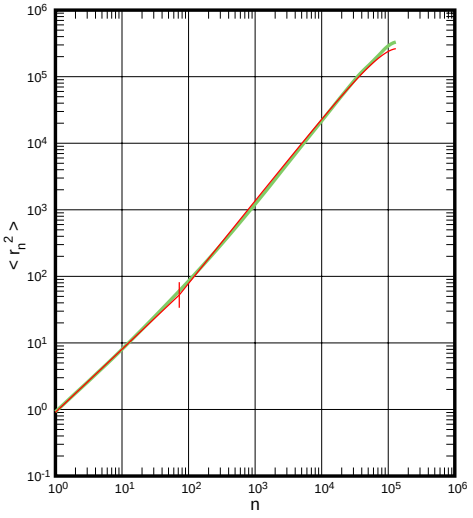

manon




terre_lune

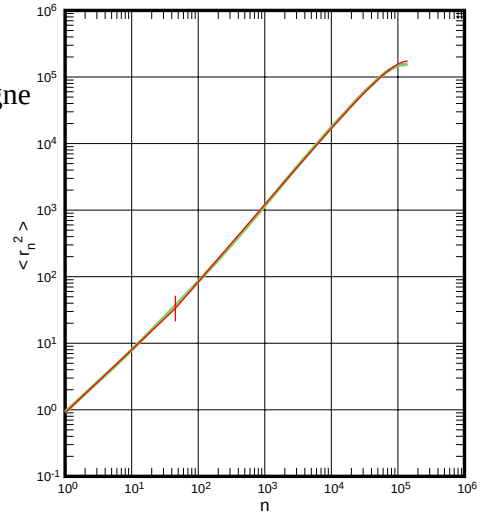



robida

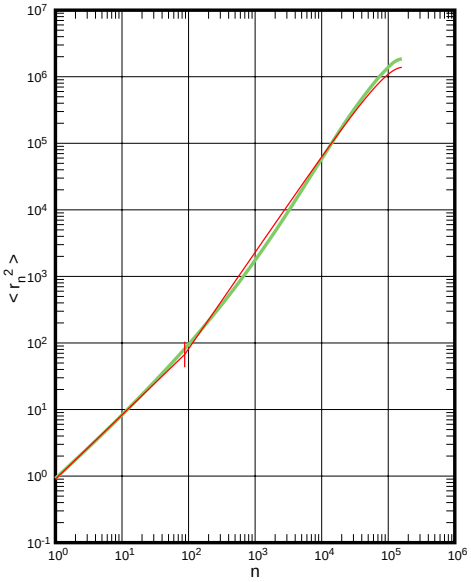


French texts: 285,000 < N < 372,000

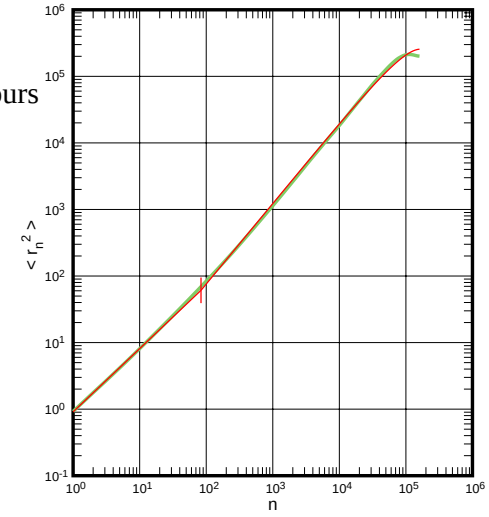
Espagne



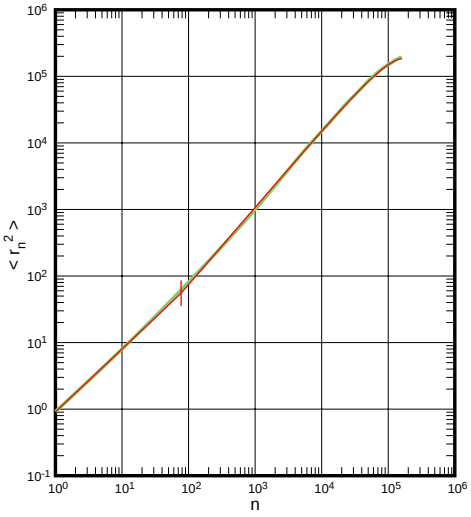

hokousai



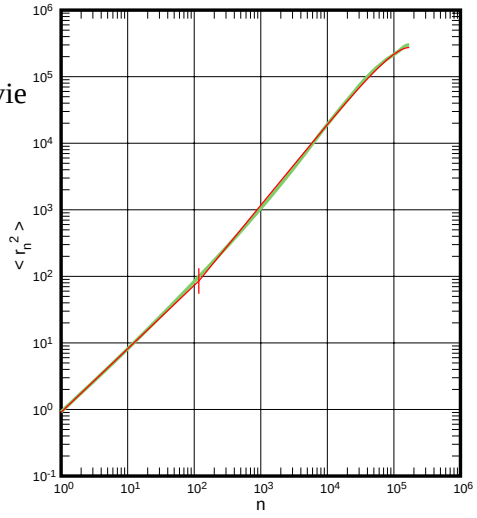

80jours



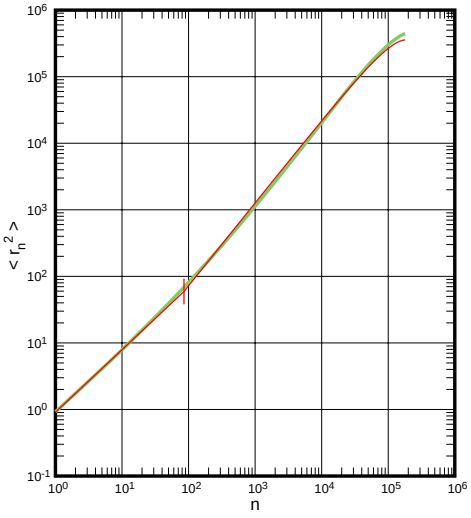

dominique




une_vie

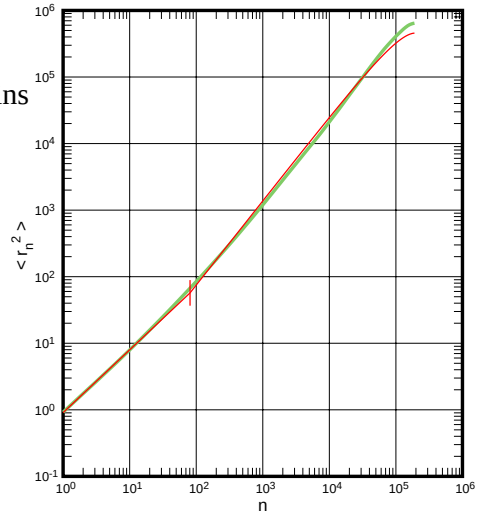



telephone

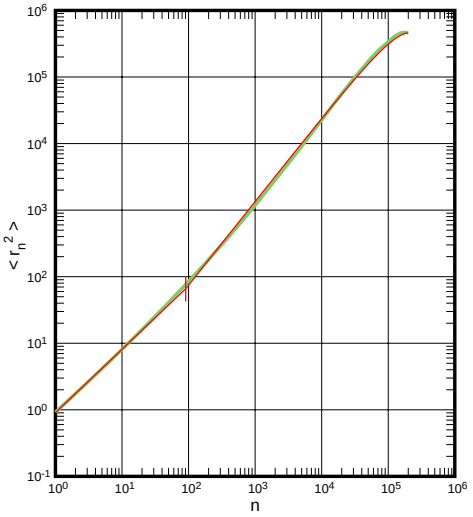


French texts: 393,000 < N < 520,000

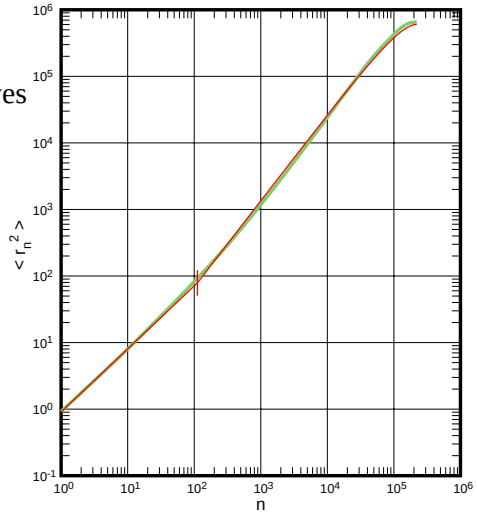
pingouins



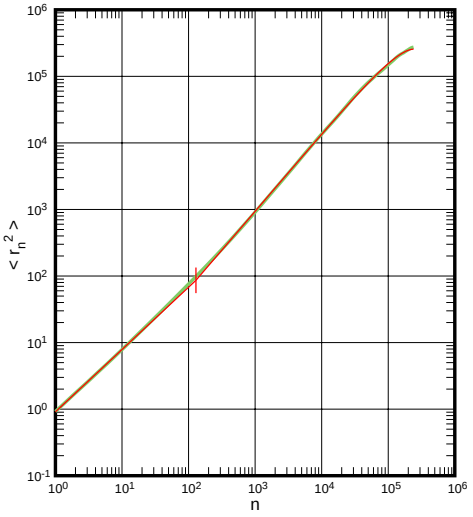

goriot



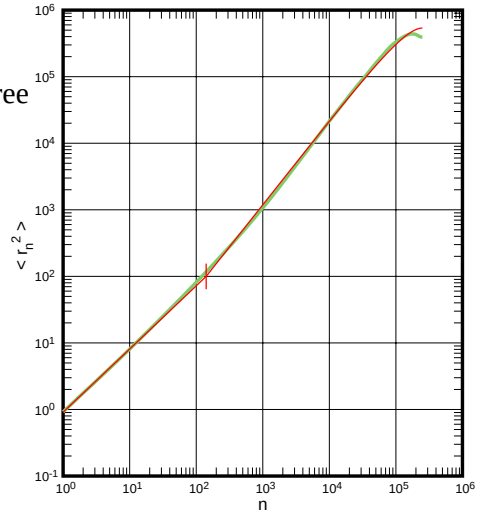

preuves



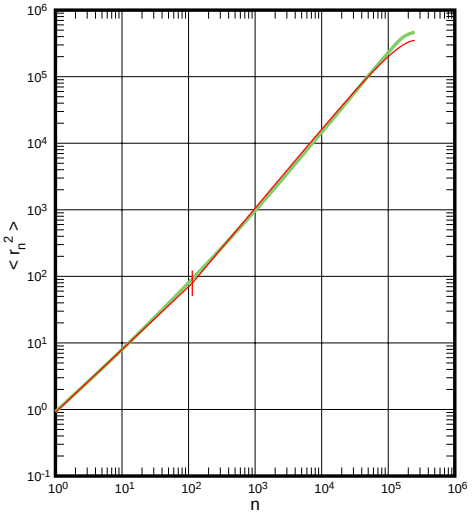

salammbo




curee

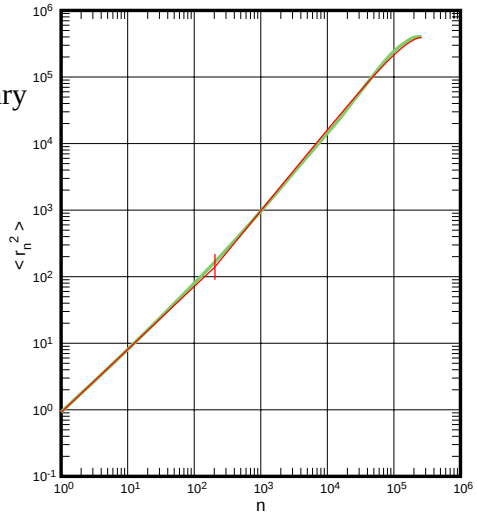



young

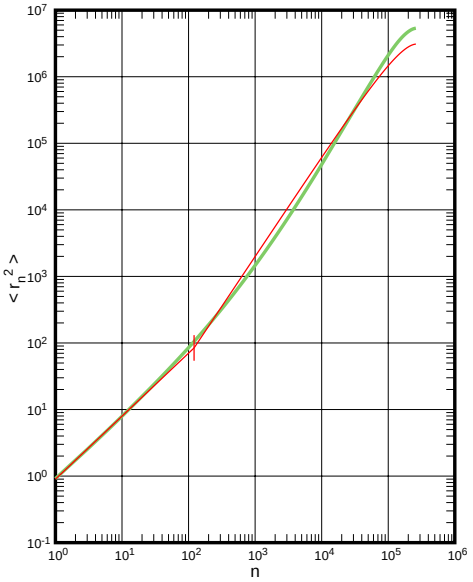


French texts: 534,000 < N < 710,000

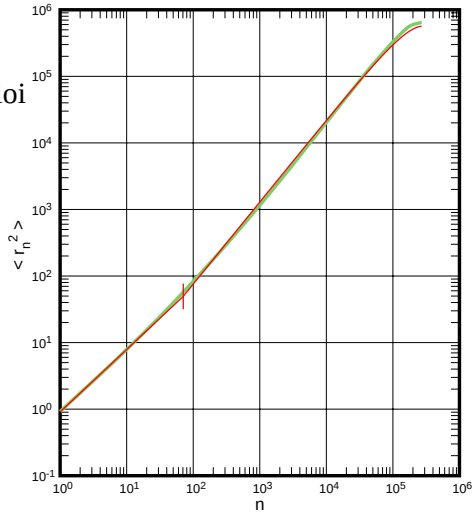
bovary



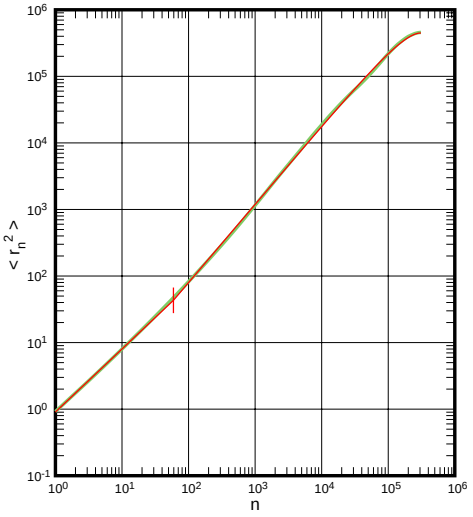

methode



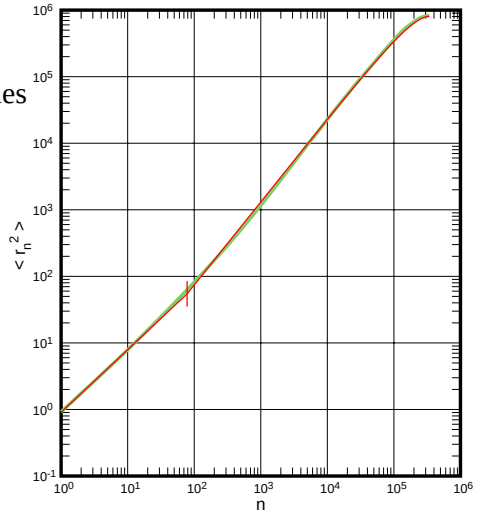

loi



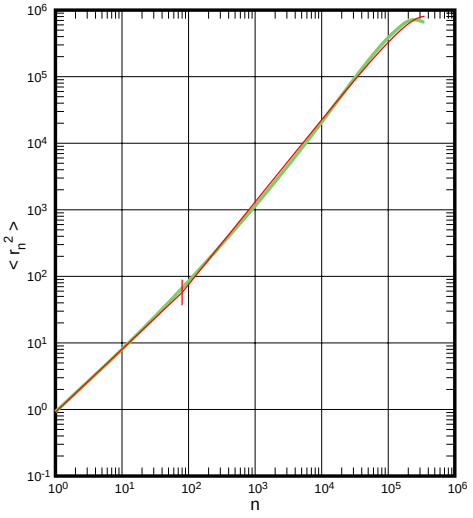

liaisons




20000lieues

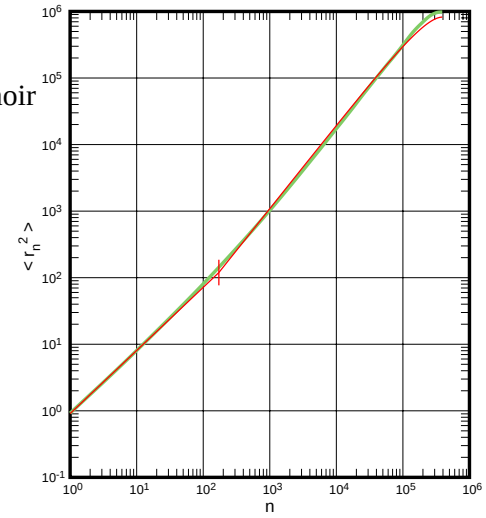



arc

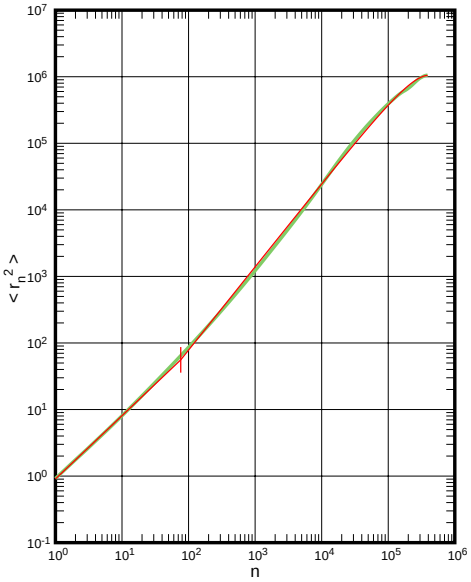


French texts: 780,000 < N < 1,050,000

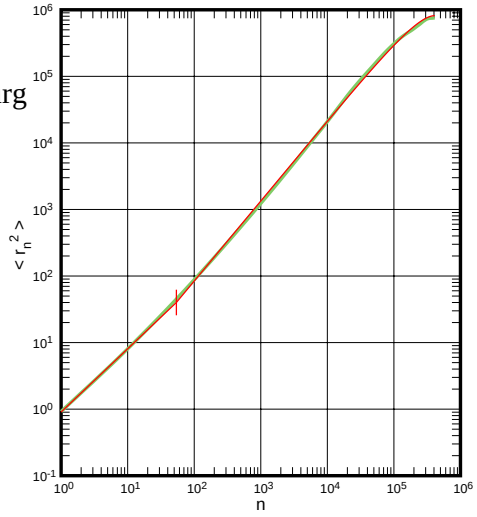
rouge_noir



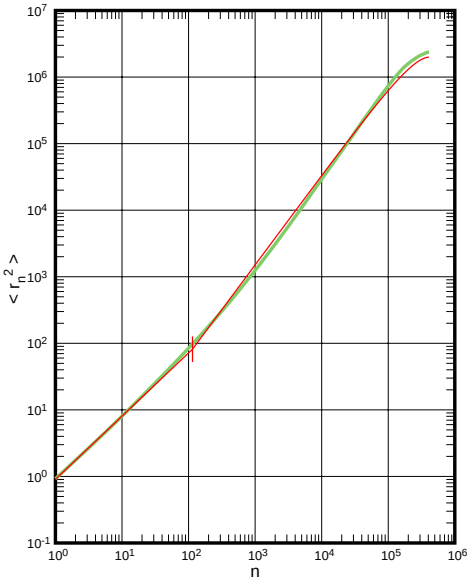

notre_dame



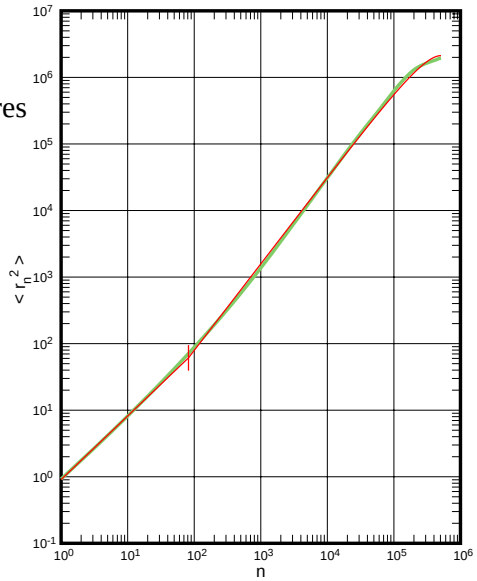

petersbourg



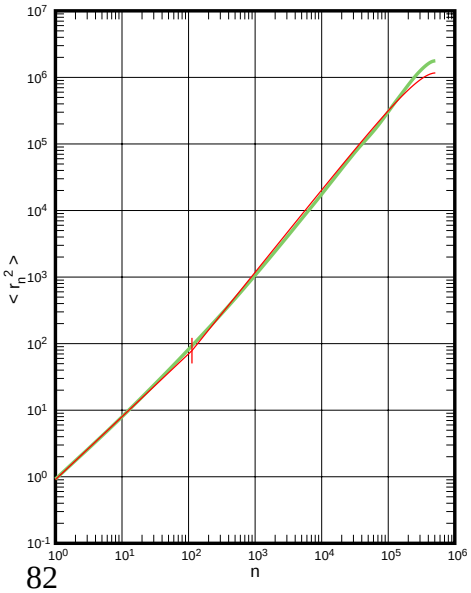

ethiopie




3mousquetaires

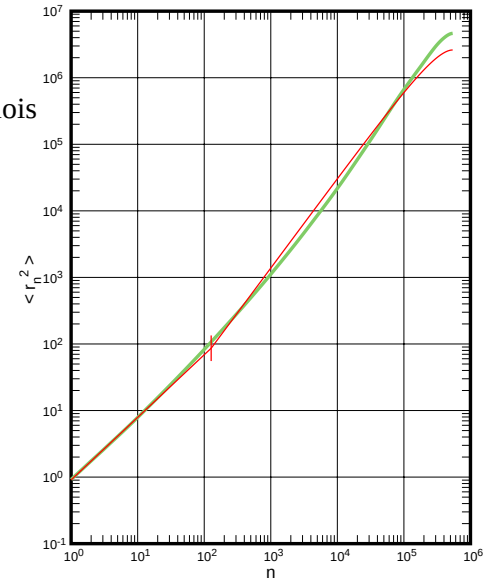



berlioz

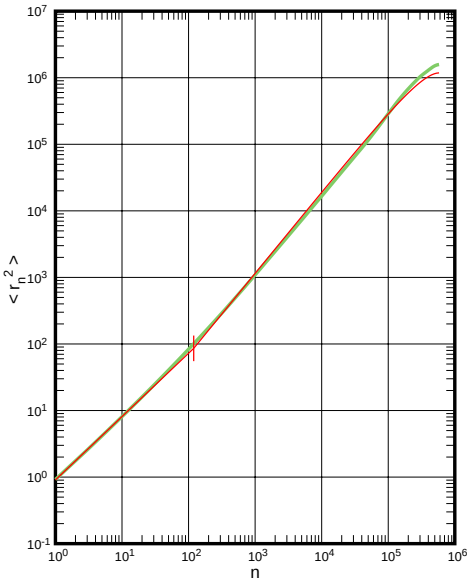


French texts: 1,090,000 < N < 2,300,000

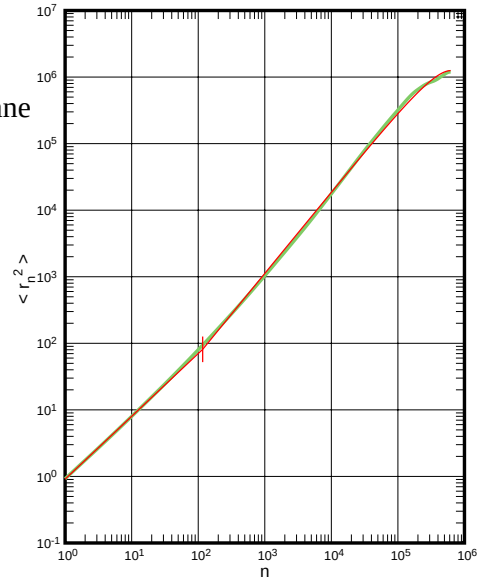
gaulois



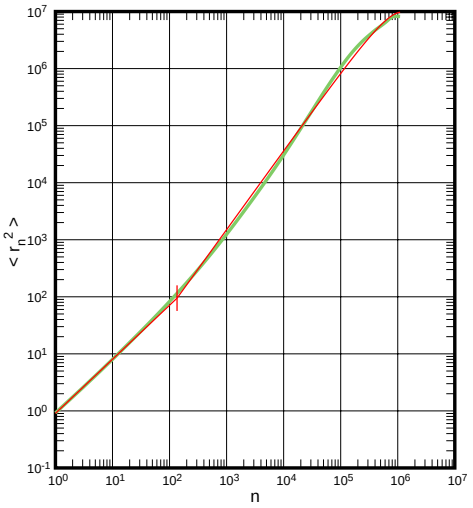

confessions




santillane

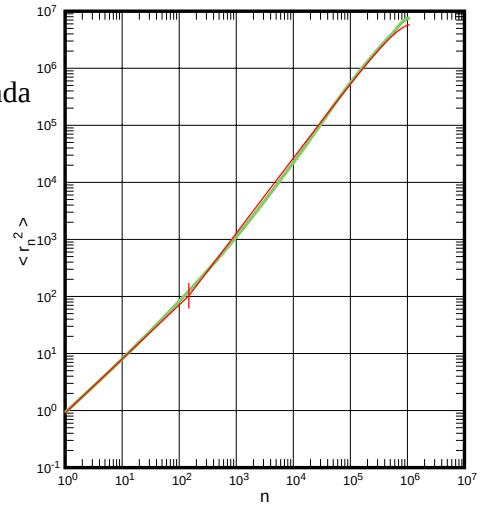



constant

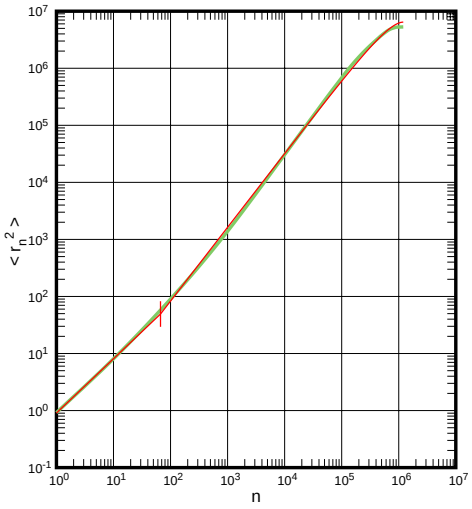


French texts:  2,300,000 < N < 3,500,000

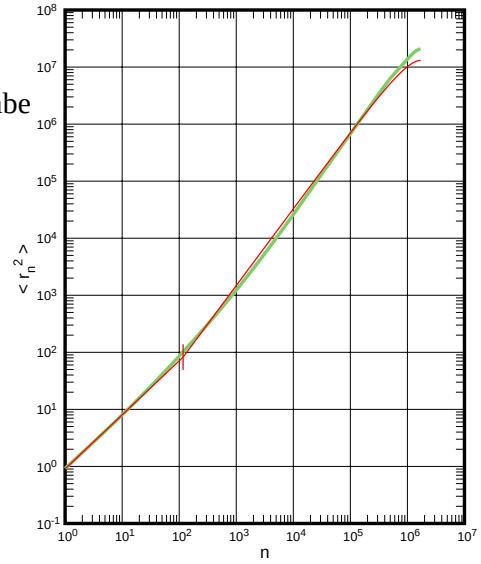
canada




lesmis

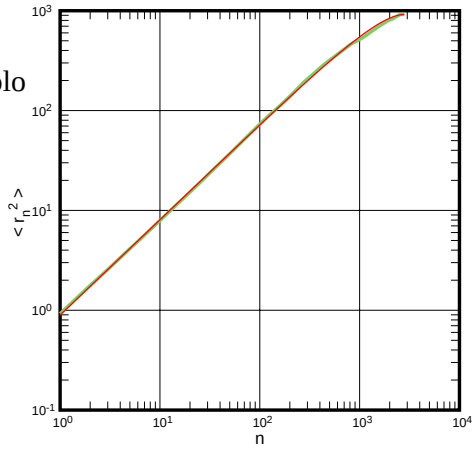


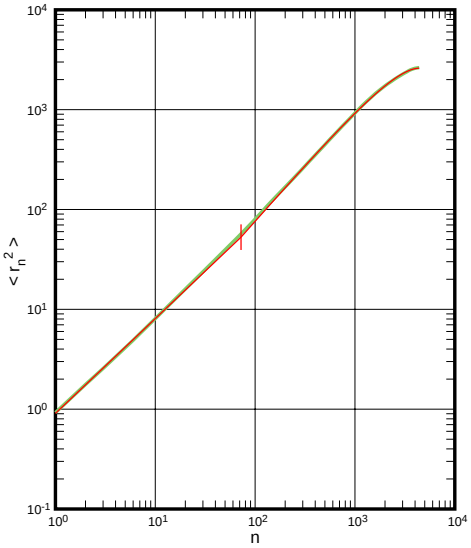

outre_tombe



German texts: 5600 < N < 27,000

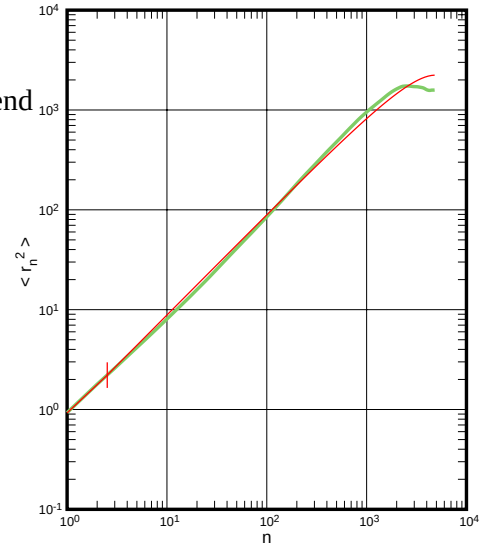
lolo

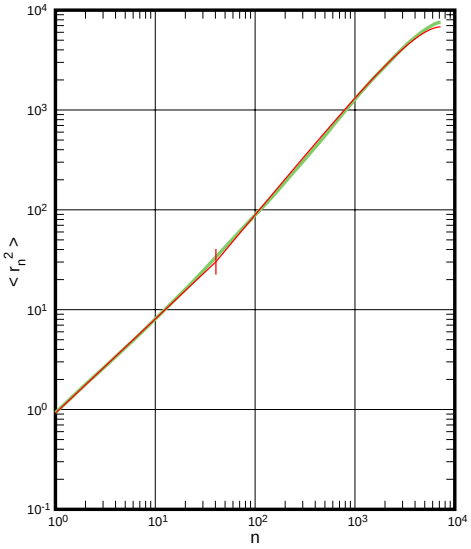



mozart

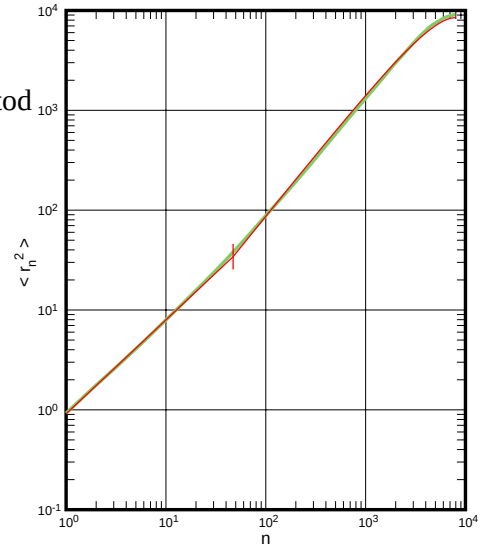
abend

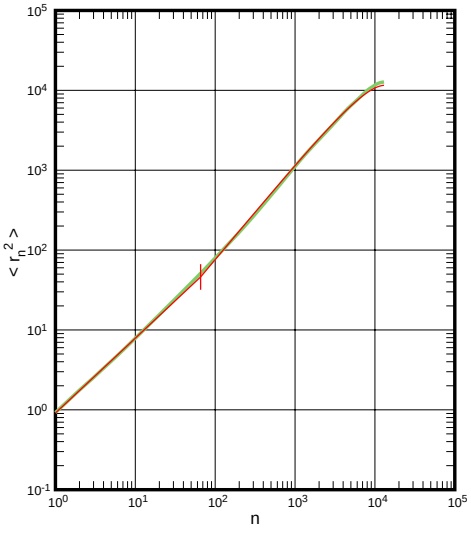



struwwelpeter

tod

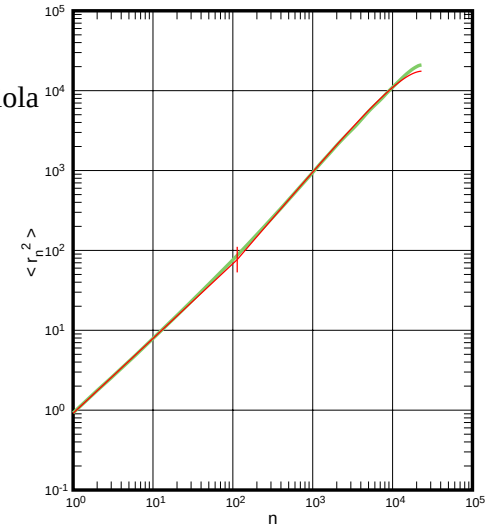


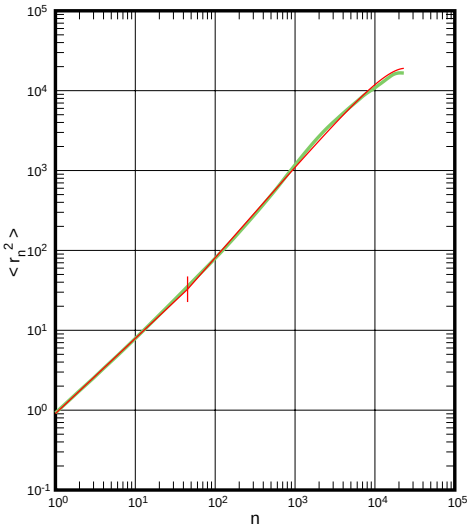

licht

German texts: 47,000 < N < 77,000

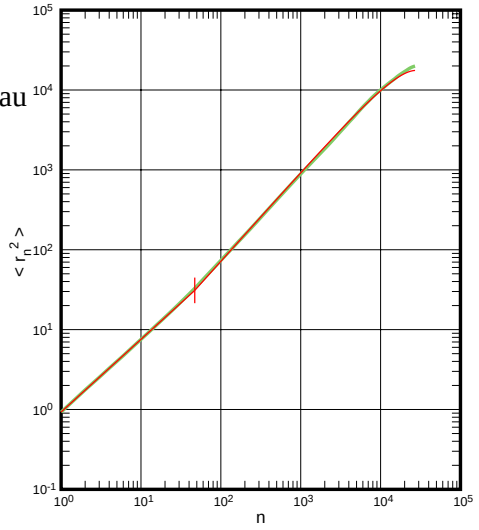
viola



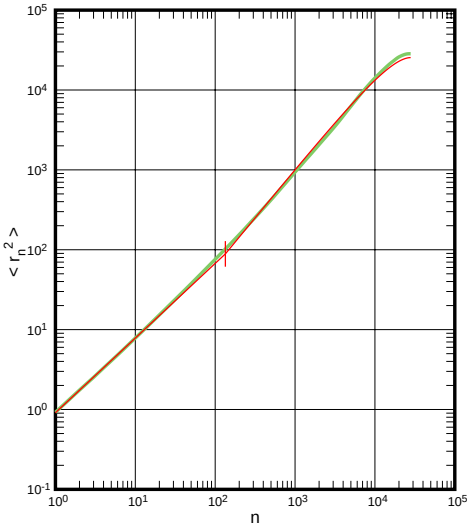

aegypter



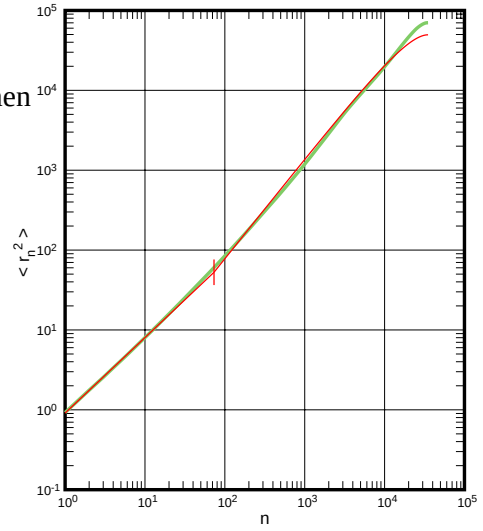

jungfrau



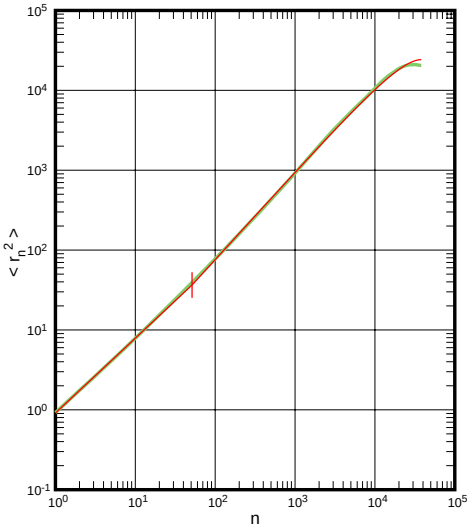

freundin




mythen

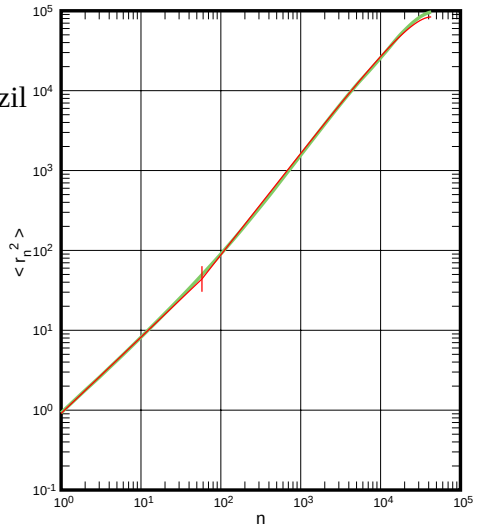



schuss

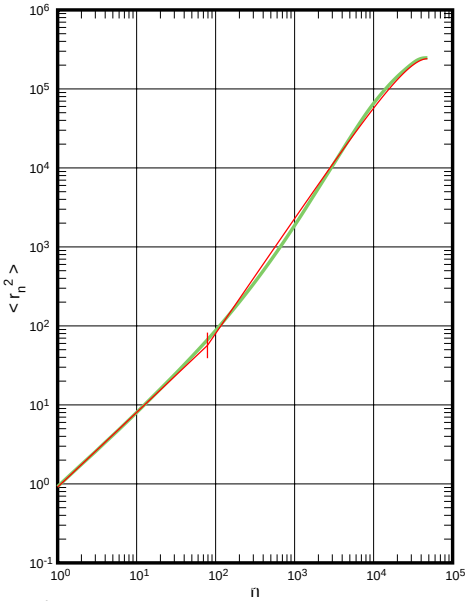


German texts: 86,000 < N < 112,000

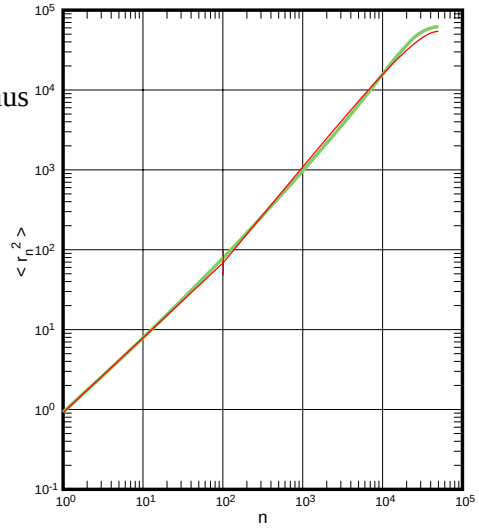
liebeskonzil



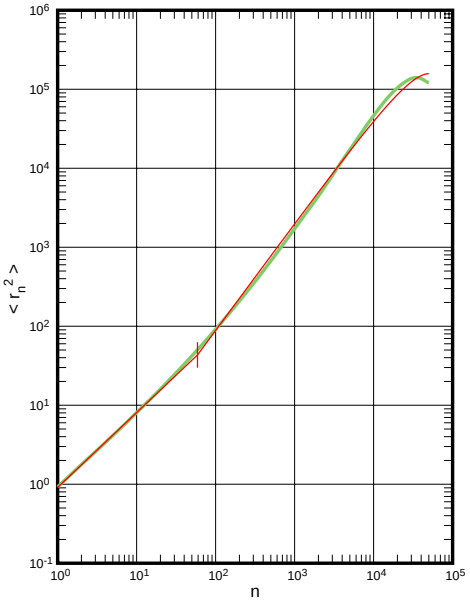

wright



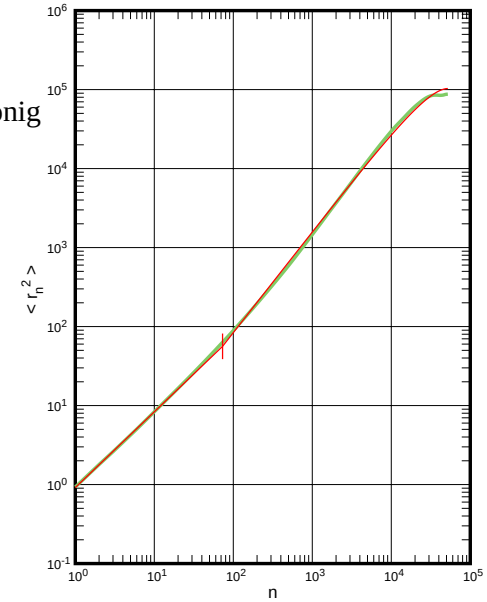

darwinismus



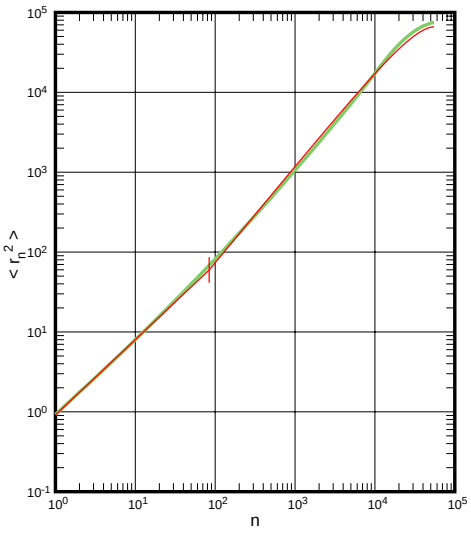

gastmahl




alpenkonig

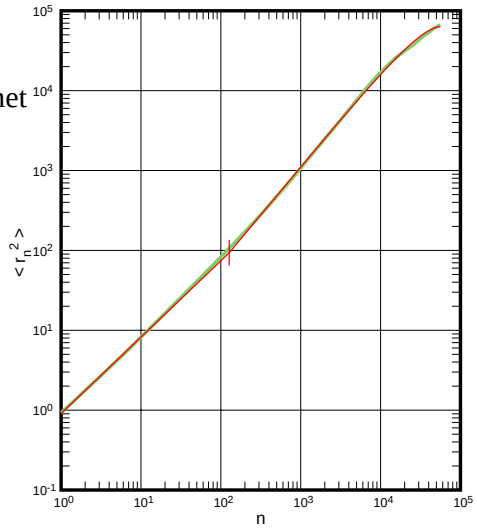



tonio

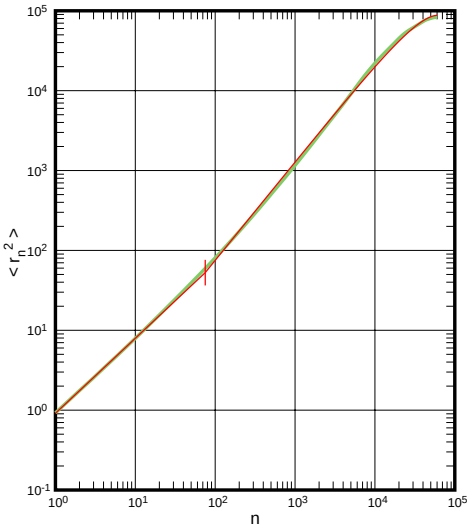


German texts: 115,000 < N < 160,000

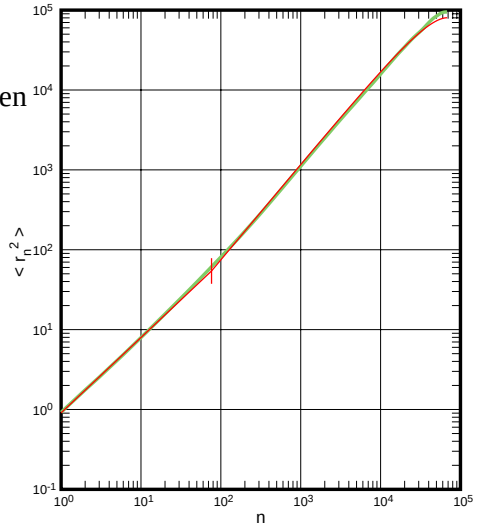
komet



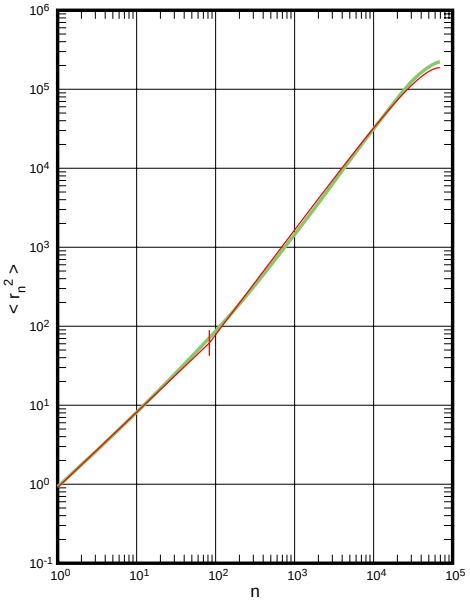

abendmahl



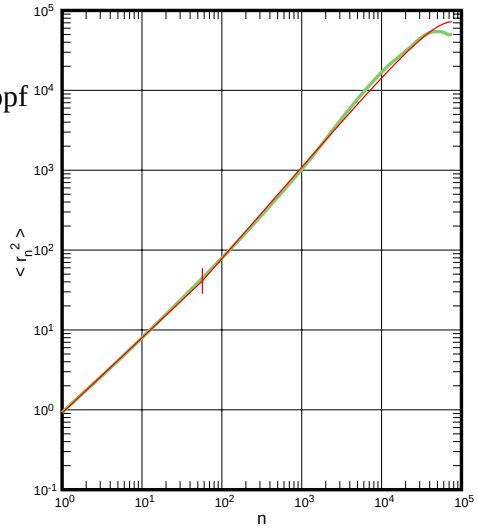

geschichten



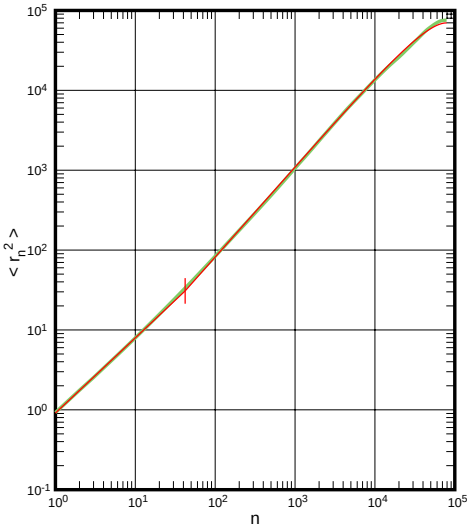

homo




topf

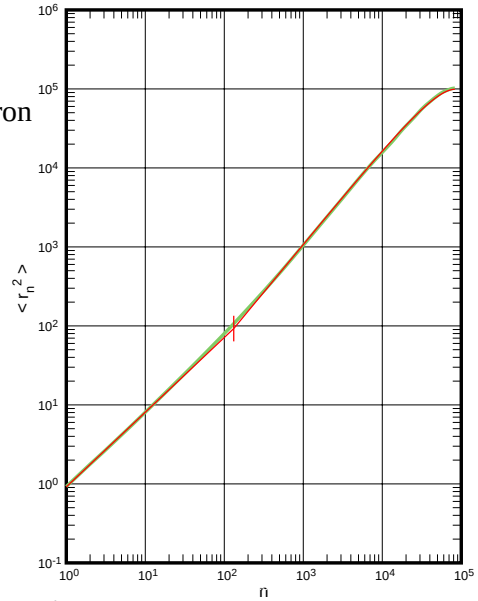



EcceHomo

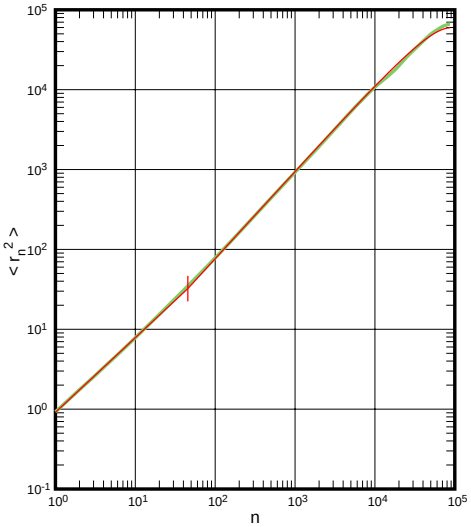


German texts: 167,000 < N < 189,000

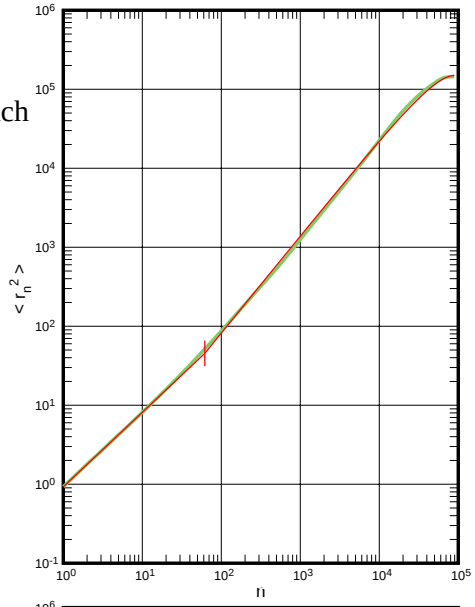
redbaron



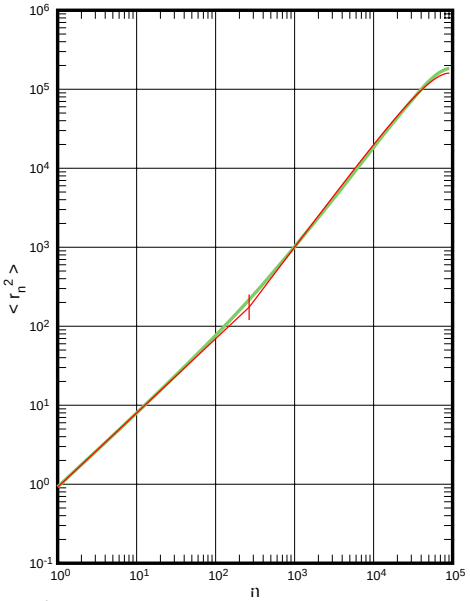

rembrandt



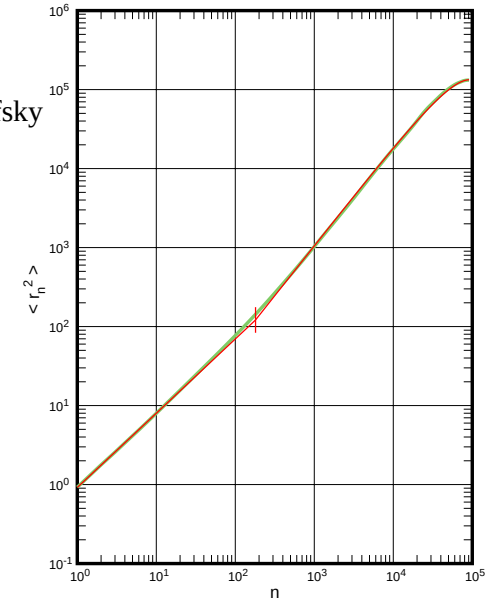

kochbuch



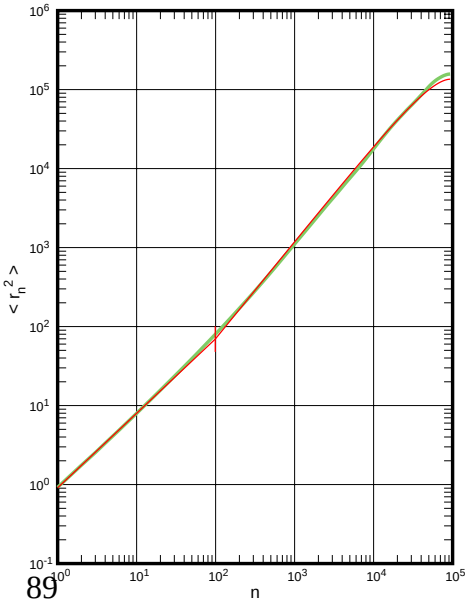

wunderlampe




Schelmuffsky

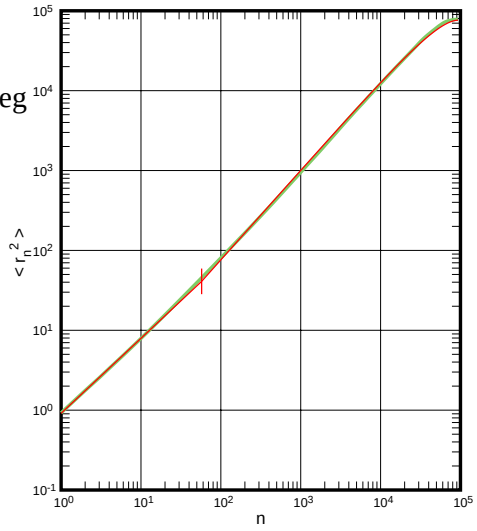



freud

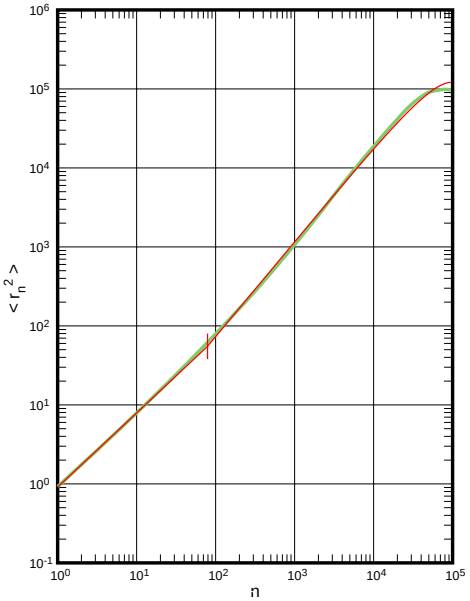


German texts: 189,000 < N < 275,000

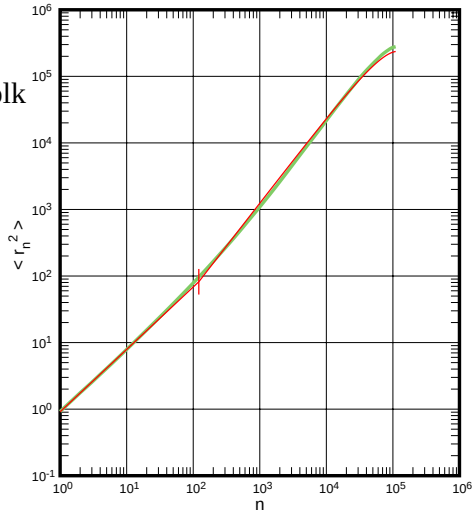
weg



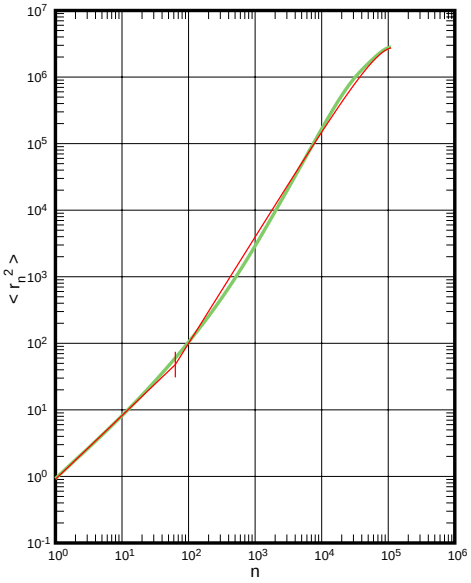

rousseau



