

罗纳德·布雷斯韦尔或许是第一位敢于对未来太阳黑子活动做出明确预测的作者，我们兴味盎然地等待着太阳的判决。

Brewster, D. (1855), *Memoirs of the Life, Writings, and Discoveries of Sir Isaac Newton*, 2 vols., Thomas Constable, Edinburgh.

Brigham, E. & Morrow, R. E. (1967), 'The fast Fourier transform', *Proc. IEEE Spectrum* 4, 63-70.

Brillouin, L. (1956), *Science and Information Theory*, Academic Press, New York.

Bross, I. D. J. (1963), 'Linguistic analysis of a statistical controversy', *Am. Stat.* 17, 18.

已出版的对贝叶斯方法最激烈的论争性批评之一——没有丝毫尝试检查贝叶斯方法给出的实际结果！所有想要了解推断进展为何以及以何种方式被拖延这么久的人都应该阅读本文。最初写于1963年的杰恩斯的文章（Jaynes, 1976）就是对布罗斯的回复，具体场景在杰恩斯的著作（Jaynes, 1983, 第149页）中有解释。

Brown, E. E. & Duren, B. (1986), 'Information integration for decision support', *Decision Support Syst.*, 4, (2), 321-329.

Brown, R. (1828), 'A brief account of microscopical observations', *Edinburgh New Phil. J.* 5, 358-371.

布朗运动的第一份报告。

Burg, J. P. (1967), 'Maximum entropy spectral analysis', Proceedings of the 37th Meeting of the Society of Exploration Geophysicists.

Burg, J. P. (1975), 'Maximum entropy spectral analysis', Ph.D. Thesis, Stanford University.

Busnel, R. G. & Fish, J. F., eds. (1980), *Animal Sonar Systems*, NATO ASI Series, Vol. A28, Plenum Publishing Corp., New York.

1979年在英国泽西岛举行的一次会议上的非常长的（1082页）报告。

Cajori, F. (1928), in *Sir Isaac Newton 1727-1927*, Waverley Press, Baltimore, pp. 127-188.

Cajori, F. (1934), *Sir Isaac Newton's Mathematical Principles of Natural Philosophy and his System of the World*, University of California Press, Berkeley.

Carnap, R. (1950), *Logical Foundations of Probability*, Routledge and Kegan Paul Ltd., London.

Chen, Wen-chen, & de Groot, M. H. (1986), 'Optimal search for new types', in Goel, P. & Zellner, A., eds. (1986), *Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti*, Elsevier Science Publishers, Amsterdam, pp. 443-458.

Childers, D., ed. (1978), *Modern Spectrum Analysis*, IEEE Press, New York.

最大熵频谱分析早期工作的重印集。

Chow, Y., Robbins, H. & Siegmund, D. (1971), *Great Expectations: Theory of Optimal Stopping*, Houghton Mifflin & Co., Boston.

Cobb, L. & Watson, B. (1980), 'Statistical catastrophe theory: an overview', *Math. Modelling*, 1, 311-317.

我们对这篇文章没有异议,但想补充两个历史性说明。(1) 他们的“随机微分方程”是大约 1917 年以来物理学家所称的“福克-普朗克方程”。然而,我们习惯于将我们的统计工作归功于数学家柯尔莫哥洛夫。(2) 今天,与勒内·汤姆名字相关的多值“折叠”函数的稳定性考虑等价于单值熵函数的凸性,这些结果是由吉布斯在 1873 年给出的。

Cohen, T. J. & Lintz, P. R. (1974), 'Long term periodicities in the sunspot cycle', *Nature*, **250**, 398.

Cooley, J. W. & Tukey, J. W. (1965), 'An algorithm for the machine calculation of complex Fourier series', *Math. Comp.*, **19**, 297-301.

Cooley, J. W., Lewis, P. A. & Welch, P. D. (1967), 'Historical notes on the fast Fourier transform', *Proc. IEEE* **55**, 1675-1677.

Cook, A. (1994), *The Observational Foundations of Physics*, Cambridge University Press.

注意,物理量是根据用于测量它们的实验装置来定义的。当然,这只是尼尔斯·玻尔在 1927 年强调的老生常谈。

Cox, D. R. & Hinkley, D. V. (1974), *Theoretical Statistics*, Chapman & Hall, London; reprints 1979, 1982.

主要是重复旧的抽样理论方法,使用一种奇怪的记号法,可能会使最简单的方程变得不可读。然而,它有许多有用的历史总结和旁注,指出了统计理论的局限性或扩展,这是在其他地方找不到的。贝叶斯方法仅在倒数第 10 章中介绍,然后作者继续对贝叶斯方法重复所有陈旧、错误的反对意见,表明他们不理解这些早就被杰弗里斯 (Jeffreys, 1939)、萨维奇 (Savage, 1954) 和林德利 (Lindley, 1956) 纠正了的古老误解。一位著名的统计学家注意到这一点,认为考克斯和欣克利“使得统计学倒退了 25 年”。

Cox, D. R. (1970), *The Analysis of Binary Data*, Methuen, London.