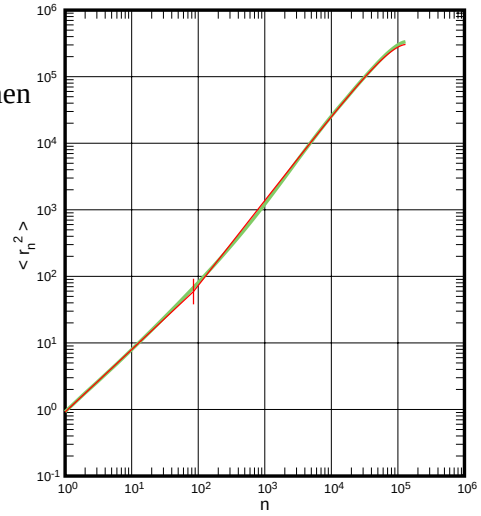

himmelsvolk




dichtungen

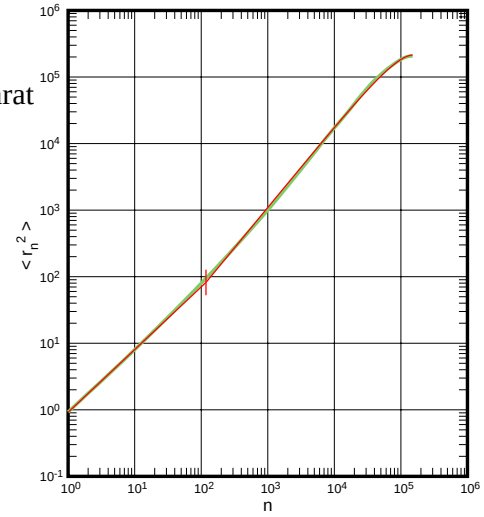


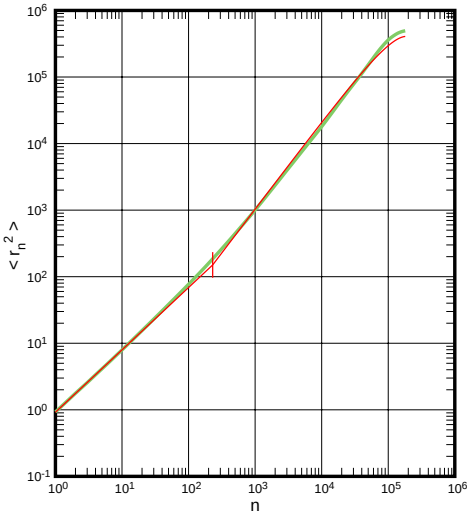

marchen



German texts: 306,000 < N < 422,000

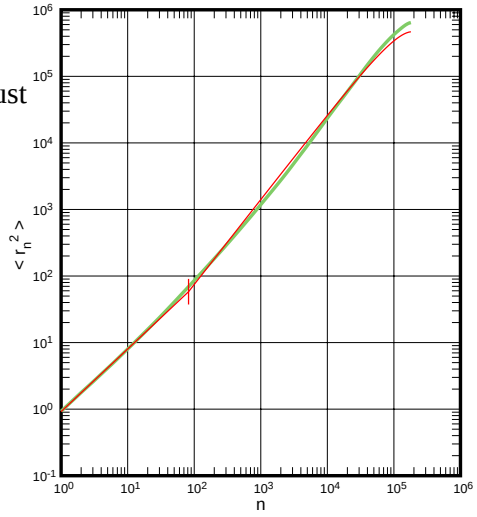
unrat

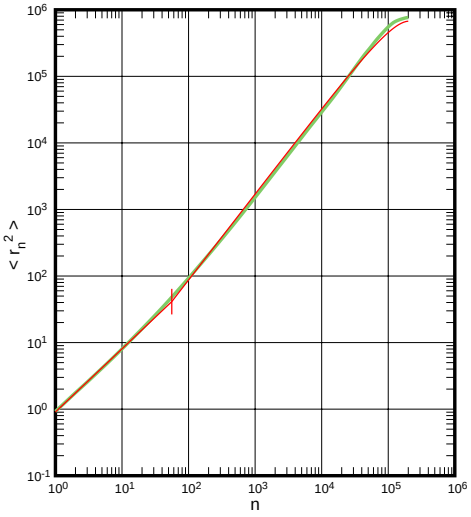



celsissimus

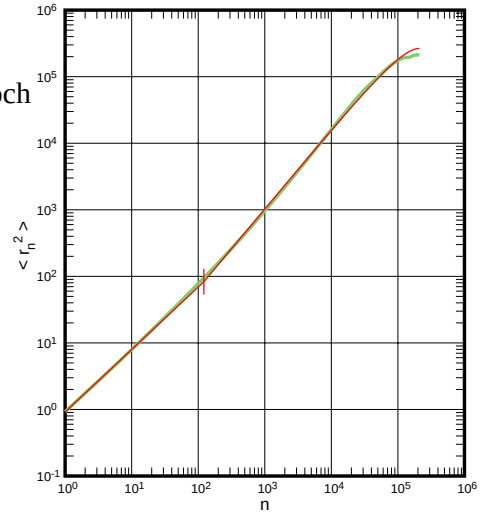
faust



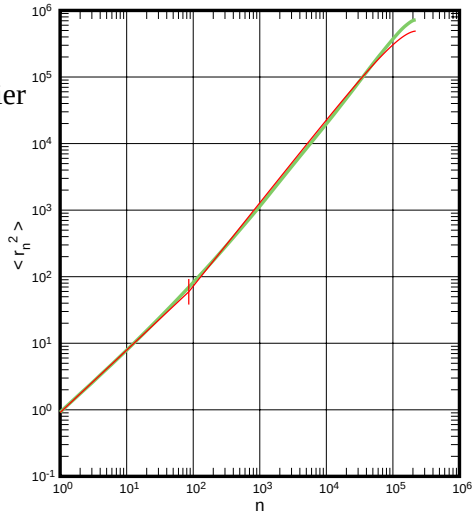

zarathustra

moloch

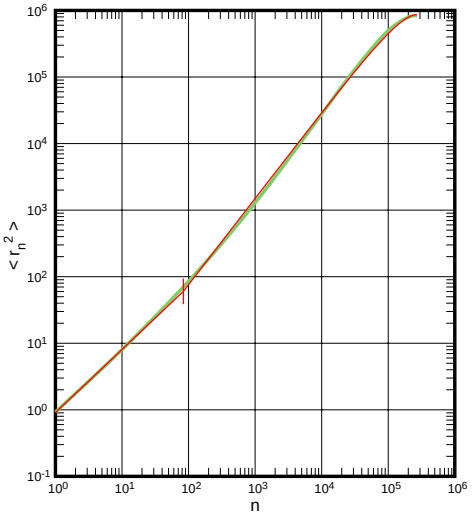


German texts: 458,000 < N < 640,000

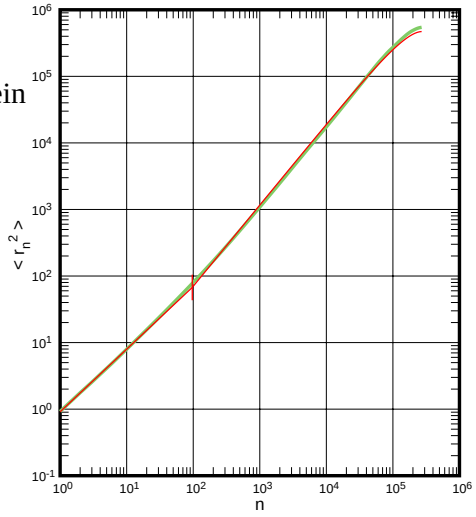
fourier



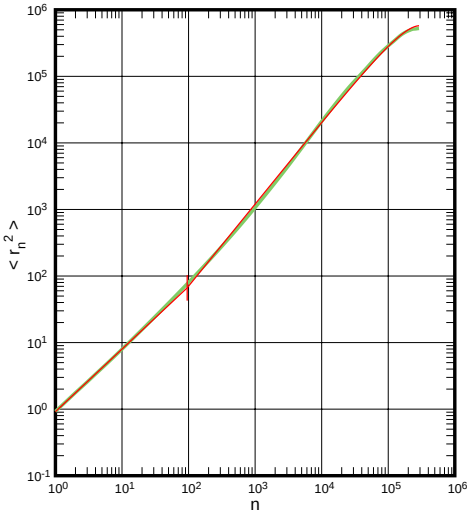

jagor



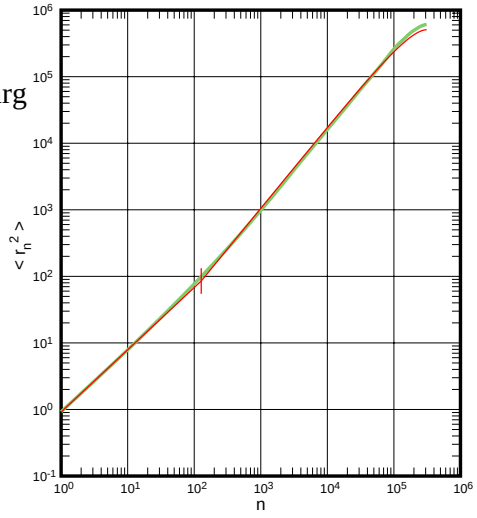

lichtenstein




hoheit

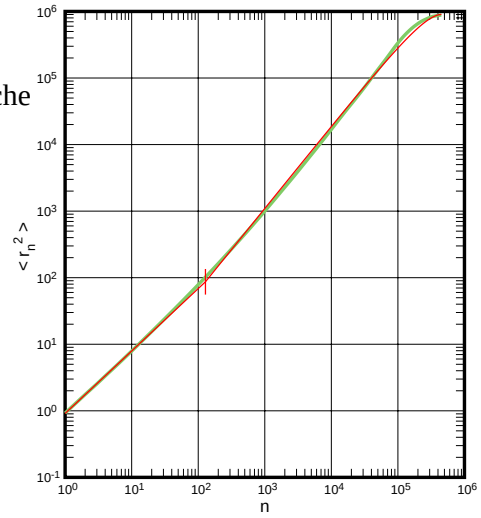


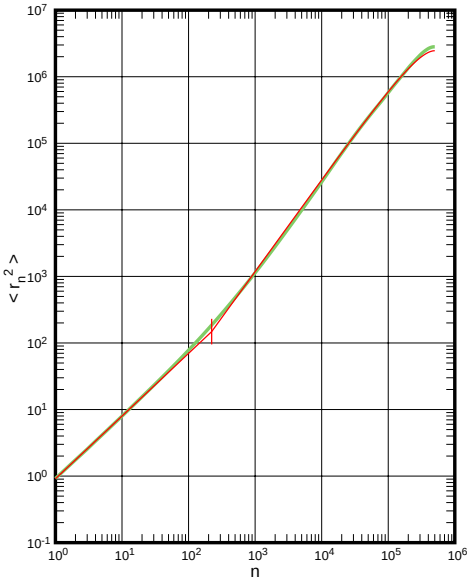

hindenburg



German texts: 915,000 < N < 1,850,000

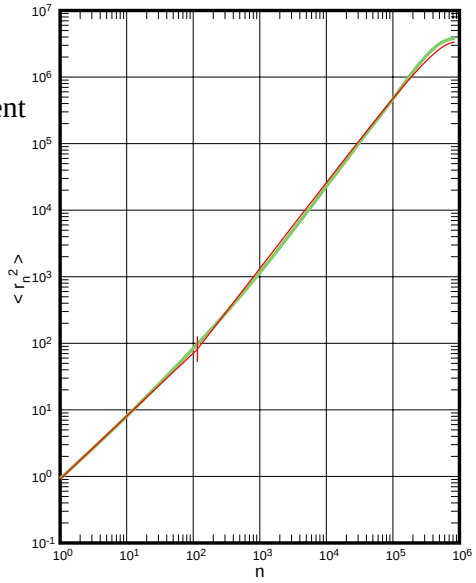
italienische

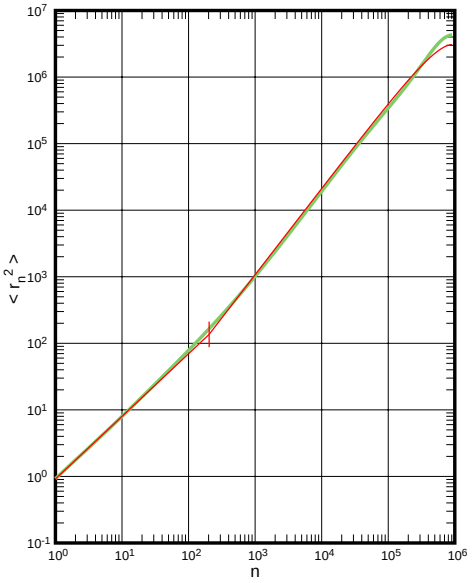



kritik

continent

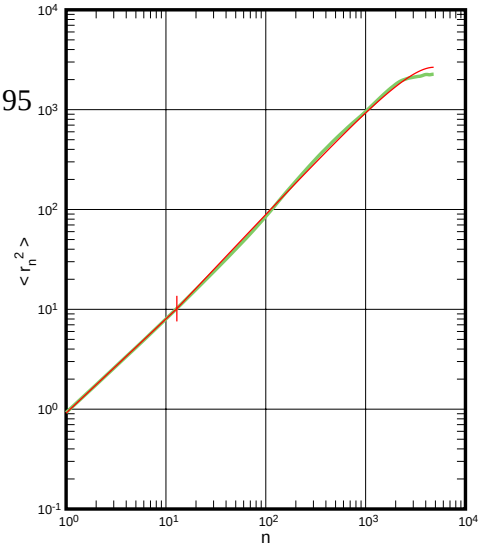


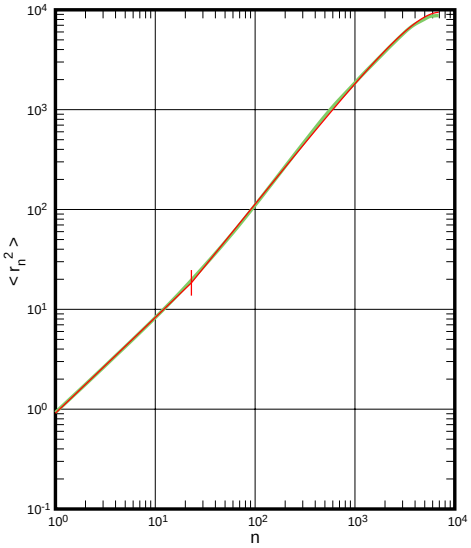




amerika_bang

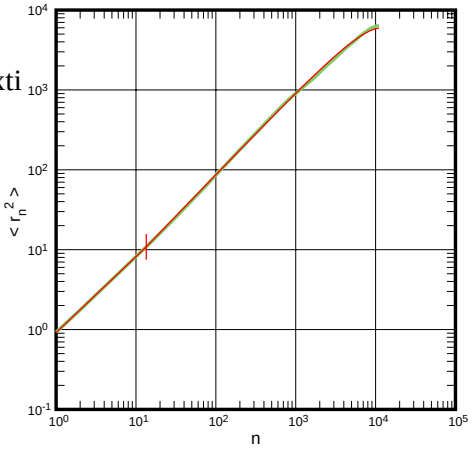
Latin texts: 9600 < N < 35,000

95

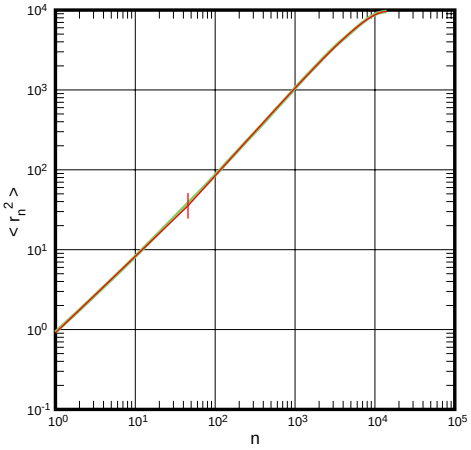


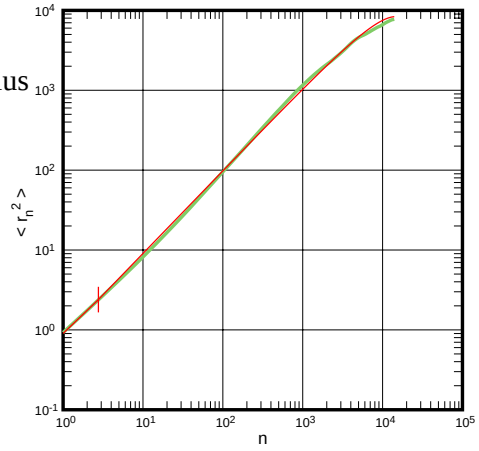

magna



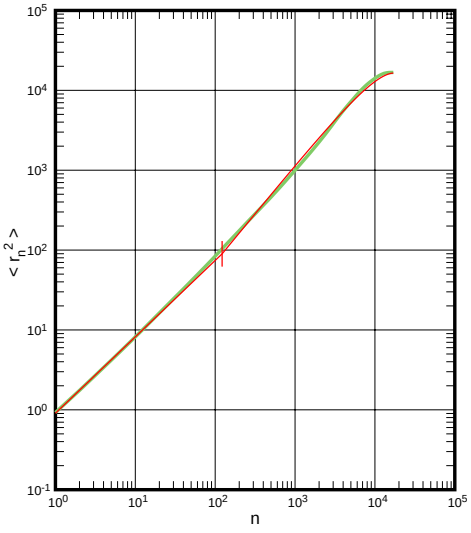

sexti




encomium



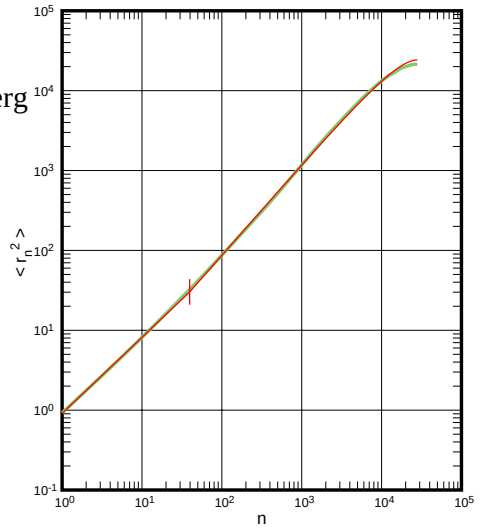

caius



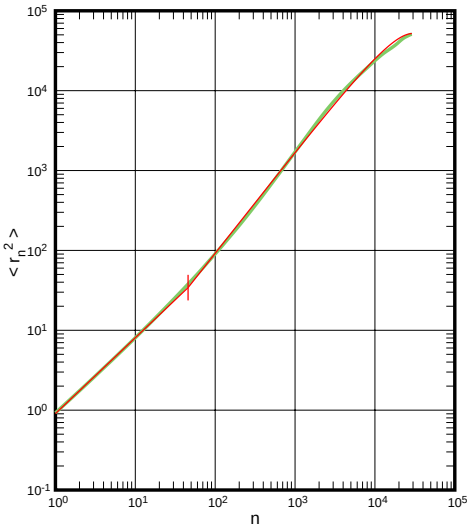

esther

Latin texts: 56,000 < N < 170,000

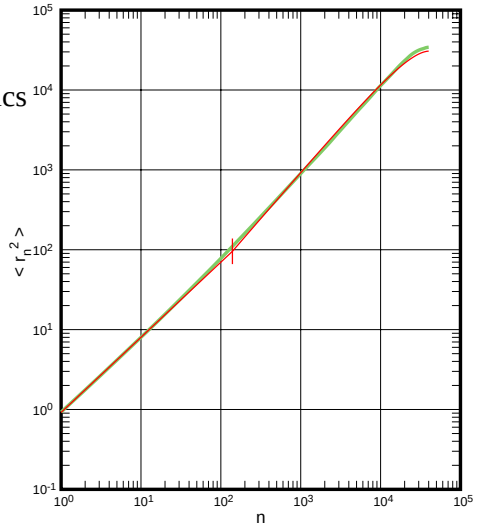
augsberg



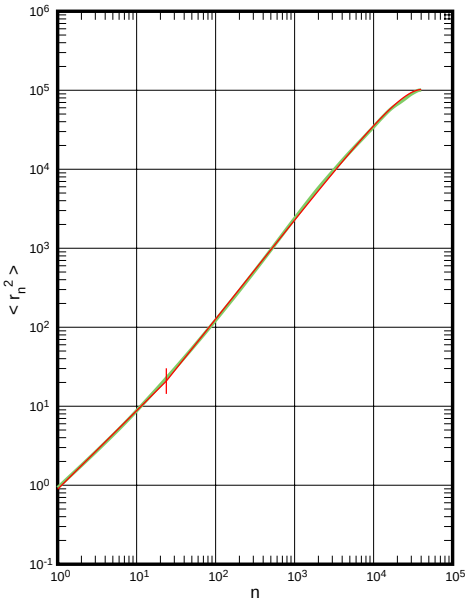

danielis



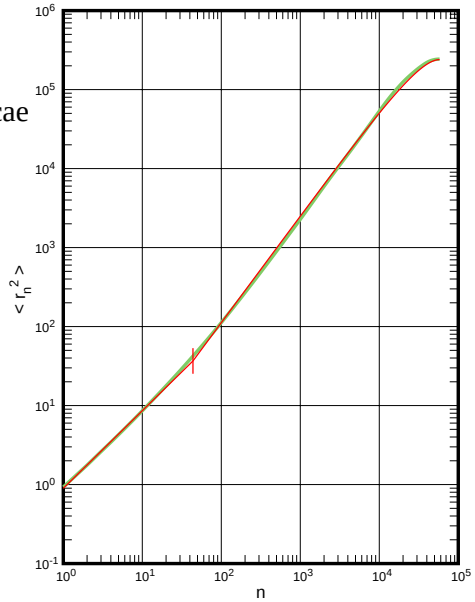

georgics



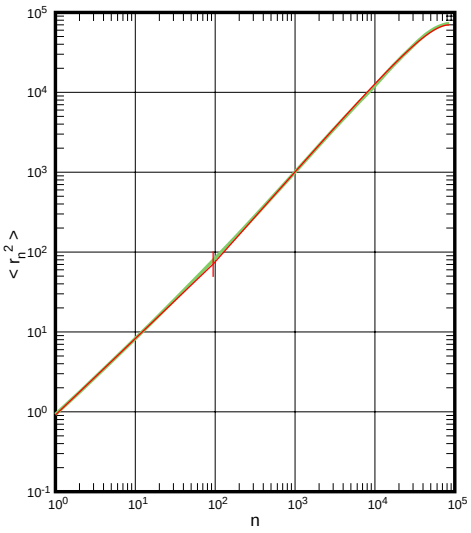

beze




iaponicae

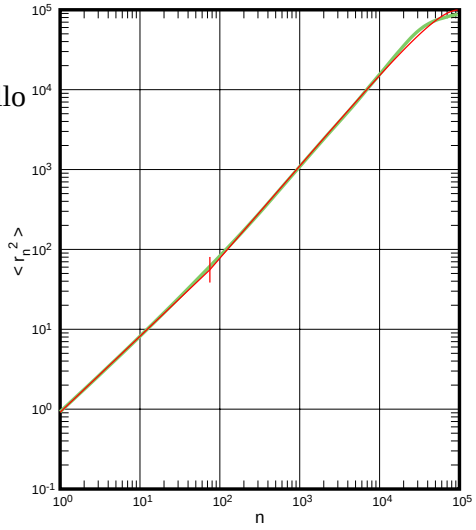



fasti

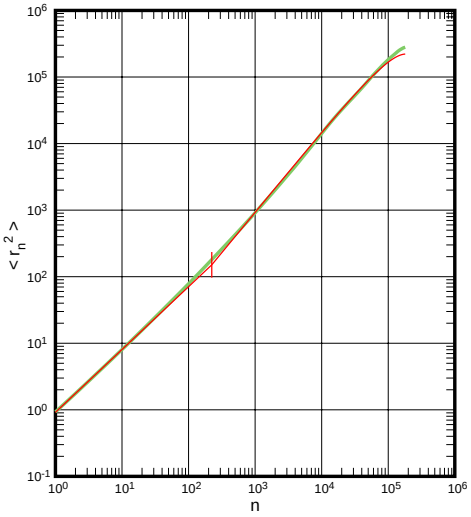


Latin texts: 194,000 < N < 2,100,000

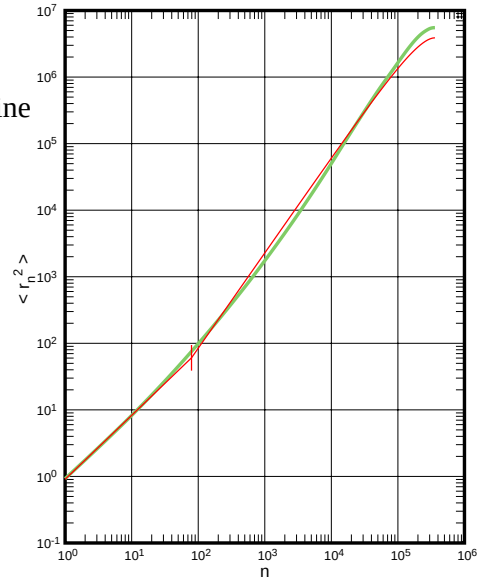
bello



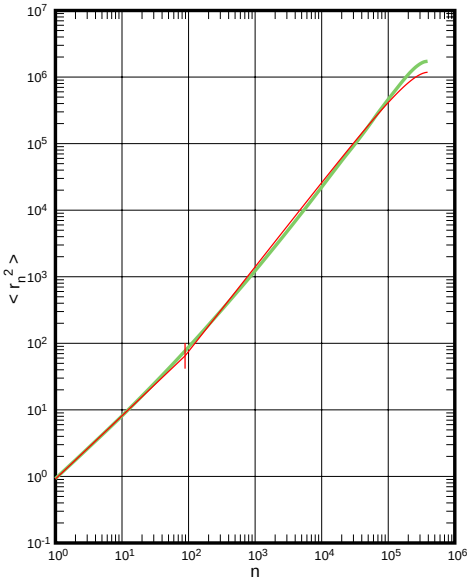

aeneidos



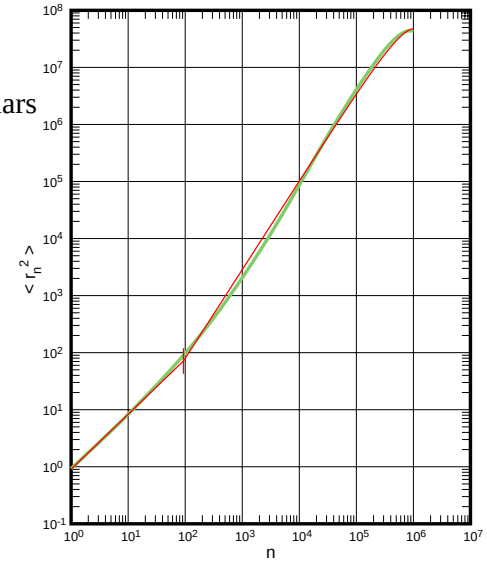

augustine




bodin

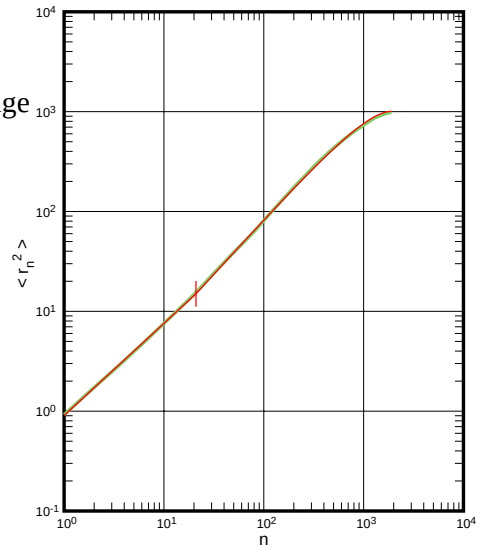



templars

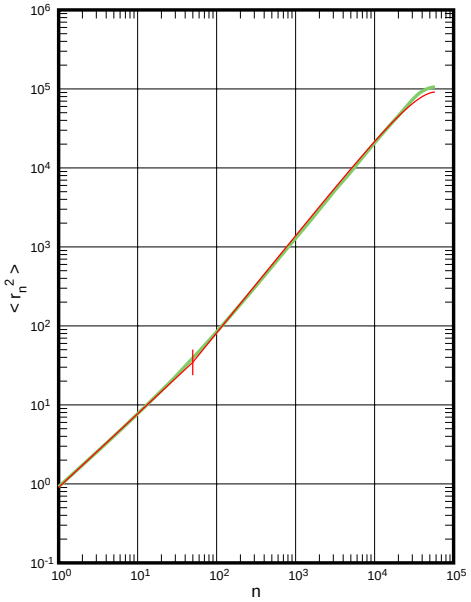


Other languages:  3,800 < N < 452,000

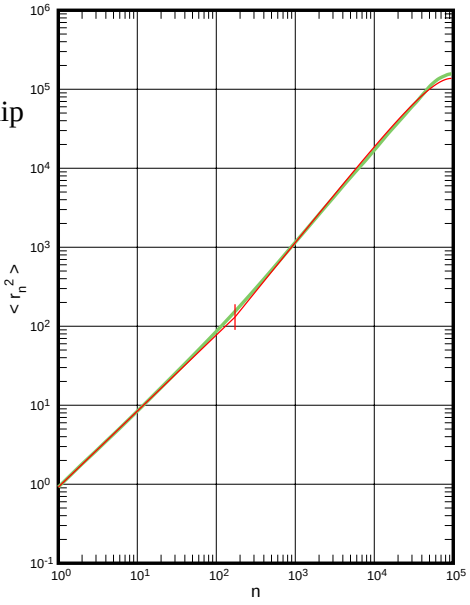
conge



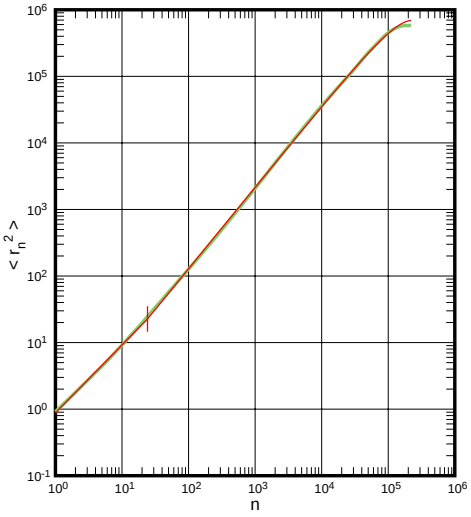

roland




luchtschip

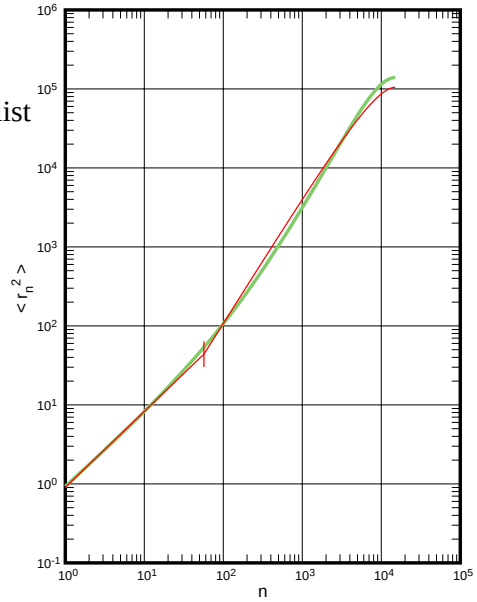



kalevala

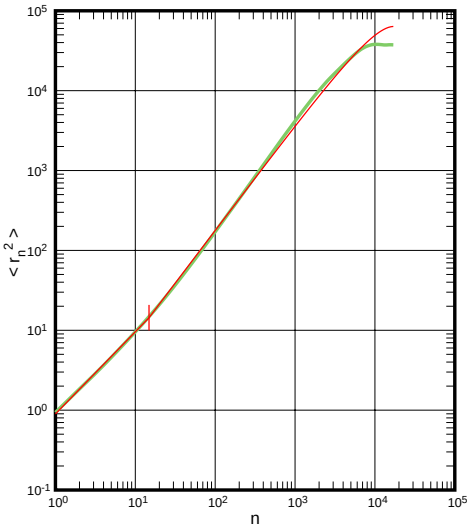


Mixed-language texts: 30,000 < N < 43,000

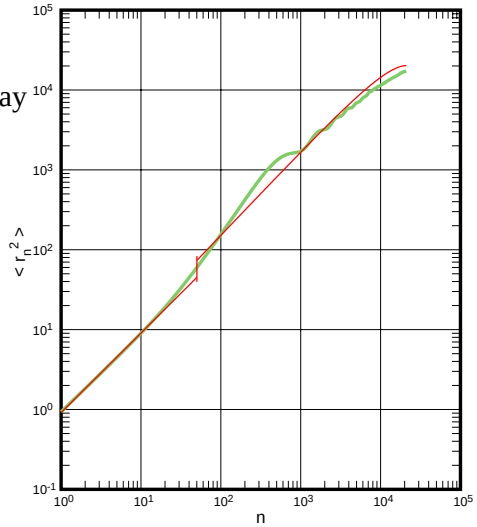
blacklist




chinook

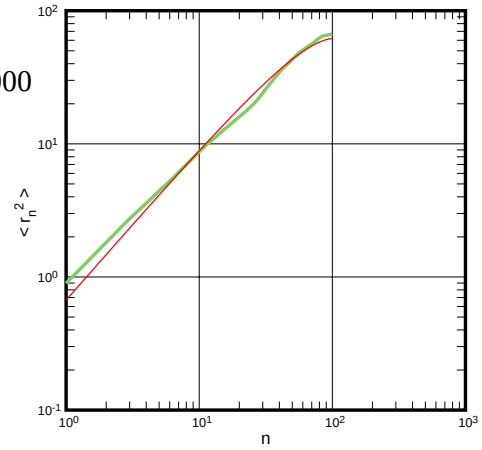


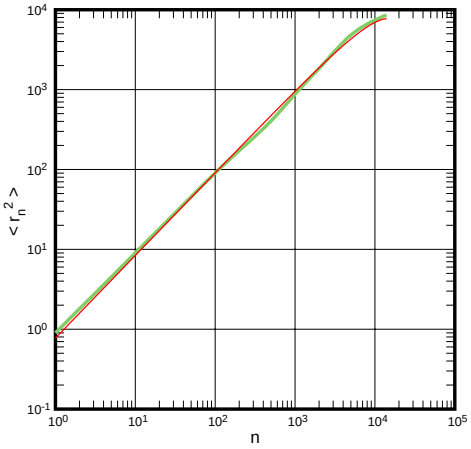

ojibway



Random numerical sequences:  200 < N < 228,000

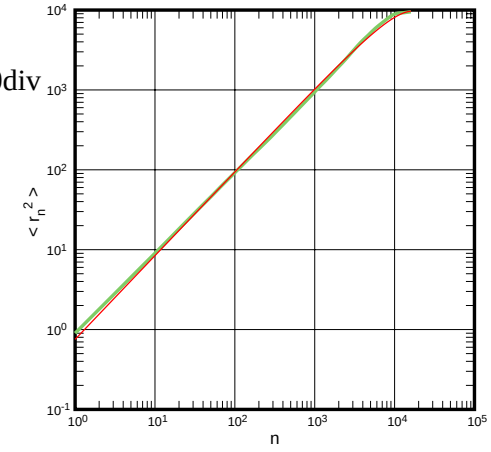
Fib1000

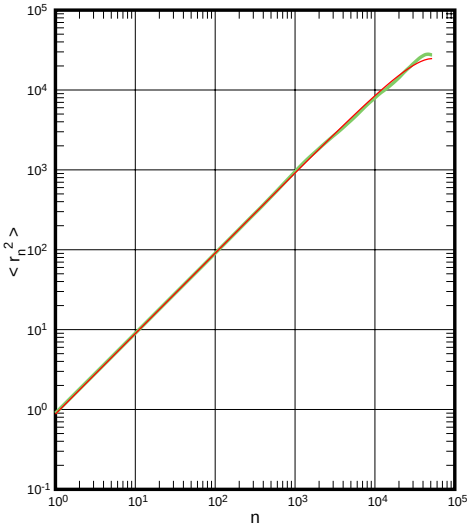



fac9000div

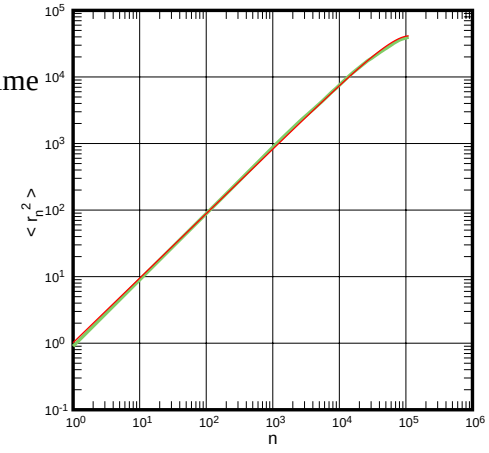
fac10000div



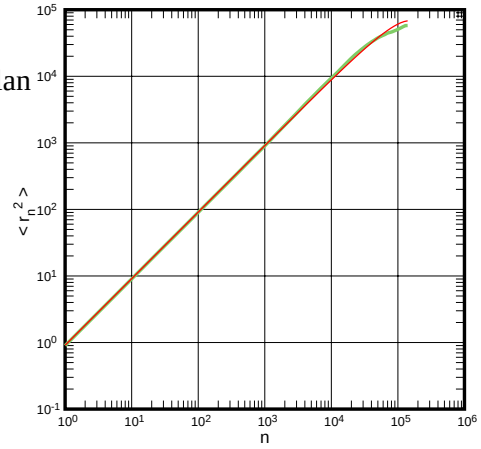

fiblist

32prime

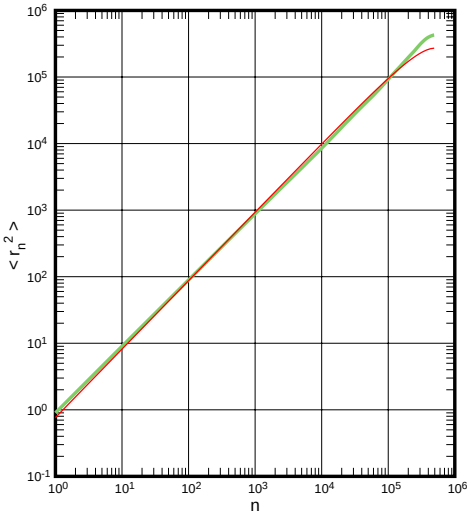


Random numerical sequences:  300,000 < N < 5,000,000

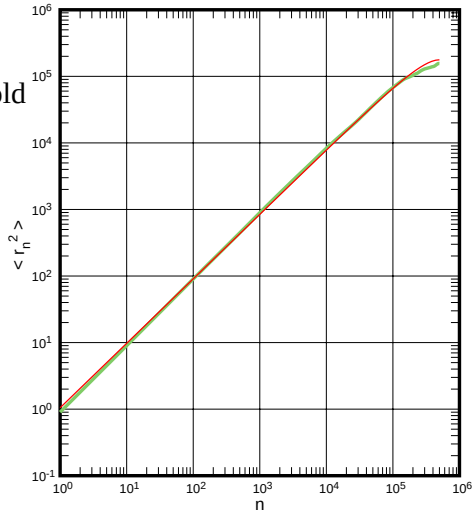
catalan



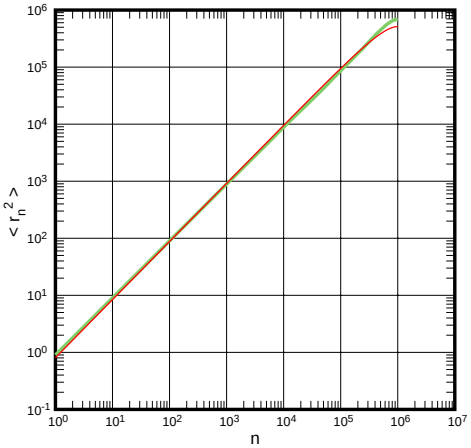

pi



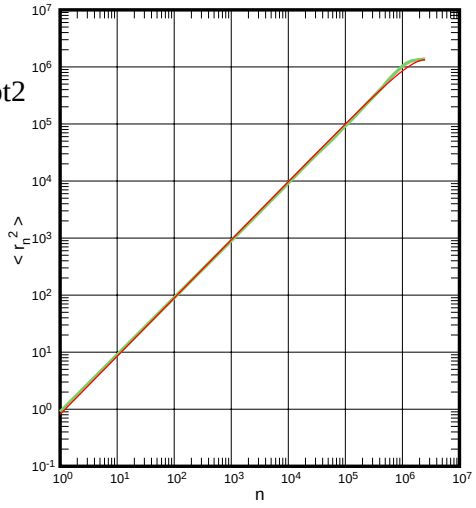

gold




e

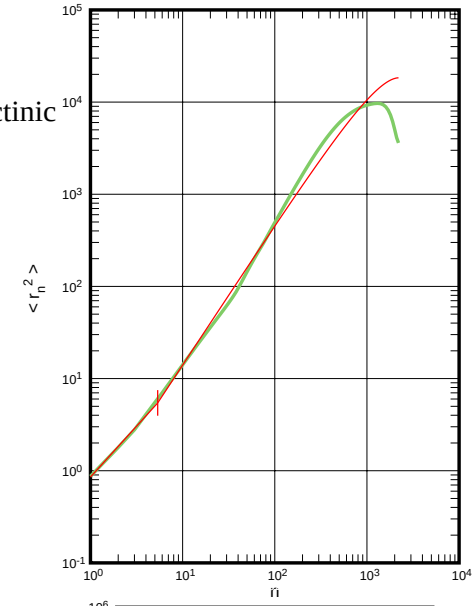



root2

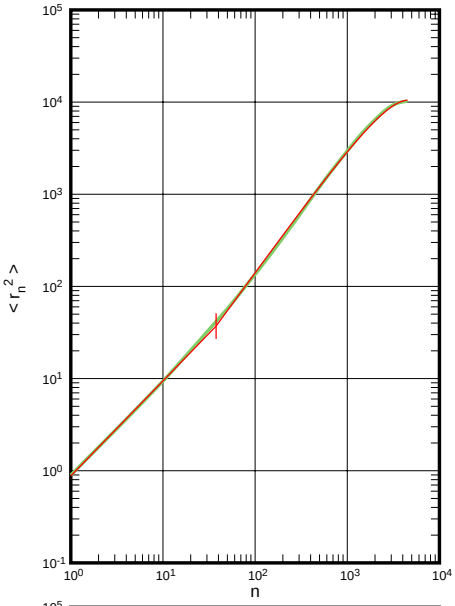


Non-random numerical sequences:  4,000 < N < 32,000

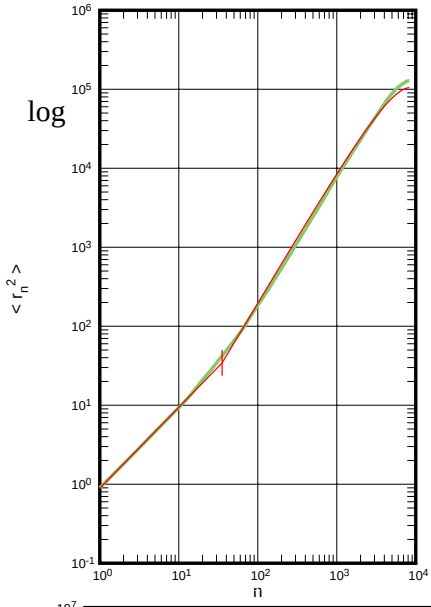
actinic



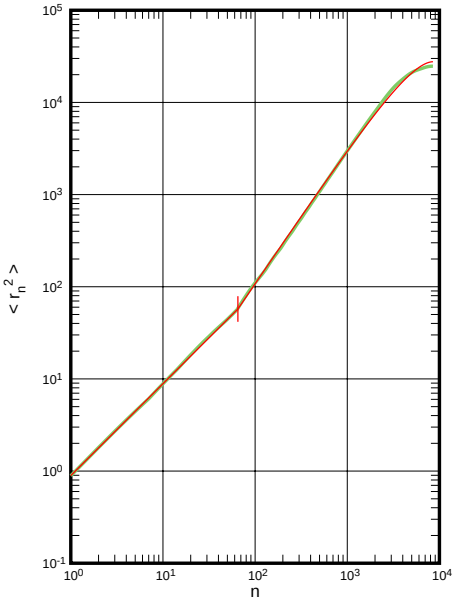

pascal



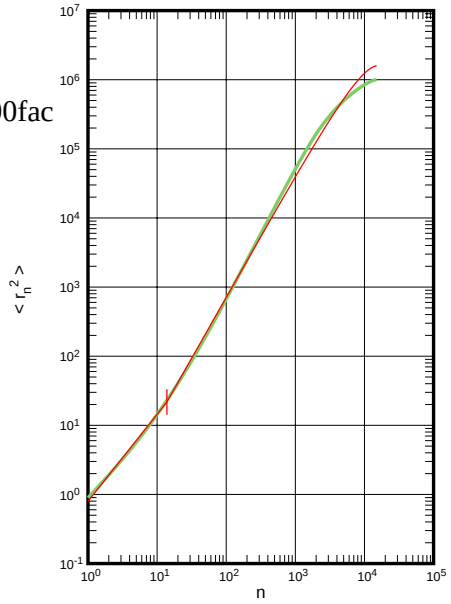

log




trig

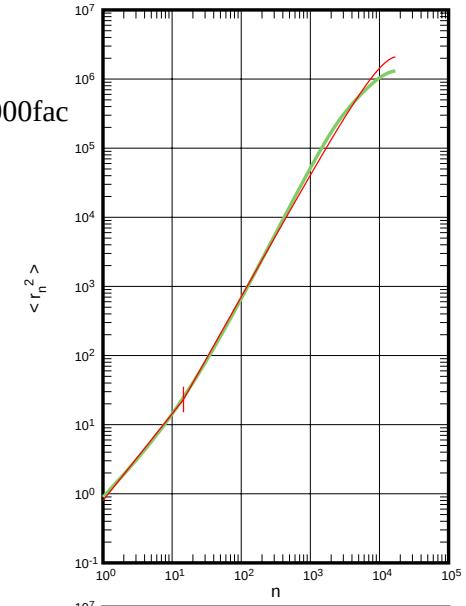



9000fac

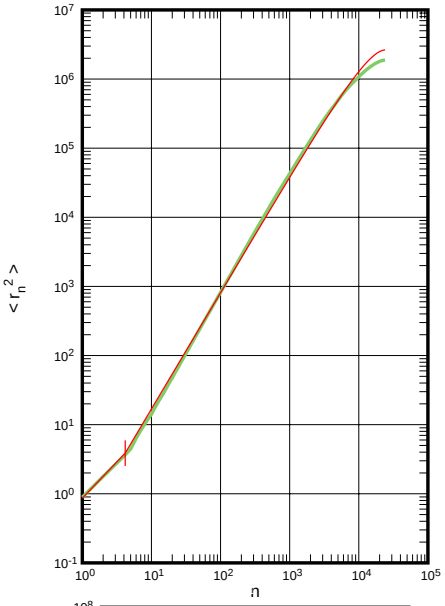


Non-random numerical sequences: 35,000 < N < 1,060,000

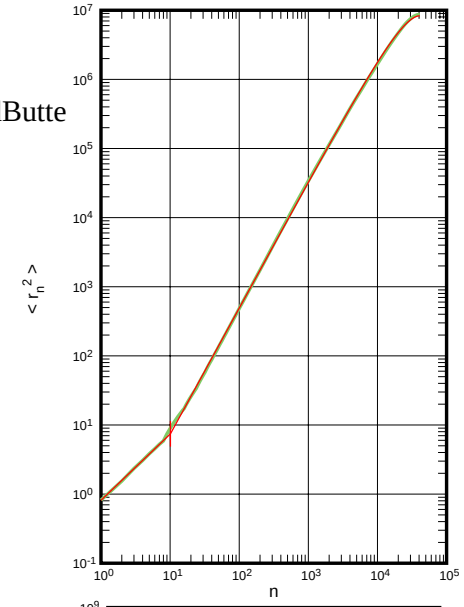
10000fac



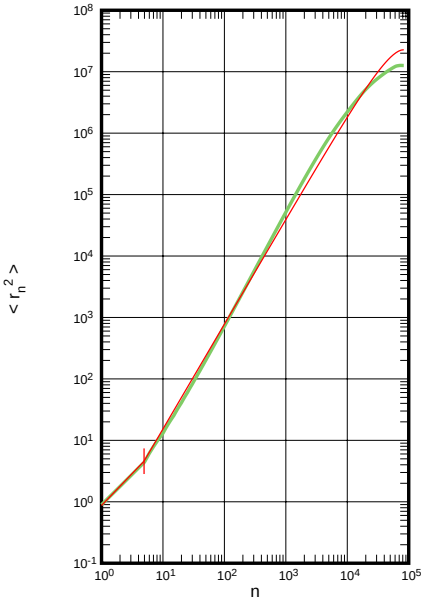

primes



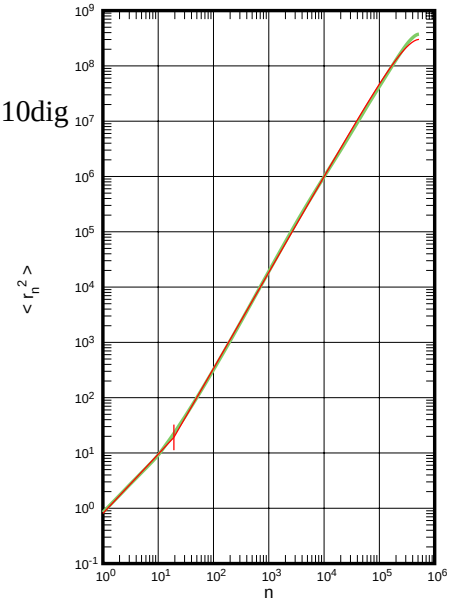

RedButte




zc_dig

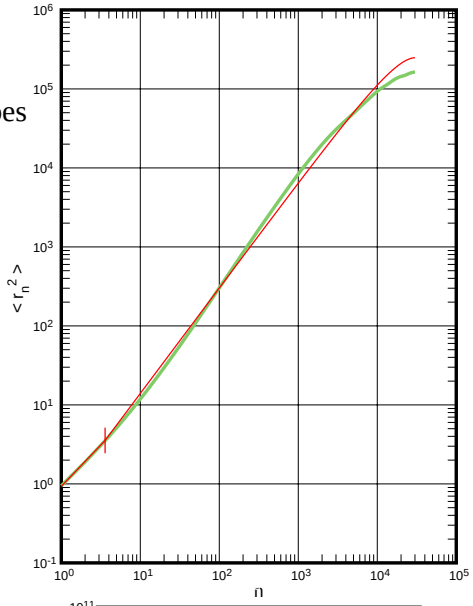



CIA2010dig

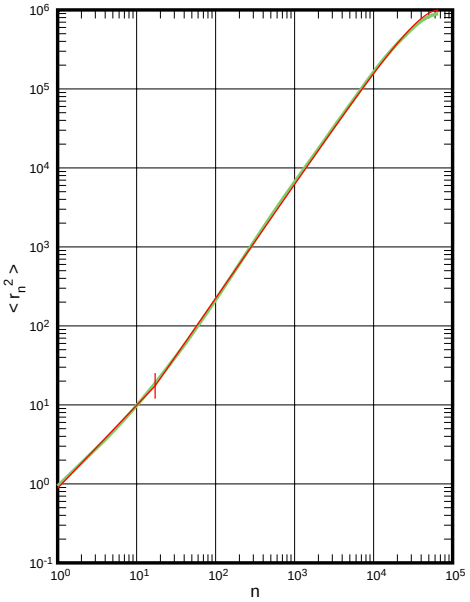


Computer lists:  60,000 < N < 2,700,000

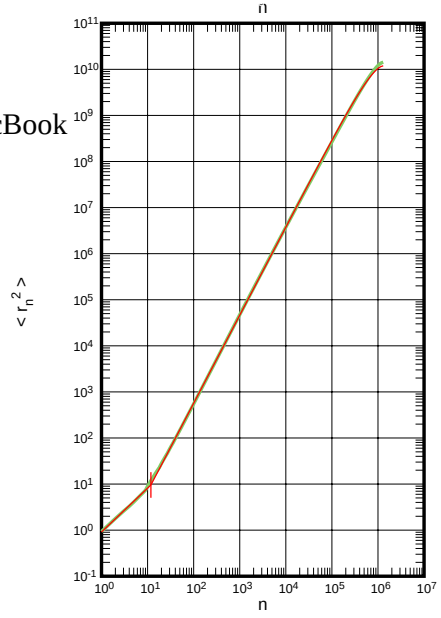
recipes




zeno

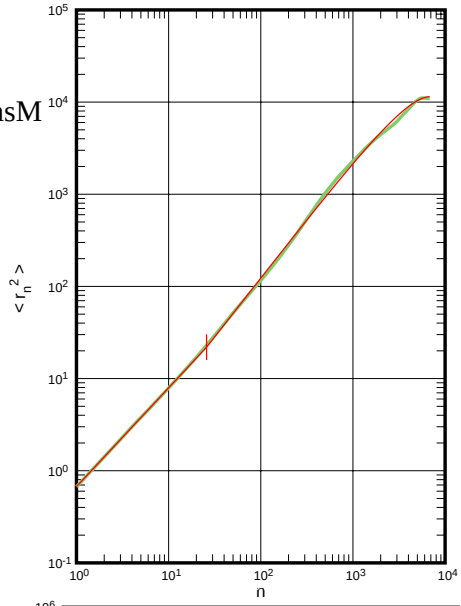



lsRMacBook

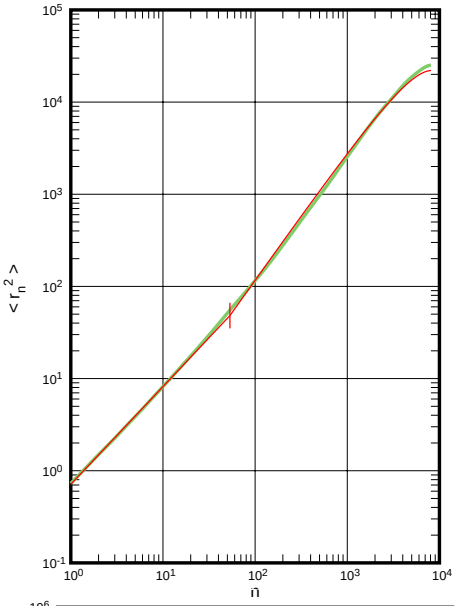


Genetic Codes: 13,000 < N < 270,000

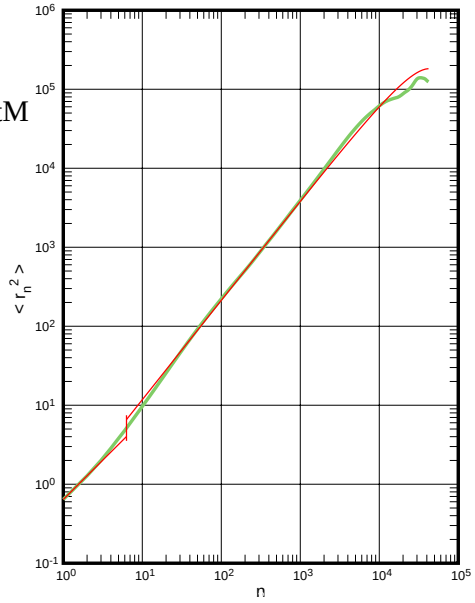
CelegansM



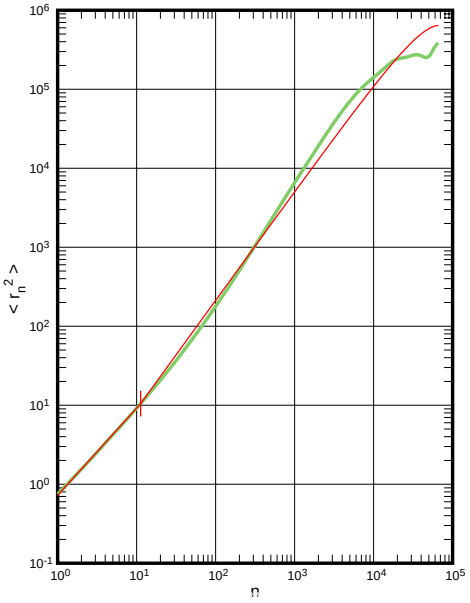

PanMT



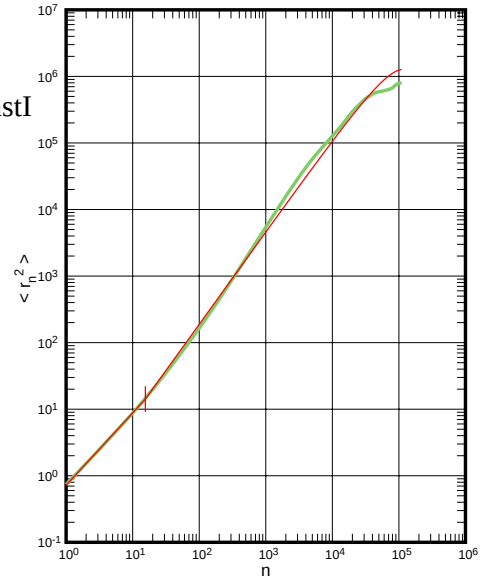

yeastM



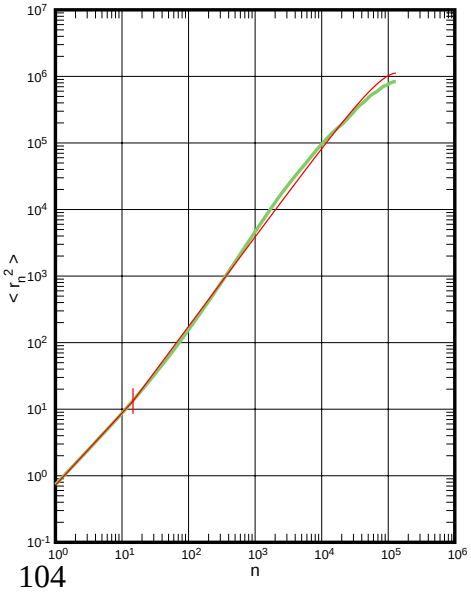

herpes2BJ




yeastI

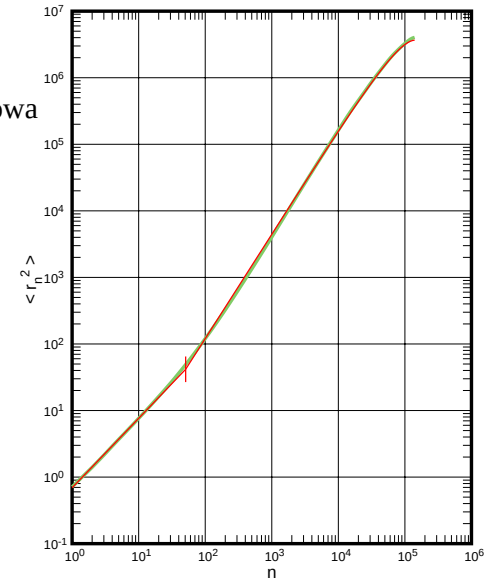



yeastVI

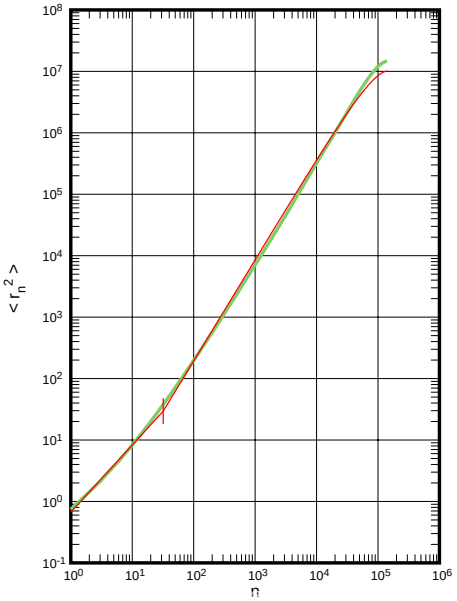


Genetic Codes:  288,000 < N < 580,000

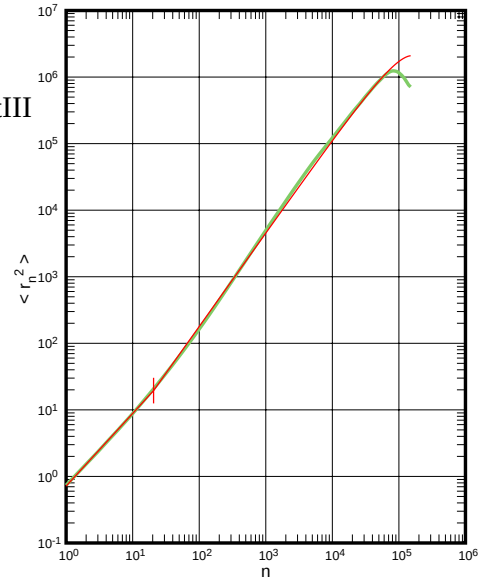
Iowa



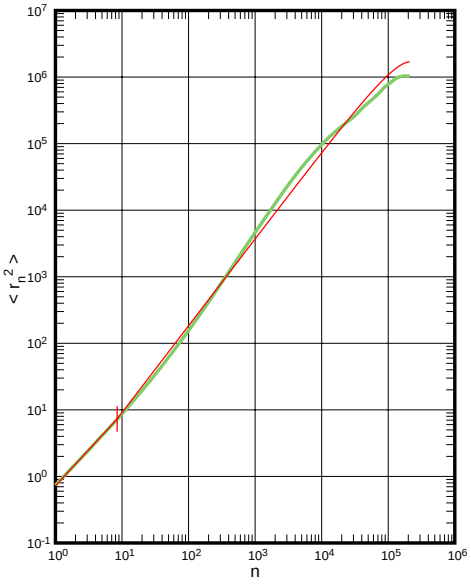

koi



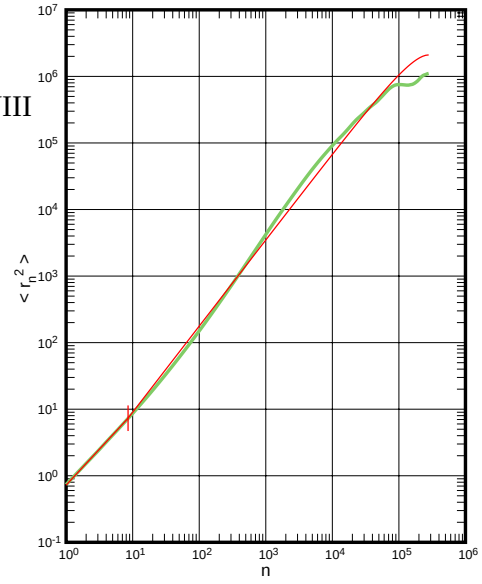

yeastIII



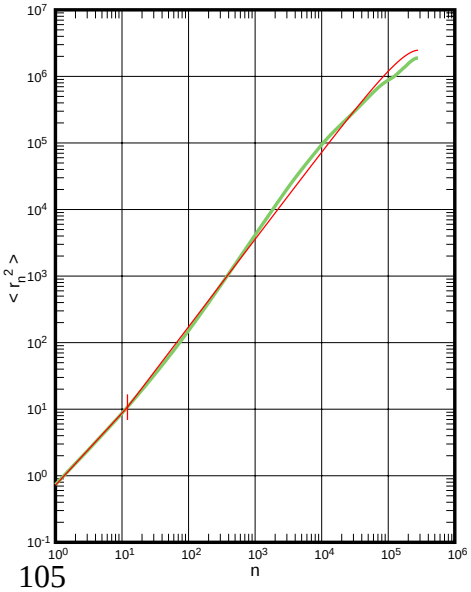

yeastIX




yeastVIII

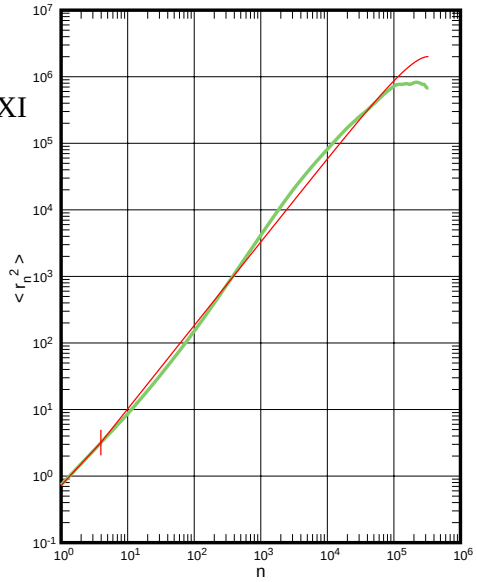



yeastV

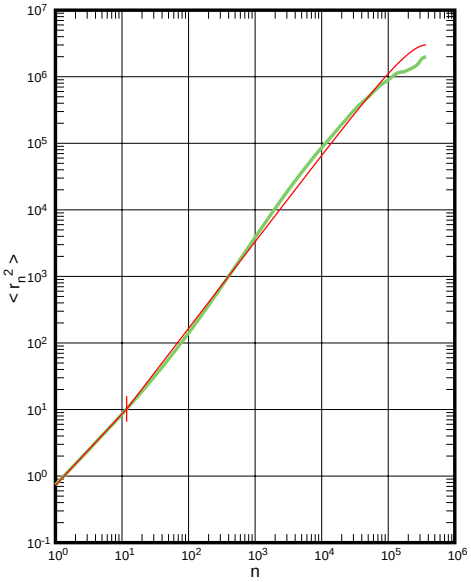


Genetic Codes: 666,000 < N < 950,000

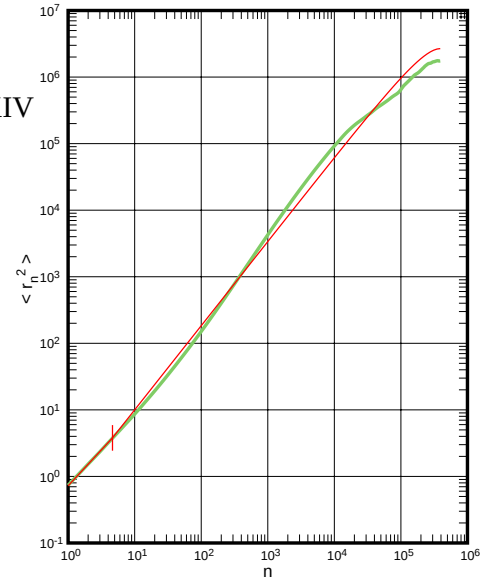
yeastXI



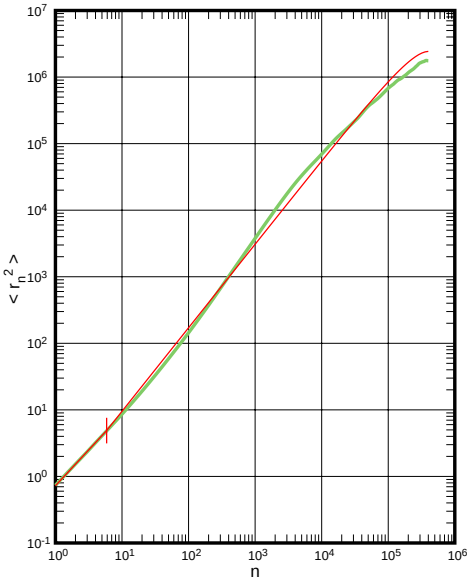

yeastX



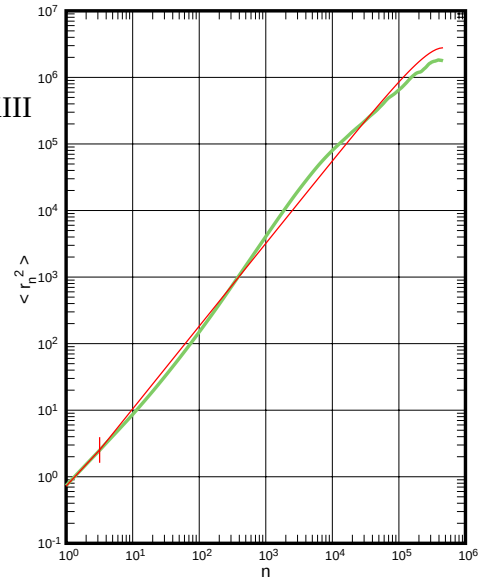

yeastXIV



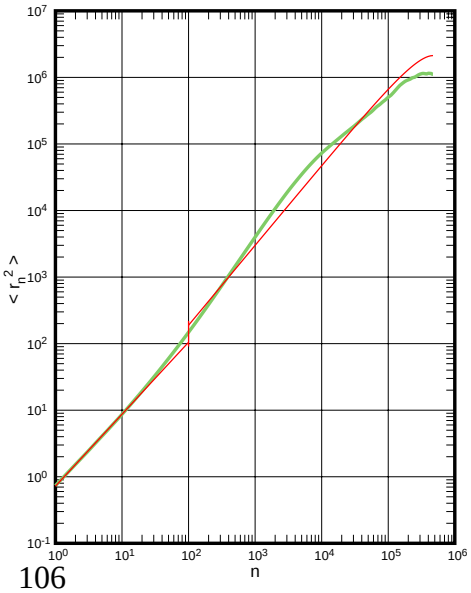

yeastII




yeastXIII

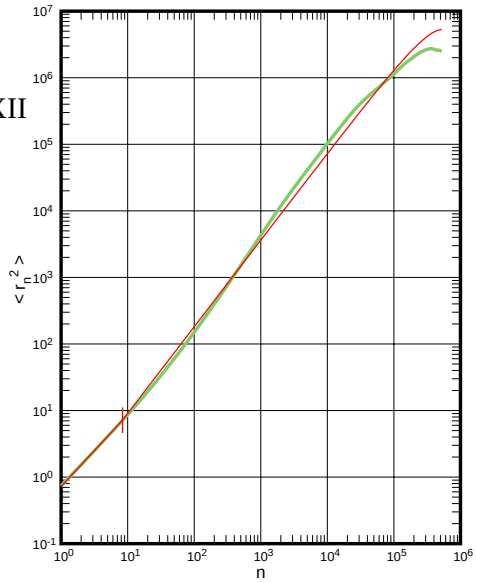



yeastXVI

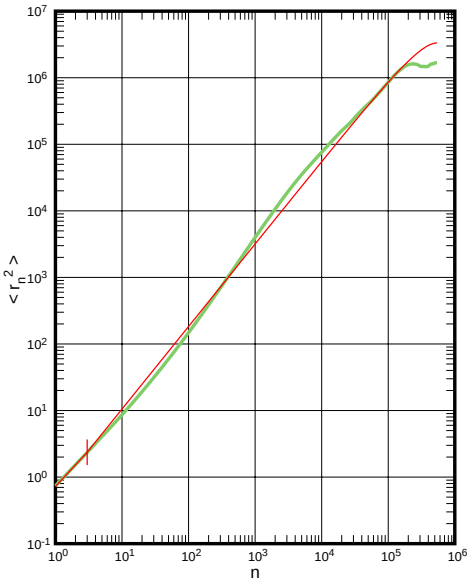


Genetic Codes: 1,080,000 < N < 14,000,000

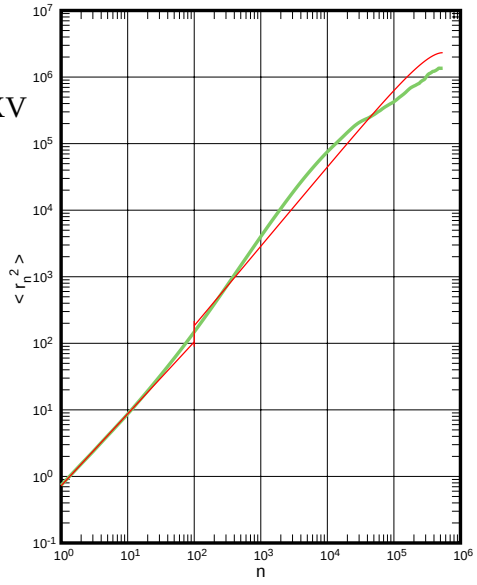
yeastXII



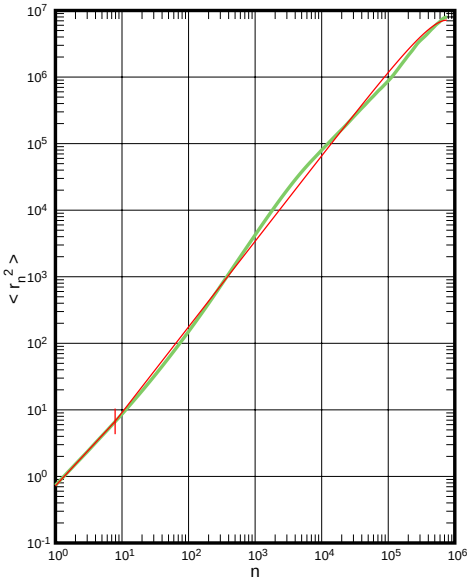

yeastVII



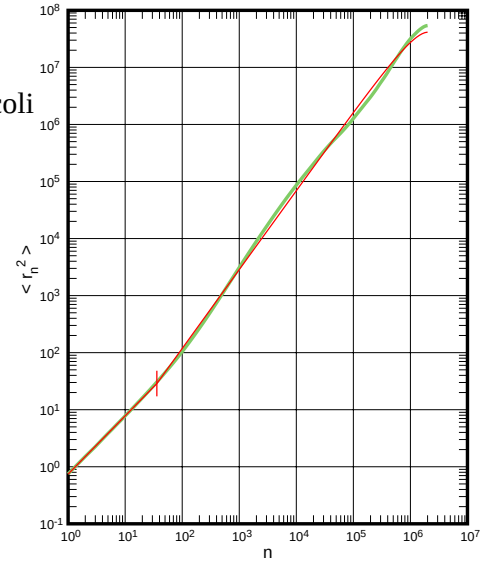

yeastXV



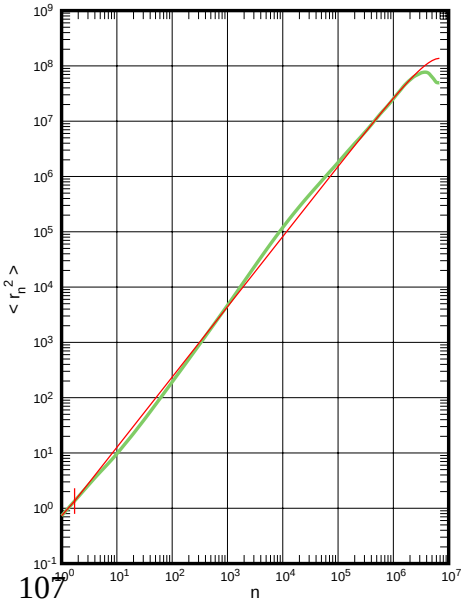

yeastIV




Ecoli

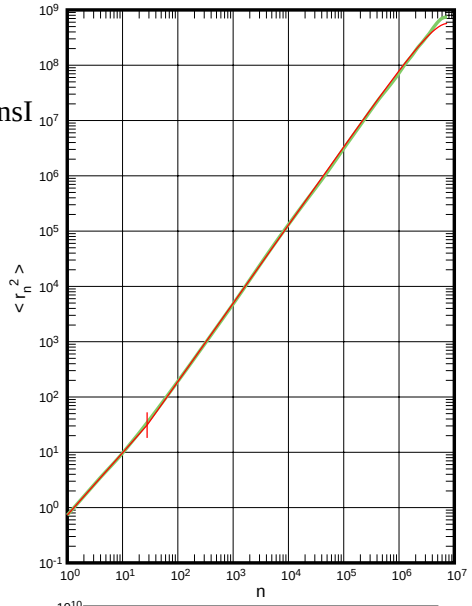



CelegansIII

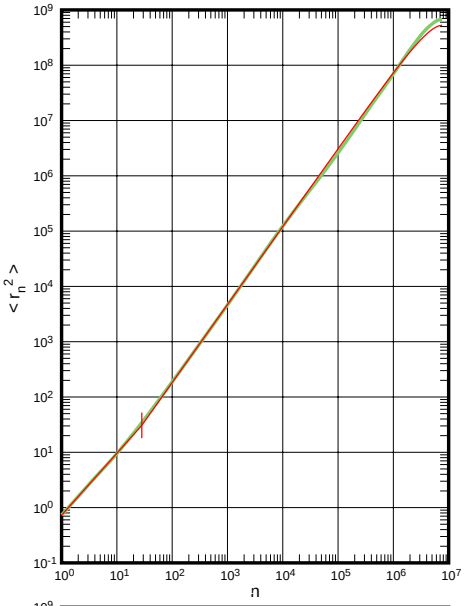


Genetic Codes:  15,000,000 < N < 20,900,000

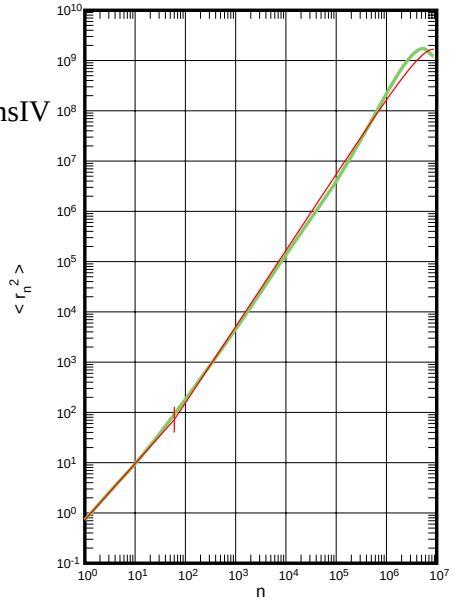
CelegansI



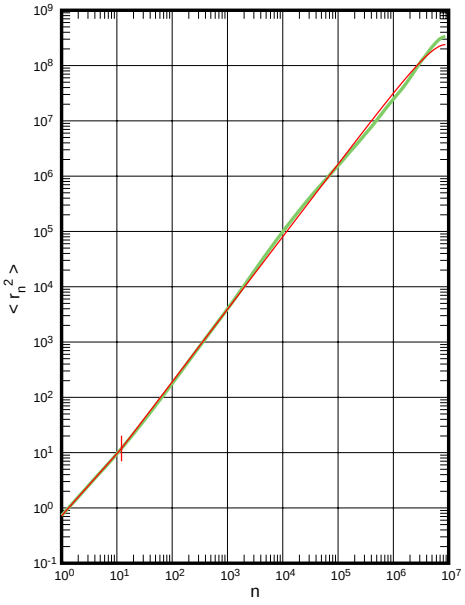

CelegansII



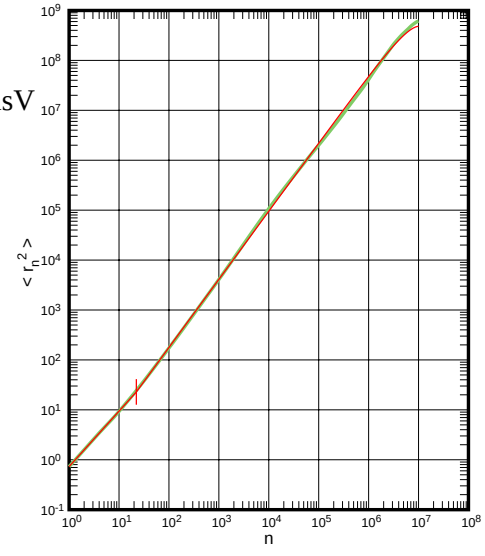

CelegansIV




CelegansX

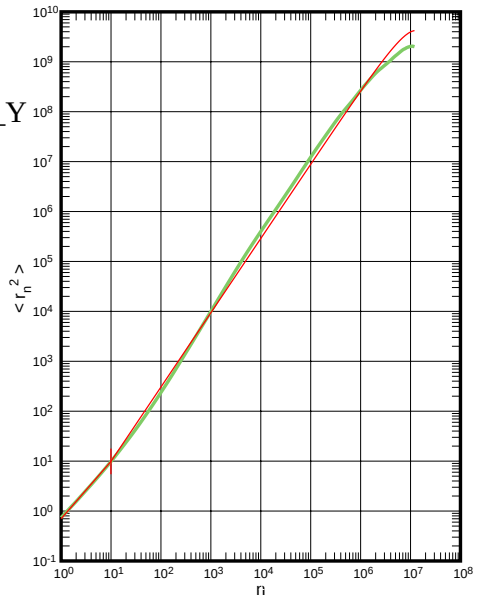



CelegansV

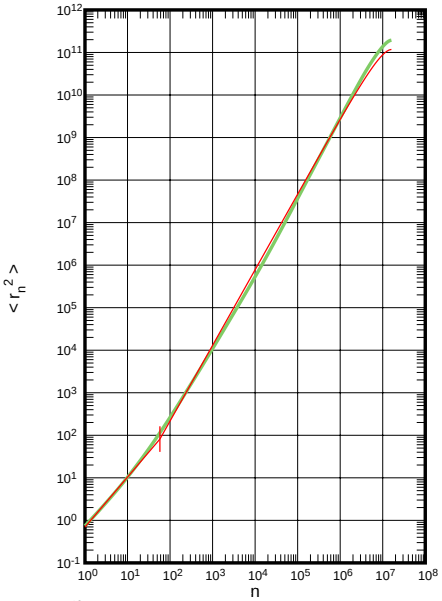


Genetic Codes:  25,000,000 < N < 54,000,000

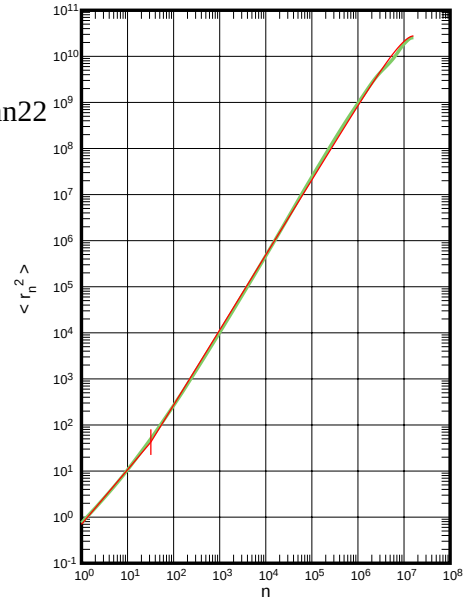
Pan_Y



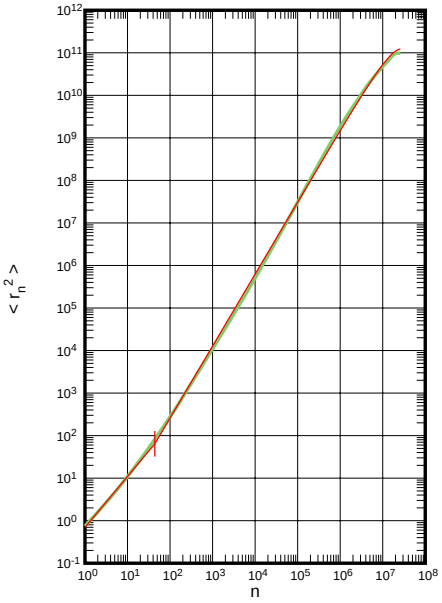

Pan21




Pan22

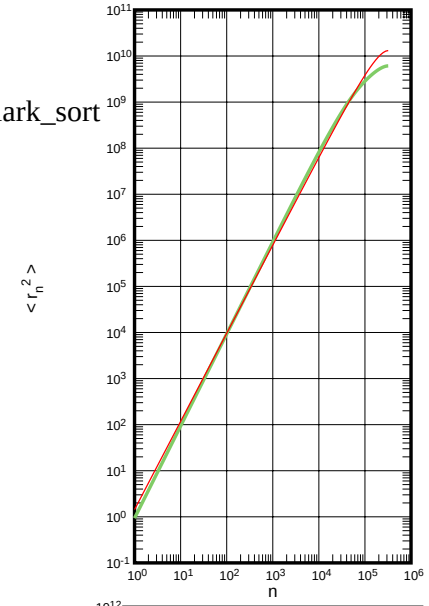



Pan19

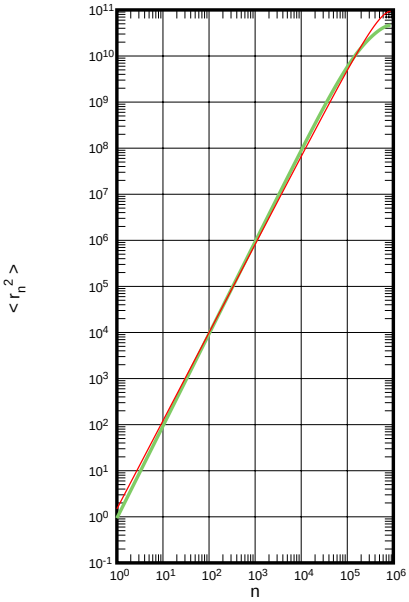


Letter-sorted texts:  64,000 < N < 3,300,000

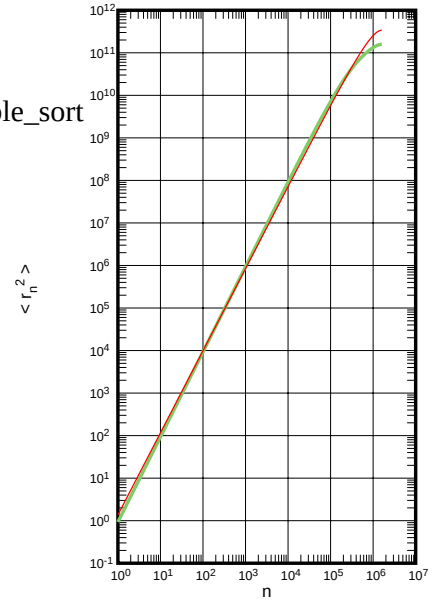
songlark_sort




wealth_sort




bible_sort



**PLOTS OF $g_n$**

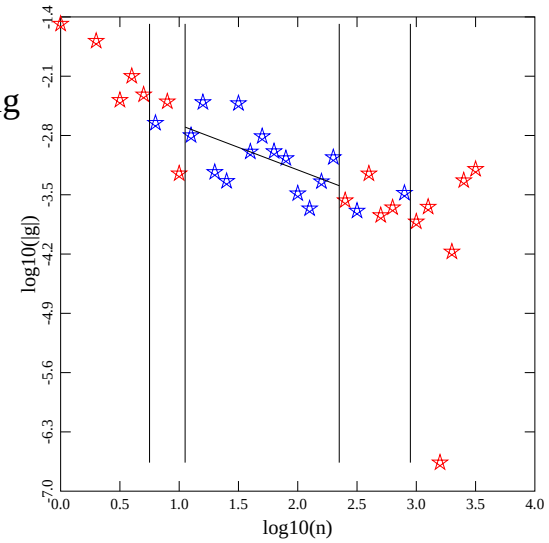
Introduction: Log-log plots of $g_n$ as a function of $n$ appear on the following pages.

Each symbol represents the absolute value of a coarse-grained average. The coarse-graining was taken over logarithmic intervals of width one-tenth of a decade in $n$. Blue and red colors indicate when the coarse-grained average is positive and negative, respectively. In most cases, $g_n$ is positive and behaves as $n^{\beta}$ with $\beta < 0$. (See the section on Fitting Parameters for the values of $\beta$.)

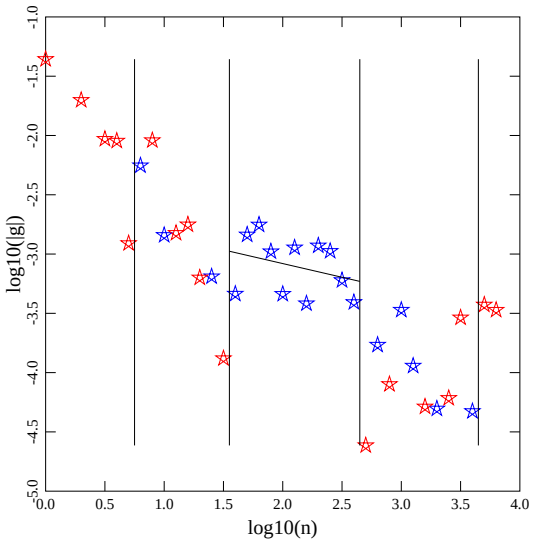
As many as four vertical lines have been drawn in each plot to guide the eye. Because of the correlation hole at small $n$, the first few points are almost always negative, and the first vertical line on the left indicates the first appearance of a positive point when reading from the left. Because of the central correlation hole, the last few points are also almost always negative, and the first vertical line on the right indicates the first appearance of a positive point when reading from the right. In each plot, we also identified the longest train of positive points in the data, and the limits of this train are indicated by the two central vertical lines. When fewer than four vertical lines are seen, it is either because some lines coincide, or because the data do not fit the general pattern of the two correlation holes.

We calculated $\beta$ as the slope of the least-squares line through the positive train. This least-squares line also appears in each plot. In a minority of cases, including especially the shorter texts, the positive train contained fewer than 10 points, i.e., it extended over less than one decade in $n$. In such cases, no attempt was made to compute $\beta$, and the plot was not generated. This is why many shorter texts do not appear below.