Cox, R. T. (1978), 'Of inference and inquiry', in Levine, R. D. & Tribus, M., eds., *The Maximum Entropy Formalism*, MIT Press, Cambridge, MA, pp. 119-167.

注意,对应于命题的逻辑,存在对于问题的双重逻辑。正如杰恩斯在他的著作 (Jaynes, 1983, 第 382~388 页) 中讨论的那样,随着进一步的发展,这可能变得非常重要。

Cozzolino, J. M. & Zahner, M. J. (1973), 'The maximum-entropy distribution of the future market price of a stock', *Operations Res.*, **21**, 1200-1211.

Creasy, M. A. (1954), 'Limits for the ratio of means', *J. Roy. Stat. Soc. B* **16**, 175-185.

Csiszar, I. (1984), 'Sanov property, generalized I-projection and a conditional limit theorem', *Ann. Prob.*, **12**, 768-793.

Czuber, E. (1908), *Wahrscheinlichkeitsrechnung und Ihre Anwendung auf Fehlerausgleichung*, 2 vols., Teubner, Berlin.

这里可以找到沃尔夫的一些著名的扔骰子数据。大约在 1850~1890 年期间,苏黎世天文学家沃尔夫进行并报告了大量“随机”实验,此处给出了对这些情况的说明。

Daganzo, C. (1977), *Multinomial Probit: The Theory and its Application to Demand Forecasting*, Academic Press, New York.

- Dale, A. I. (1982), 'Bayes or Laplace? An examination of the origin and early applications of Bayes' theorem', *Arch. Hist. Exact Sci.* **27**, 23-47.
- Daniel, C. & Wood, F. S. (1971), *Fitting Equations to Data*, John Wiley, New York.
- Daniell, G. J. & Pottton, J. A. (1989), 'Liquid structure factor determination by neutron scattering - some dangers of maximum entropy', in Skilling, J., ed. (1989), *Maximum Entropy and Bayesian Methods*, Proceedings of the Eighth Maximum Entropy Workshop, Cambridge, UK, August 1988, Kluwer Academic Publishers, Dordrecht, pp. 151-162.
- 这里的“危险”是指，新手第一次尝试在复杂问题上使用最大熵方法时结果可能并不令人满意，因为他们在回答与用户的想法不同的问题。所以最初的努力实际上是一个“训练”，让人知道如何正确地定义问题。
- Davenport, W. S. & Root, W. L. (1958), *Random Signals and Noise*, McGraw-Hill, New York.
- David, F. N. (1962), *Games, Gods and Gambling*, Griffin, London.
- 概率论最早起源的历史。注意，在考古学中，“越往早走，证据越零碎”。这里有我们在别处找不到的那种深刻洞察力。
- de Finetti, B. (1958), 'Foundations of probability', in *Philosophy in the Mid-century*, La Nuova Italia Editrice, Florence, pp. 140-147.
- de Groot, M. H., Bayarri, M. J. & Kadane, J. B. (1988), 'What is the likelihood function?' (with discussion), in Gupta, S. S. & Berger, J. O., eds., *Statistical Decision Theory and Related Topics IV*, Springer-Verlag, New York.
- de Groot, M. H. & Cyert, R. M. (1987), *Bayesian Analysis and Uncertainty in Economic Theory*, Chapman & Hall, London.
- de Groot, M. H., Fienberg, S. E. & Kadane, J. B. (1986), *Statistics and the Law*, John Wiley, New York.
- Deming, W. E. (1943), *Statistical Adjustment of Data*, John Wiley, New York.
- Dempster, A. P. (1963), 'On a paradox concerning inference about a covariance matrix', *Ann. Math. Stat.* **34**, 1414-1418.
- Dubois, D. & Prade, H. (1988), *Possibility Theory*, Plenum Publ. Co., New York.
- Dunnington, G. W. (1955), *Carl Friedrich Gauss, Titan of Science*, Hafner, New York.
- Dutta, M. (1966), 'On maximum entropy estimation', *Sankhya*, ser. A, **28**, (4), 319-328.
- Dyson, F. J. (1979), *Disturbing the Universe*, Harper & Row, New York.

中译本 弗里曼·J. 戴森. 宇宙波澜. 王一操, 左立华, 译. 重庆: 重庆大学出版社, 2018.

时间跨度为大约 50 年的个人回忆和猜想的集合，其中的 90% 与我们当前的目的无关，但是这里必须坚持将此文献列出，因为弗里曼·戴森在 20 世纪中叶理论物理学的发展中扮演了非常重要的角色，他对此的回忆具有独特的价值。但不幸的是，它们分散在多章的片段中。与一些不太深思熟虑的同事不同，戴森正确地看到了许多关于概率论和量子理论的一些基本的东西（但在我们看来，他忽视了其他一些同样基本的东西）。阅读这本书就像阅读开普勒，需要提取其中重要的一小部分真理。

Eddington, Sir A. (1935), *The Nature of the Physical World*, Dent, London.