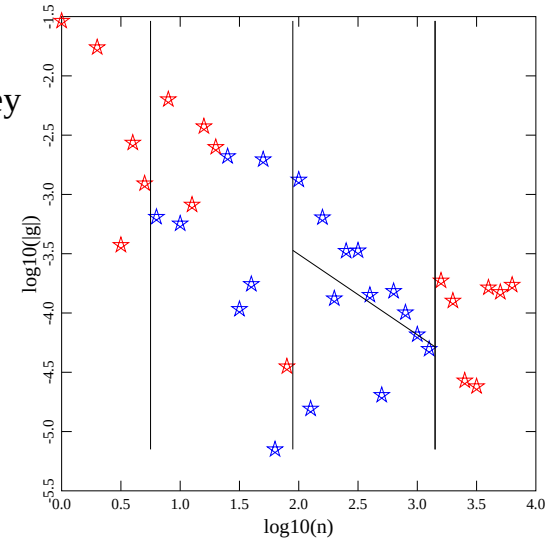
jfk_inaug



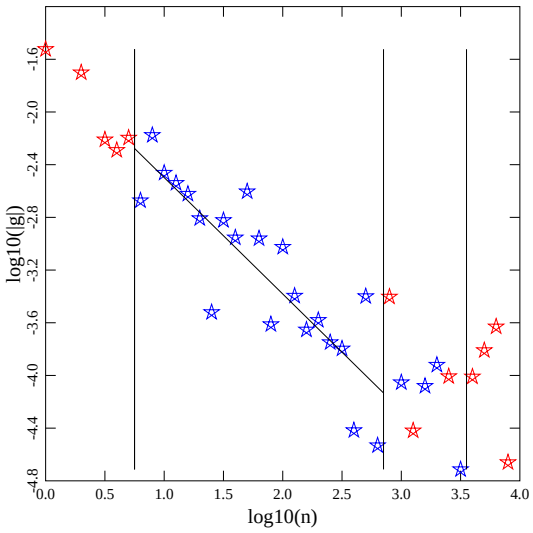

antarctica



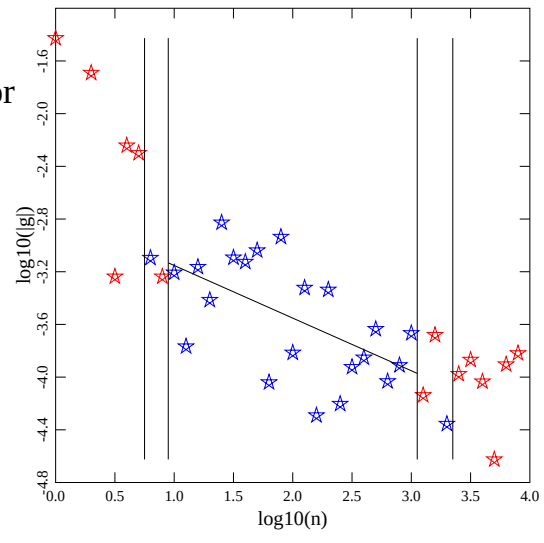

jersey



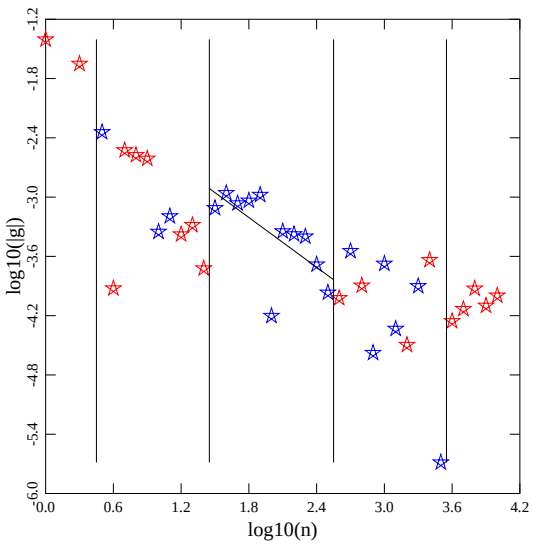

rime



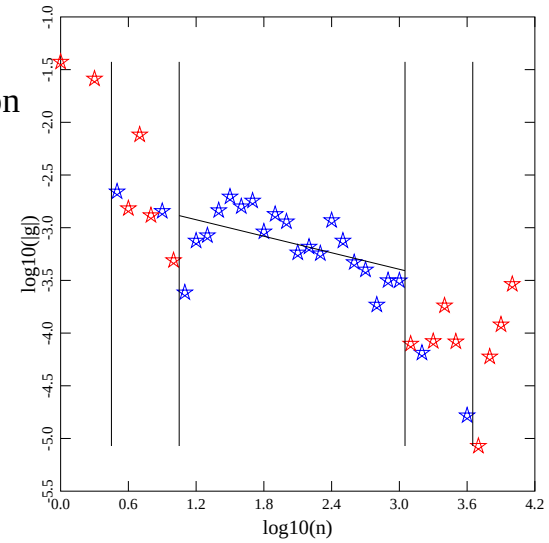

taylor



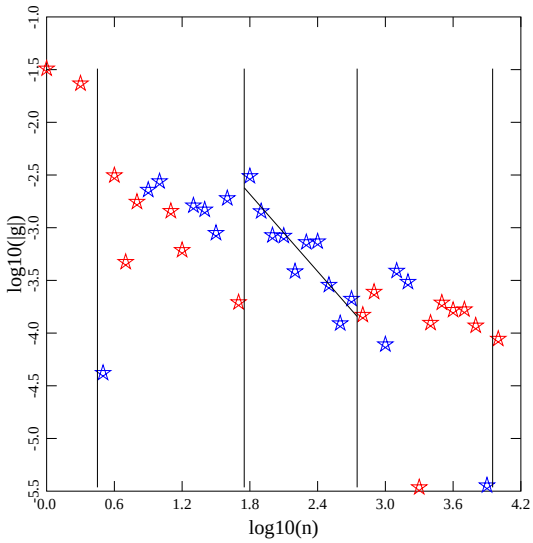

dreams



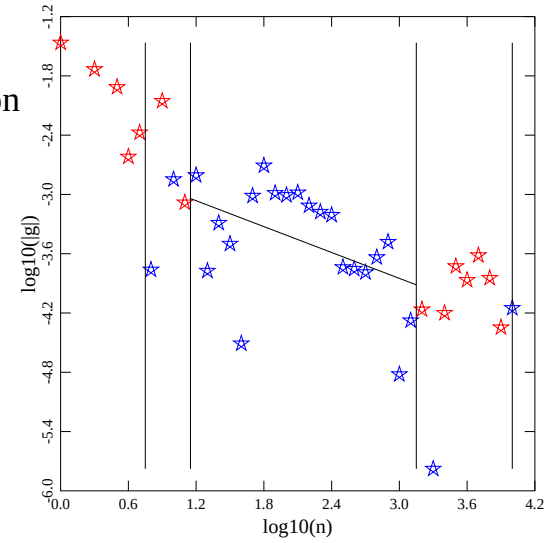
ion



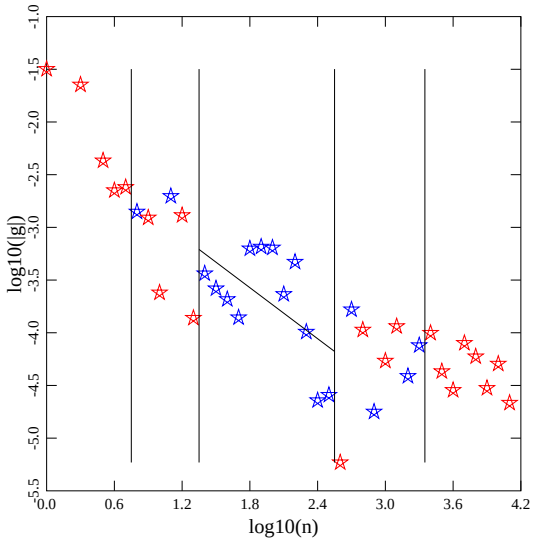

snark



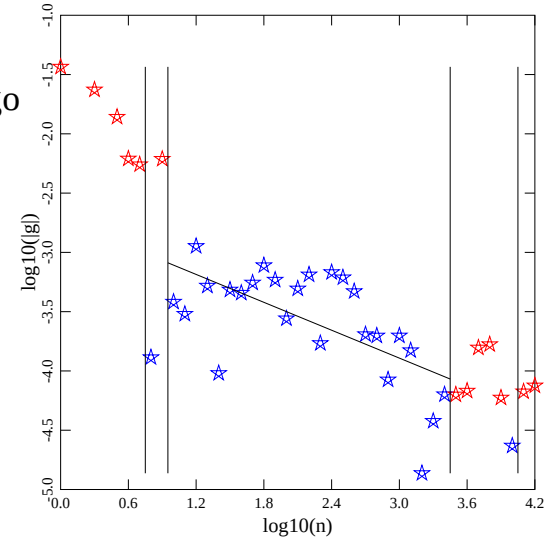

constitution



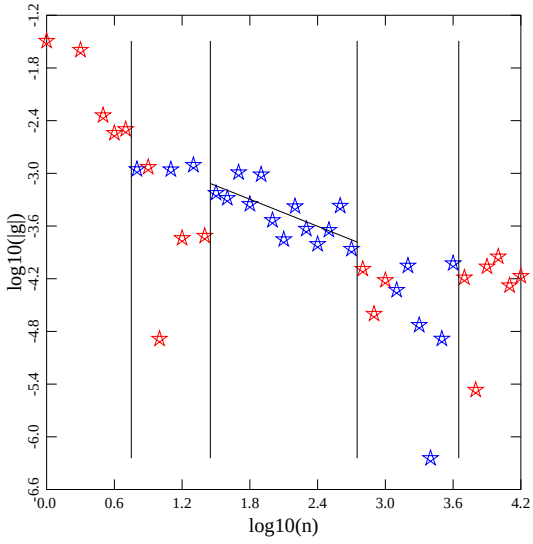

sussexvamp



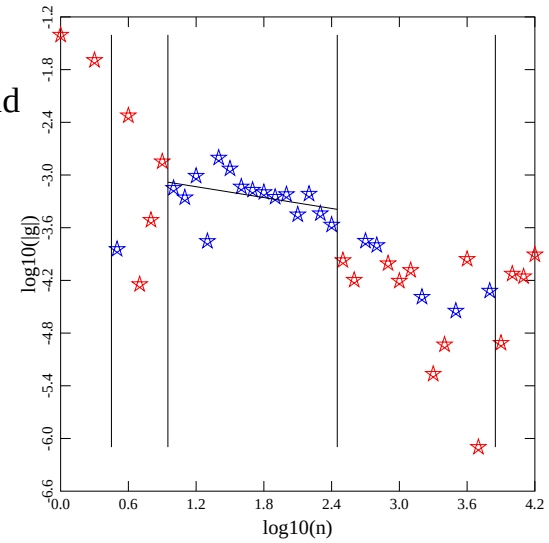

hidalgo



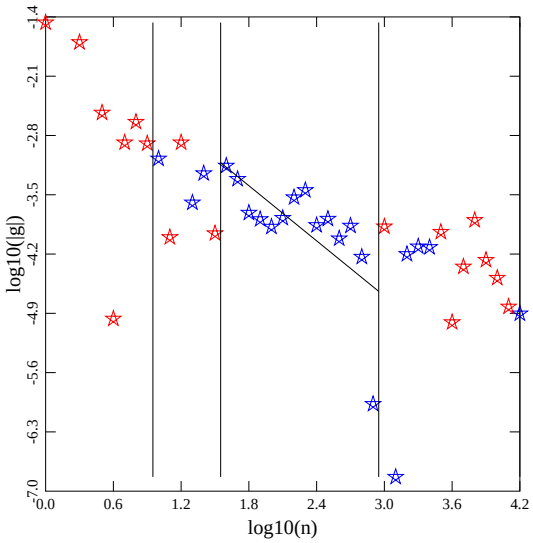

hyperion



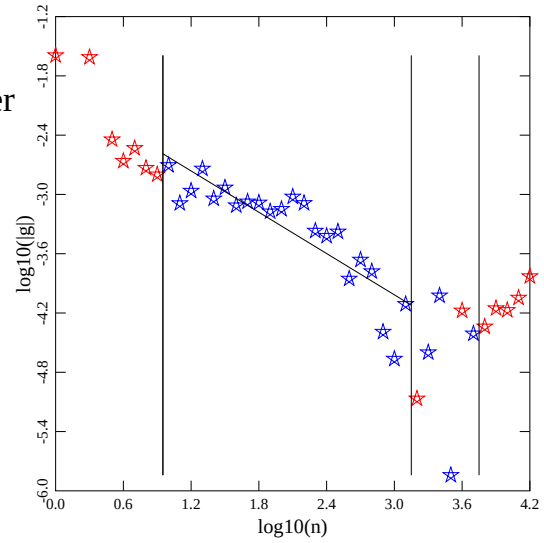
transcend



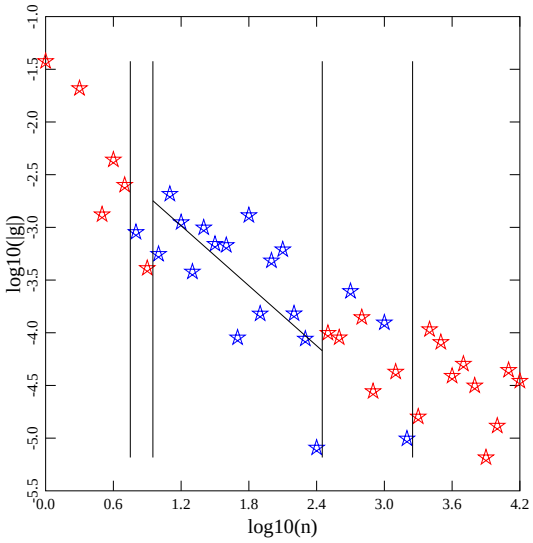

maelstrom



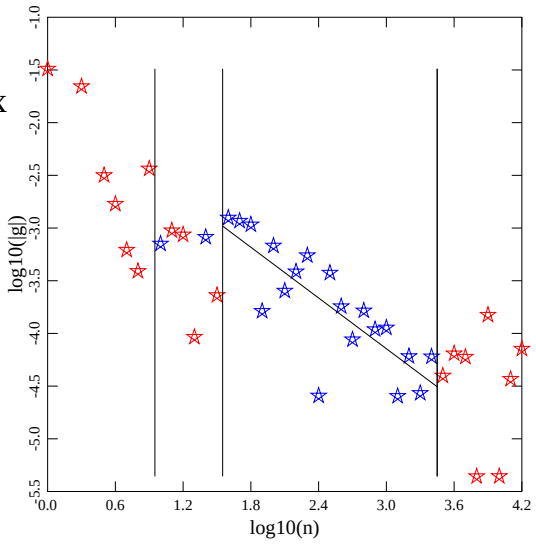

piper



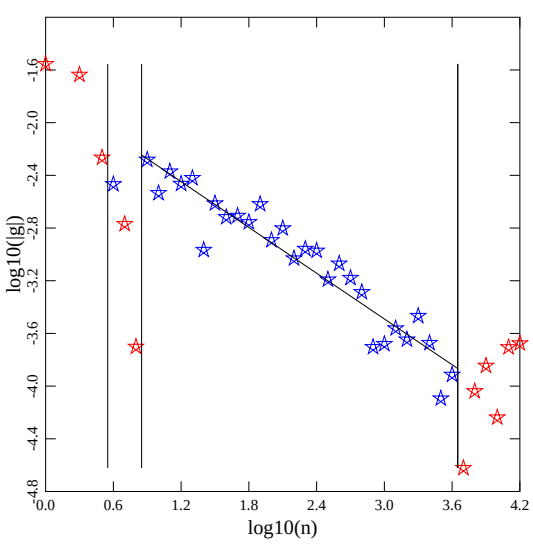

literary_ethics



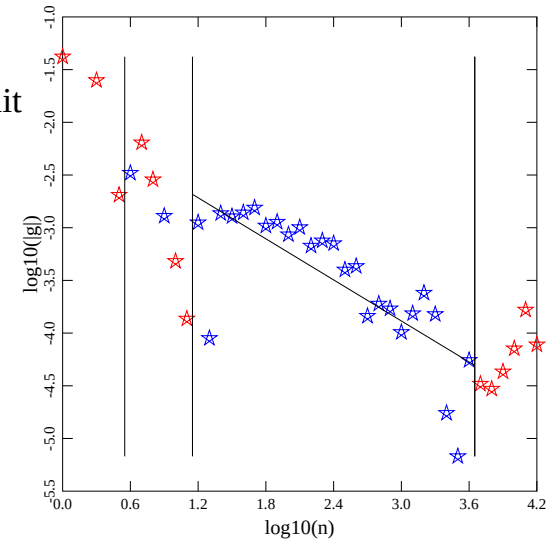

carfax



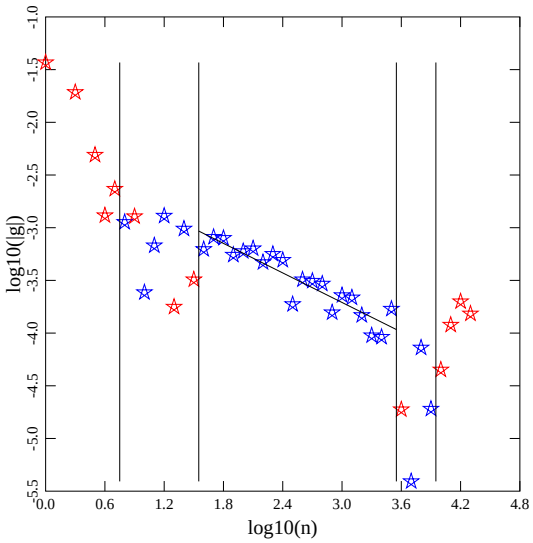

nursery



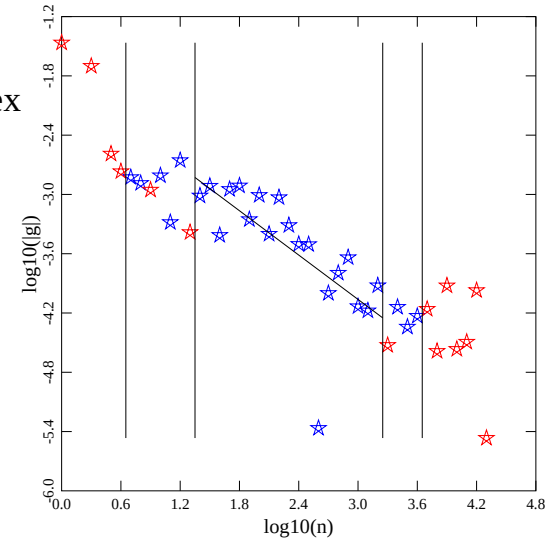
animal_gait



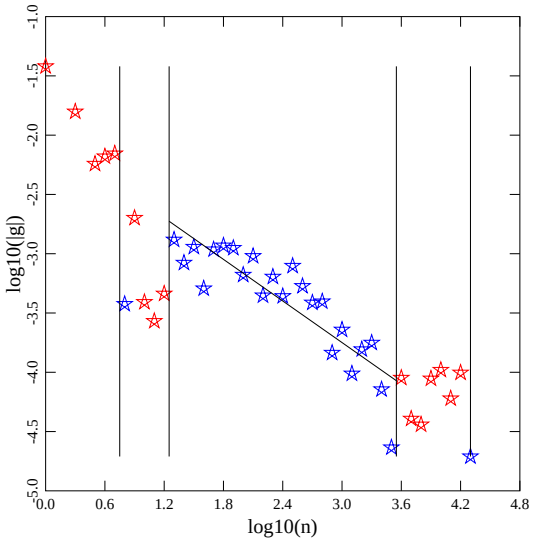

escape



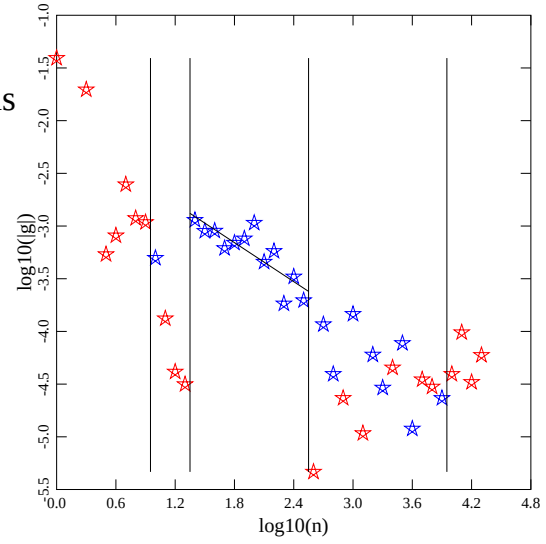

vortex



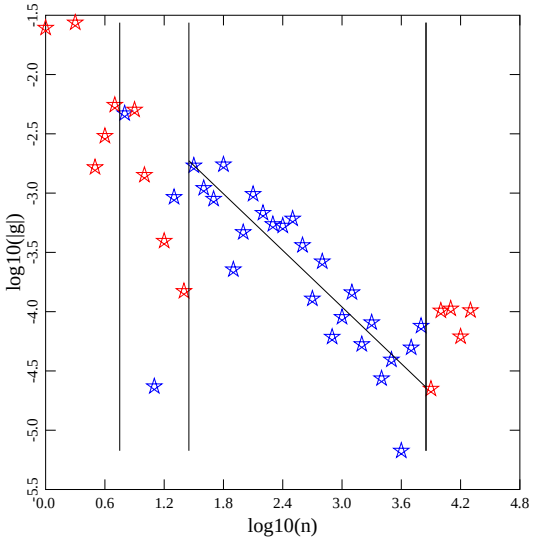

adductor



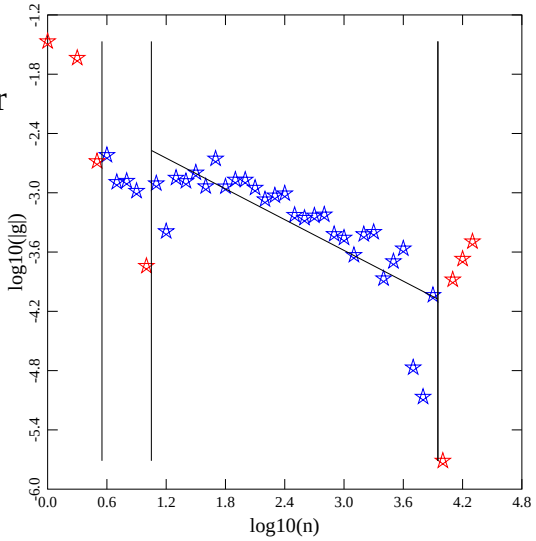

civildis



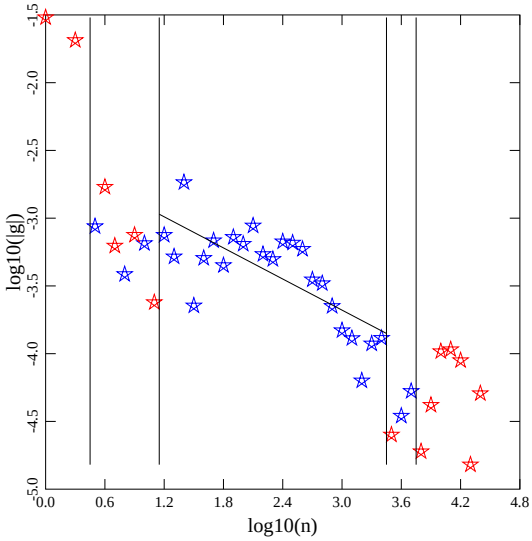

weevil



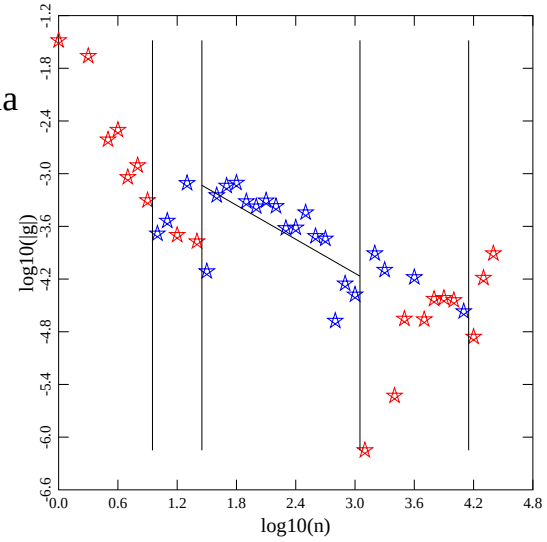
sugar



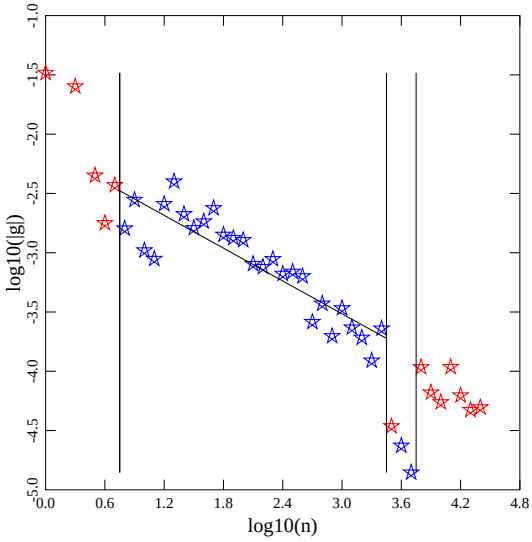

dumpling



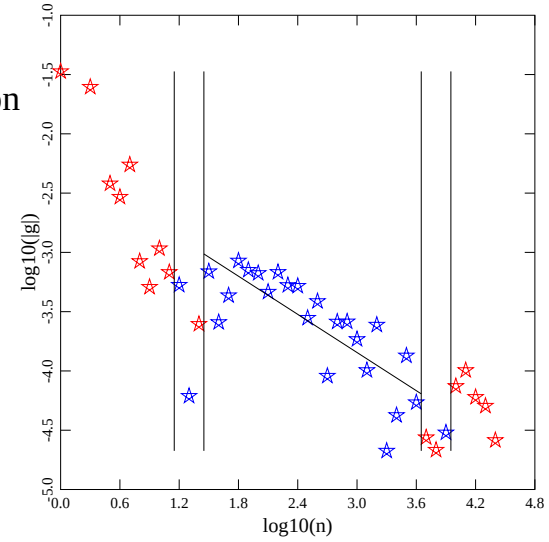

wisteria



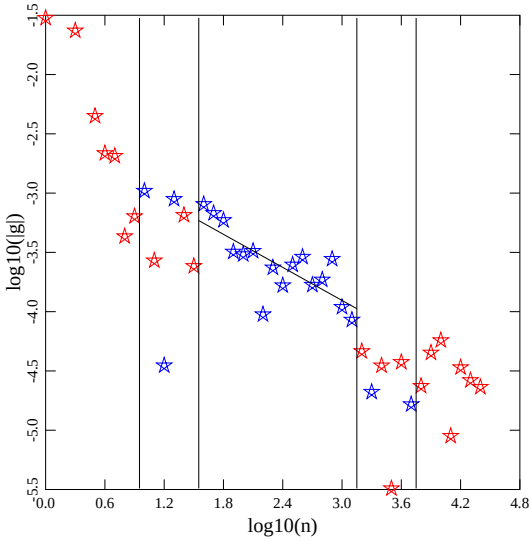

sourdough



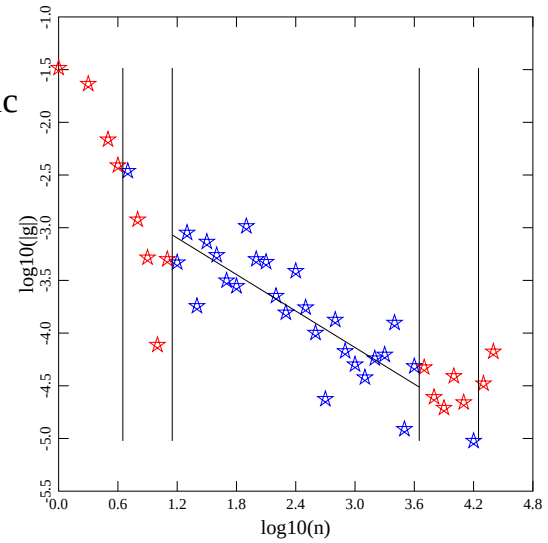

argon



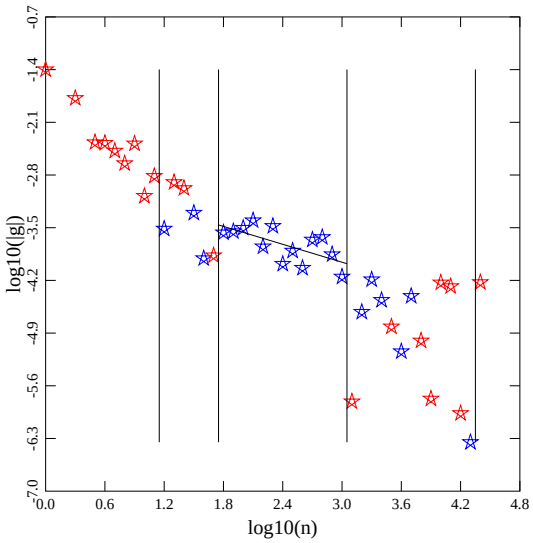

sleepy_hollow



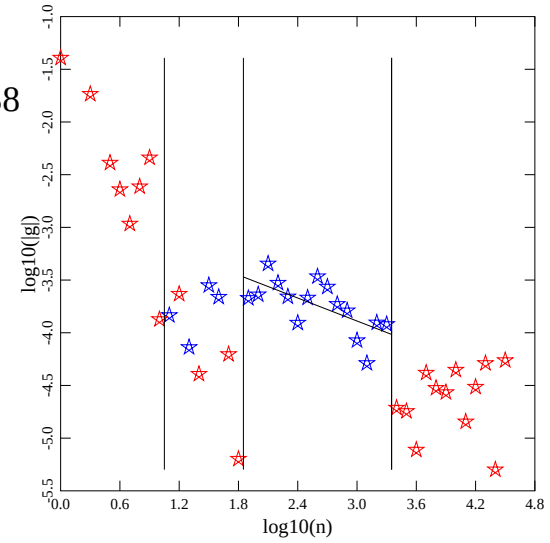
attic



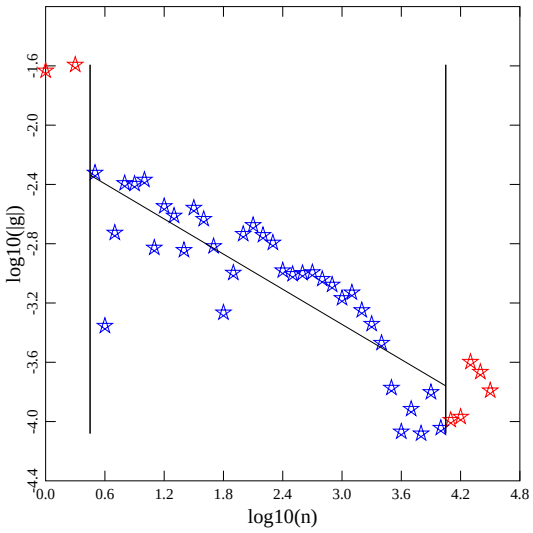

su1837



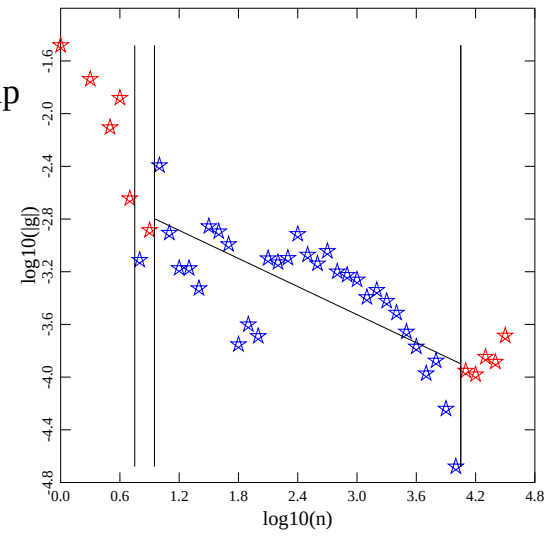

su1838



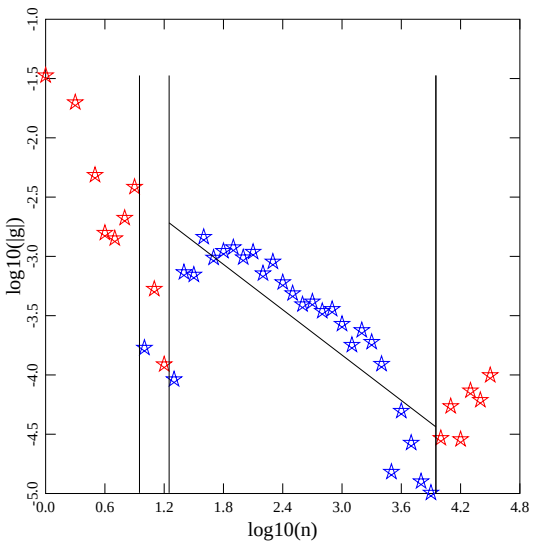

bc.birds



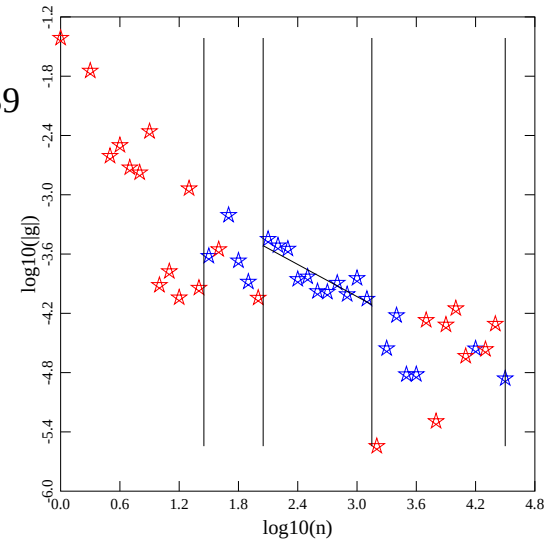

stamp



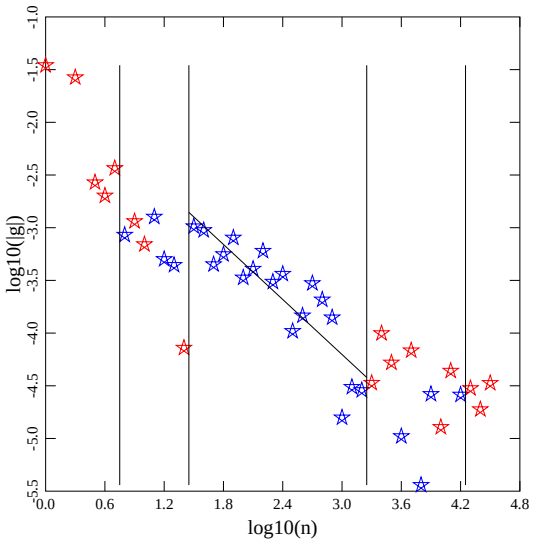

woodrat



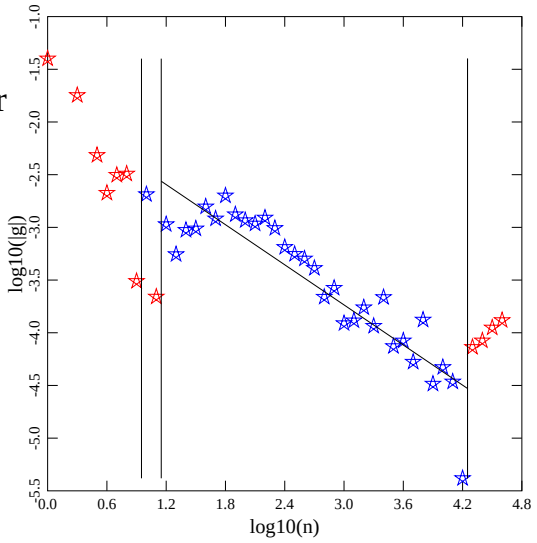
su1839



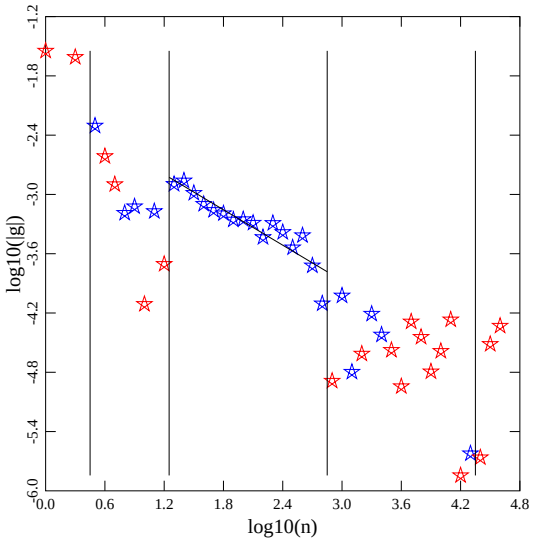

regained



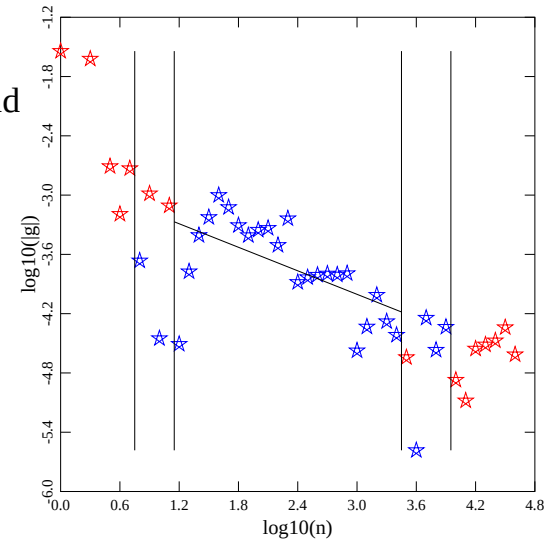

stellar



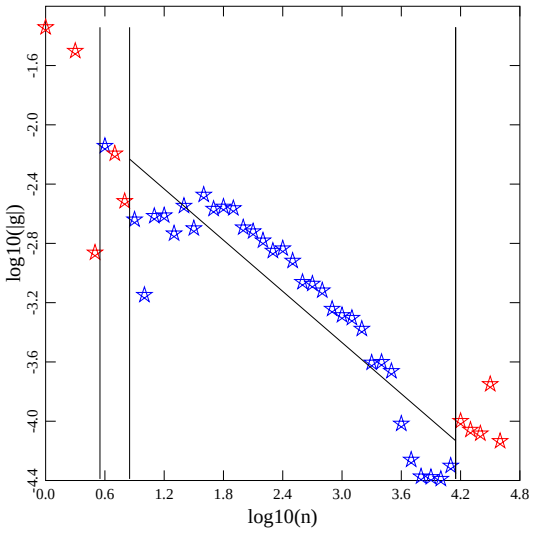

macbeth



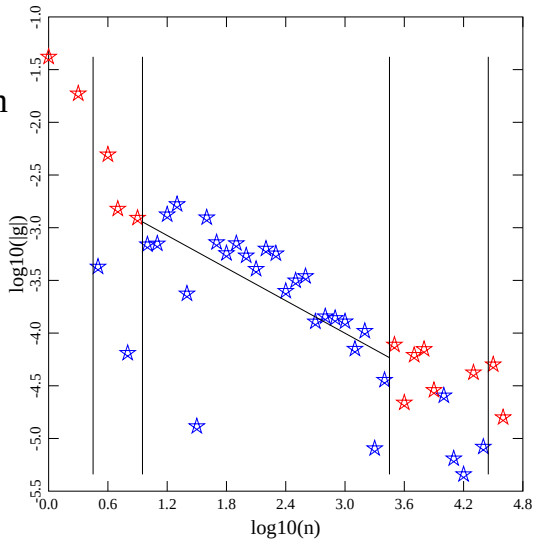

cleveland



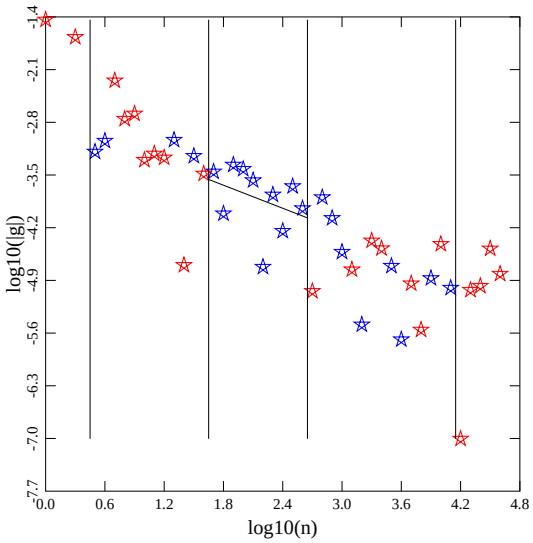

parmenides



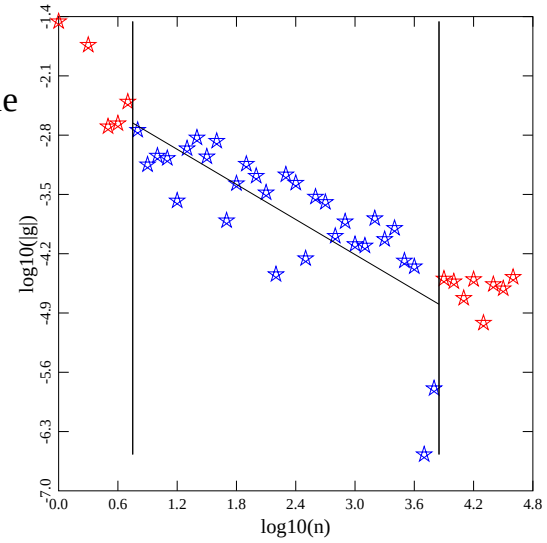
theism



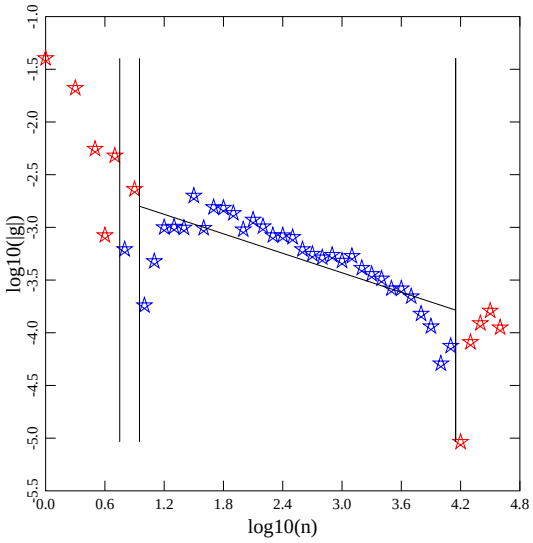

areopagitica



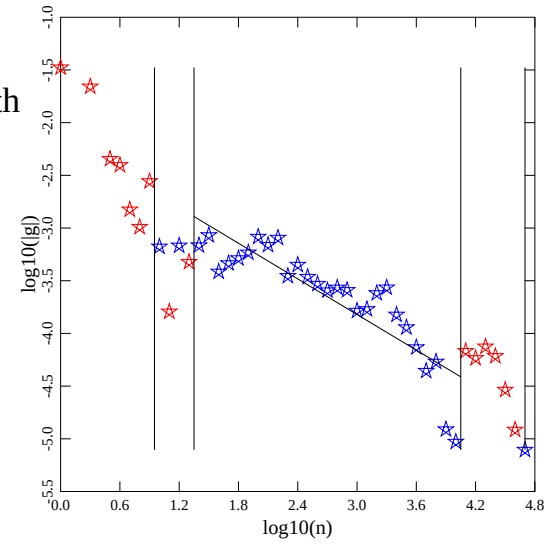

plague



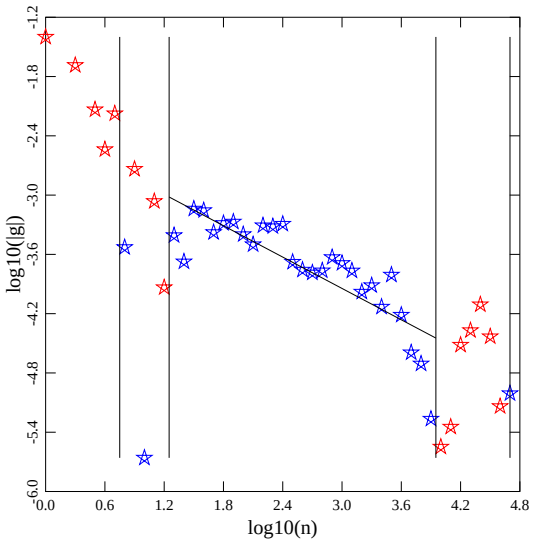

mythology



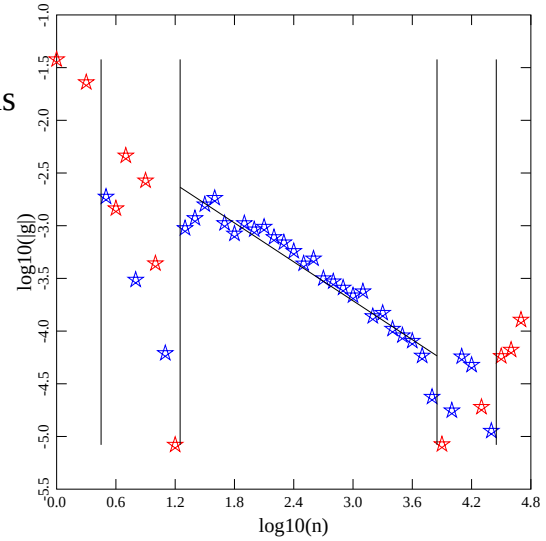

mammoth



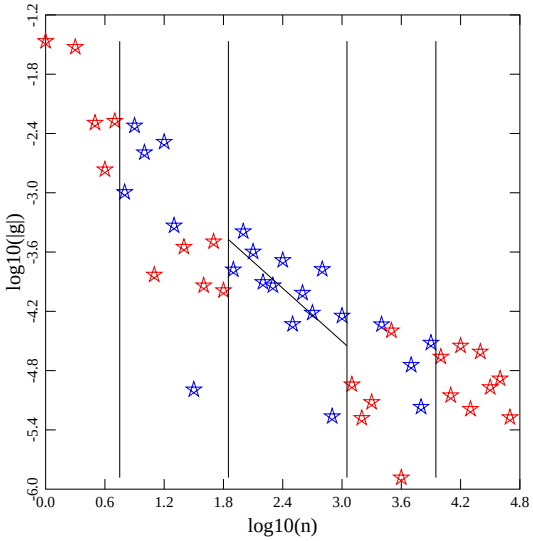

comet



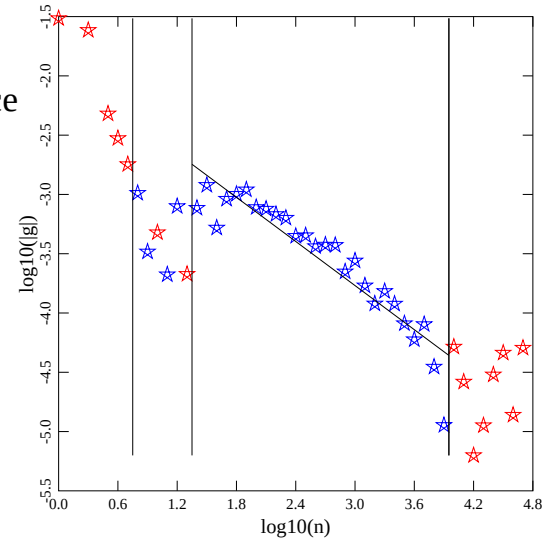
cratylus



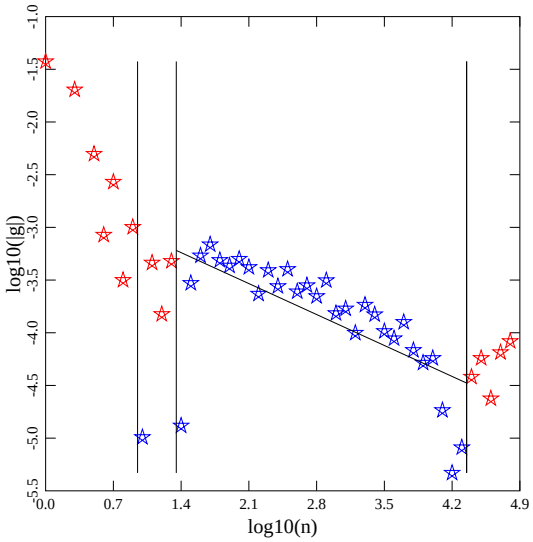

beowulf



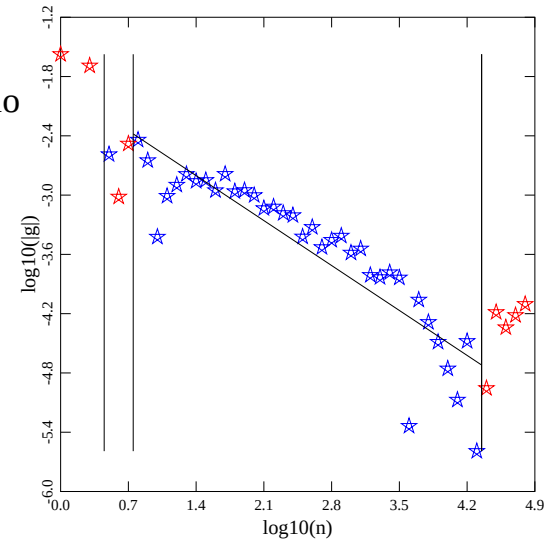

alice



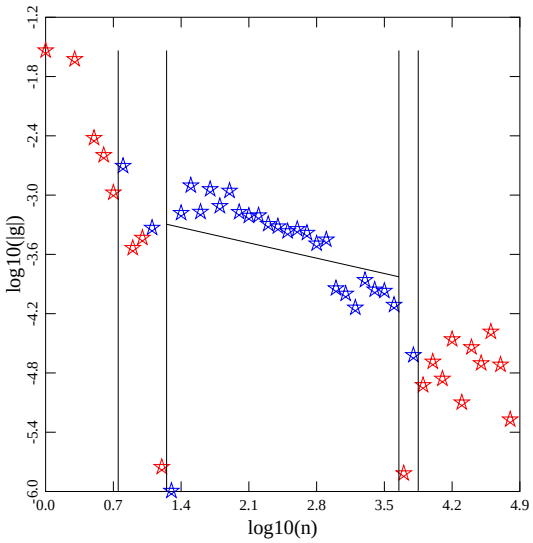

snake



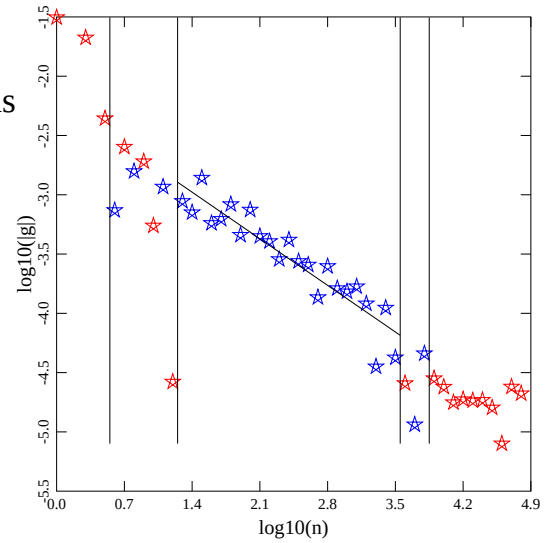

othello



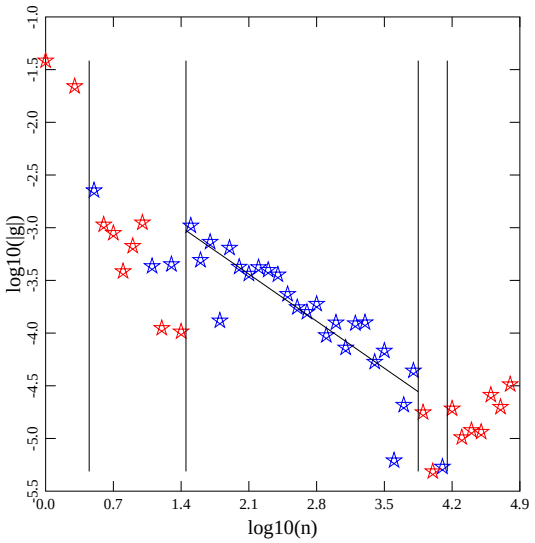

glass



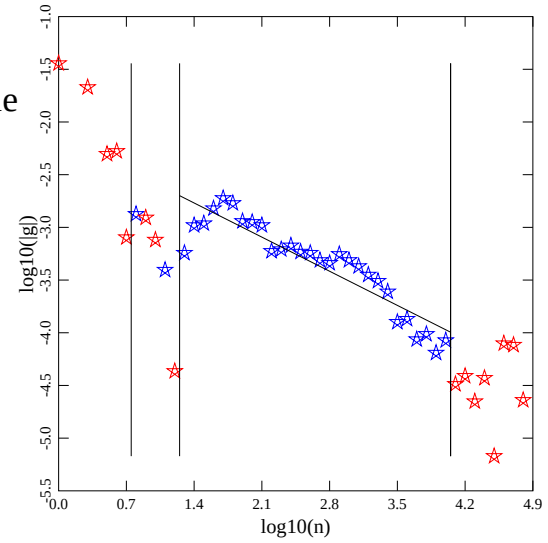
christmas



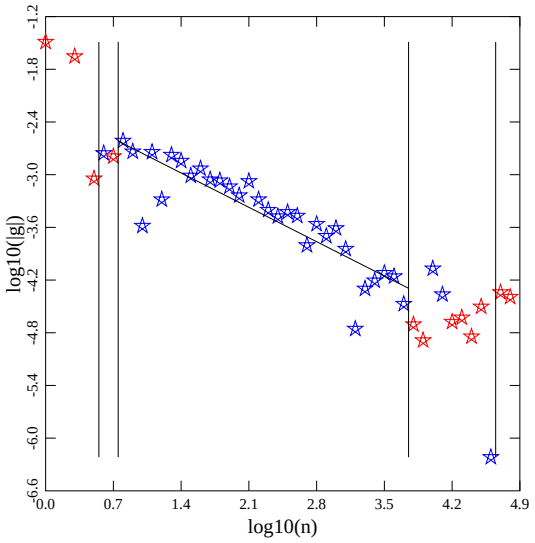

oroonoko



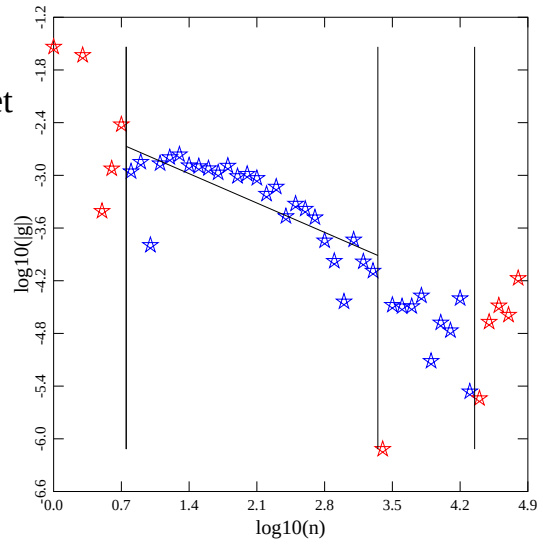

woods.hole



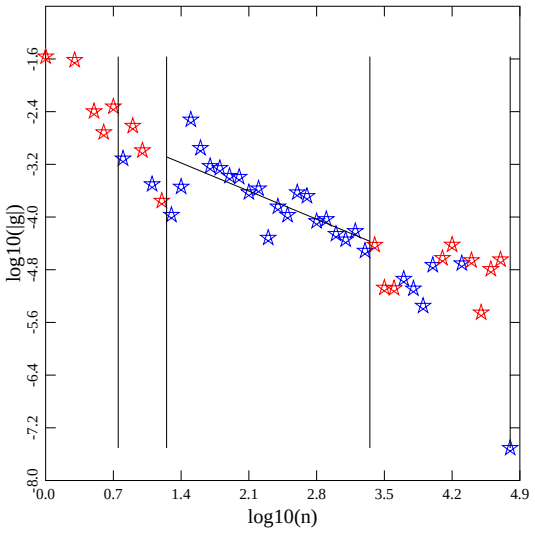

chimes



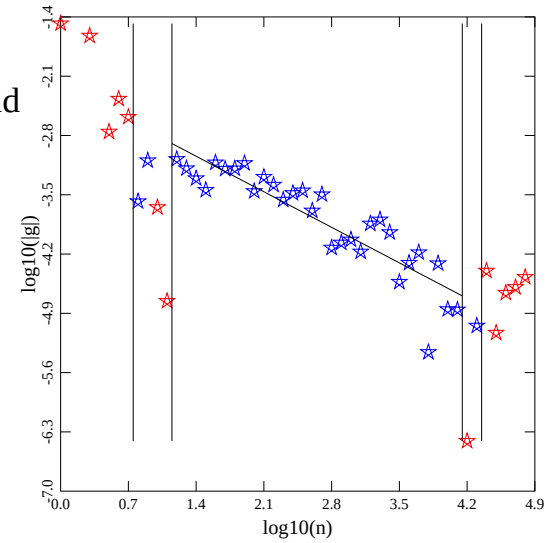

hamlet



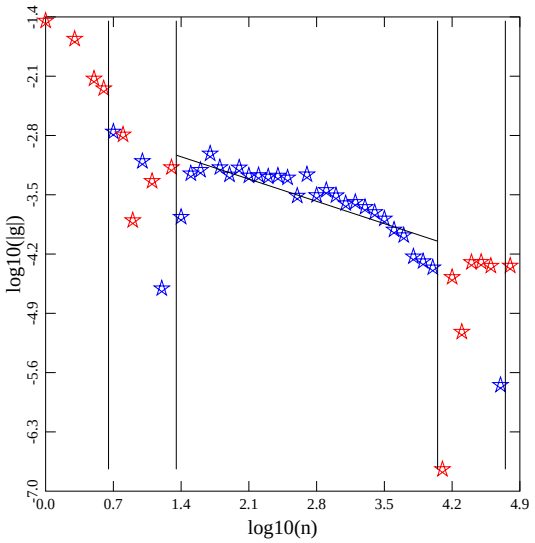

endymion



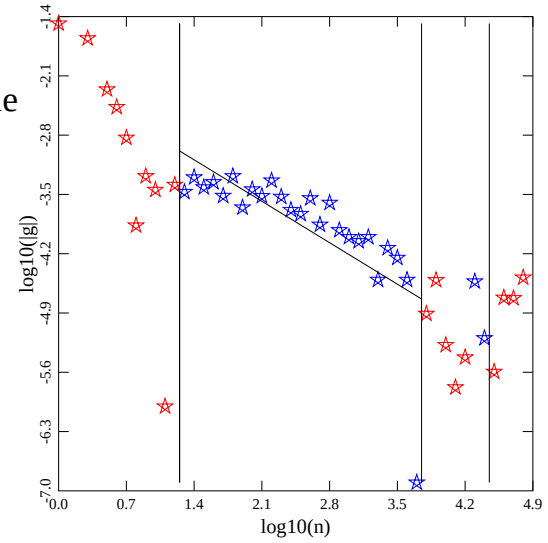
wild



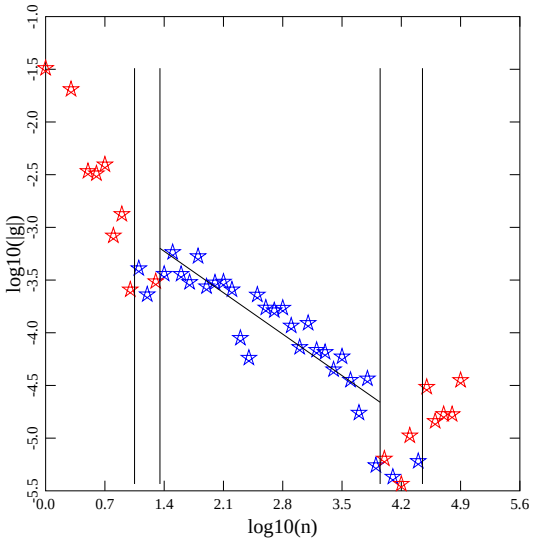

cello



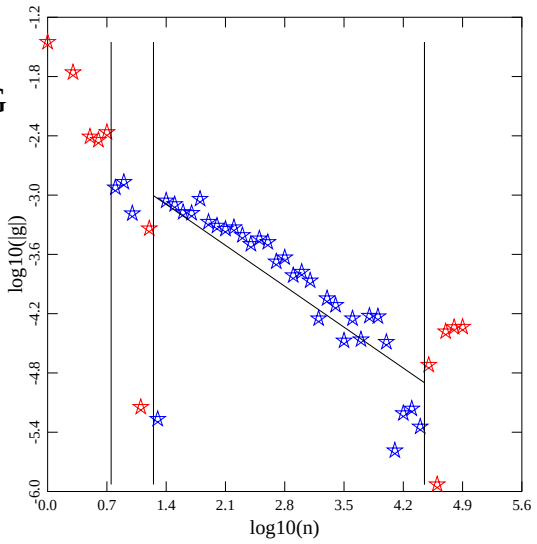

time_machine



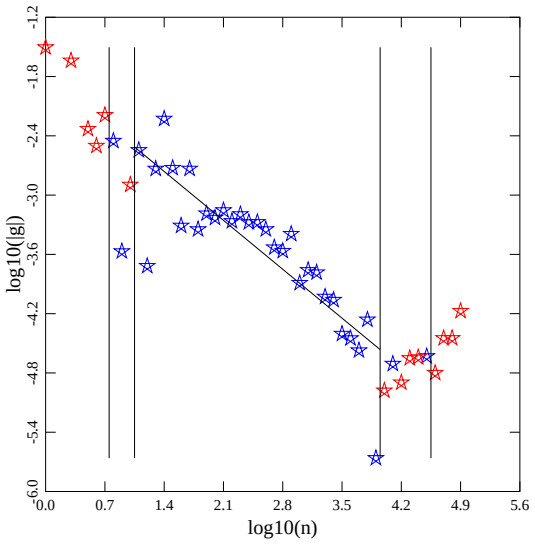

billy_budd



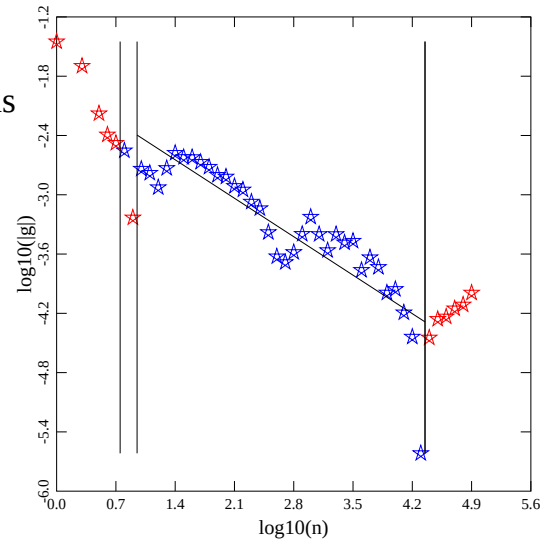

radioUKG



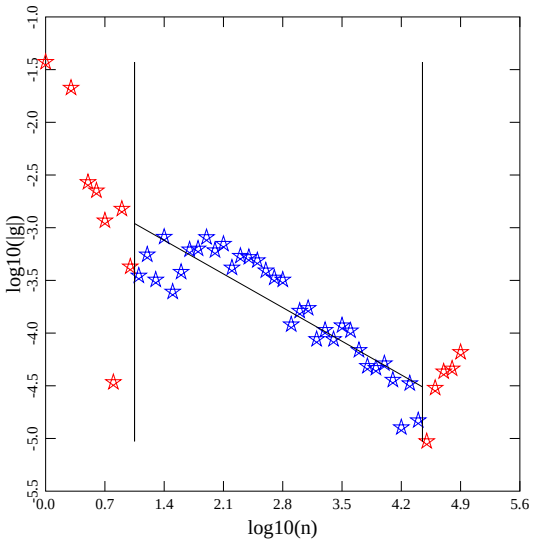

hiawatha



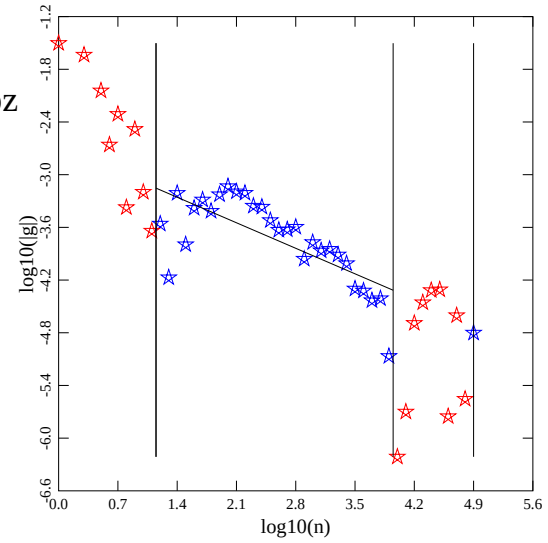
pinus



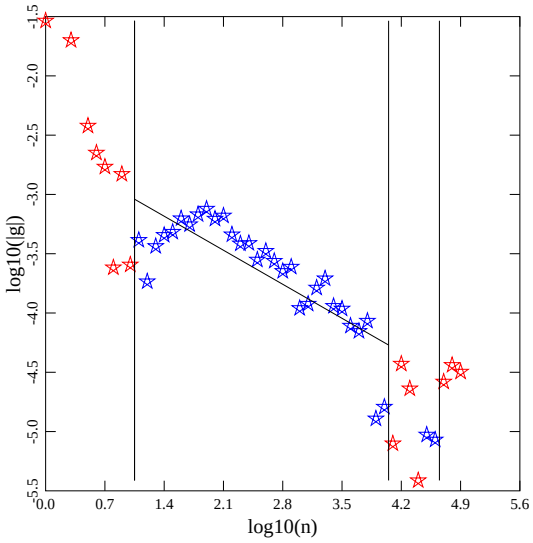

flatland



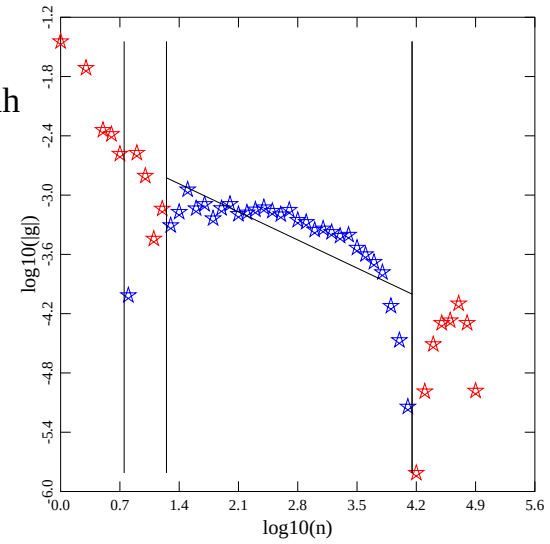

oz



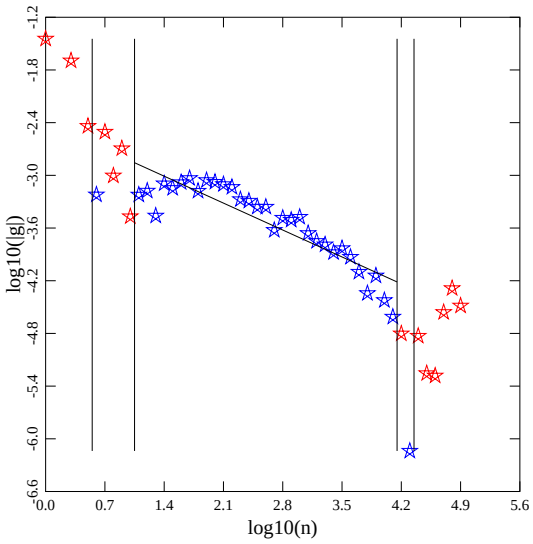

telegraph



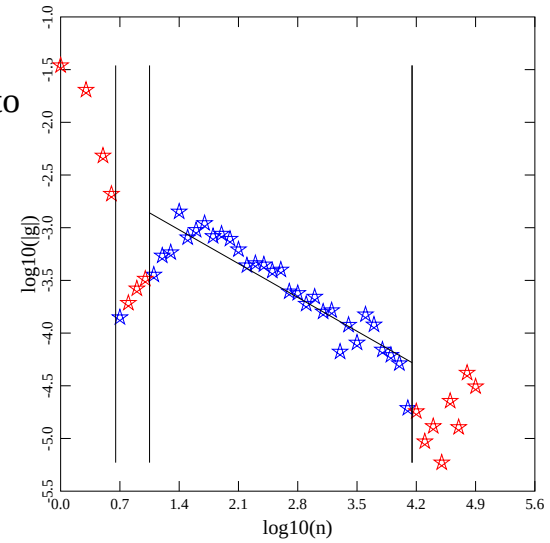

utah



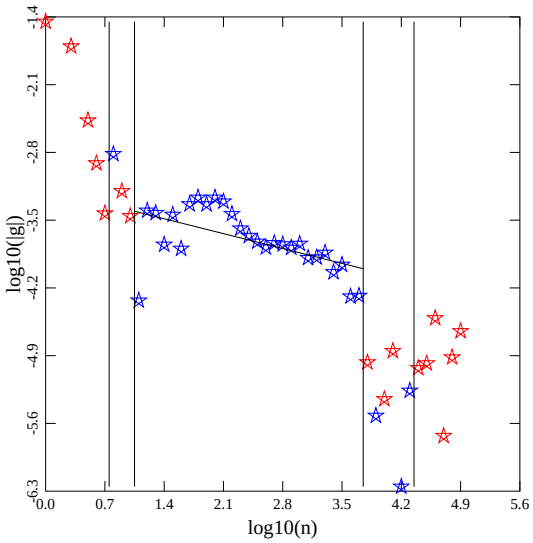

candle



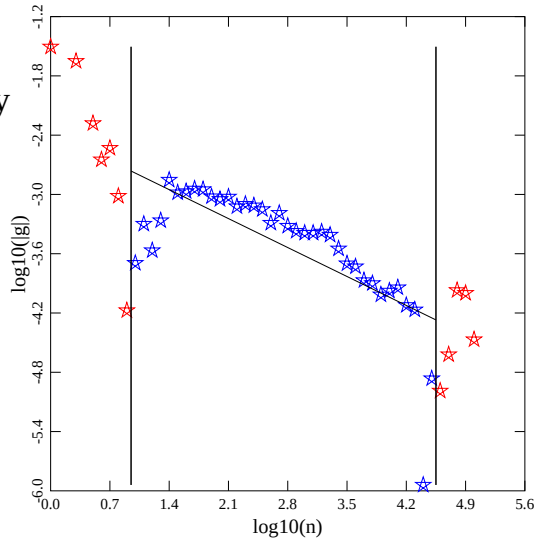
fd_auto



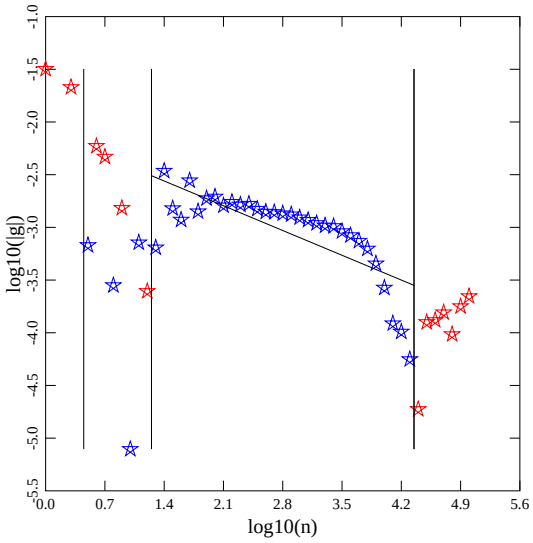

avalanche



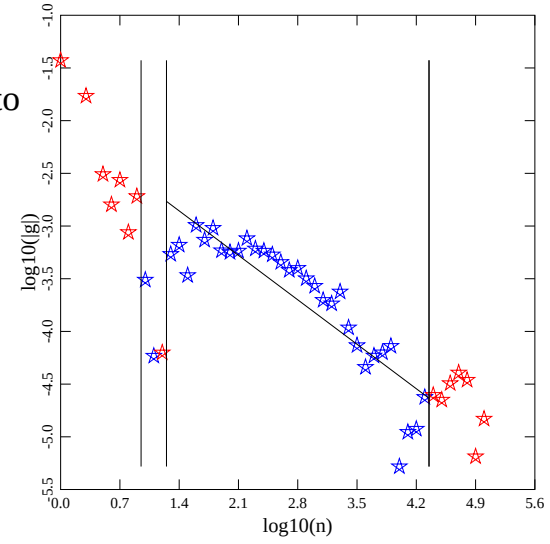

taxidermy



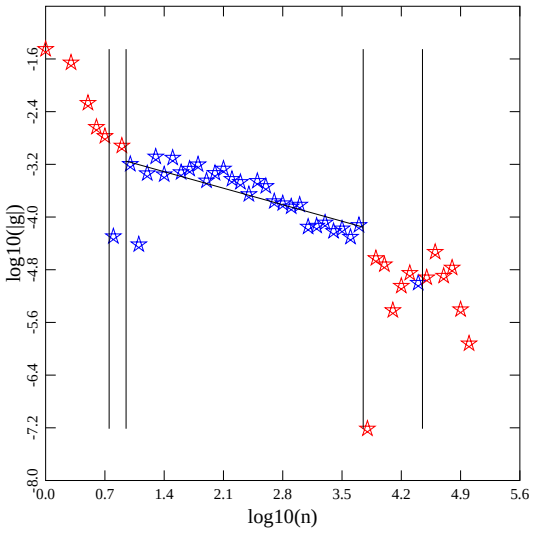

highland



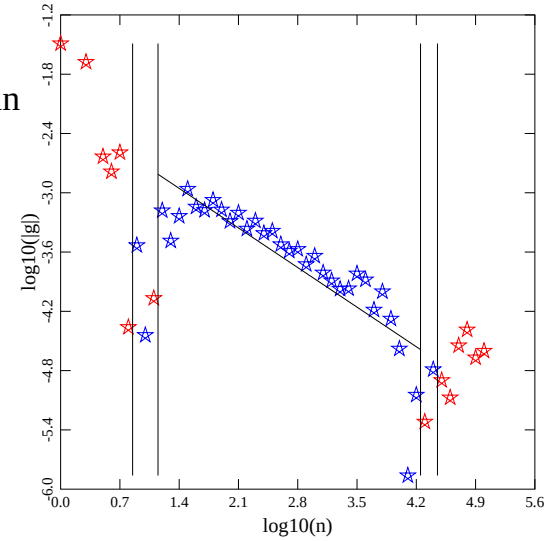

tj_auto



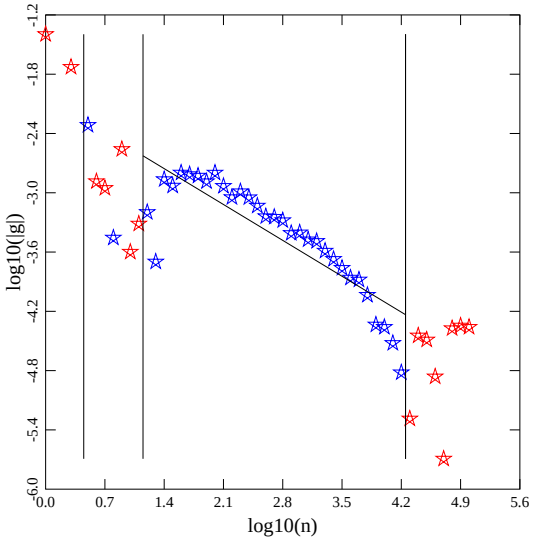

DrMoreau



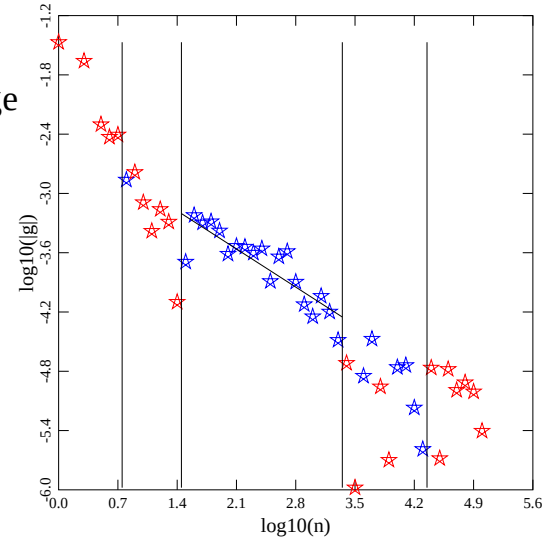
peter_pan



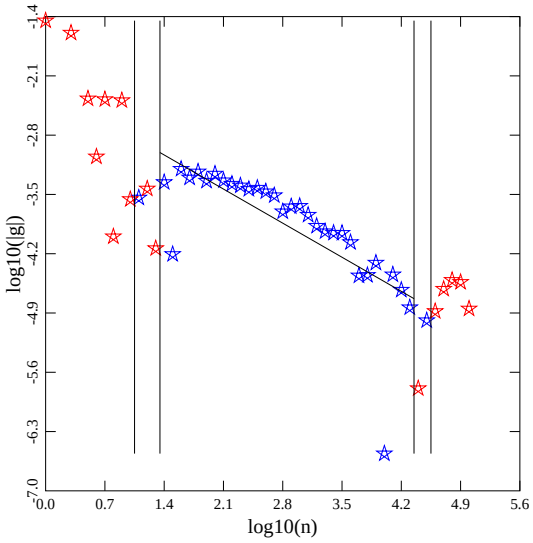

problems



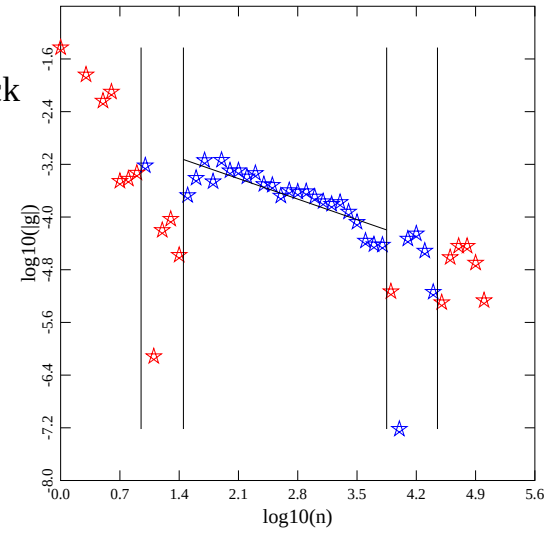

red_badge



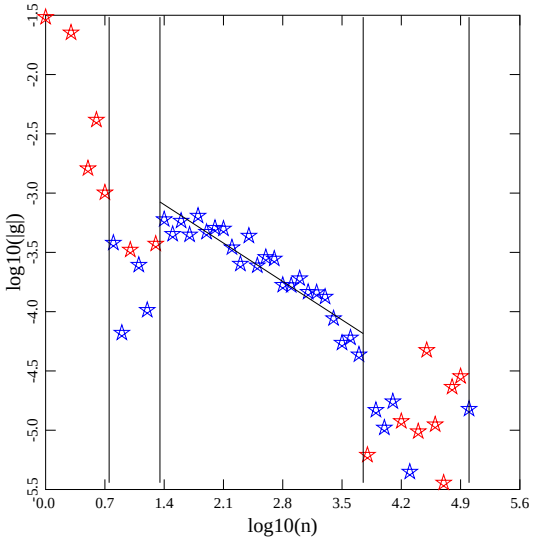

maid_mars



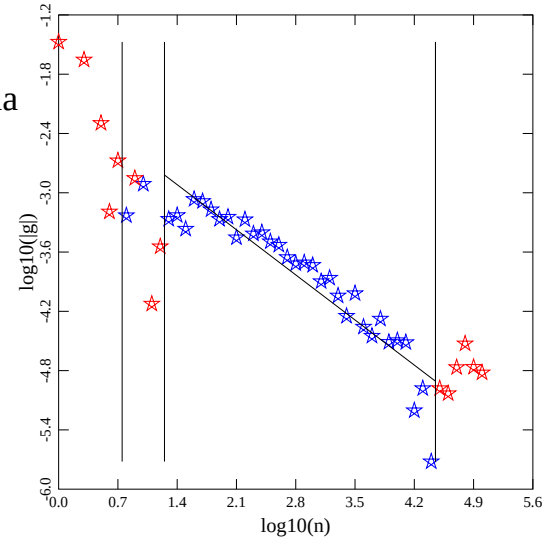

shock



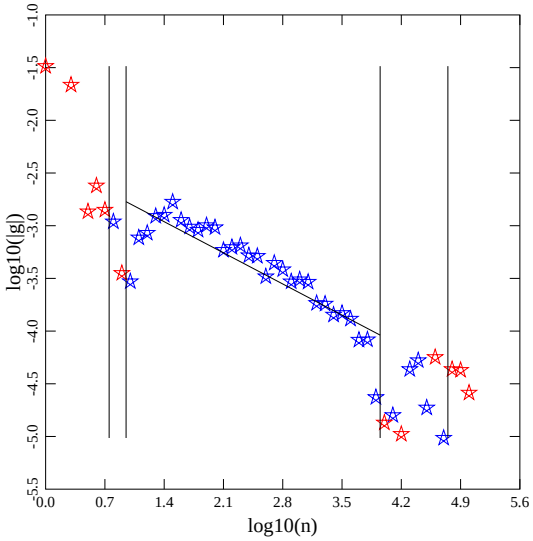

ice



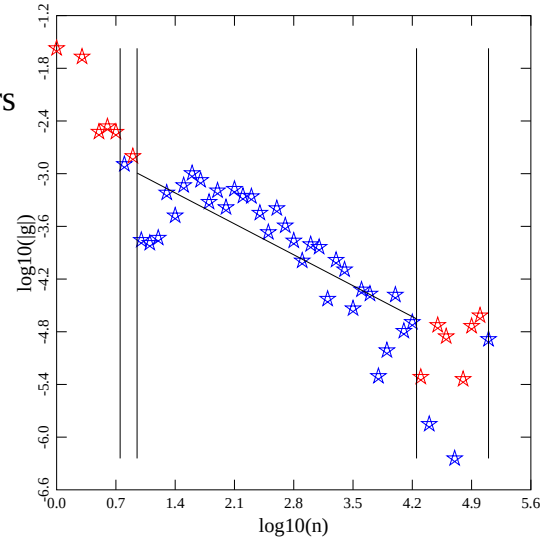
zenda



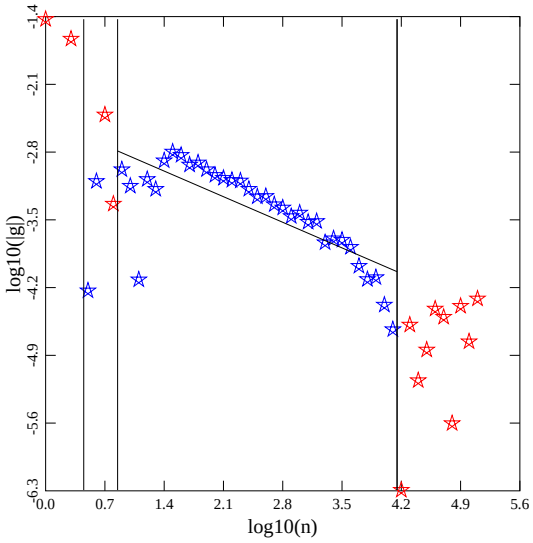

puddnhead



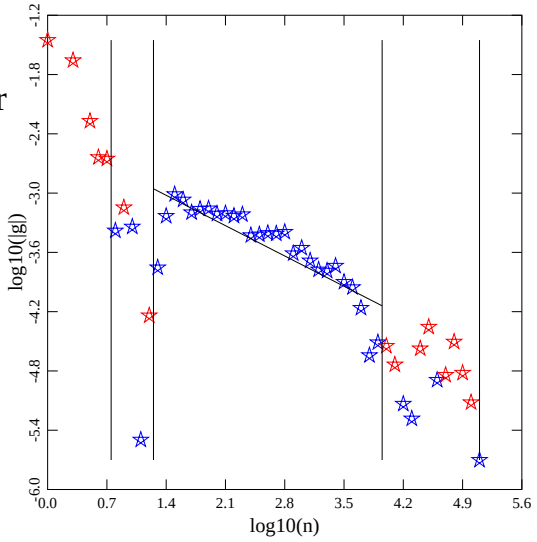

breakers



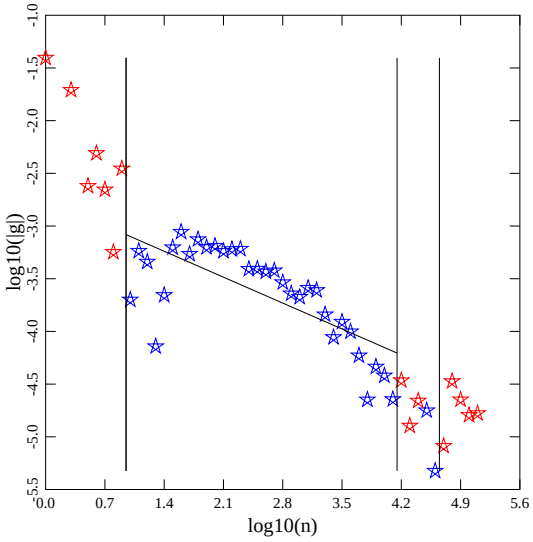

pilgrim



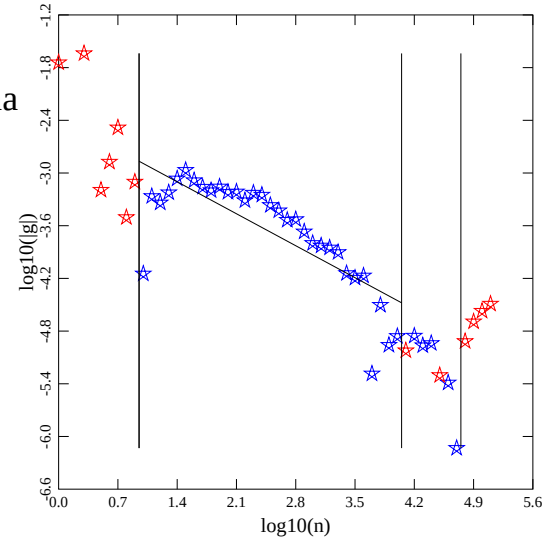

armour



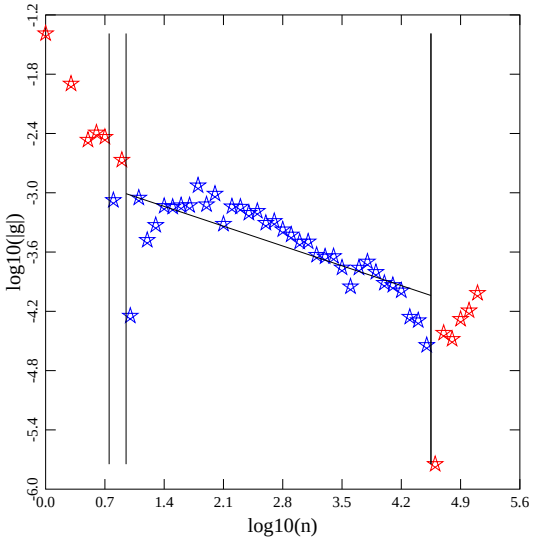

decline



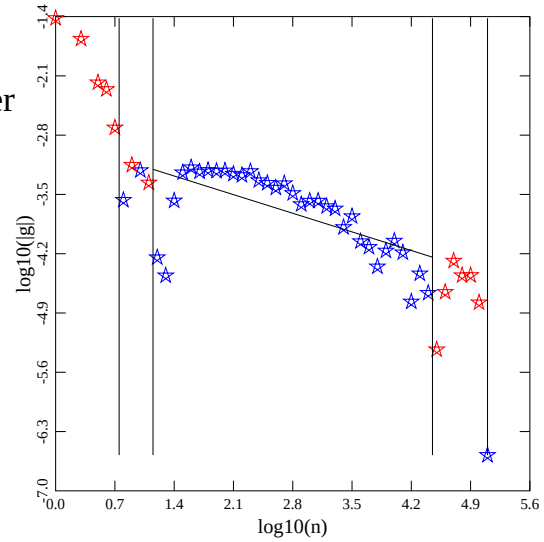
pollyanna



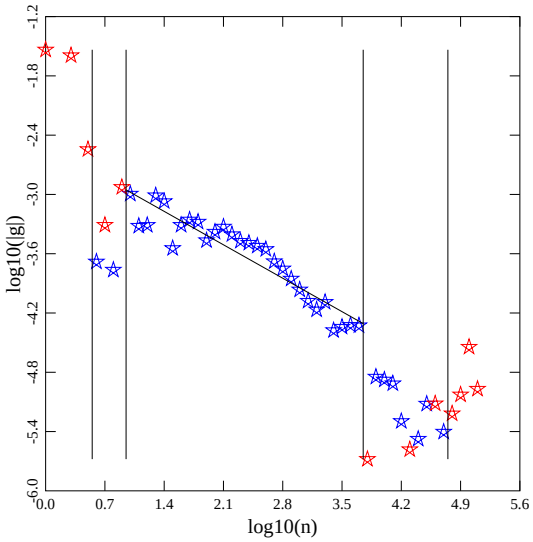

kantright



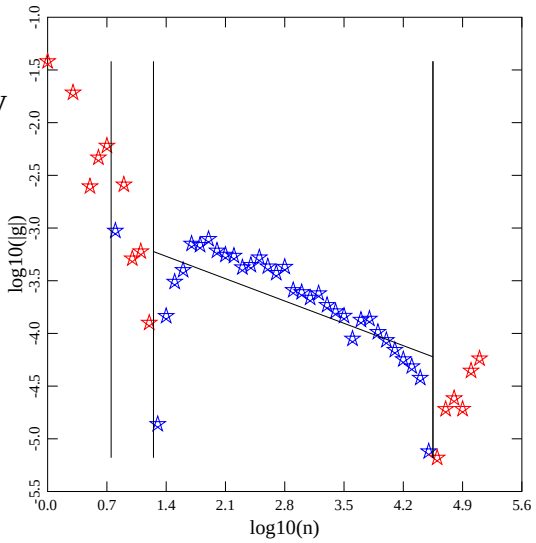

other



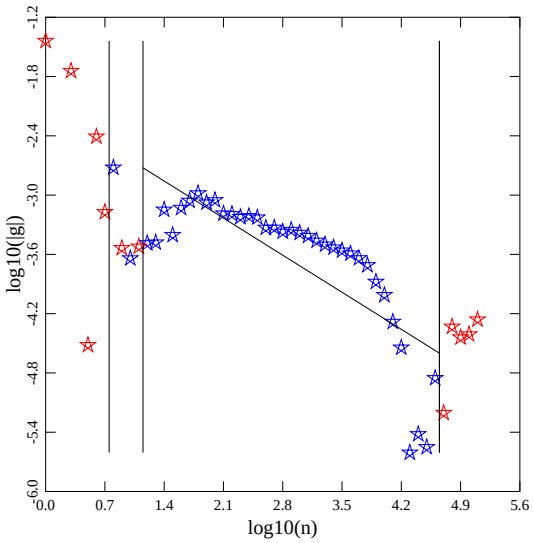

willows



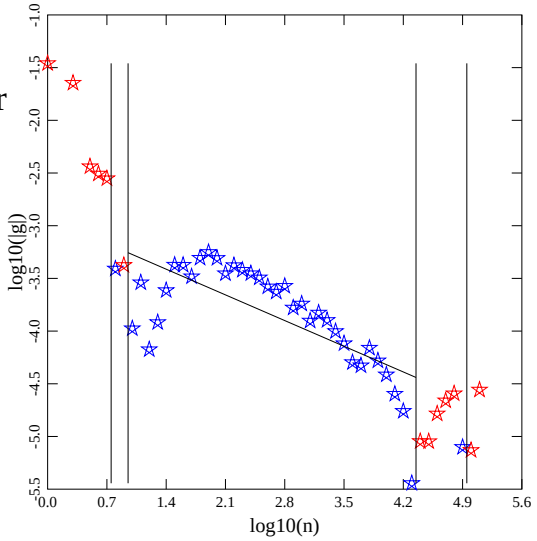

dairy



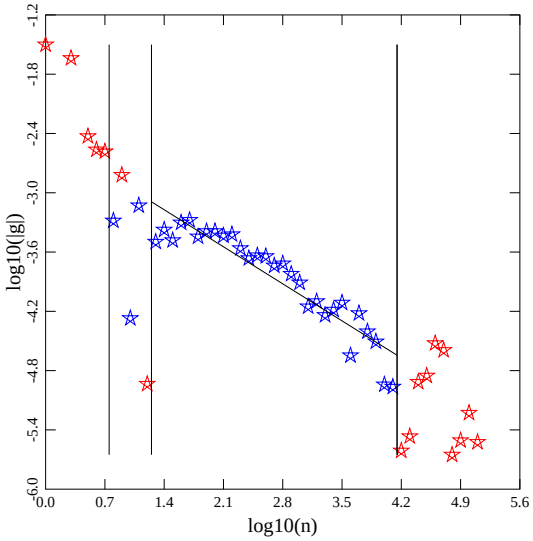

violinists



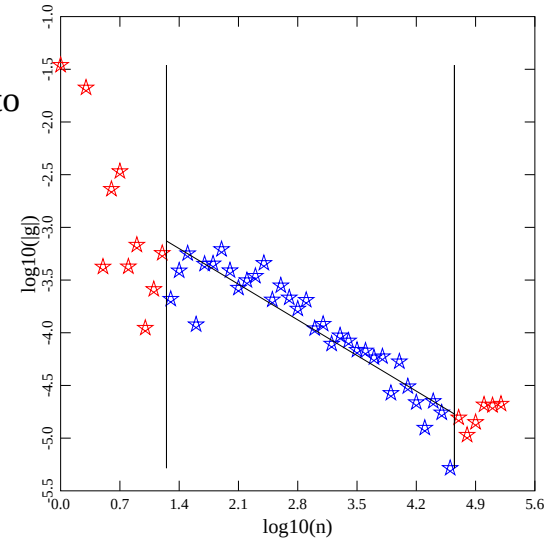
war



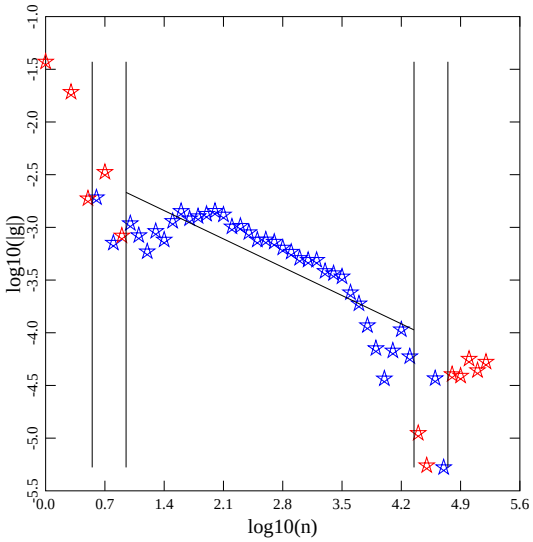

treasure



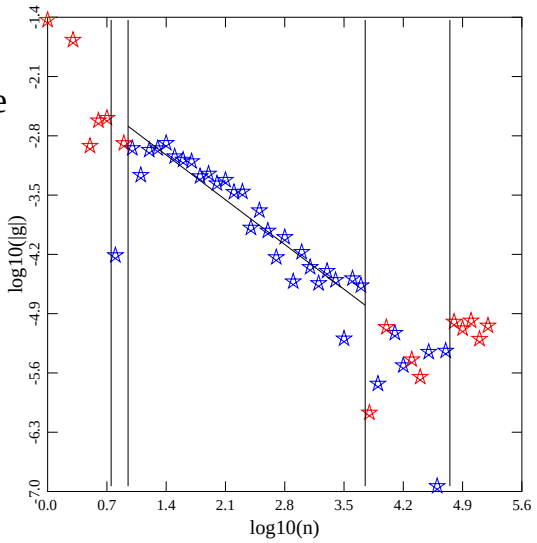

bf_auto



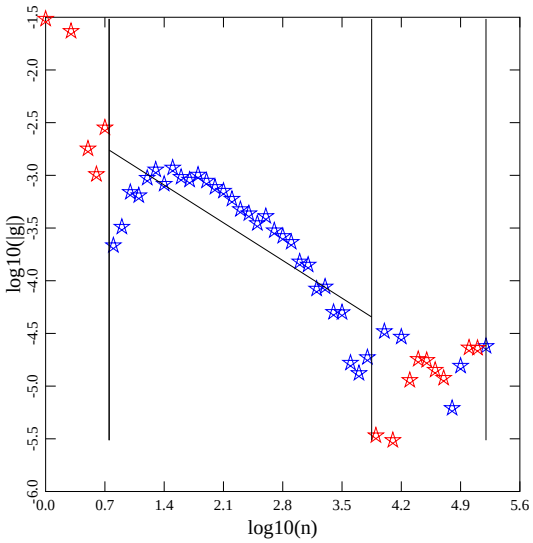

virginia



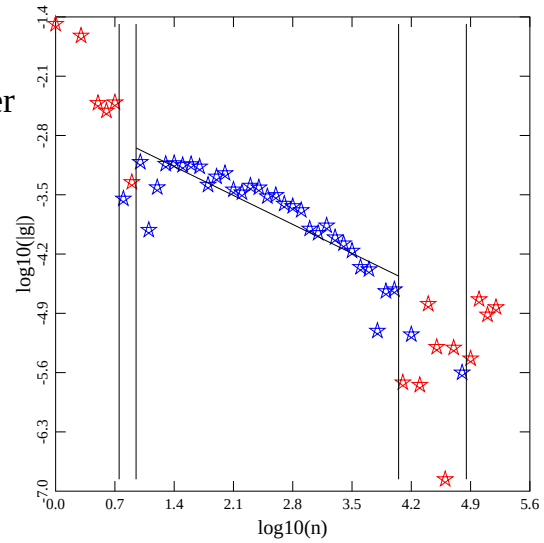

conduct_life



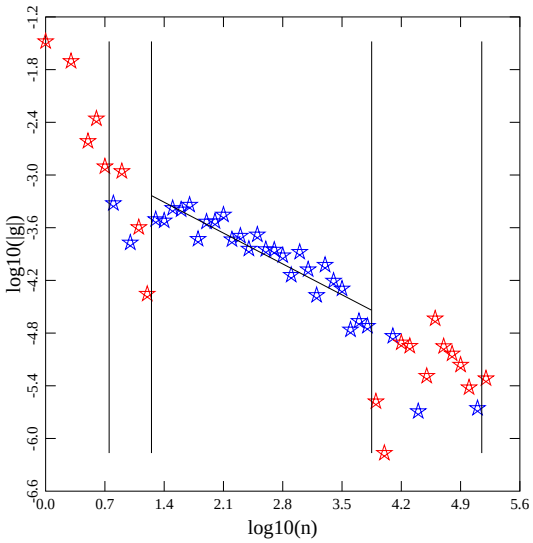

tom_sawyer



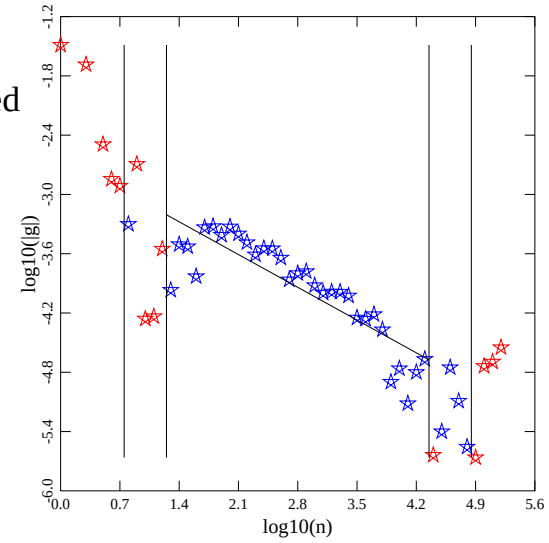
prince_pauper



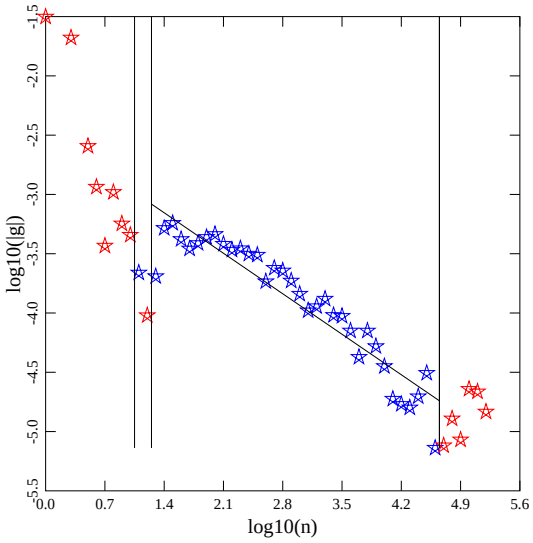

fu_manchu



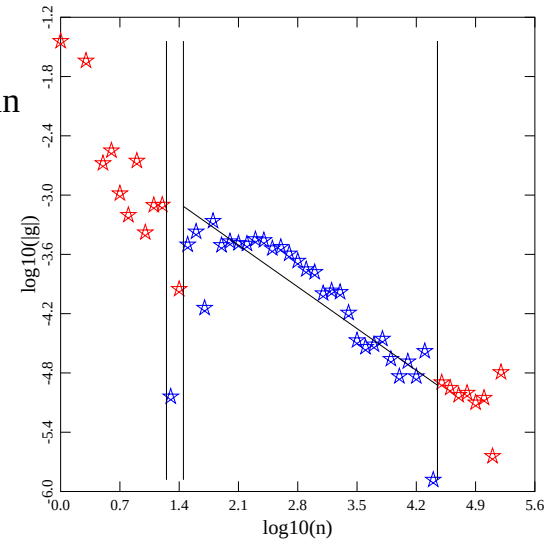

kidnapped



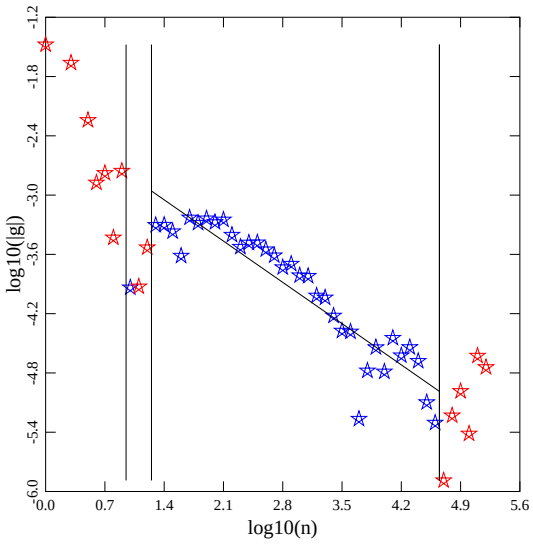

lost_world



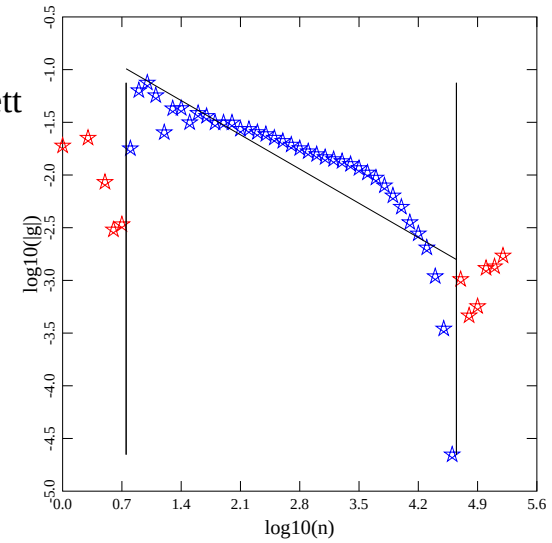

frankenstein



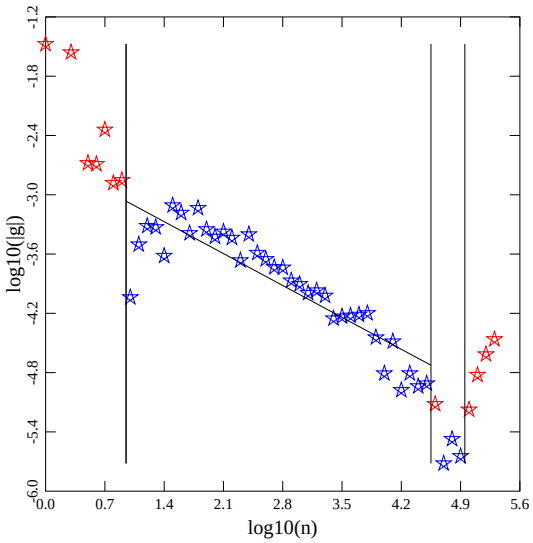

northanger



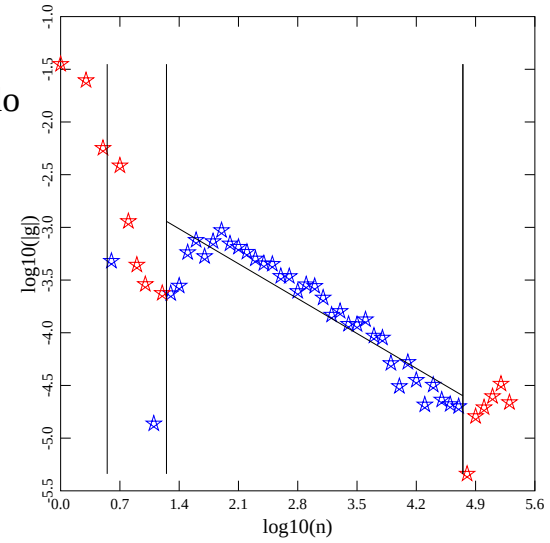
zc_lett



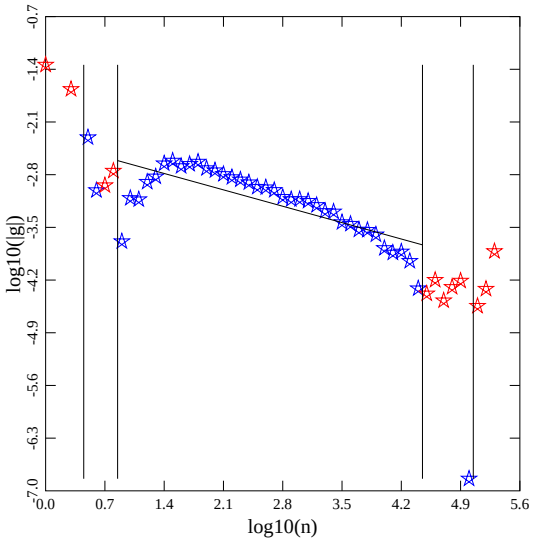

paradise



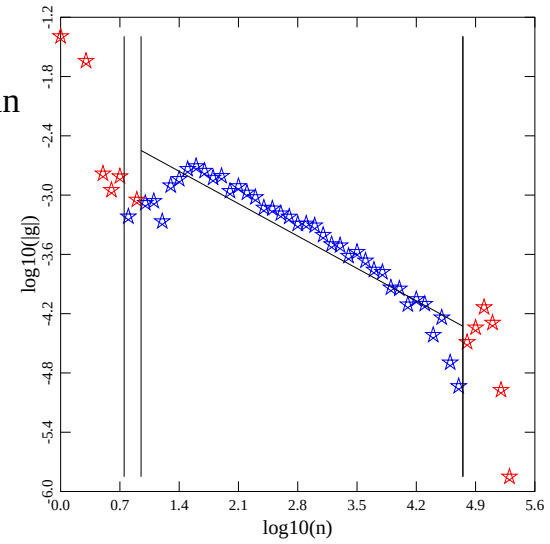

colorado



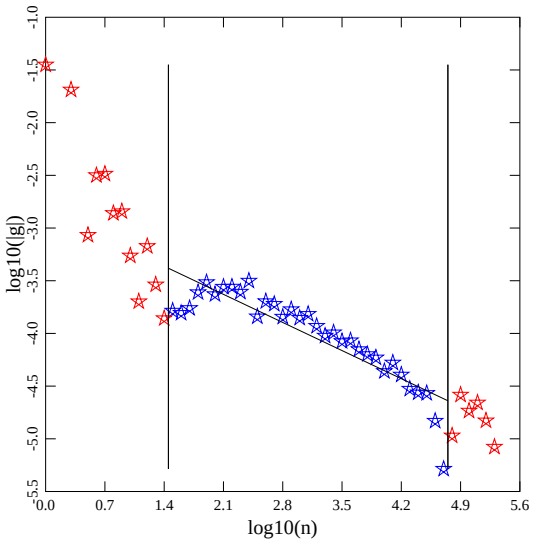

ari_physics



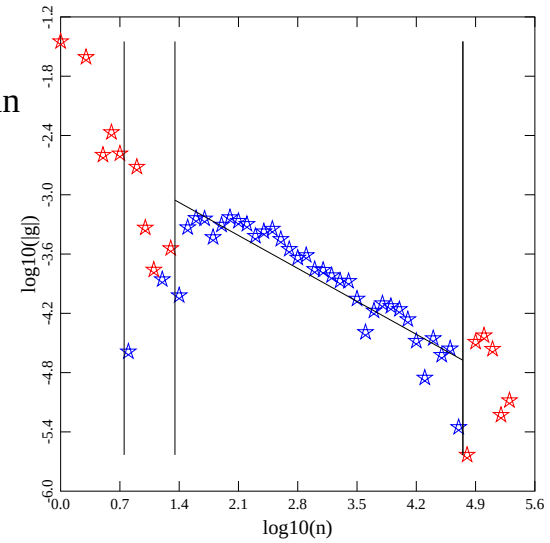

nicomachean



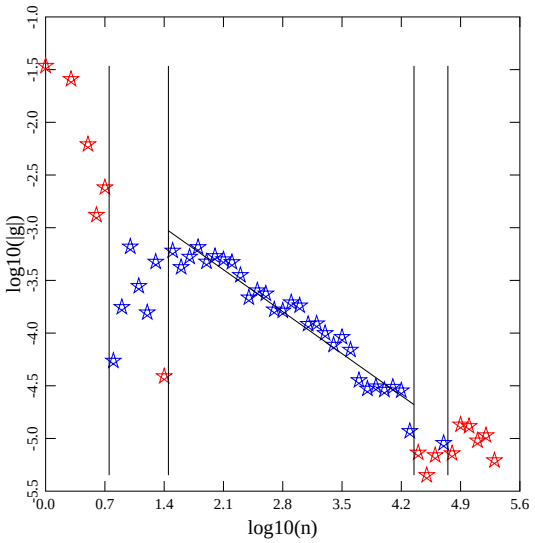

scarlet_letter



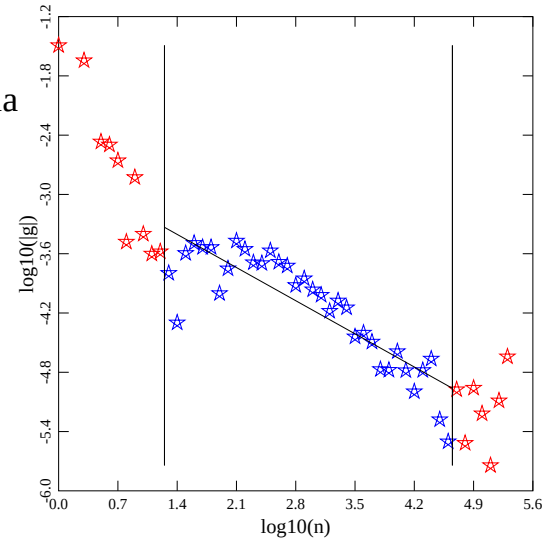
tarzan



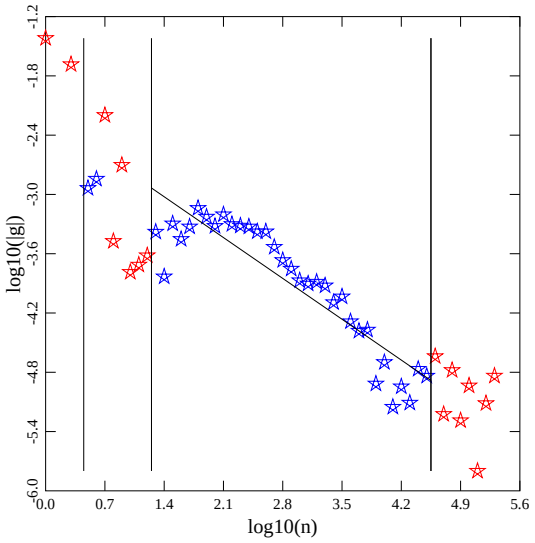

steep



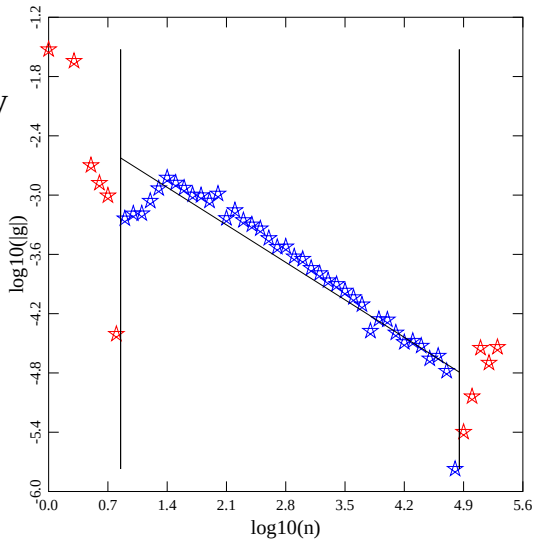

pretoria



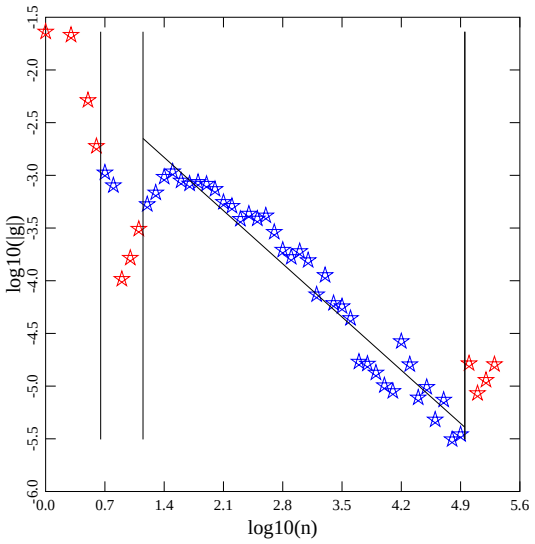

chymist



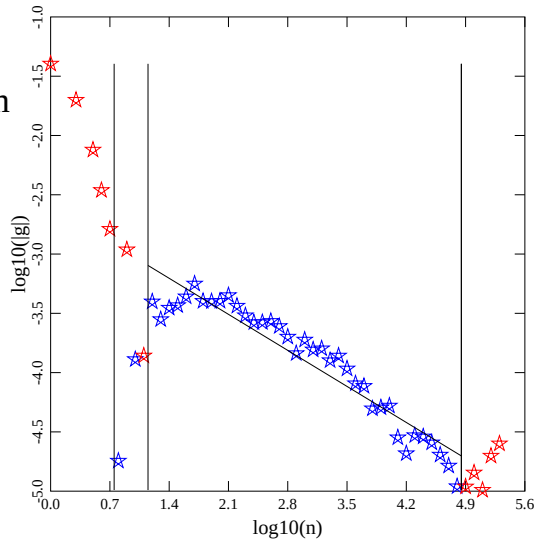

huckleberry



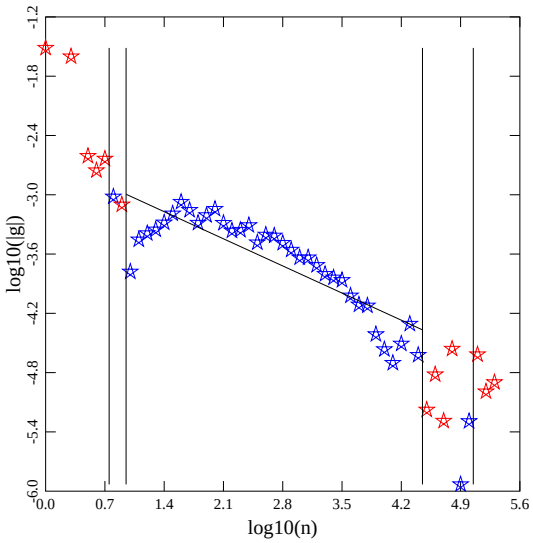

green_gables



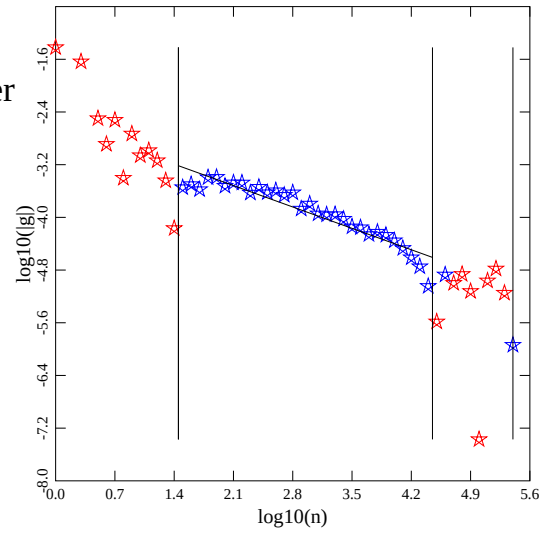
belgium



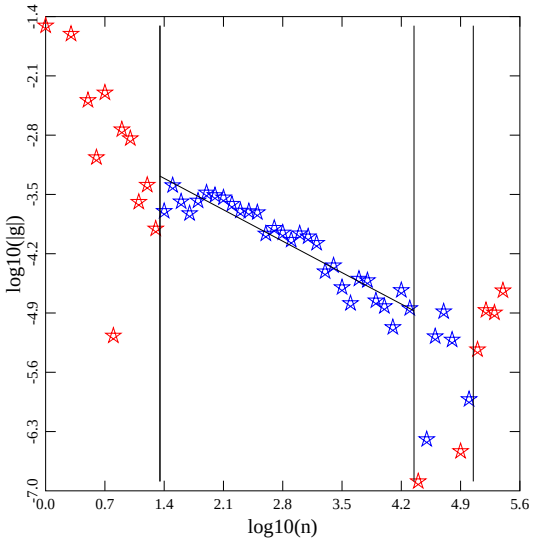

purple



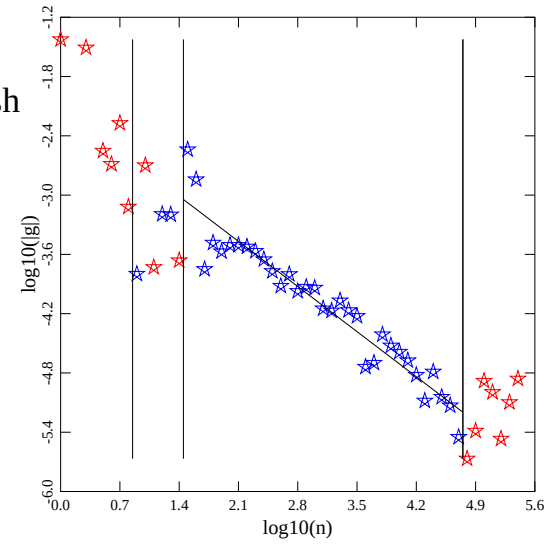

gulliver



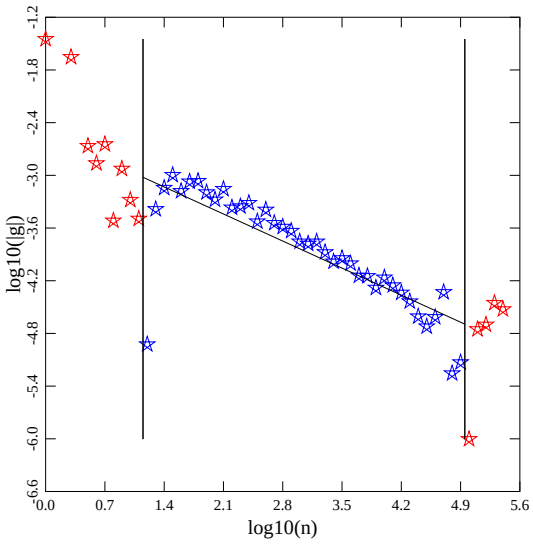

odyssey



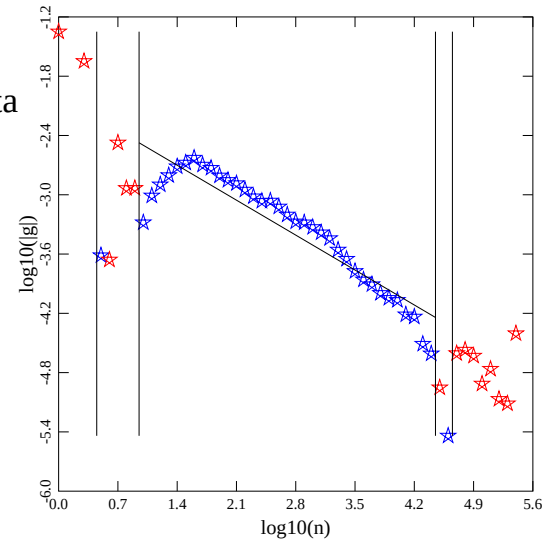

aeneid_english



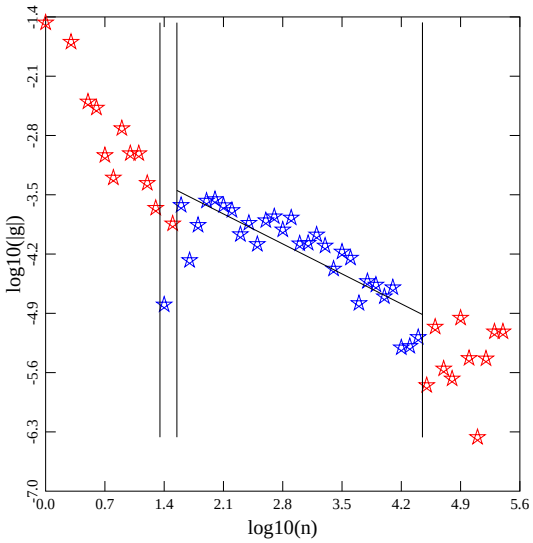

crusoe



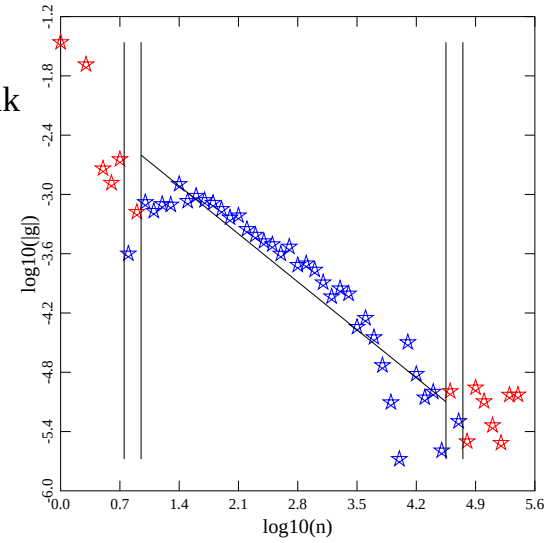
ari_meta



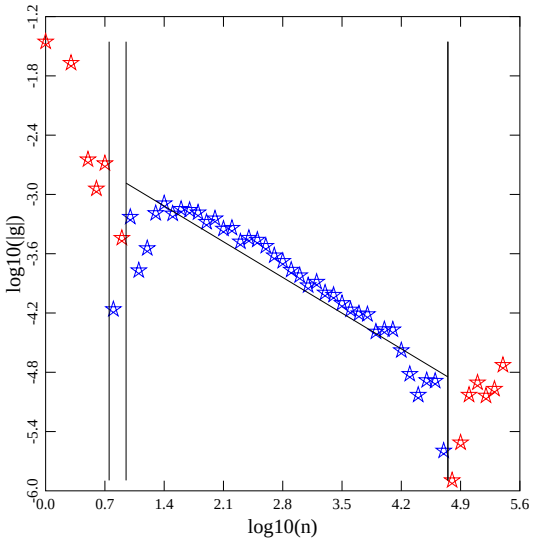

seven_gables



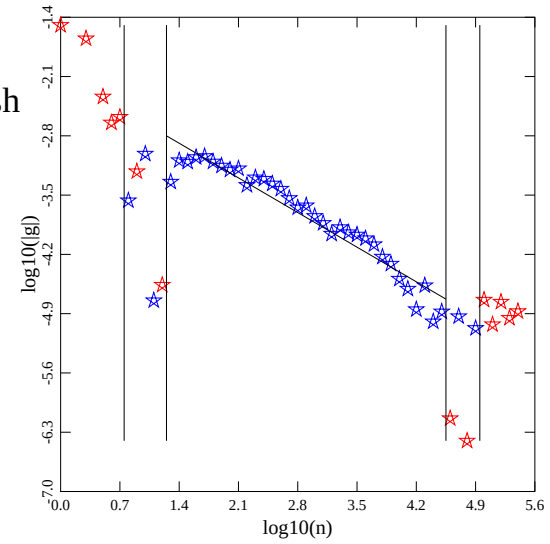

connyank



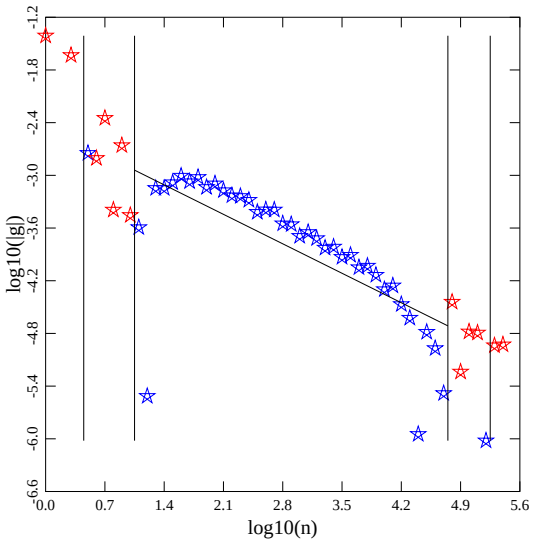

walden



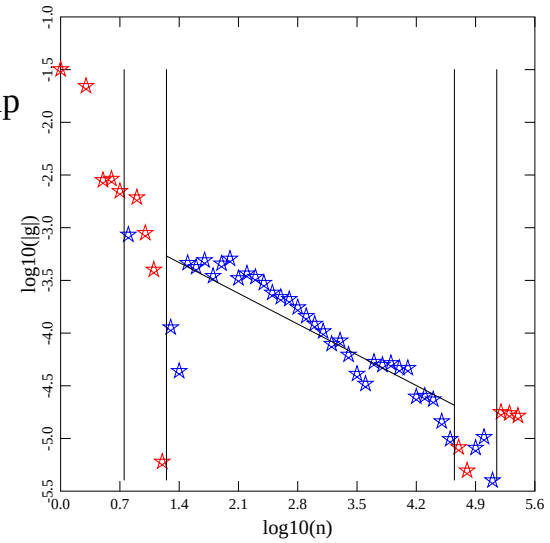

bush



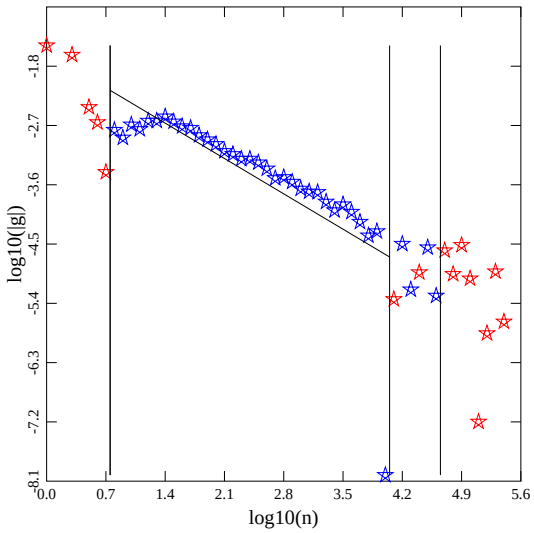

republic



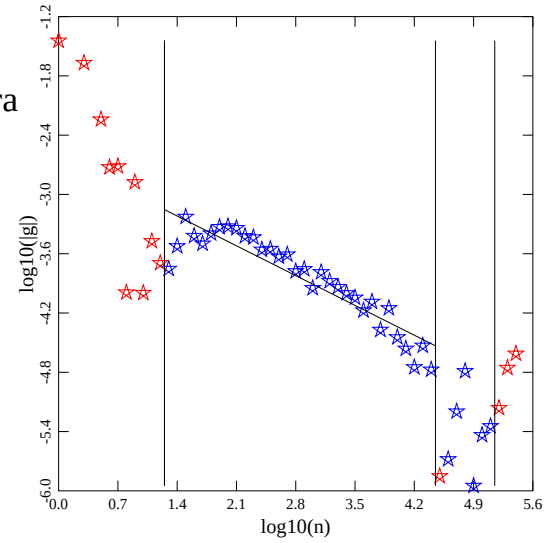
pnp



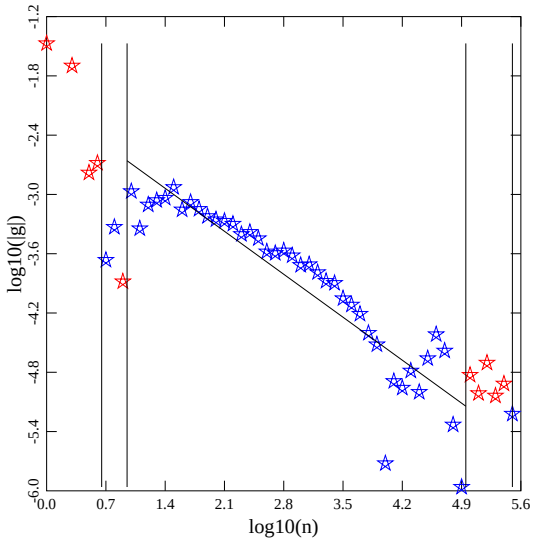

leaves



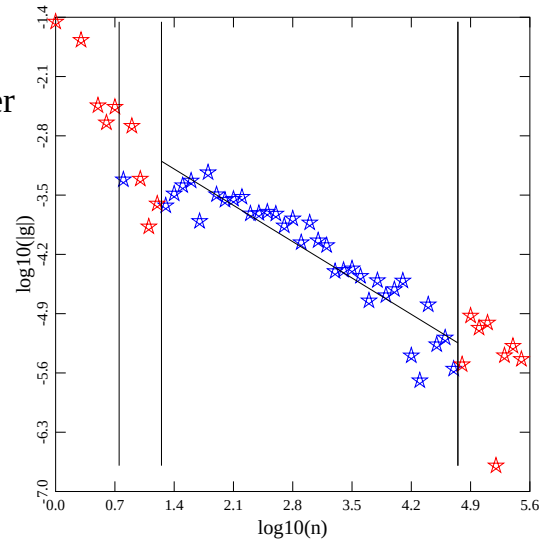

alhambra



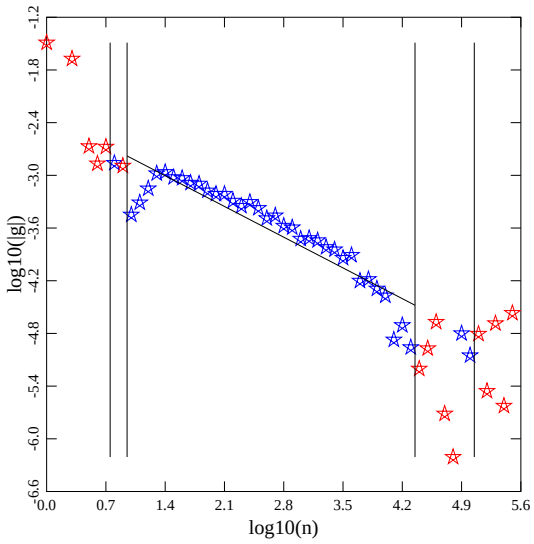

american



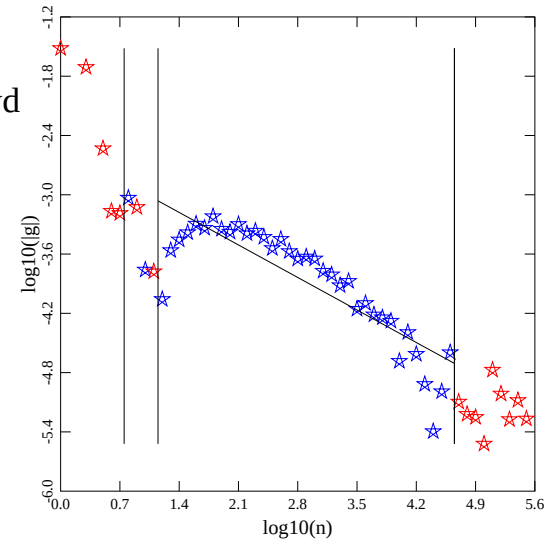

knickerbocker



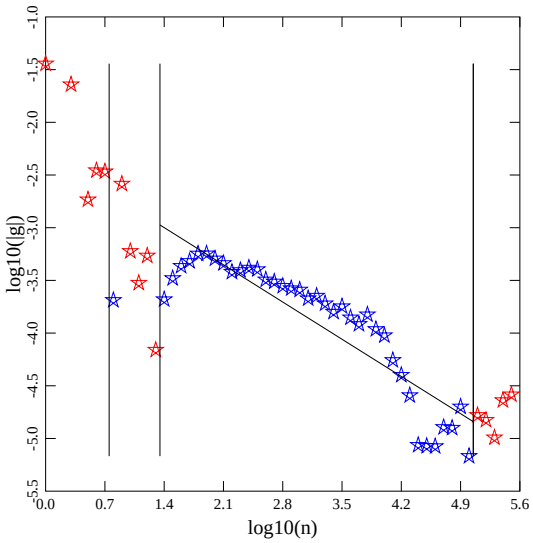

2cities



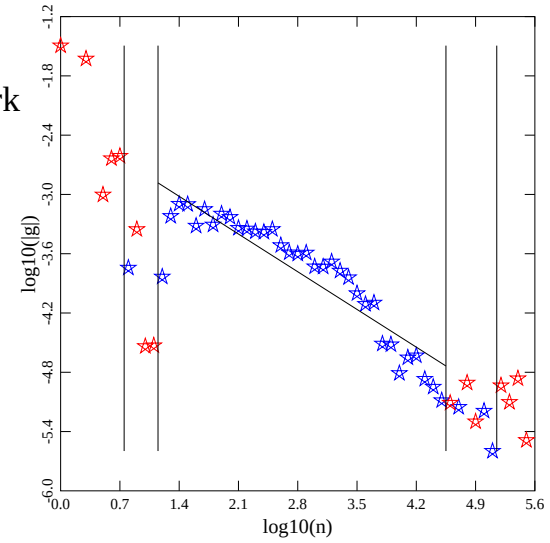
madcrowd



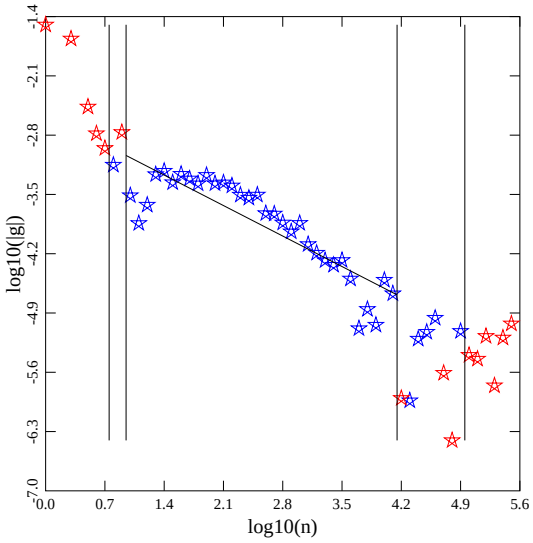

narratives



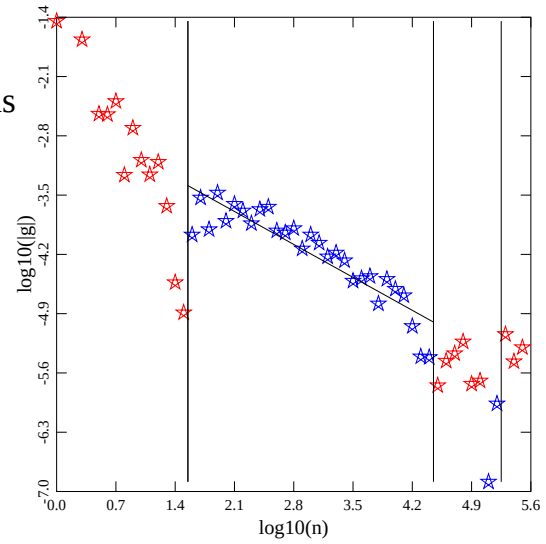

song_lark



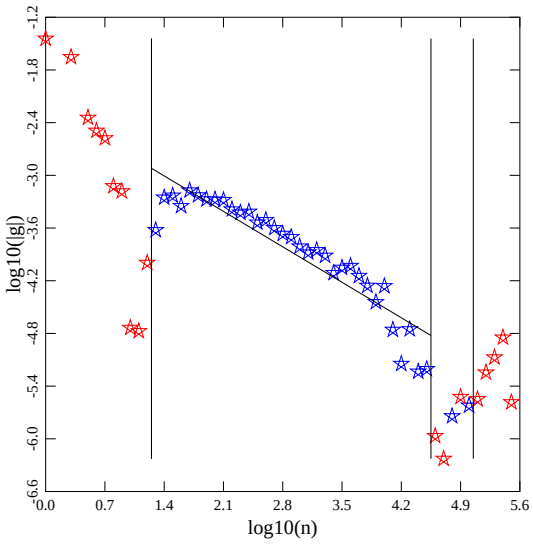

tess



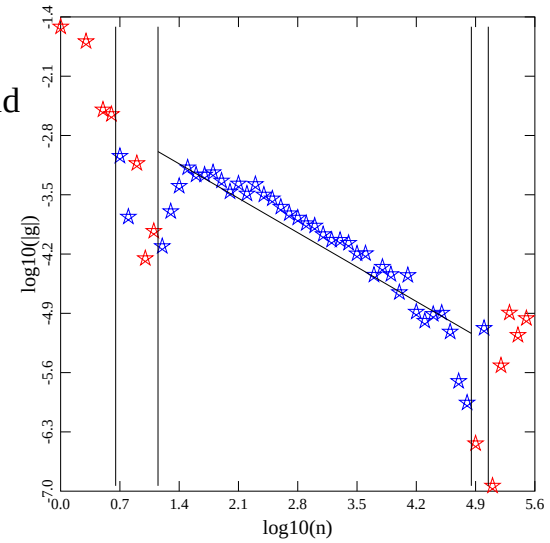

mohicans



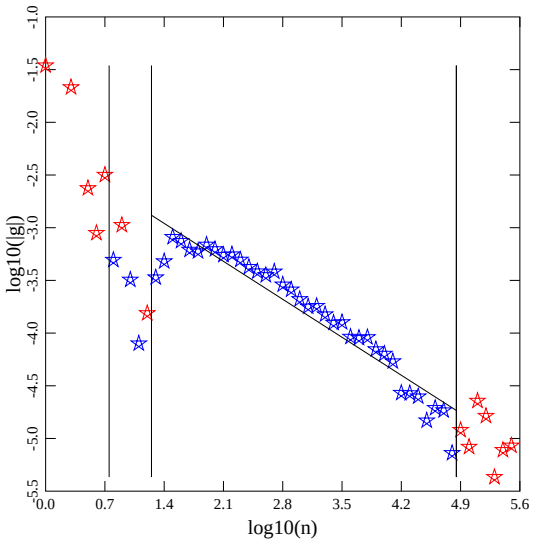

pompeii



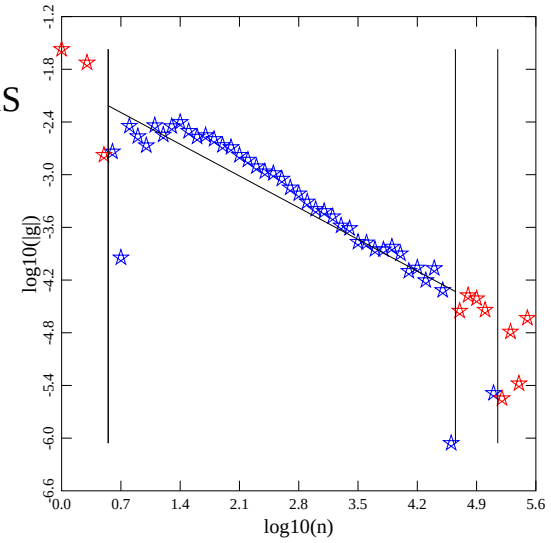
mansfield



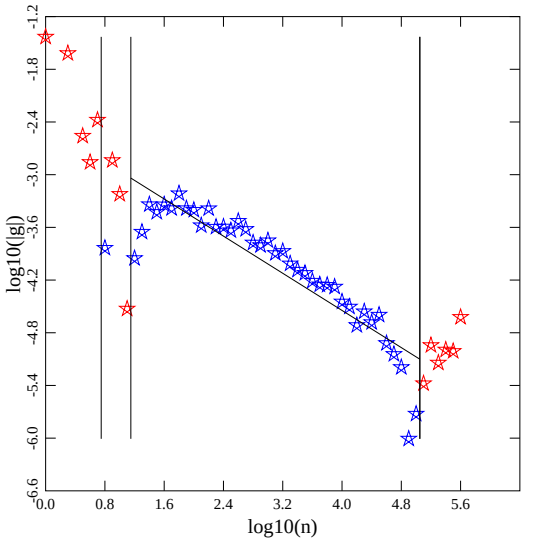

2years



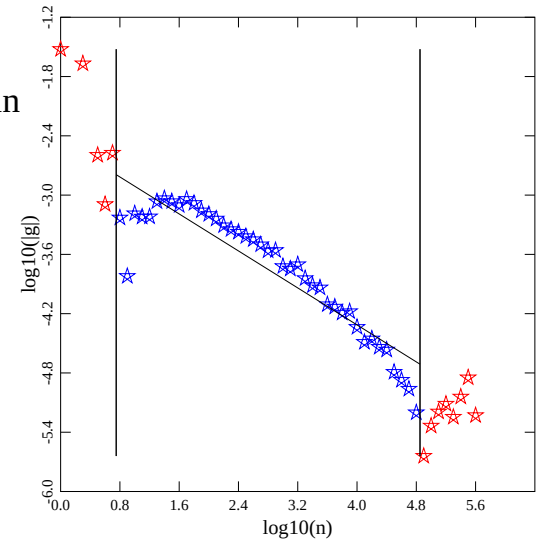

GandS



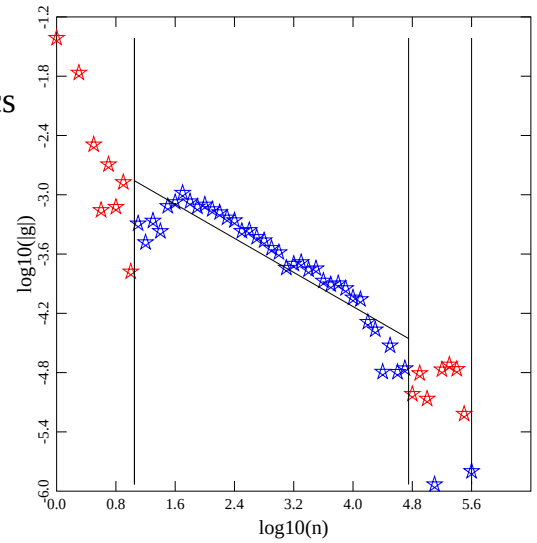

childs_history

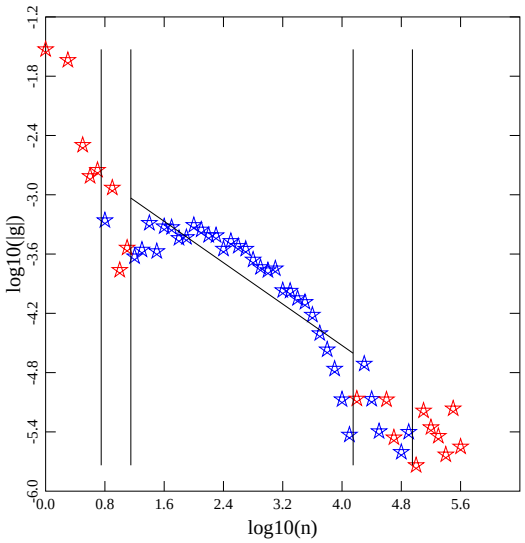



cabin



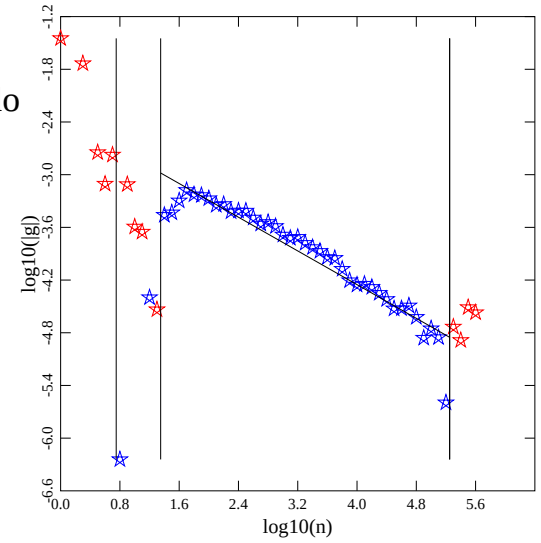
eugenics

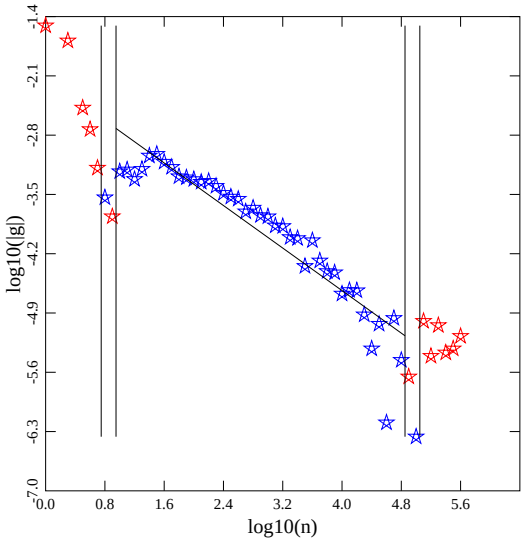



little_women

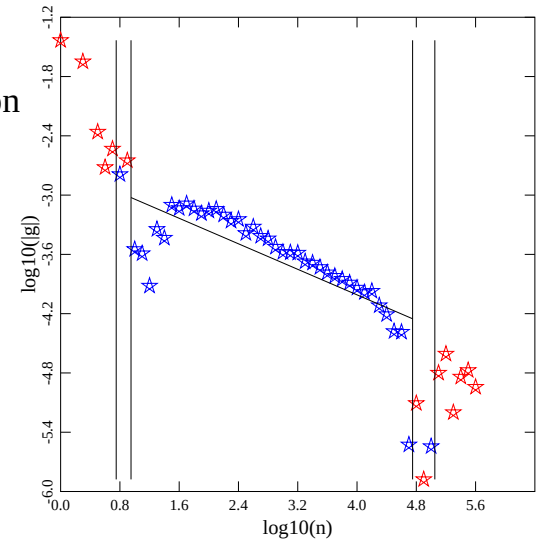
dido



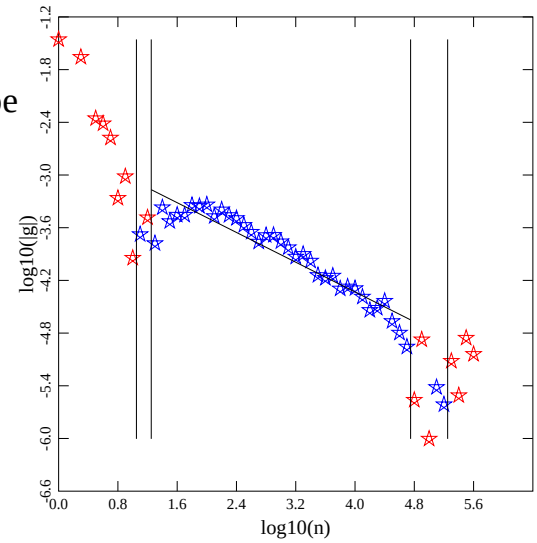




eyre

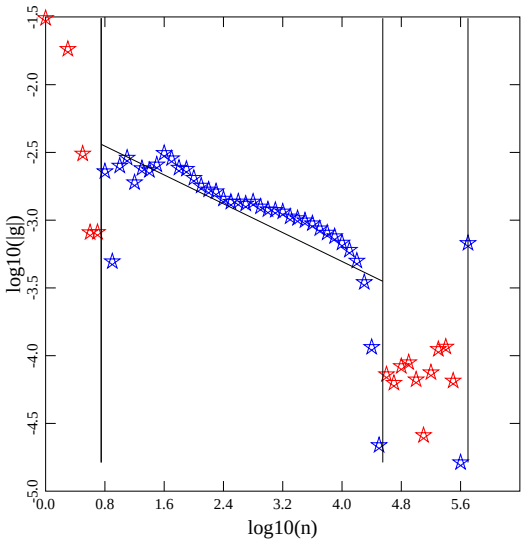
hudson



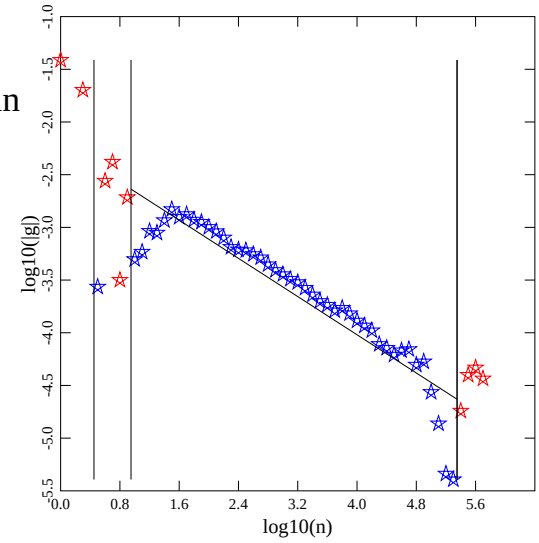
ivanhoe



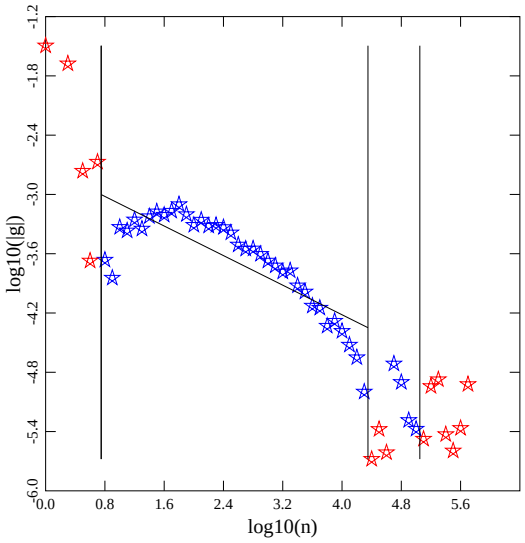

argentine



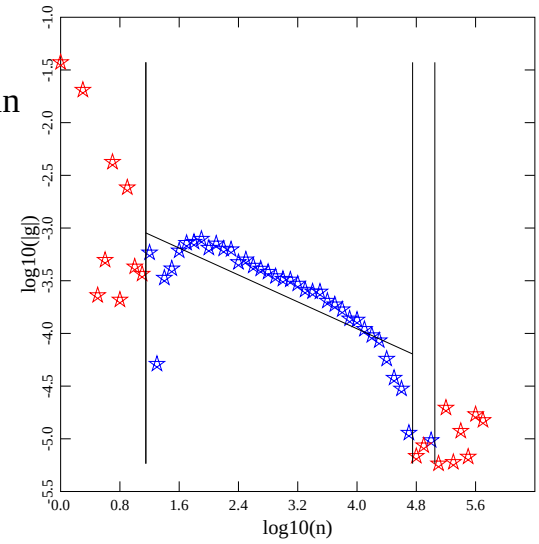

leviathan



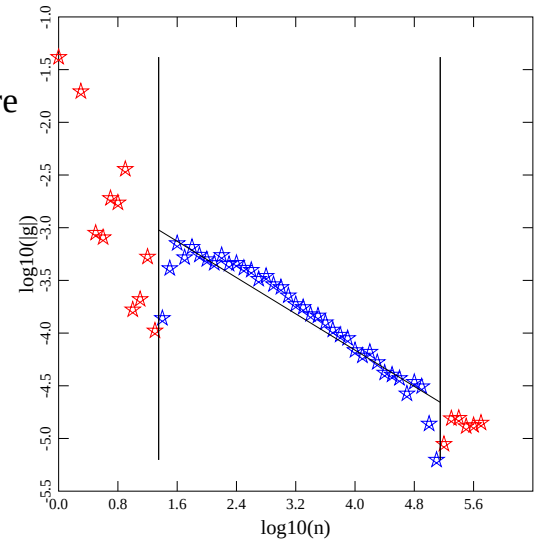

mobydick



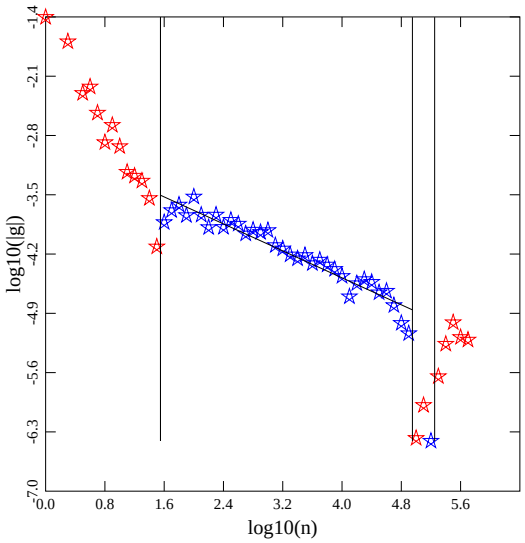

origin



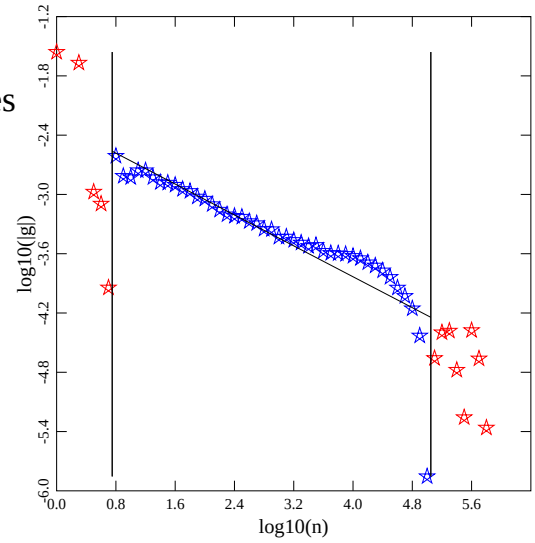
human.nature



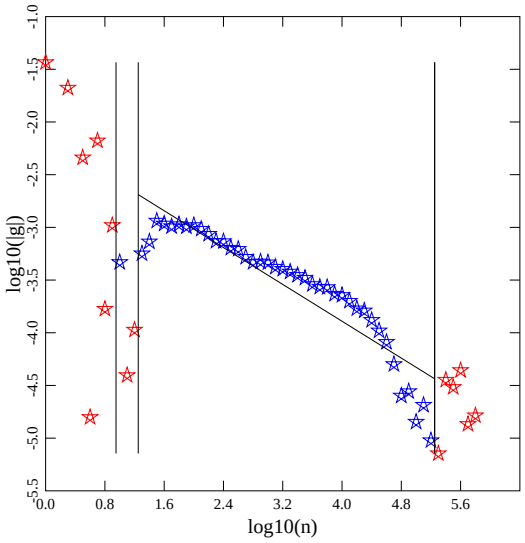

mexcon



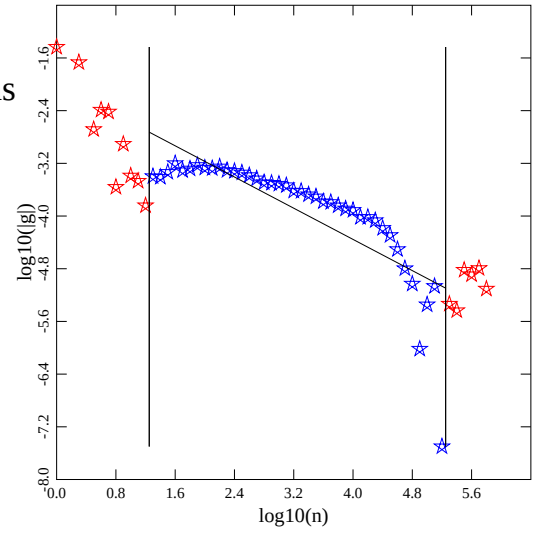

ulysses



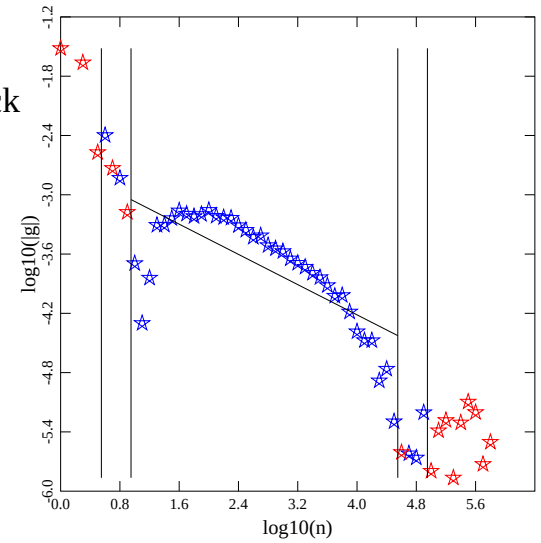

cattle



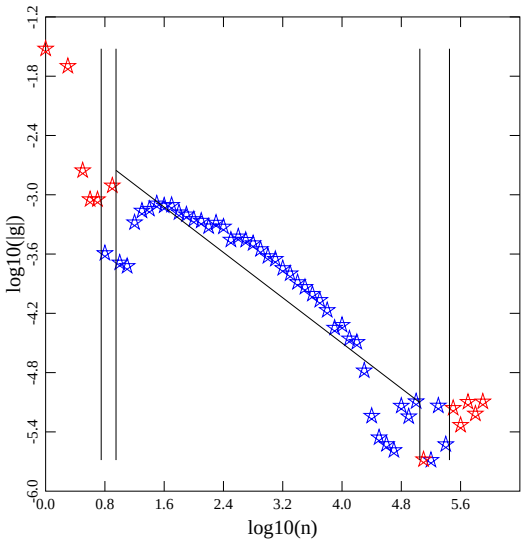

delusions



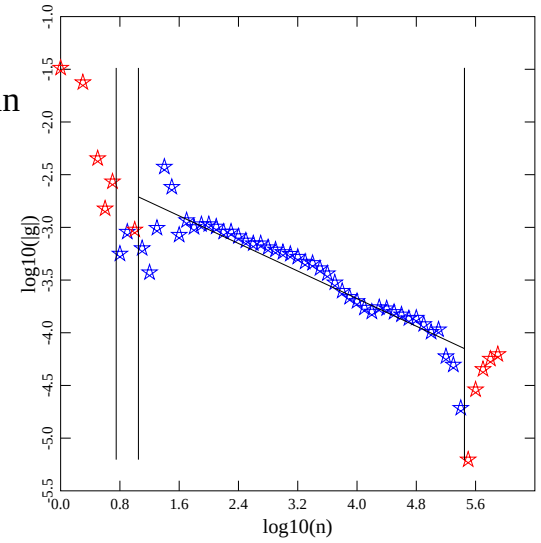
pickwick



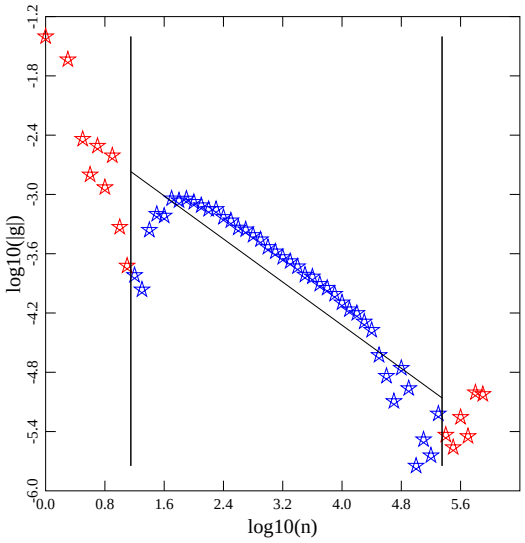

copperfield



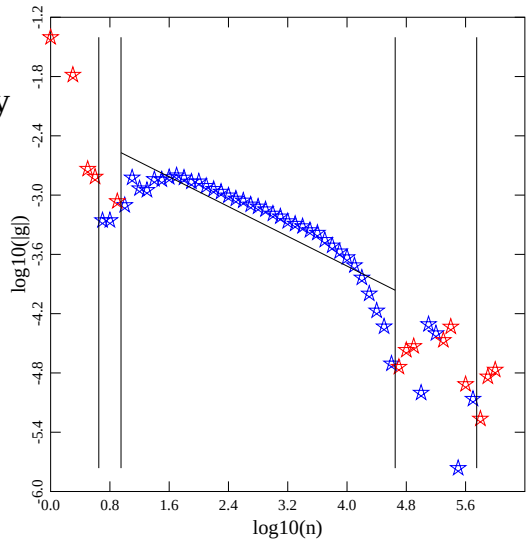

ramayan



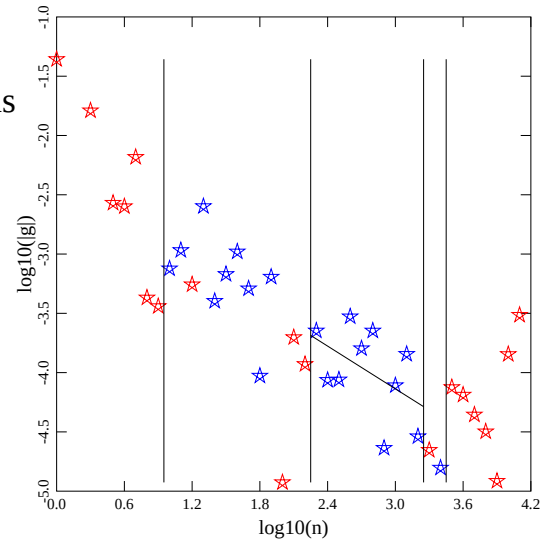

wealth



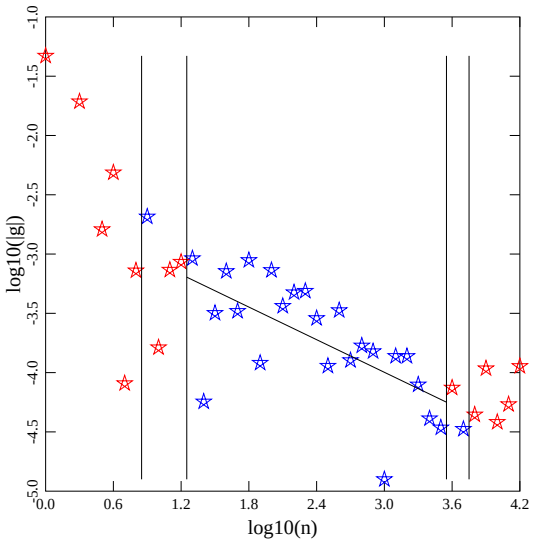

sociology



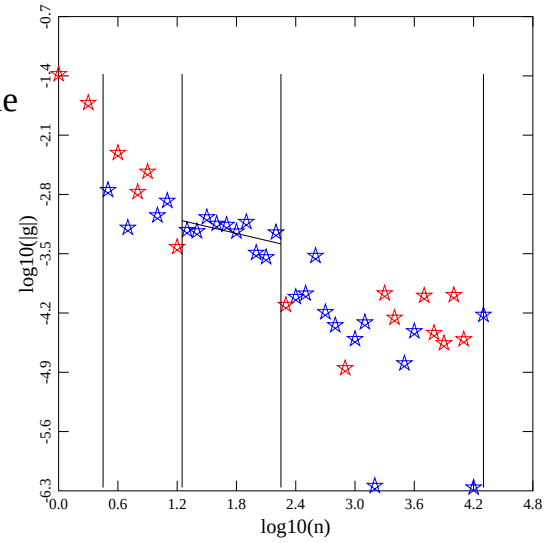
dreyfus



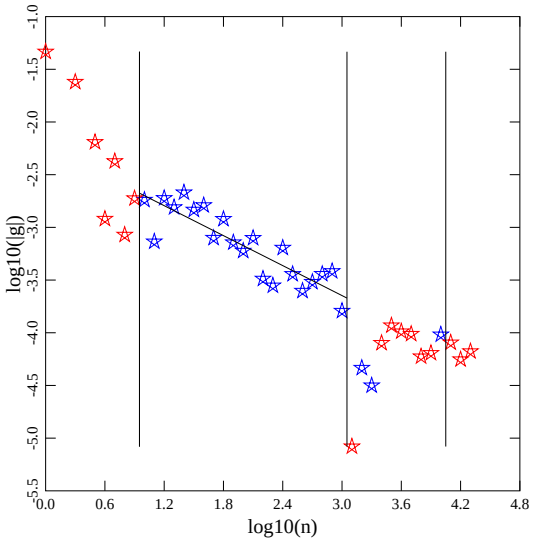

bounty_verne



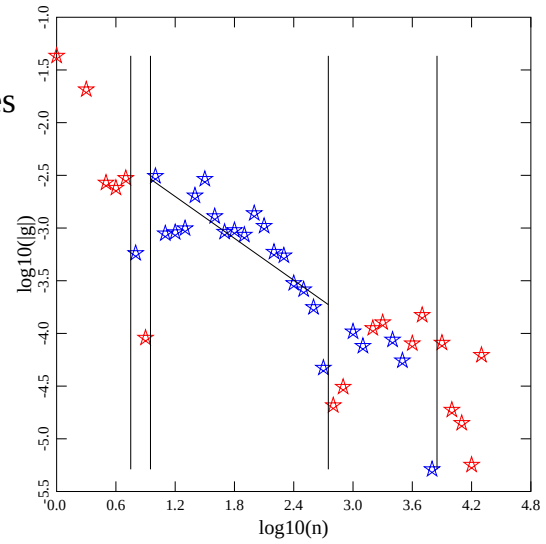

adelaide



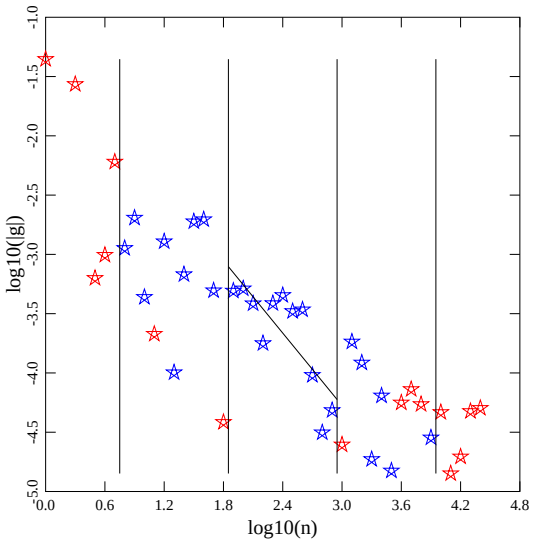

esclaves



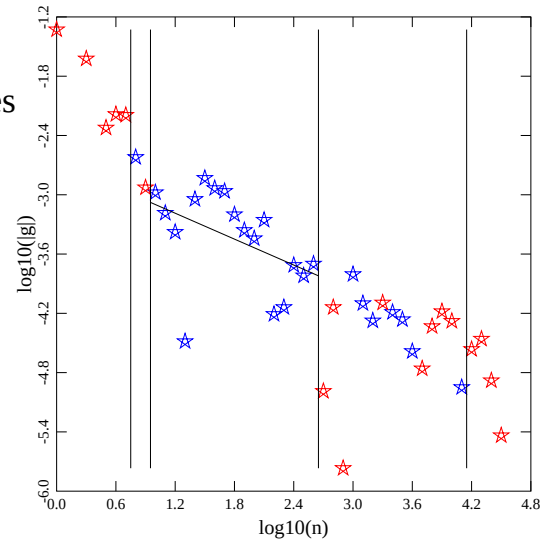

anglicismes



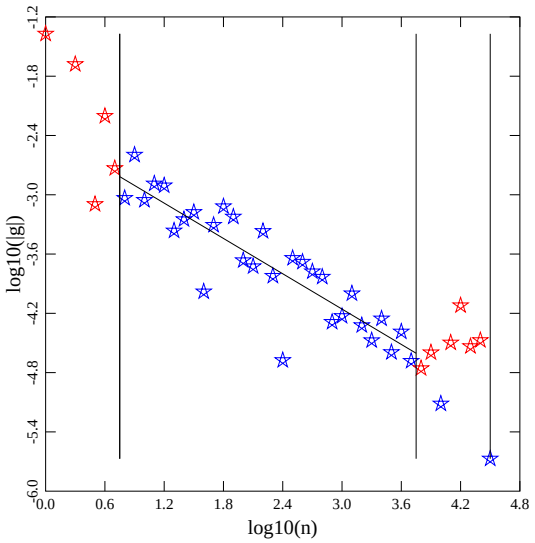

esther_racine



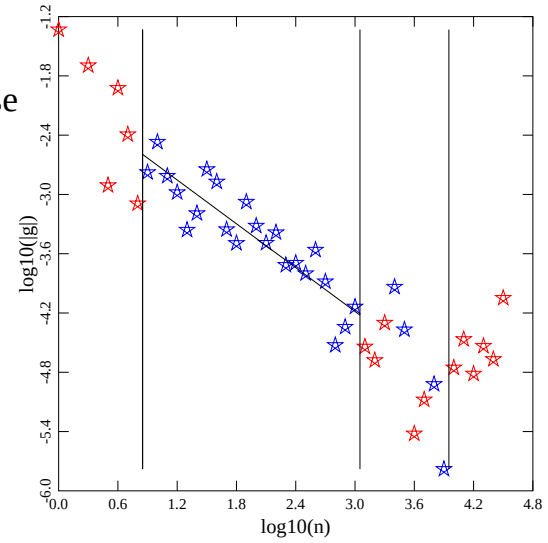
maximes



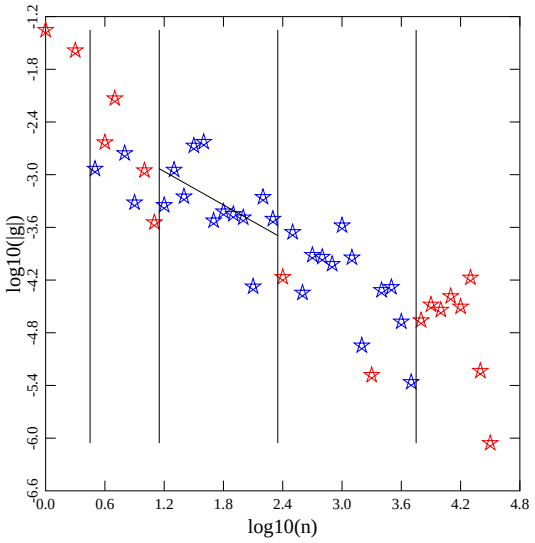

caribou



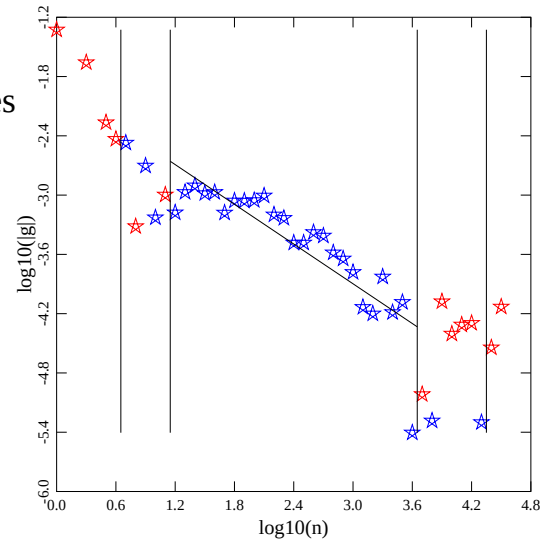

paresse



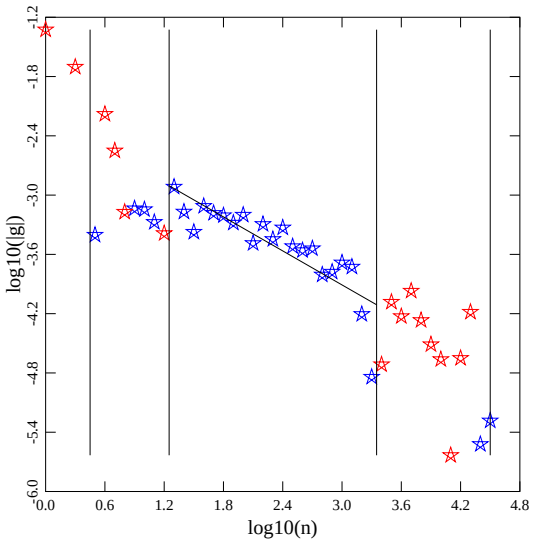

andromaque



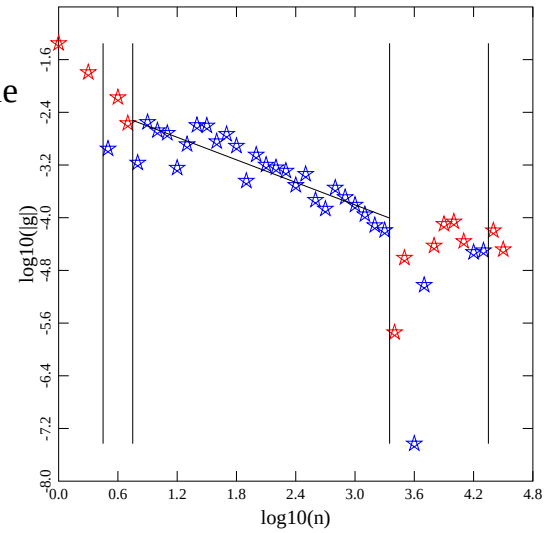

fourberies



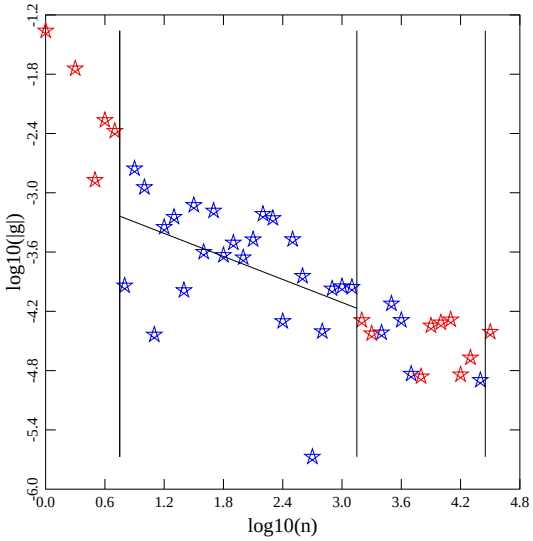

jeu



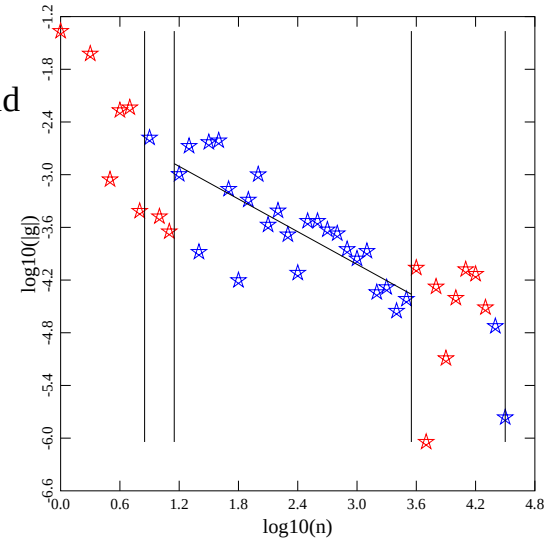
bougainville



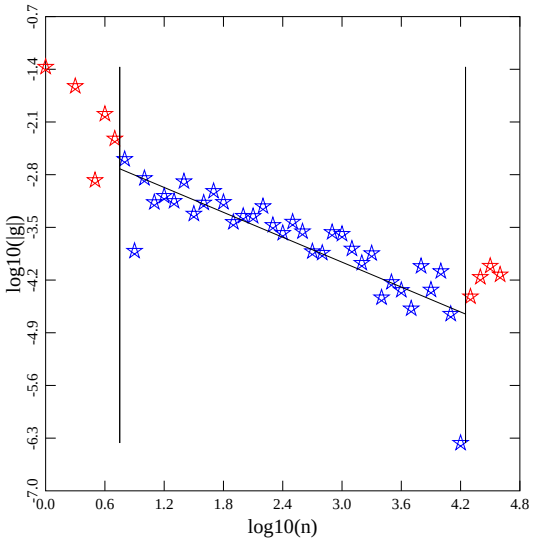

montbard



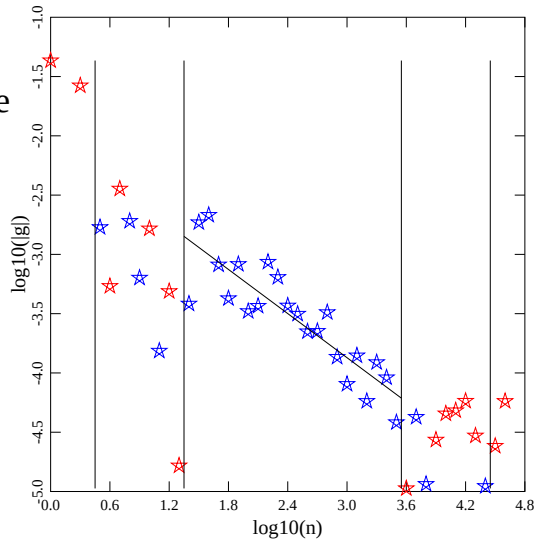

LeCid



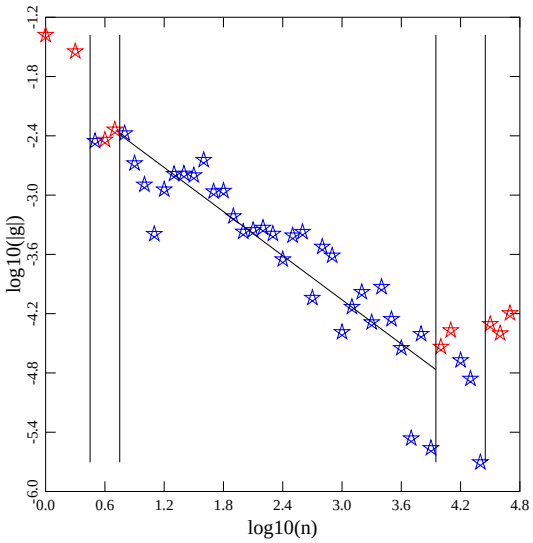

smarra



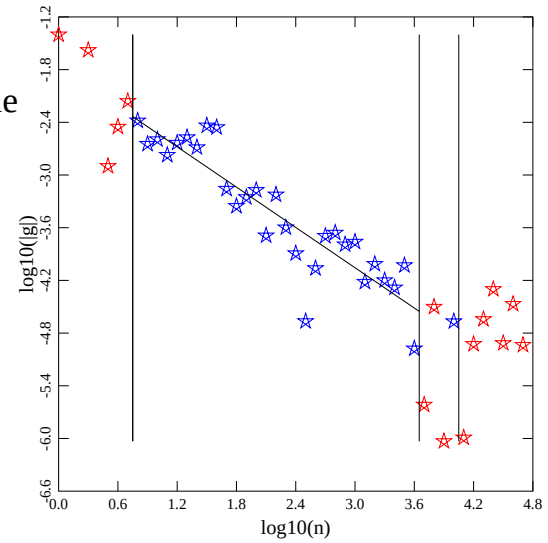

tartuffe



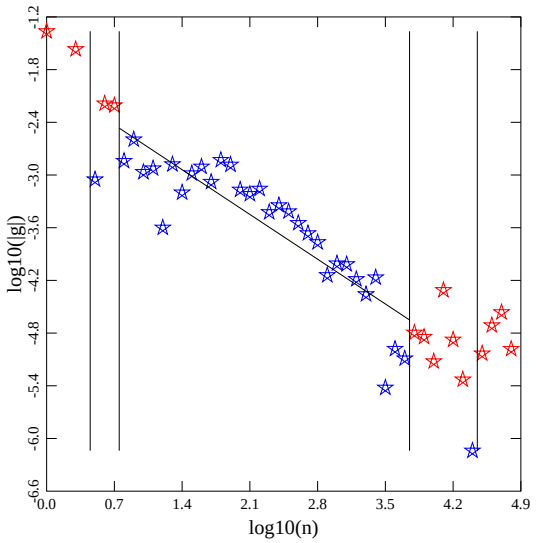

harmonies



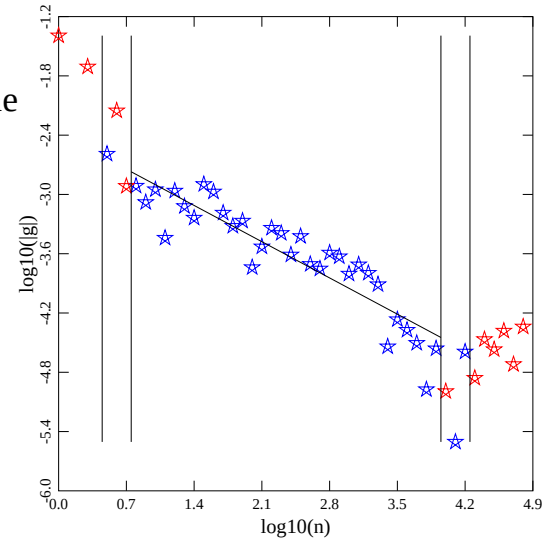
ane



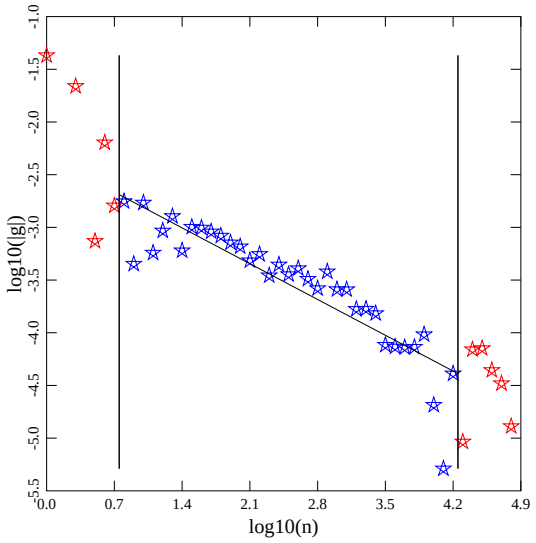

fleurs



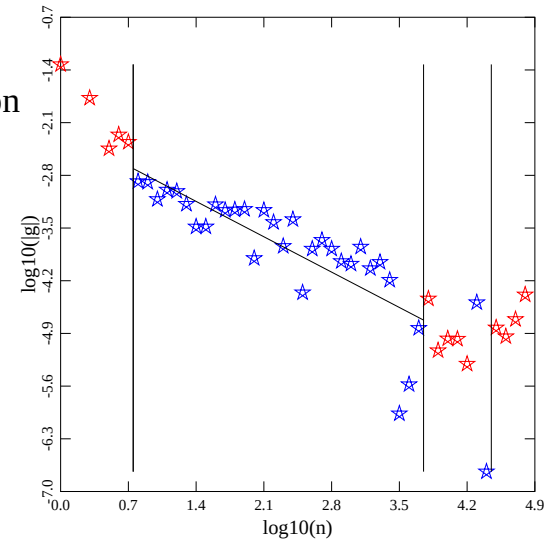

lune



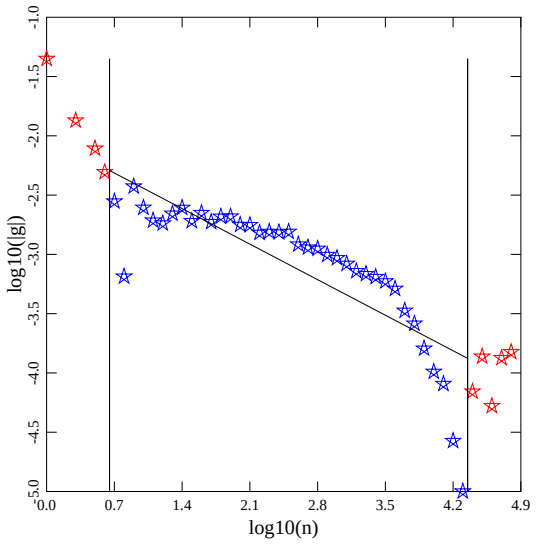

yeux_d_or



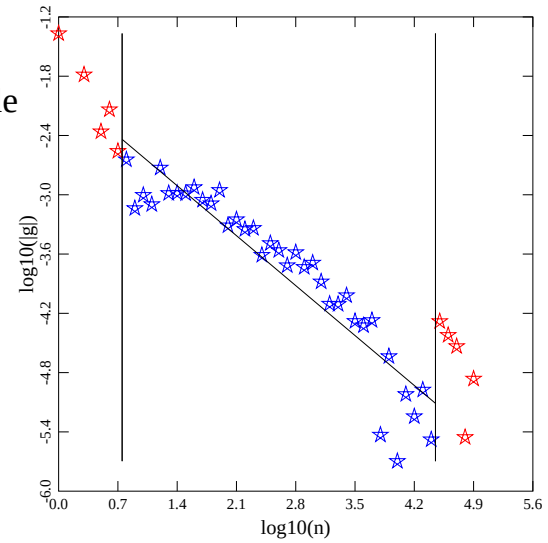

inquisition



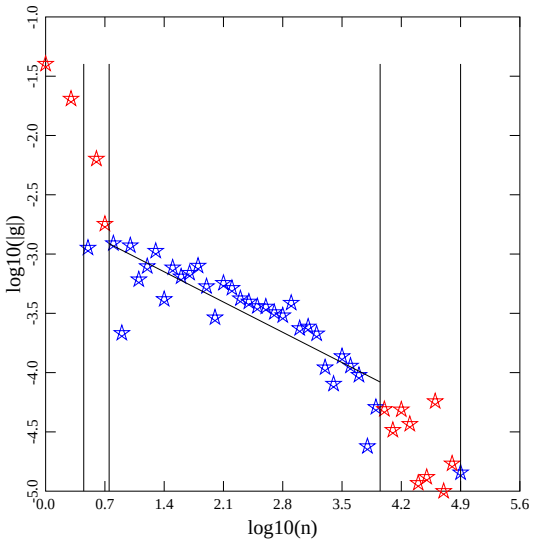

hp_bio



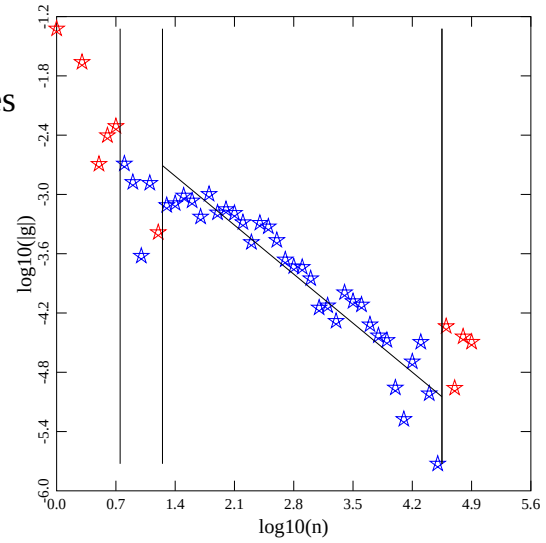
candide



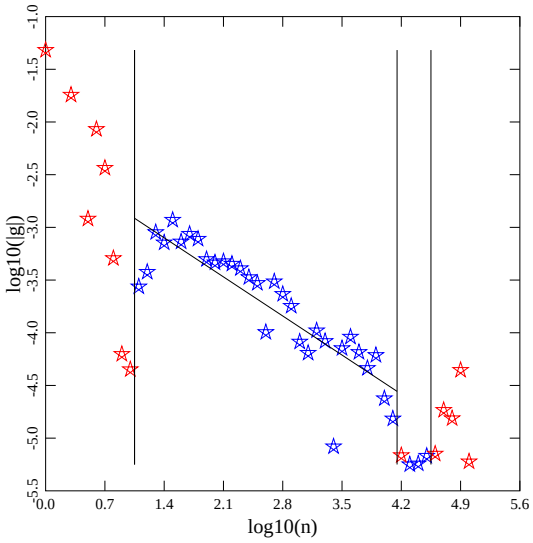

becasse



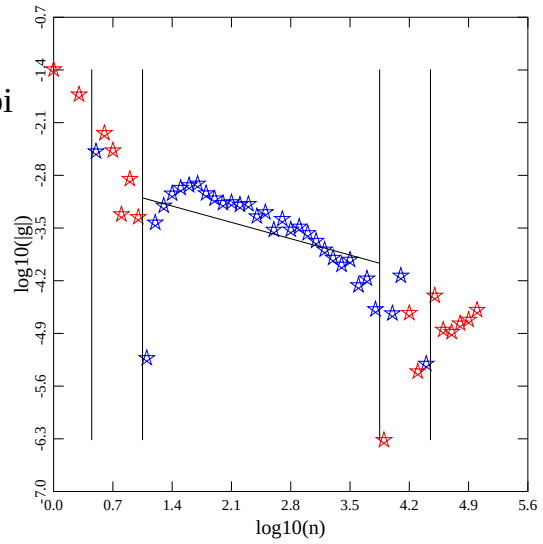

mondes



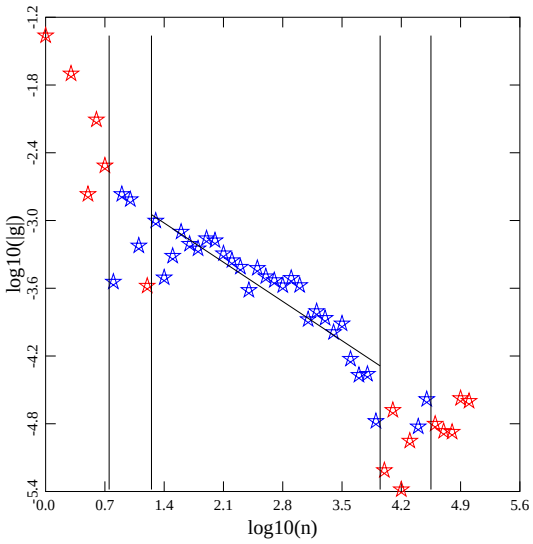

contrat1762



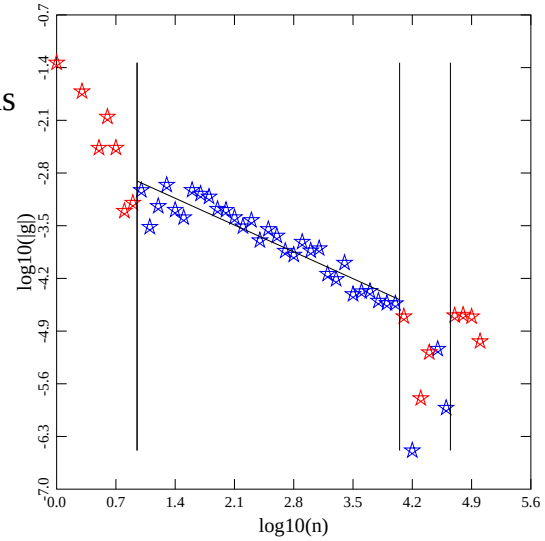

champi



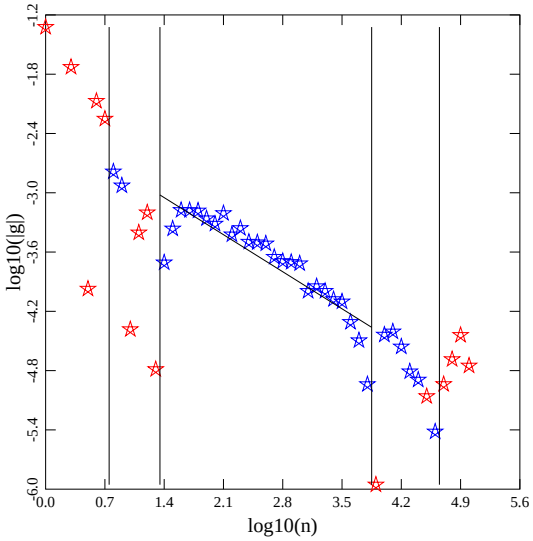

nacre



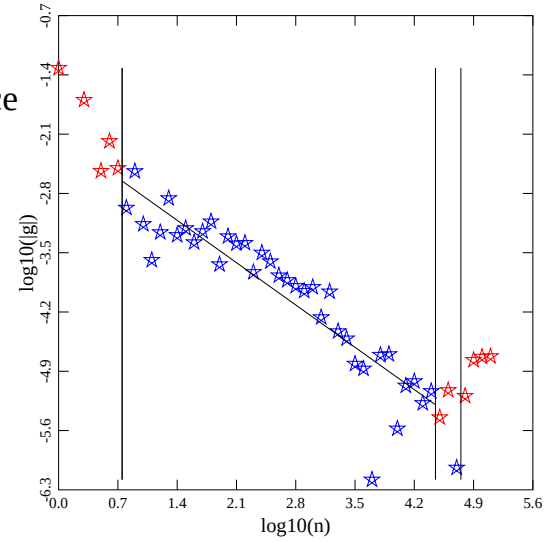
considerations



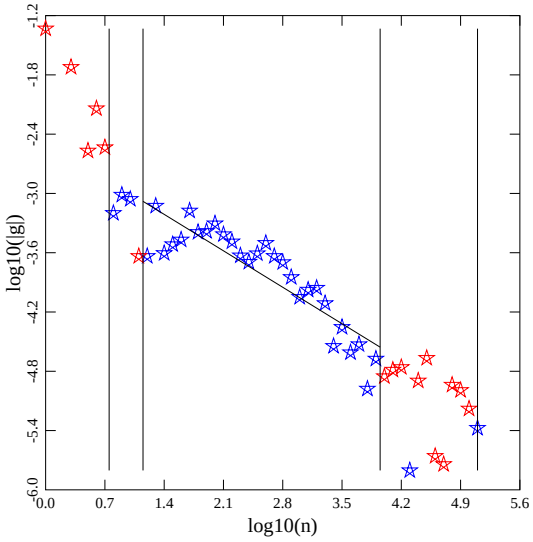

egarement



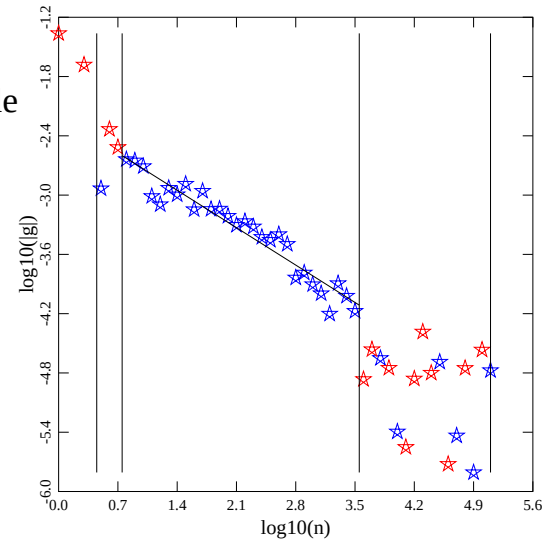

correspondance



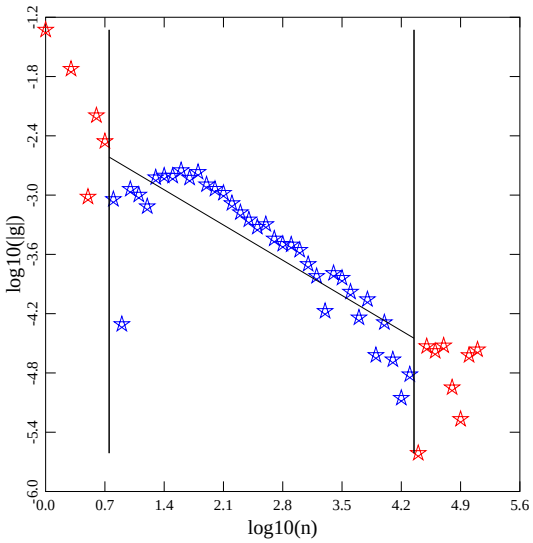

bn1837



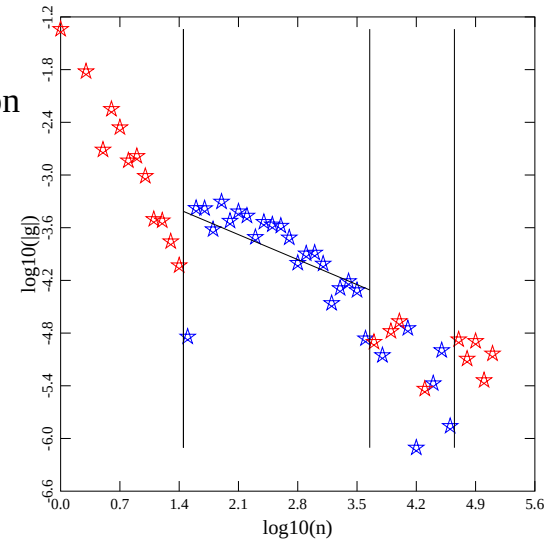

aziyade



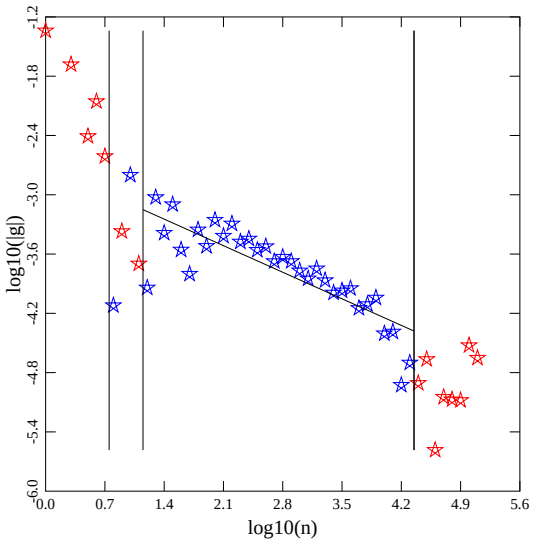

pensees



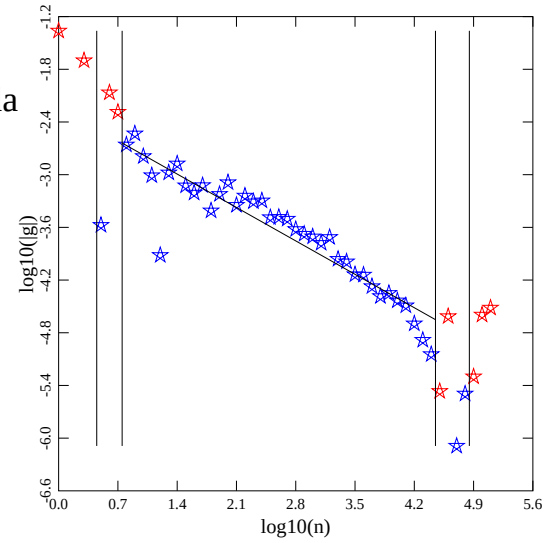
manon



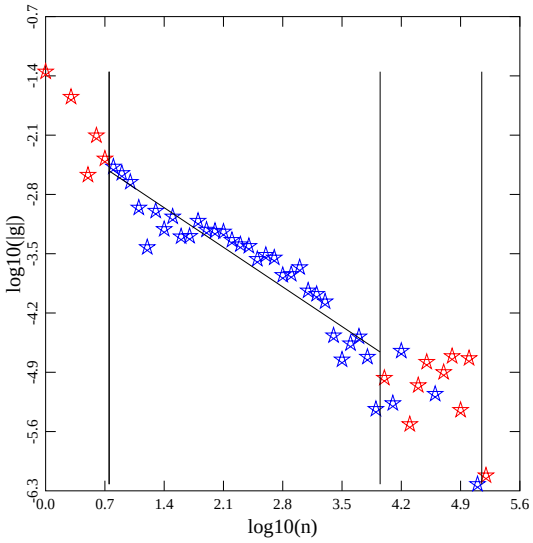

terre_lune



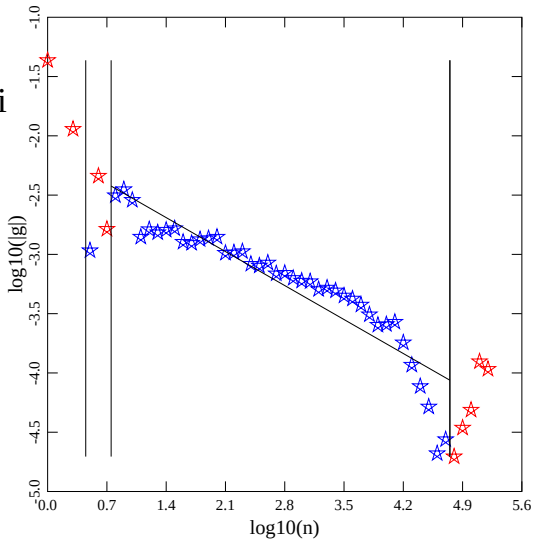

robida



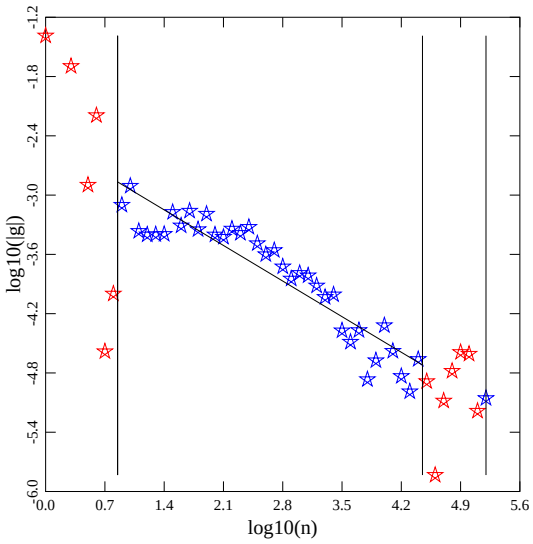

Espagne



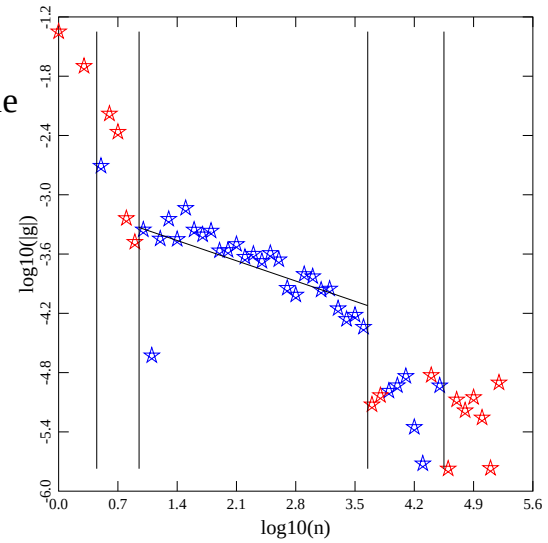

hokousai



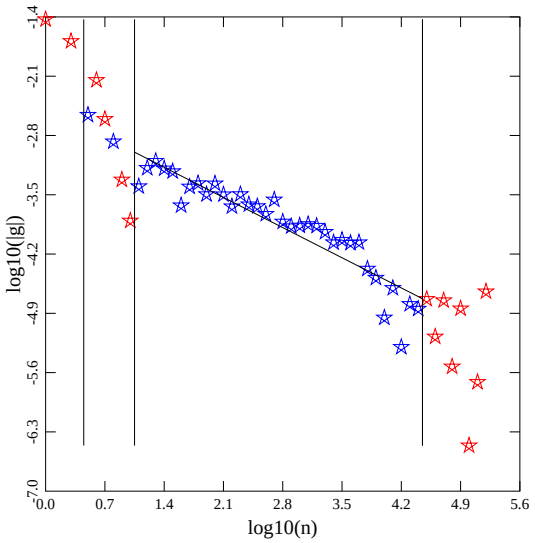

80jours



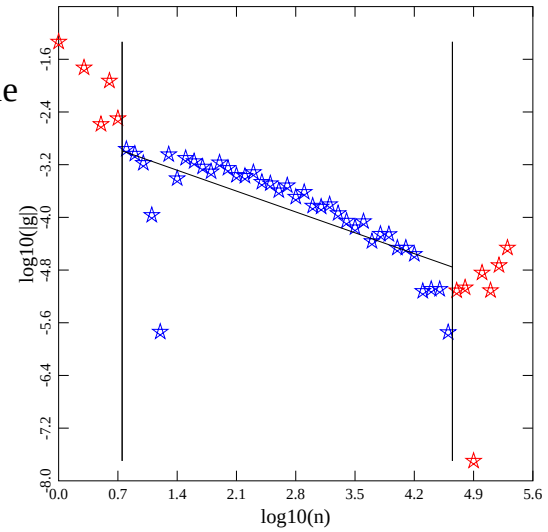
dominique



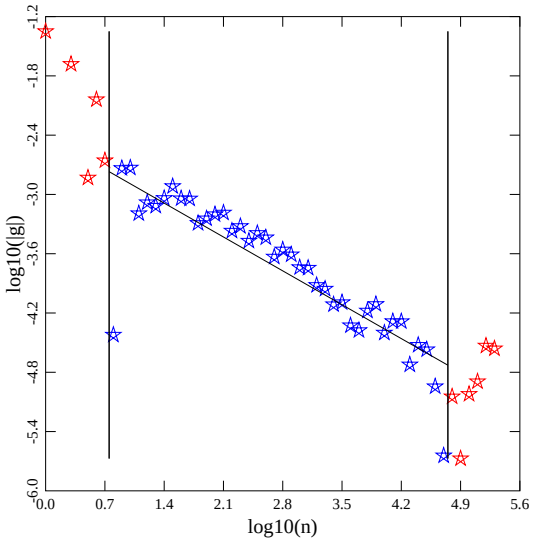

une_vie



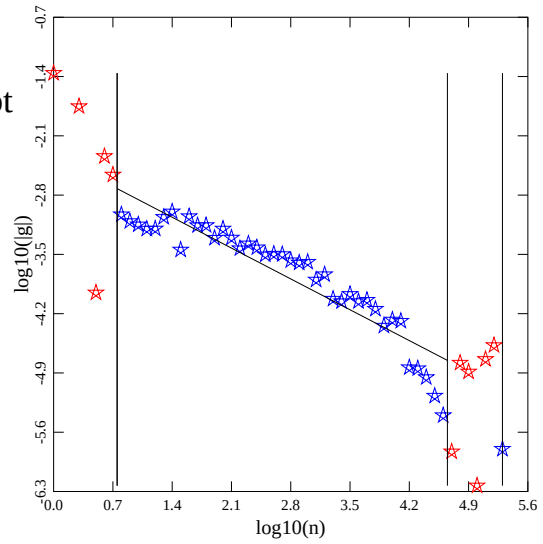

telephone



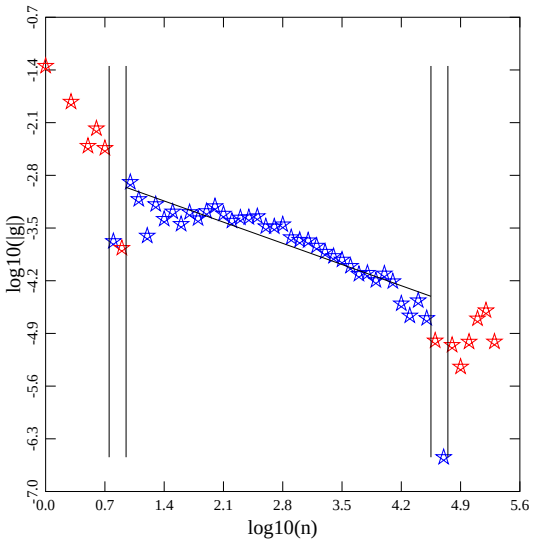

pingouins



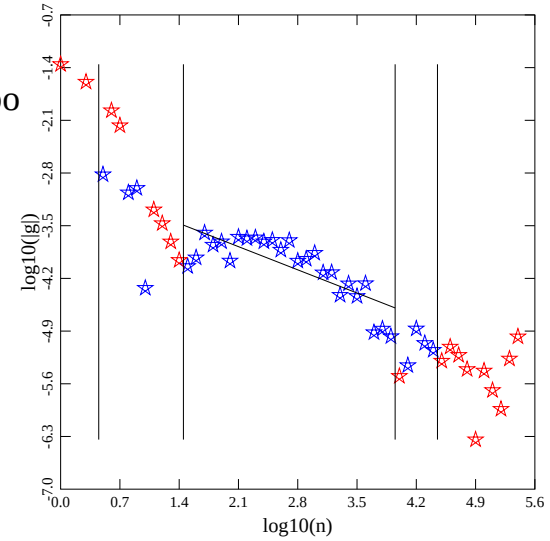

goriot



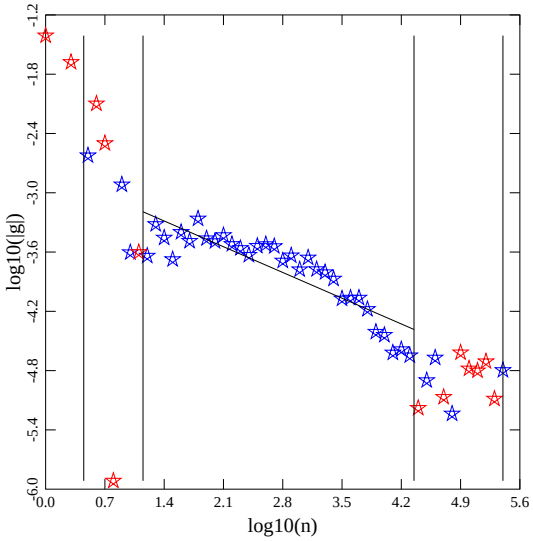

preuves



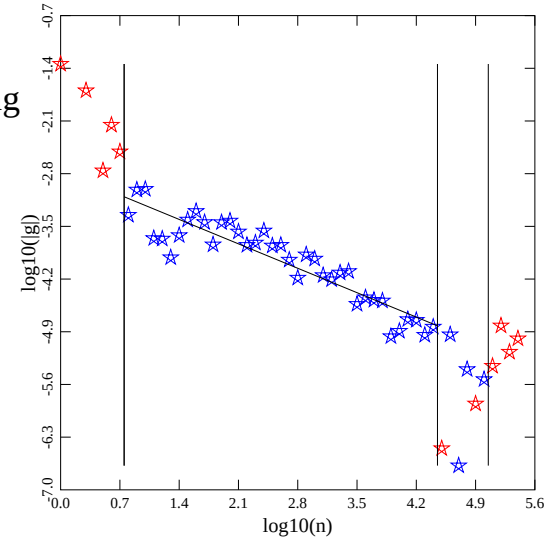
salammbo



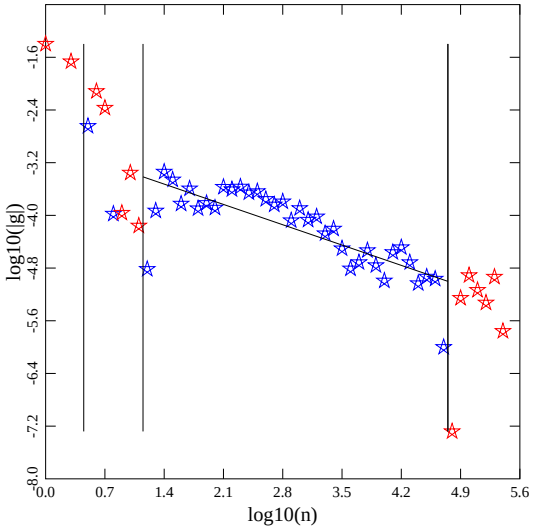

curee



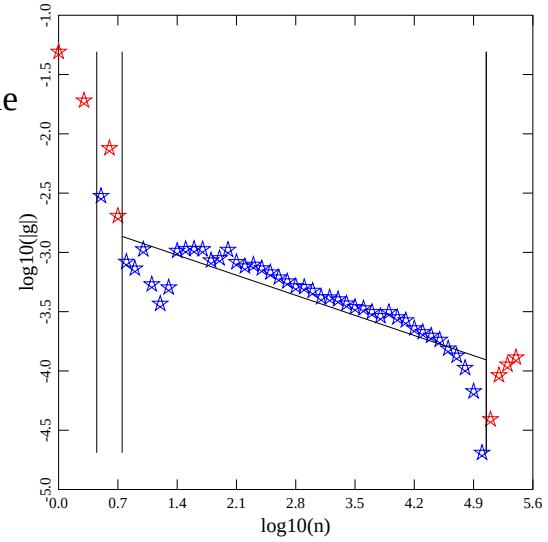

young



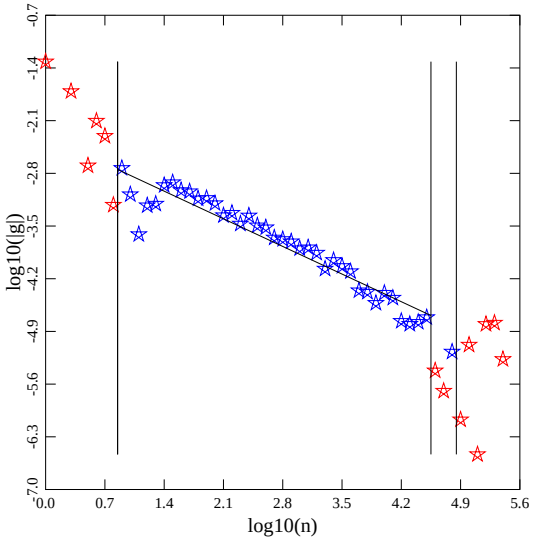

bovary



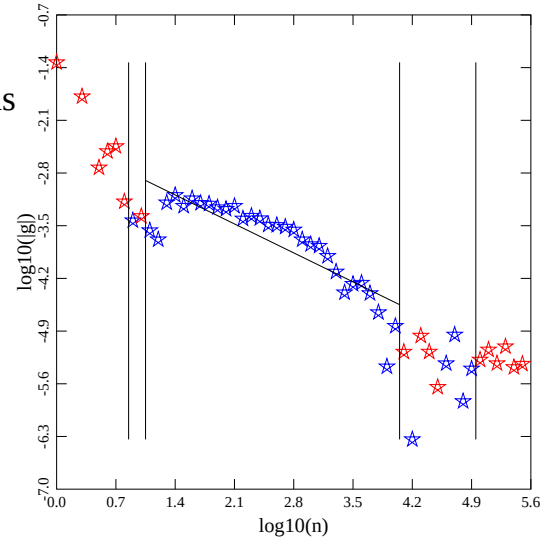

methode



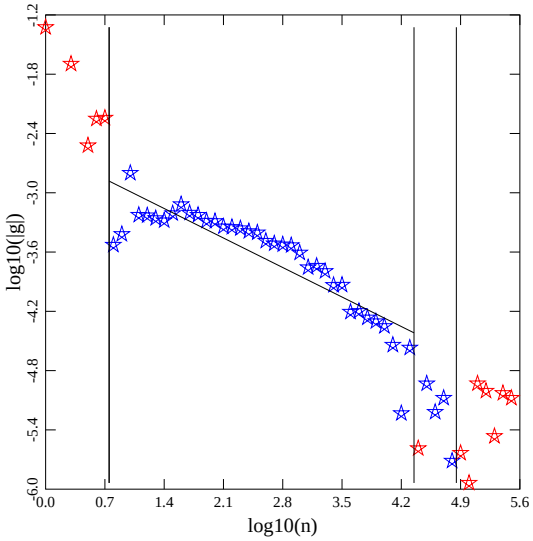

loi



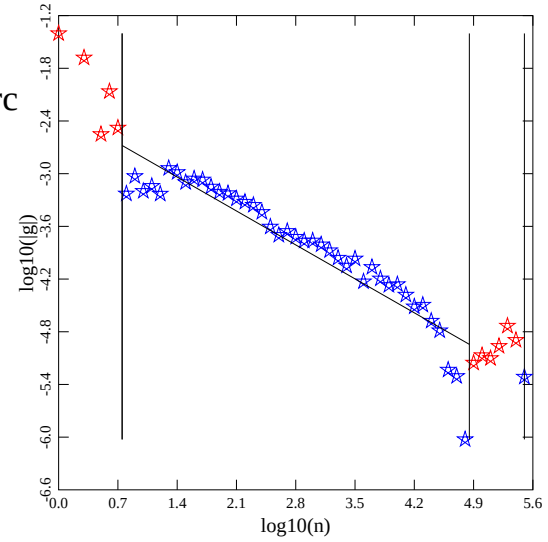
liaisons



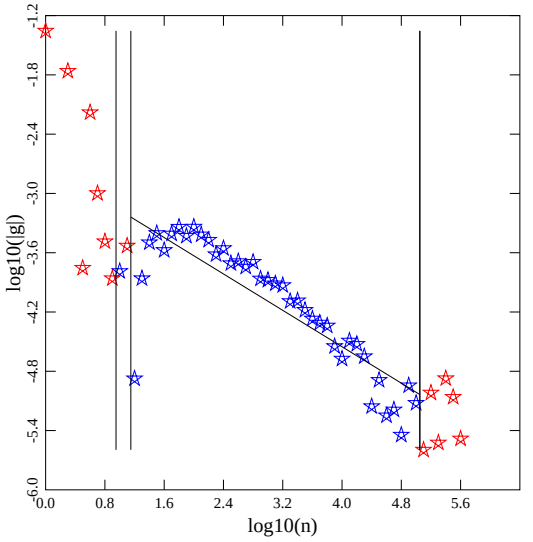

20000lieues



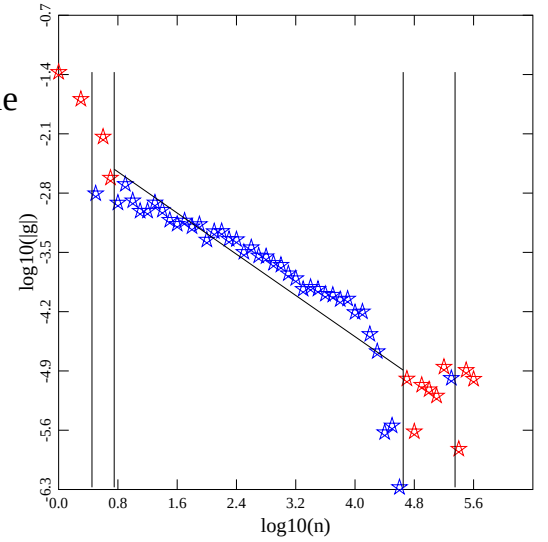

arc



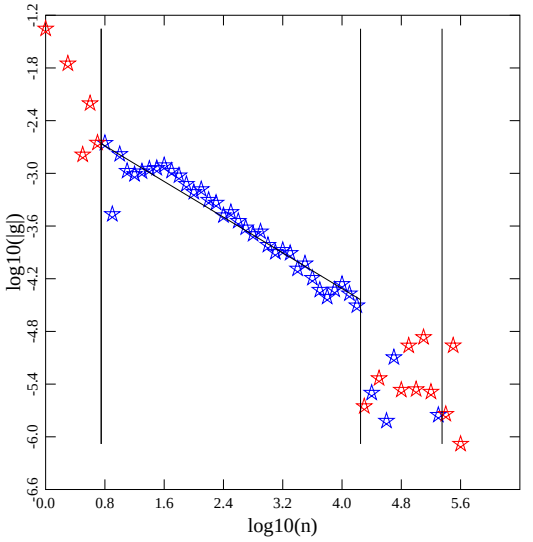

rouge_noir



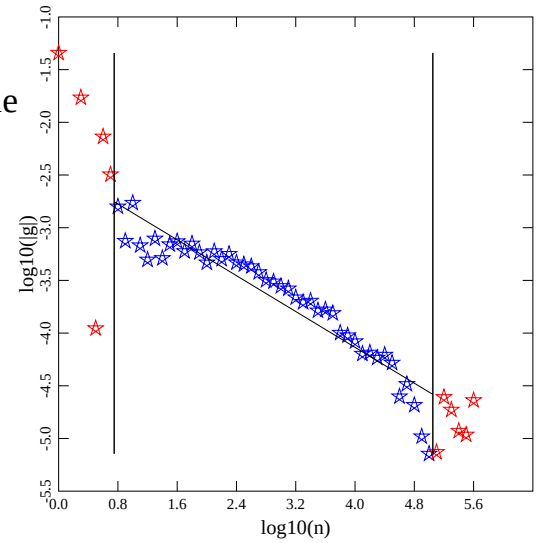

notre_dame



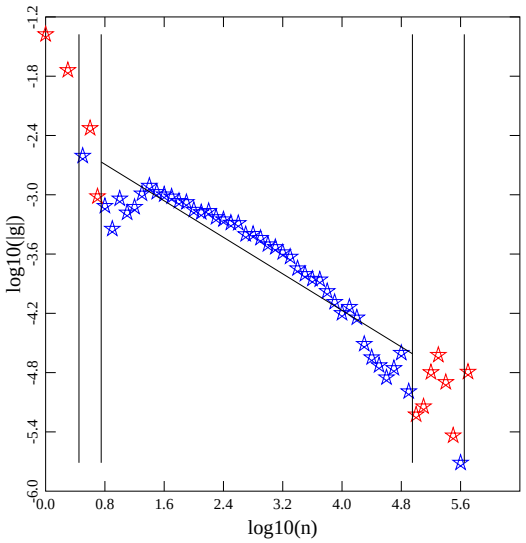

petersbourg



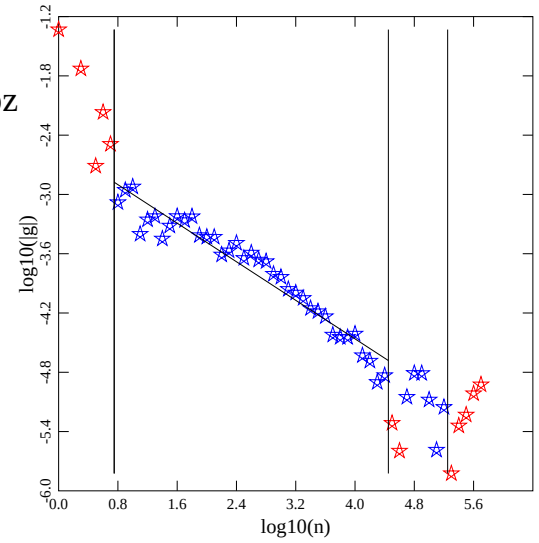
ethiopie



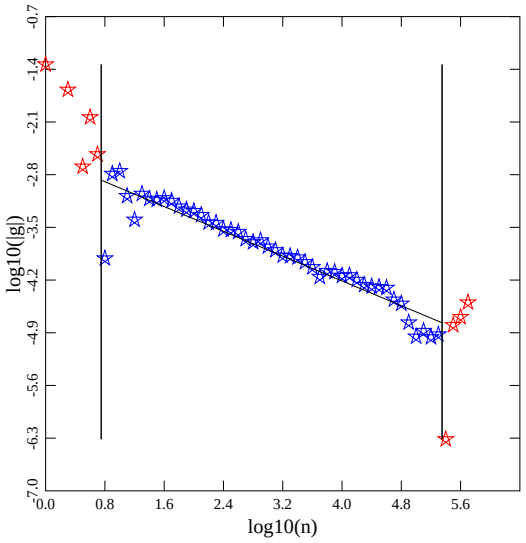

3mousquetaires



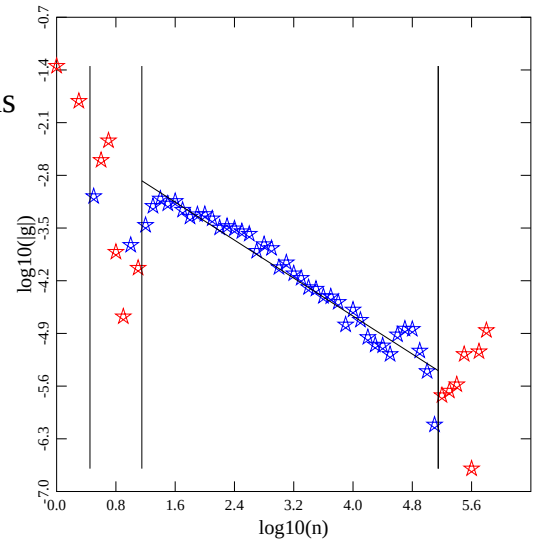

berlioz



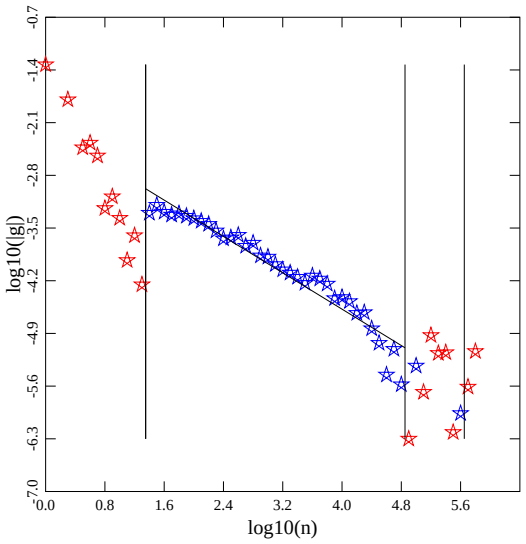

gaulois



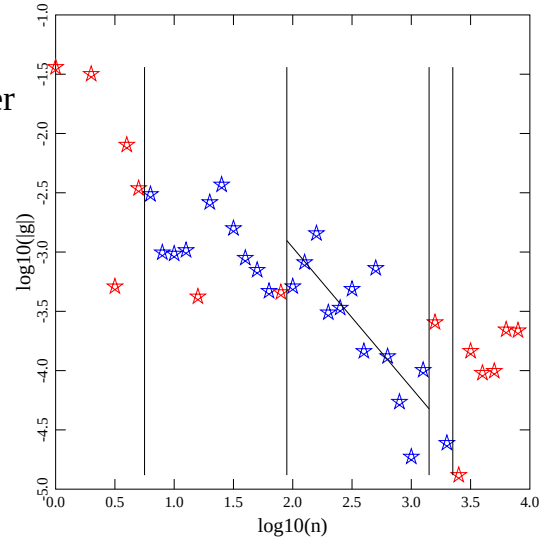

confessions



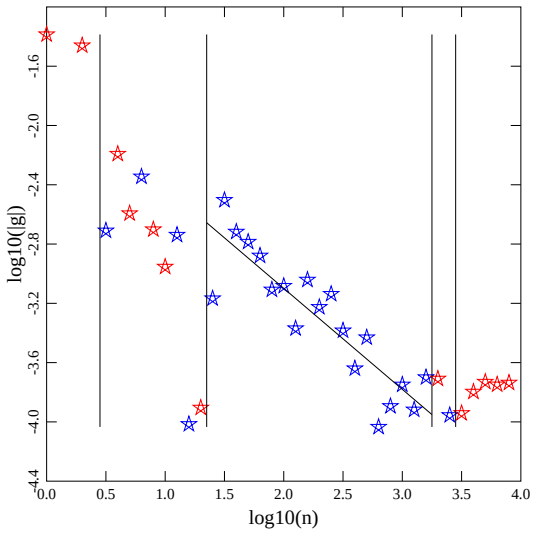

santillane



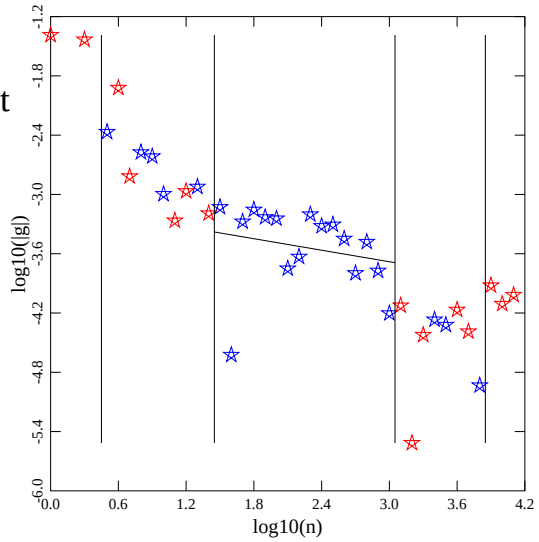
struwwelpeter



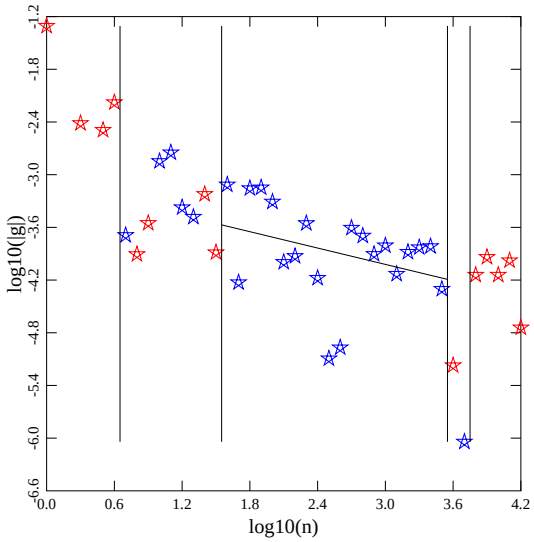

tod



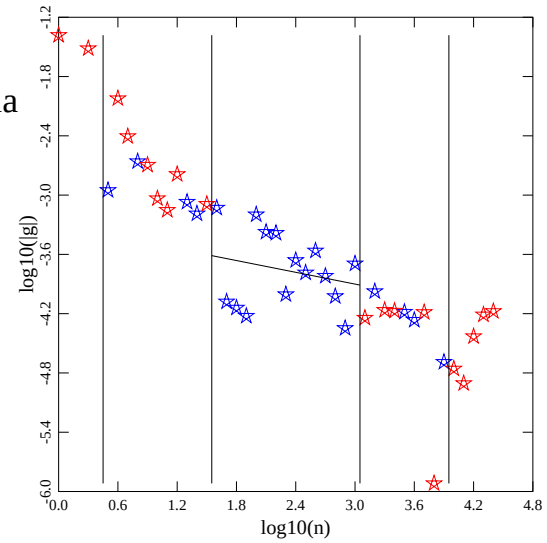

licht



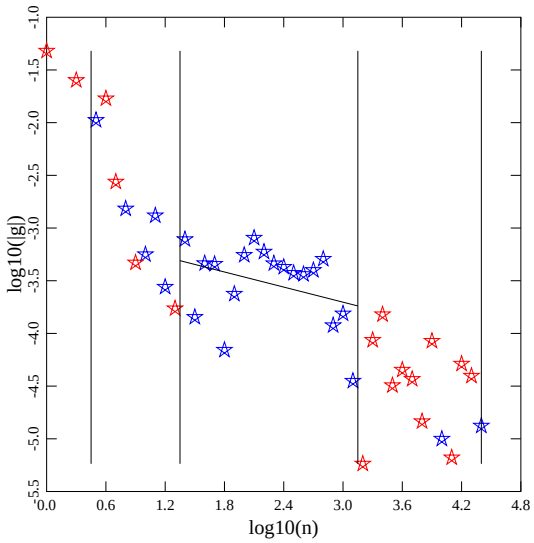

esther



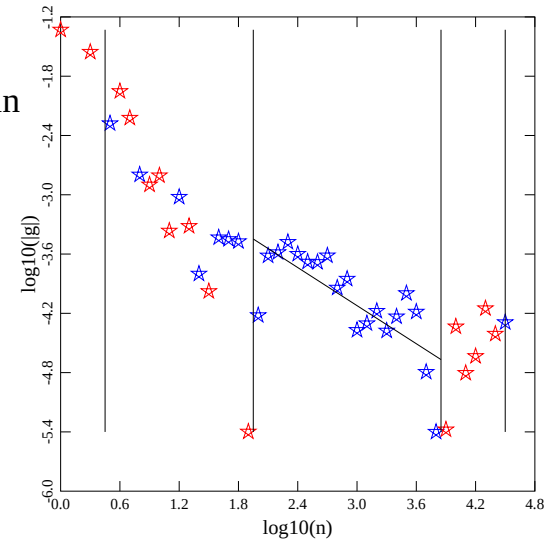

viola



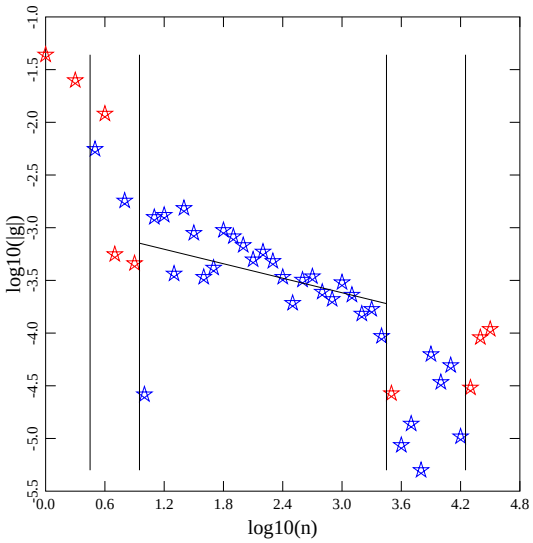

aegypter



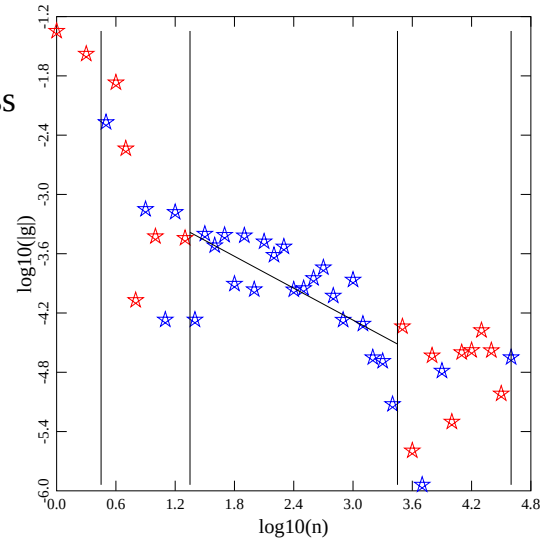
freundin



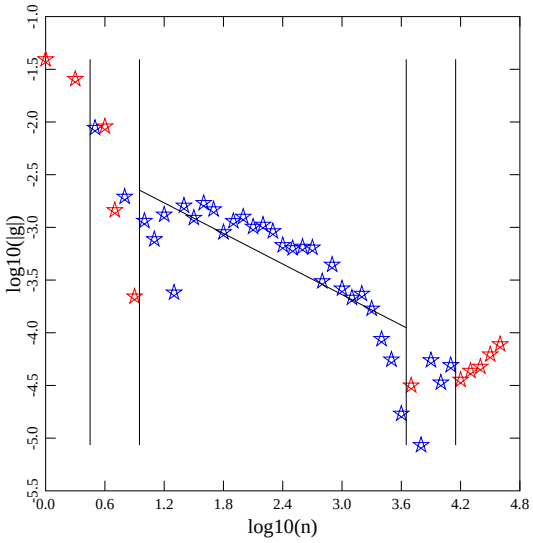

mythen



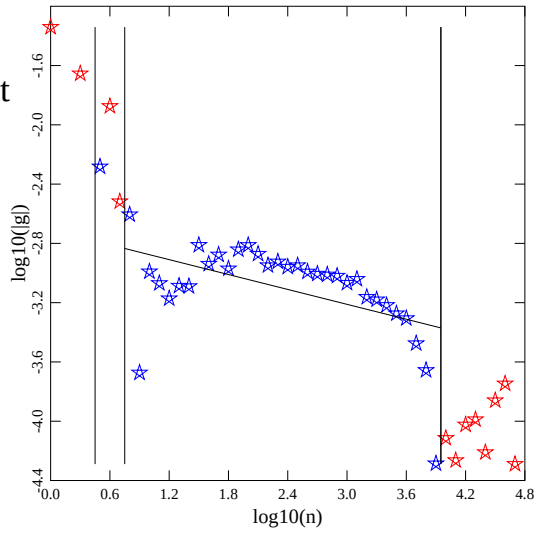

schuss



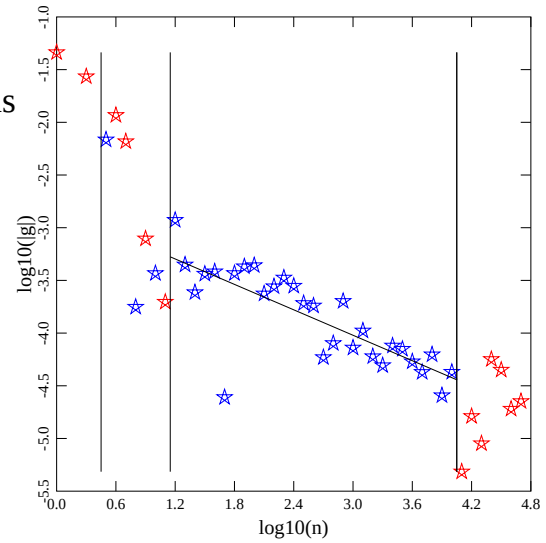

liebeskonzil



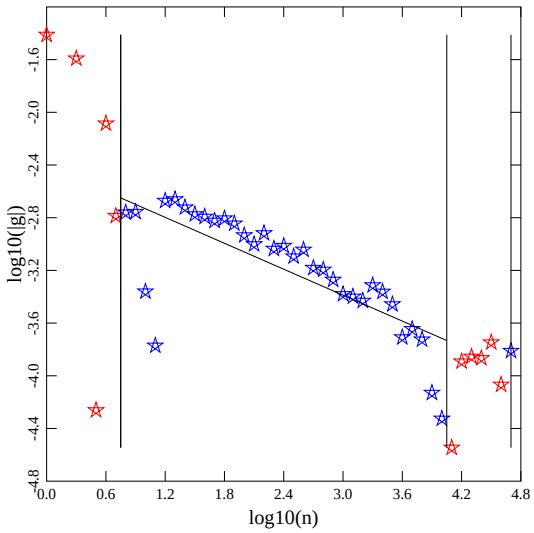

wright



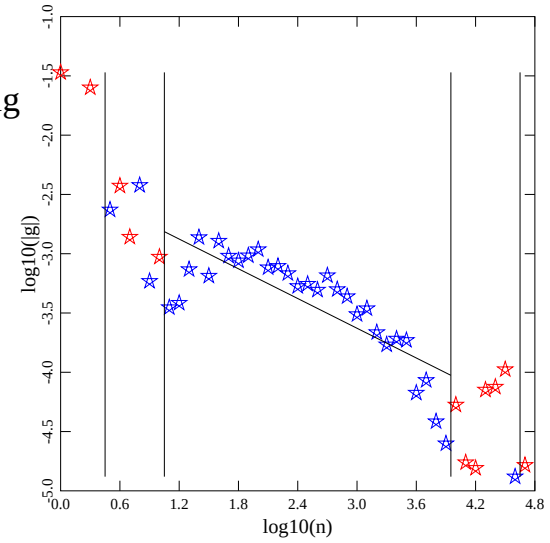
darwinismus



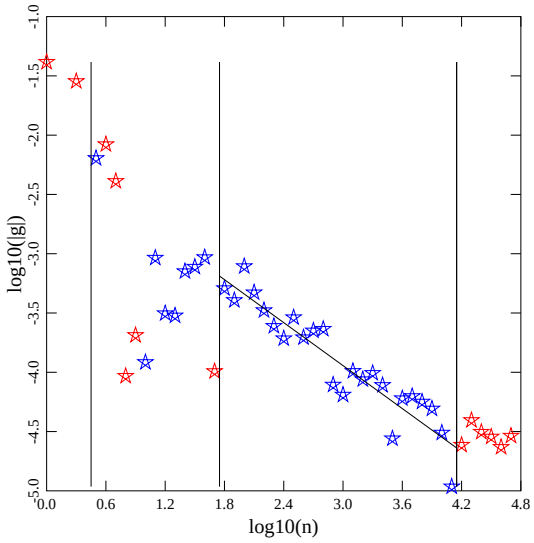

gastmahl



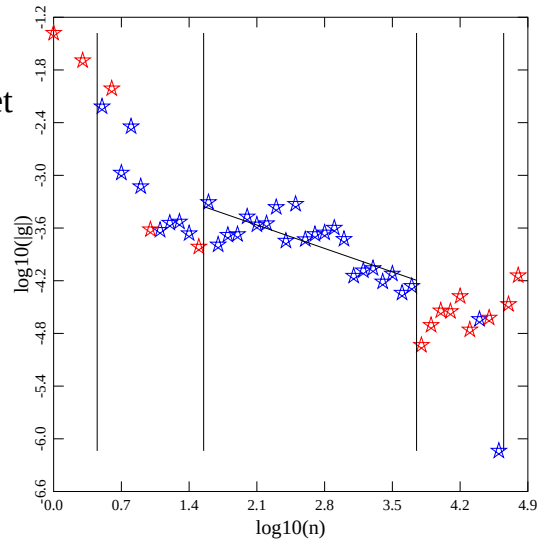

alpenkonig



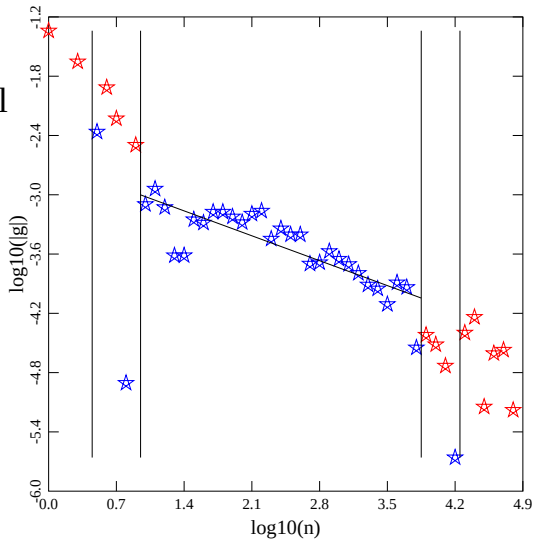

tonio



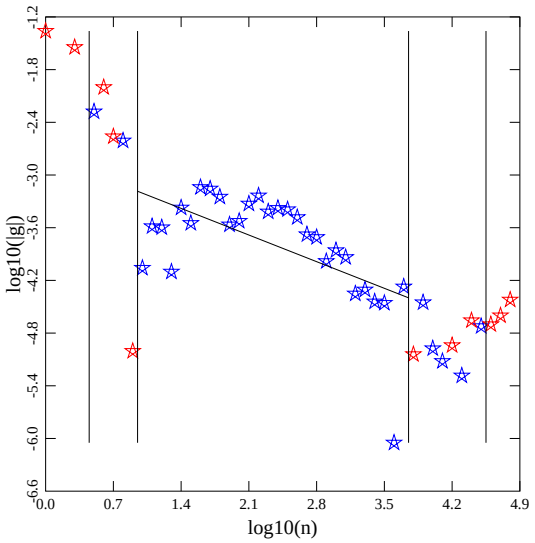

komet



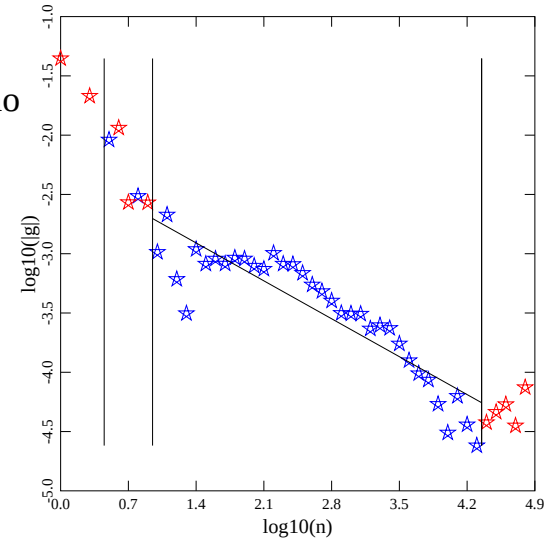
abendmahl



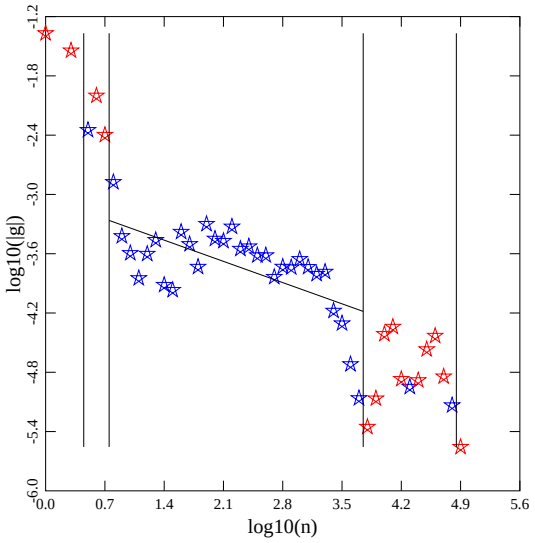

geschichten



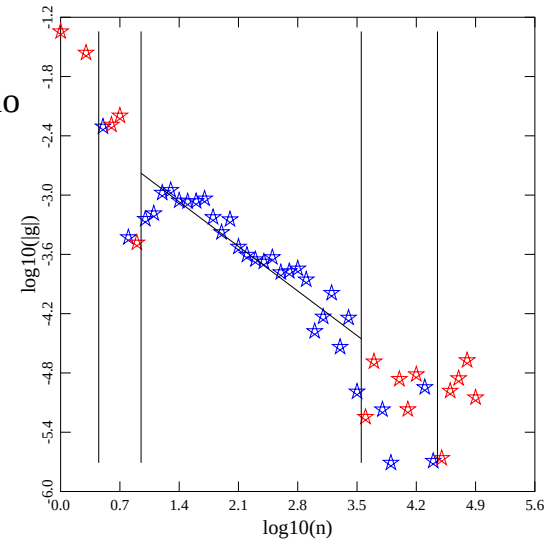

homo



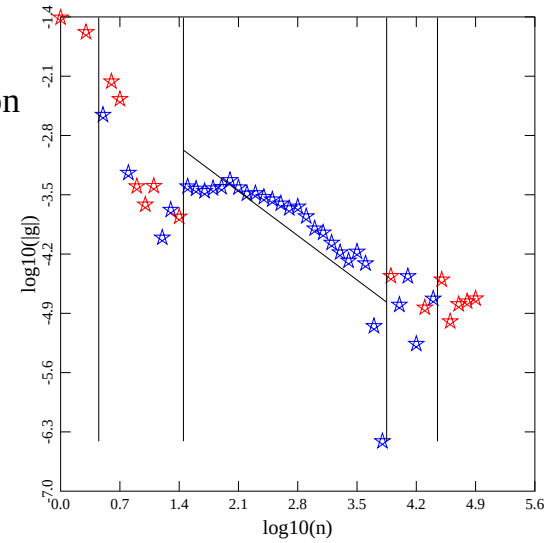

topf



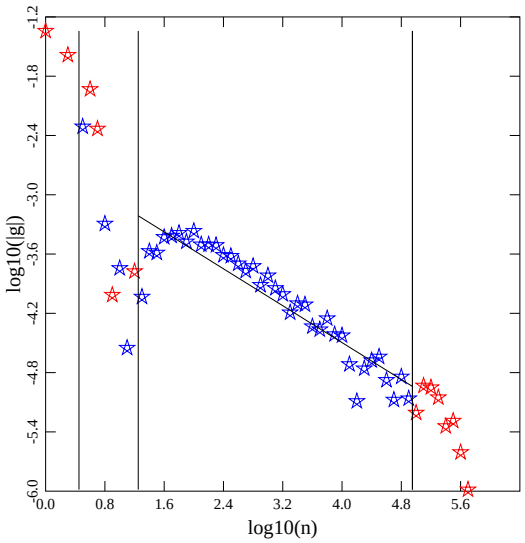

EcceHomo



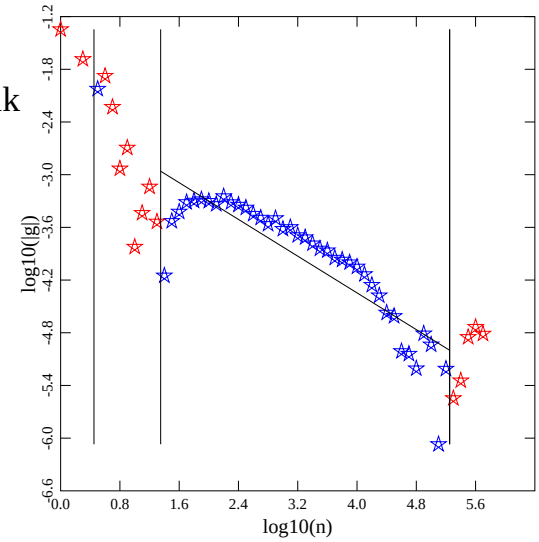
redbaron



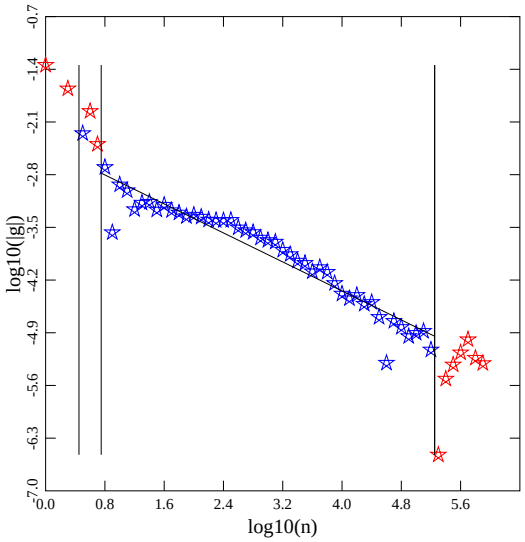

italienische



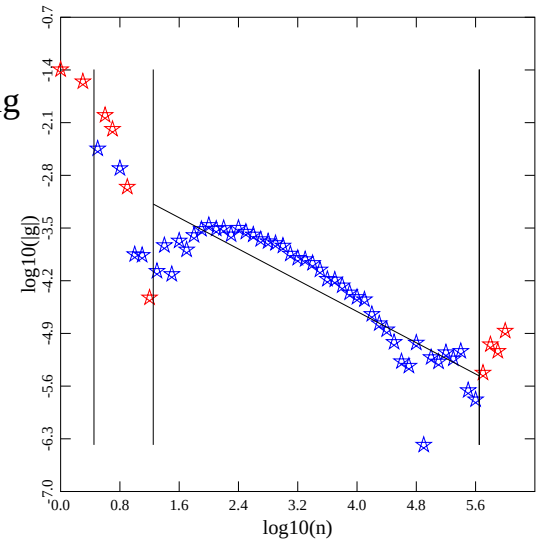

kritik



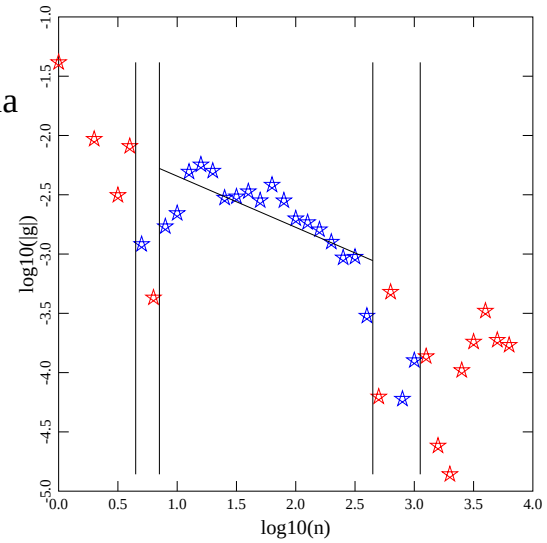

continent



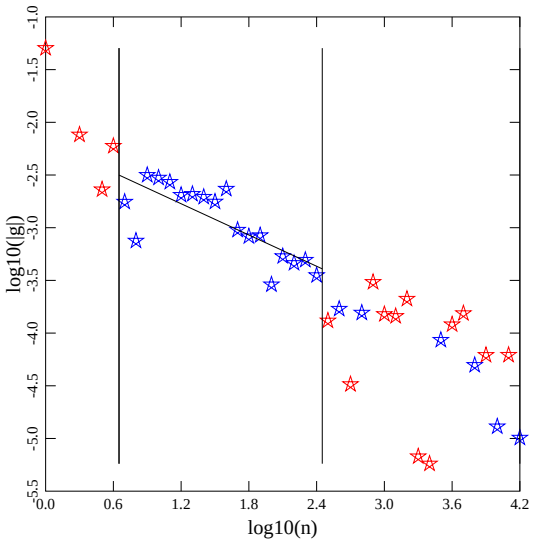

amerika_bang



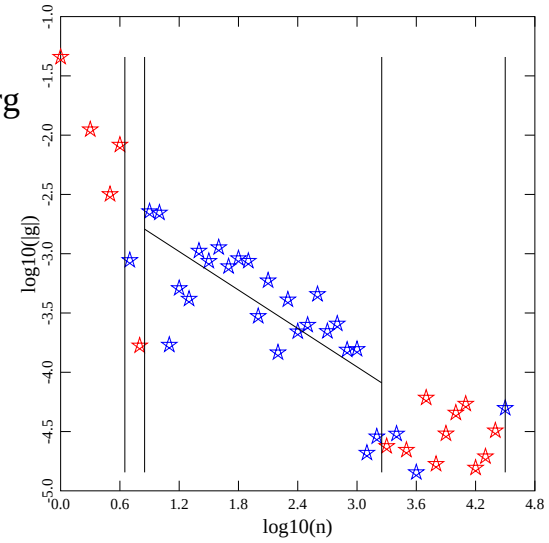
magna



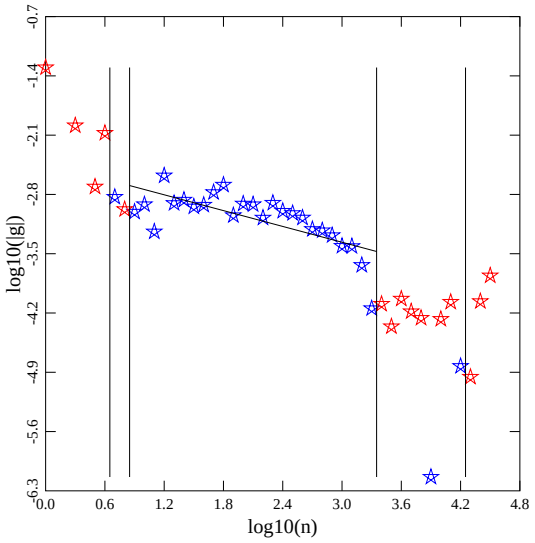

caius



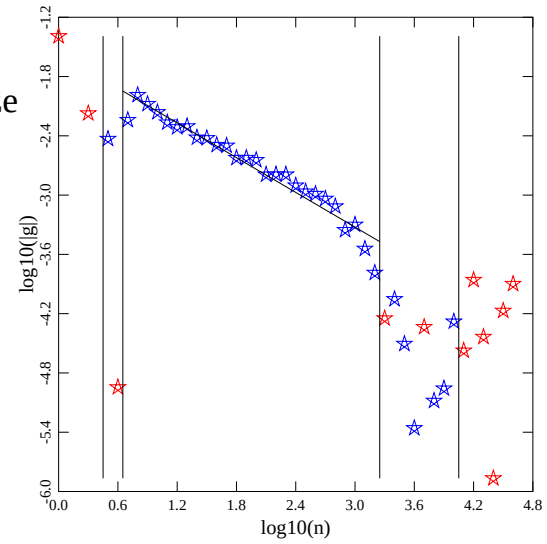

augsberg



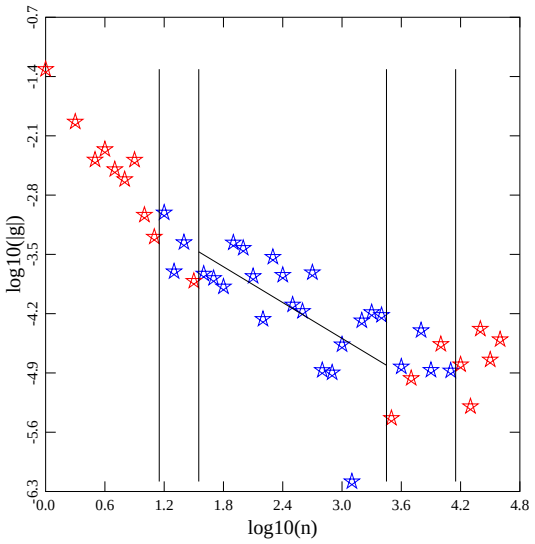

danielis



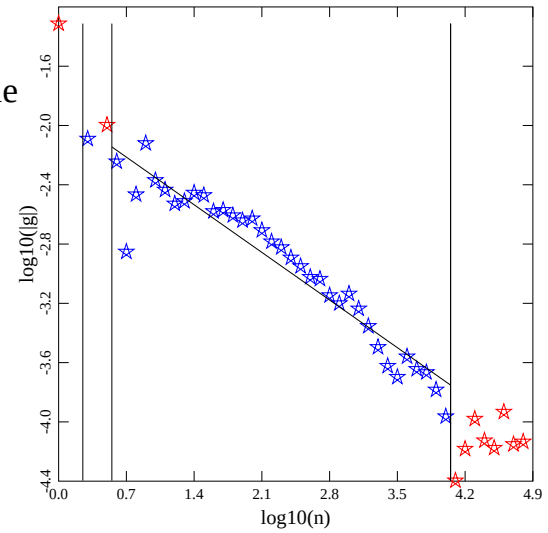

beze



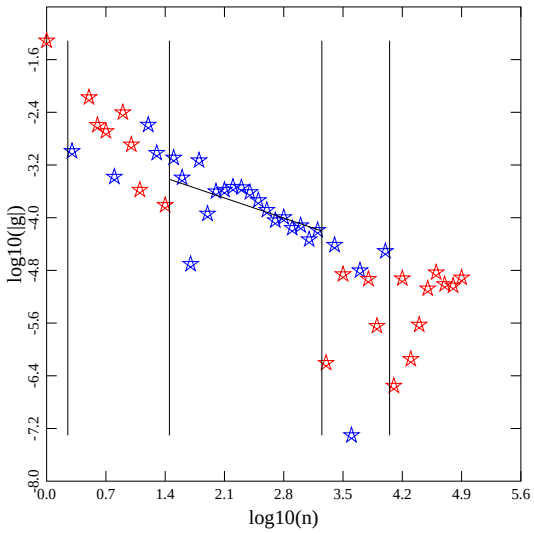

georgics



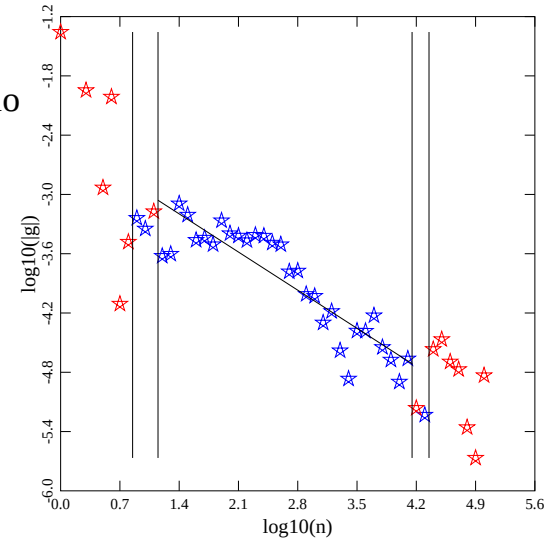
iaponicae



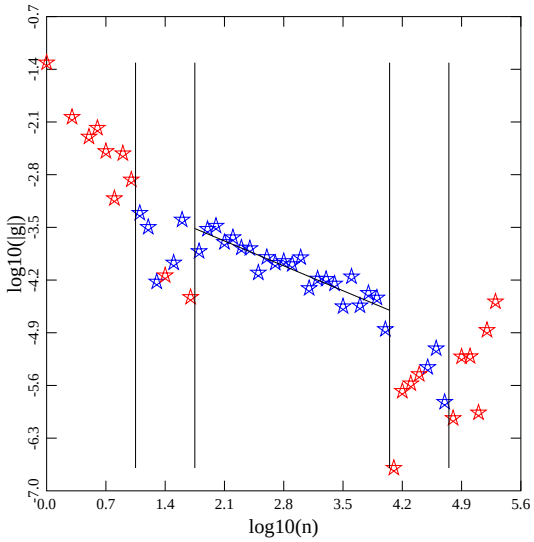

fasti



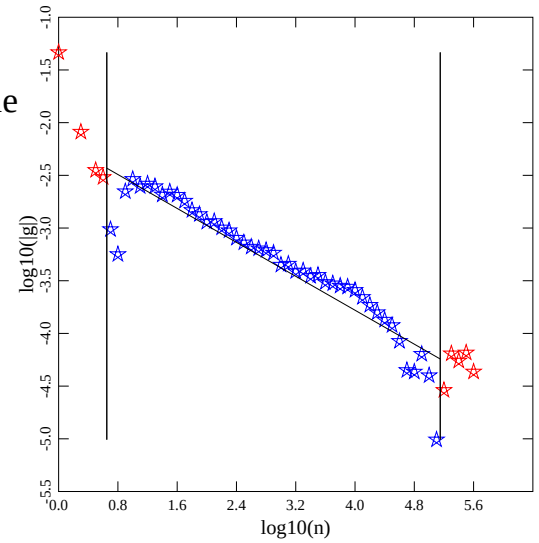

bello



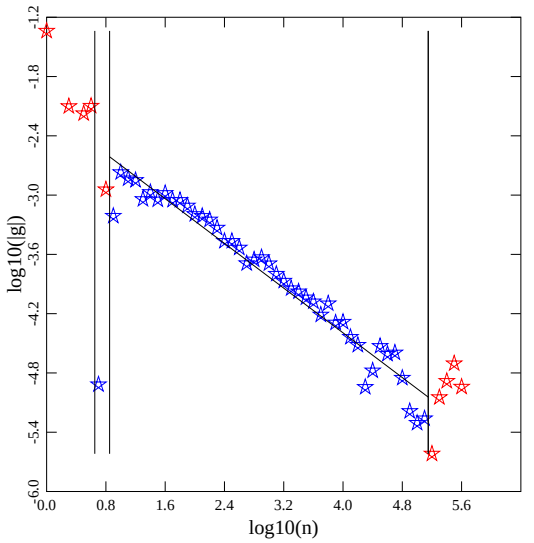

aeneidos



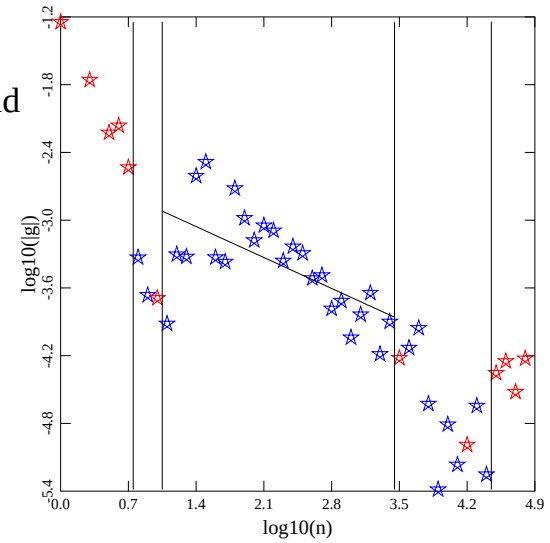

augustine



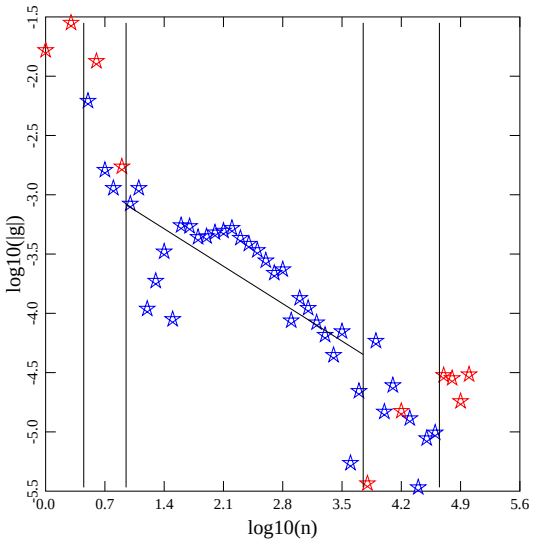

bodin



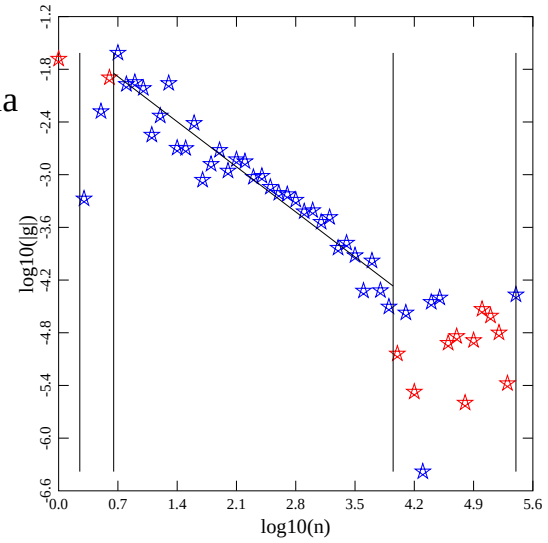
roland



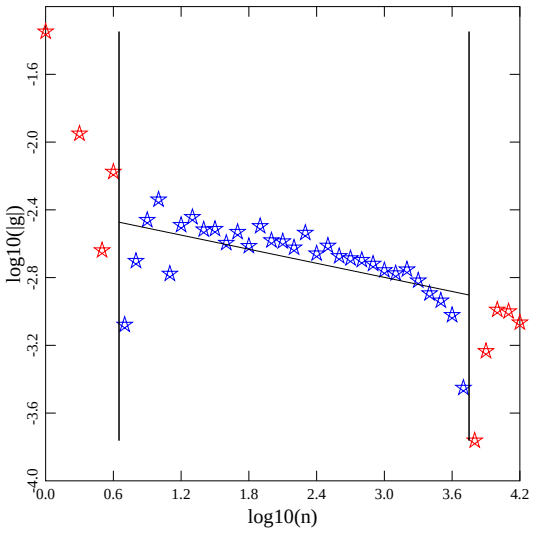

luchtschip



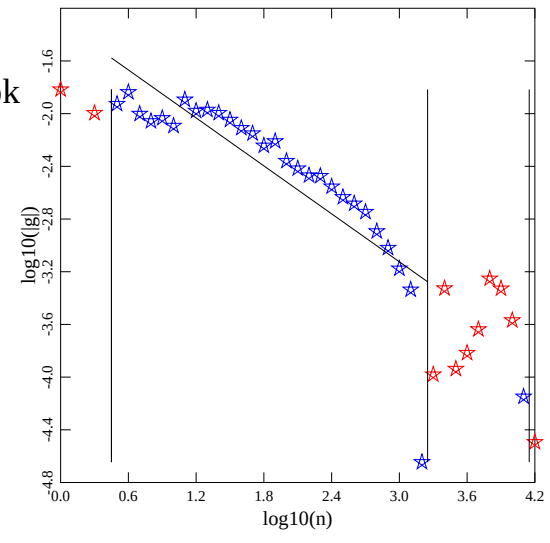

kalevala



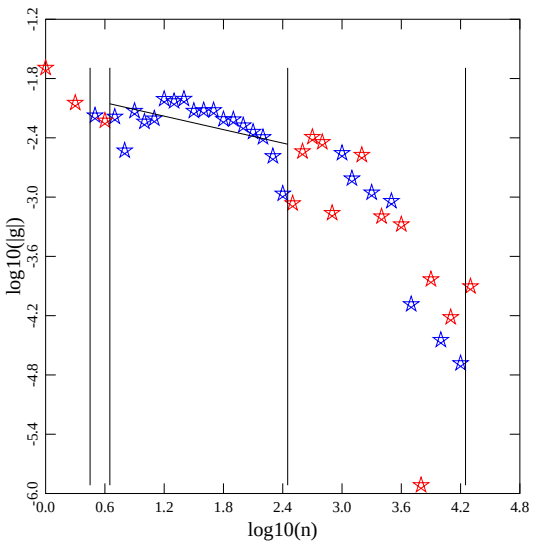

blacklist



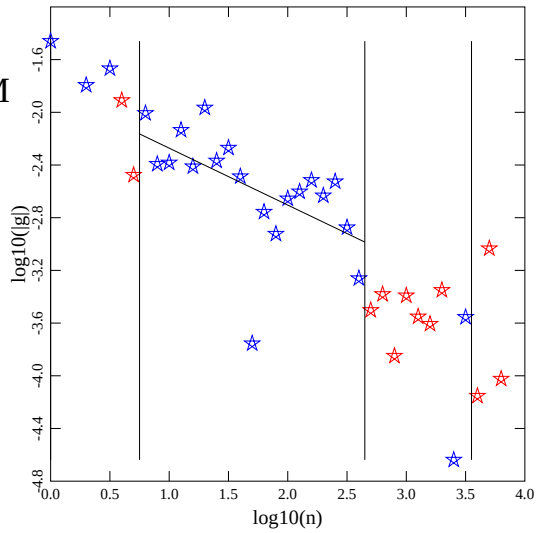

chinook



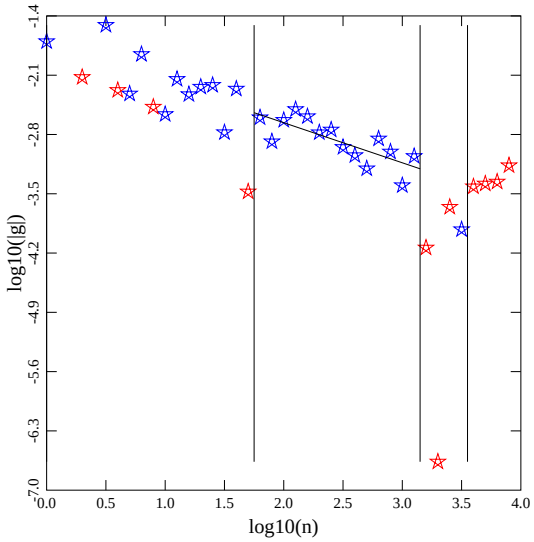

ojibway



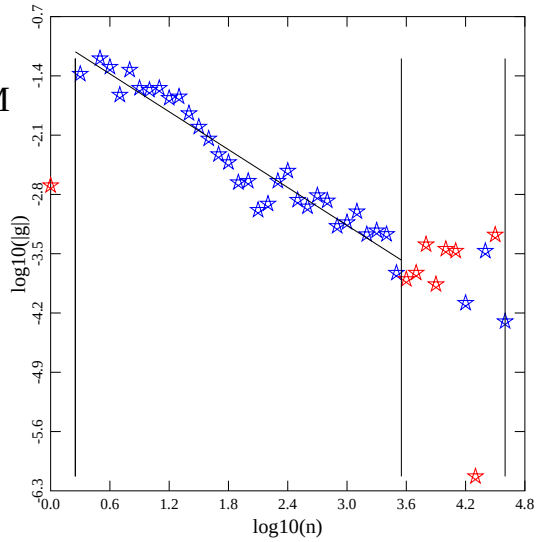
CelegansM



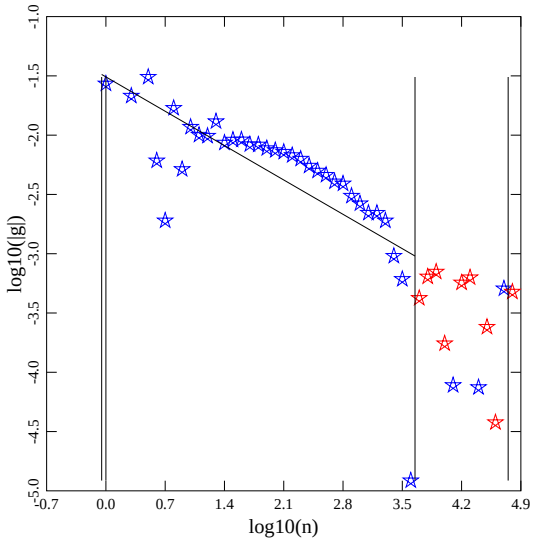

PanMT



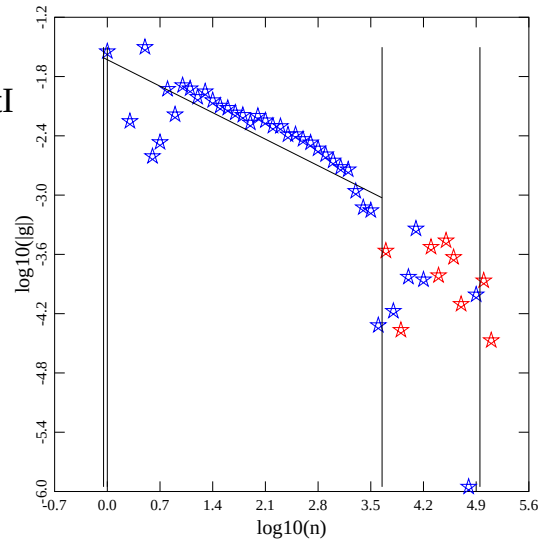

yeastM



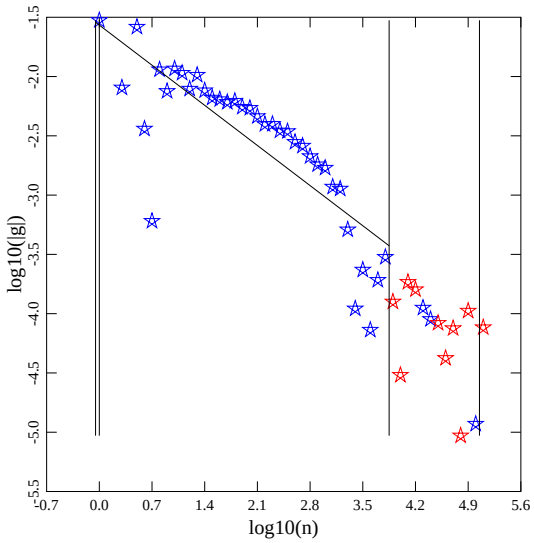

herpes2BJ



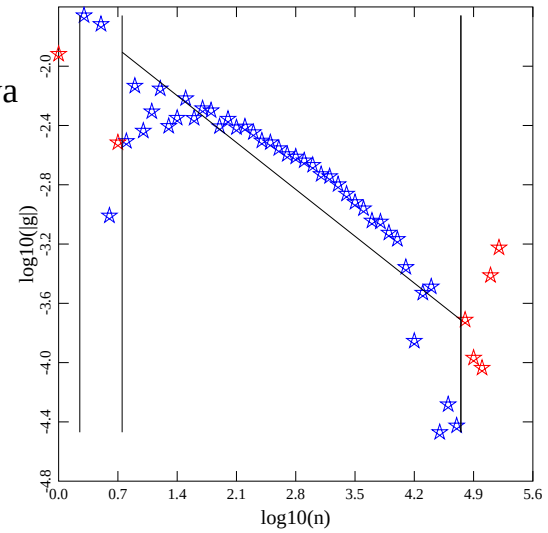

yeastI



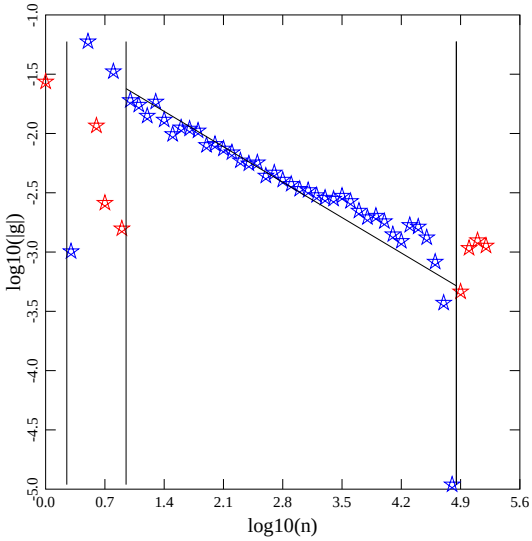

yeastVI



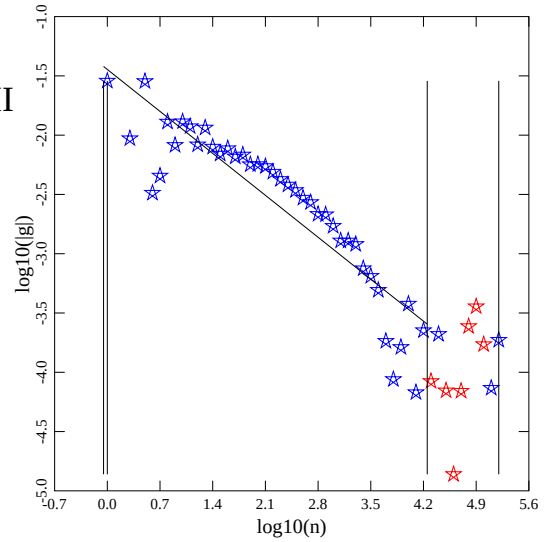
Iowa



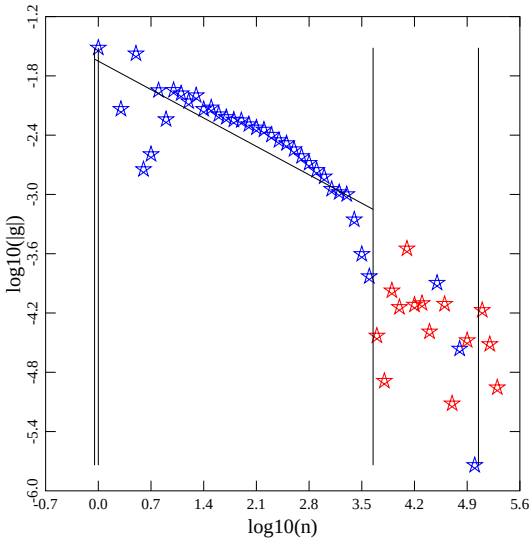

koi



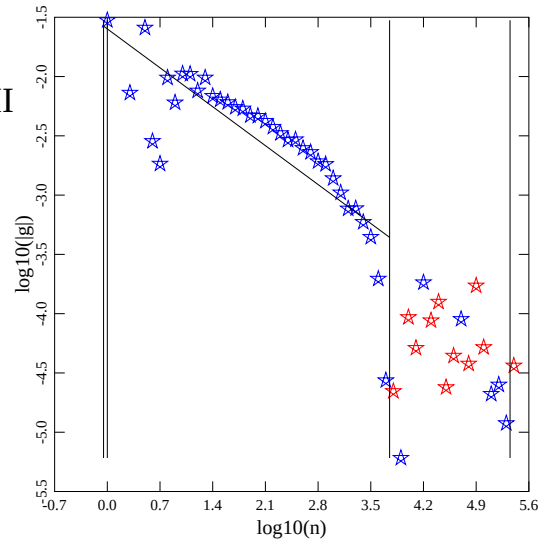

yeastIII



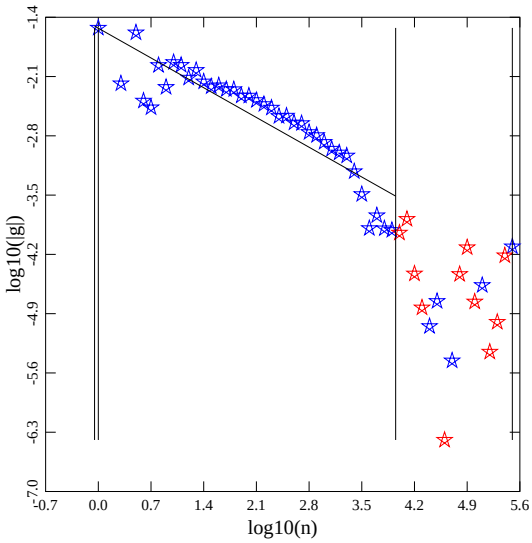

yeastIX



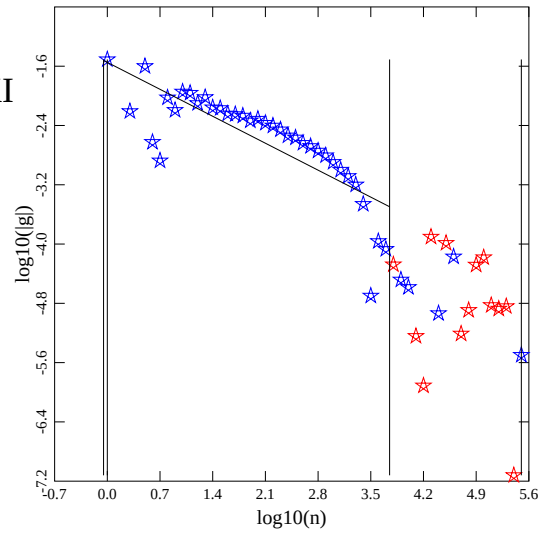

yeastVIII



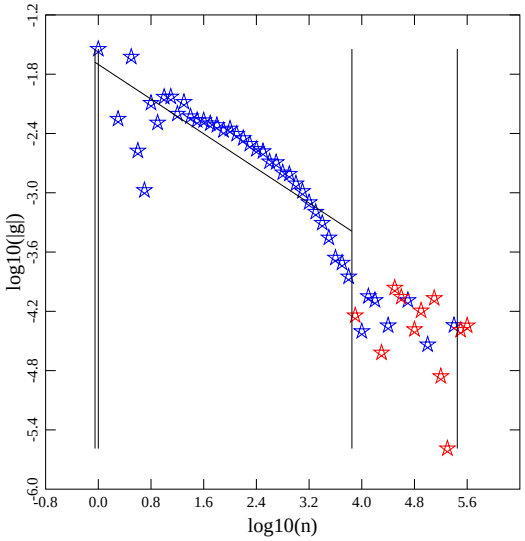

yeastV



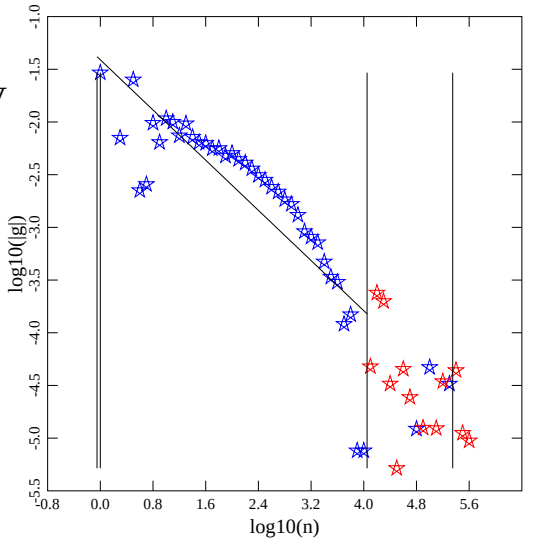
yeastXI



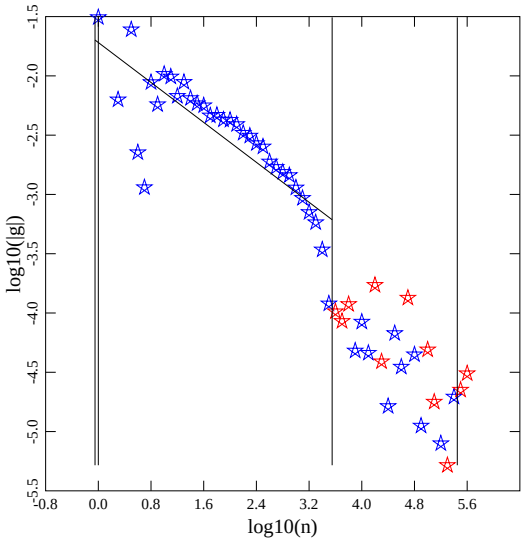

yeastX



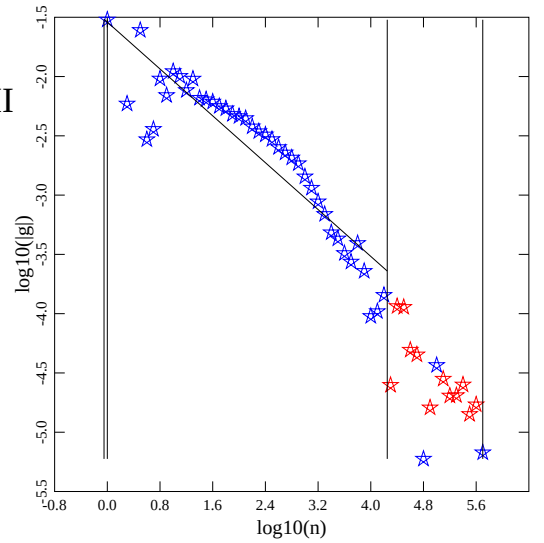

yeastXIV



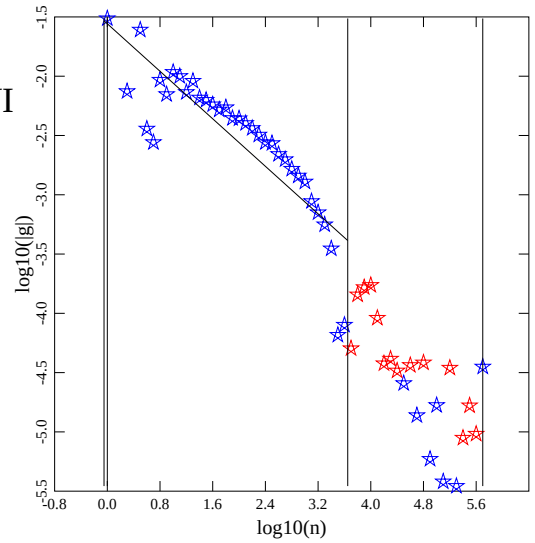

yeastII



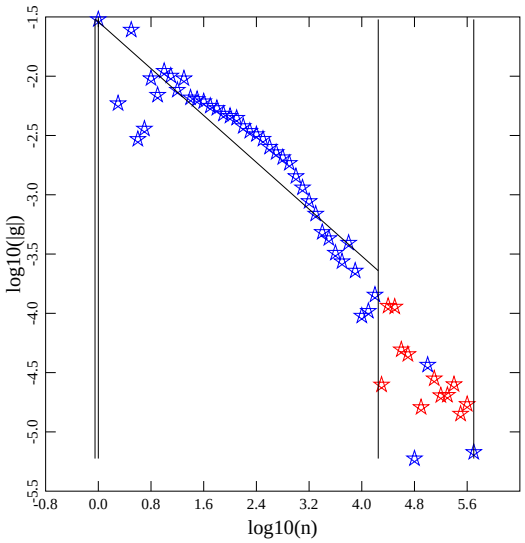

yeastXII



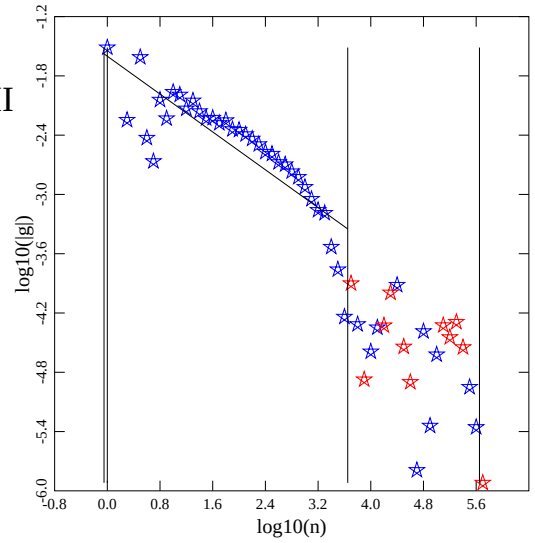
yeastXVI



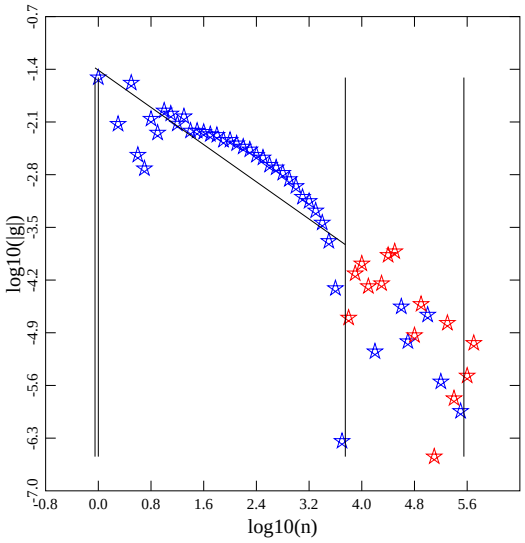

yeastXII




yeastVII




yeastXV




yeastIV