中译本 阿瑟·爱丁顿. 物理世界的本质. 王文浩, 译. 北京: 商务印书馆, 2020.

另一位与我们一样思考概率的杰出科学家.

Edwards, A. W. F. (1972), *Likelihood*, Cambridge University Press.

安东尼·爱德华兹是费希尔的最后一个学生, 尽管他与其他人一样了解与贝叶斯方法的所有技术事实, 但一些心理障碍阻止他像费希尔那样接受它们的明显后果. 因此, 遗憾的是, 我们必须分道扬镳, 在没有他的情况下继续进行推理的建设性发展.

Edwards, A. W. F. (1992), *Nature* **352**, 386–387.

贝叶斯方法的评论.

Edwards, H. M. (1987), 'An appreciation of Kronecker', *Math. Intelligencer*, **9**, 28–35.

Edwards, H. M. (1988), 'Kronecker's place in history', in Aspray, W. & Kitcher, P., eds., *History and Philosophy of Modern Mathematics*, University of Minnesota Press.

Efron, B. (1975), 'Biased versus unbiased estimation', *Adv. Math.* **16**, 259–277.

Efron, B. (1978), 'Controversies in the foundations of statistics', *Am. Math. Monthly* **85**, 231–246.

Efron, B. (1979a), 'Bootstrap methods: another look at the jackknife', *Ann. Stat.* **6**, 1–26.

Efron, B. (1979b), 'Computers and the theory of statistics: thinking the unthinkable', *SIAM Rev.* **21**, 460–480.

Efron, B. & Gong, G. (1983), 'A leisurely look at the bootstrap, the jackknife, and cross-validation', *Am. Stat.* **37**, 36–48.

正统统计学家一直试图通过发明随意的特定流程而不是应用概率论来处理推断问题. 这里解释说明了三个最近的例子. 当然, 它们都违背我们对于合理性和一致性的合情条件, 证明这一点并将他们的结果与贝叶斯方法的结果进行比较是有趣且有启发意义的.

Evans, M. (1969), *Macroeconomic Forecasting*, Harper & Row, New York.

Fechner, G. J. (1860), *Elemente der Psychophysik*, 2 vols.; Vol. 1 translated as *Elements of Psychophysics*, Boring, E. G. & Howes, D. H., eds., Holt, Rinehart & Winston, New York (1966).

中译本 古斯塔夫·费希纳. 心理物理学纲要. 李晶, 译. 北京: 中国人民大学出版社, 2015.

Fechner, G. J. (1882), *Revision der Hauptpunkte der Psychophysik*, Breitkopf u. Härtel, Leipzig.

Feinstein, A. (1958), *Foundations of Information Theory*, McGraw-Hill, New York.

就像辛钦的工作 (Khinchin, 1957) 一样, 这是一位数学家对事物的看法, 与戈德曼 (Goldman, 1953) 的物理视角几乎没有共同点.

Ferguson, T. S. (1982), 'An inconsistent maximum likelihood estimate', *J. Am. Stat. Assoc.* **77**, 831–834.

Fieller, E. C. (1954), 'Some problems in interval estimation', *J. Roy. Stat. Soc. B* **16**, 175–185.

这篇论文和克里西的论文 (Creasy, 1954; 见本参考文献) 以估计两个正态抽样分布的均值比 μ_1/μ_2 的“菲耶勒-克里西问题”而著名。这一问题产生了大量的讨论和争议, 因为正统方法没有处理它的原则——几十年来没有人肯屈尊检查一下贝叶斯解。这是估计问题的一个典型的例子。很容易说明, 只有贝叶斯方法提供了解决它所需的技术工具。若泽·贝尔纳多 (José Bernardo, 1977) 最终从贝叶斯的角度考虑了这个问题。

Fisher, R. A. (1912), 'On an absolute criterion for fitting frequency curves', *Messeng. Math.* **41**, 155–160.

Fisher, R. A. (1915), 'Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population', *Biometrika* **10**, 507–521.

Fisher, R. A. (1930), 'Inverse probabilities', *Proc. Camb. Phil. Soc.* **26**, 528–535.

Fisher, R. A. (1935), *The Design of Experiments*, Oliver & Boyd, Edinburgh; six later editions to 1966.

Fisher, R. A. (1938), *Statistical Tables for Biological, Agricultural and Medical Research* (with F. Yates), Oliver & Boyd, Edinburgh; five later editions to 1963.

Fisher, R. A. & Tippett, L. H. C. (1928), 'Limiting forms of the frequency distribution of the largest or smallest member of a sample', *Proc. Camb. Phil. Soc.* **24**, 180–190.

Fougeré, P. F. (1977), 'A solution to the problem of spontaneous line splitting in maximum entropy power spectrum analysis', *J. Geophys. Res.* **82**, 1051–1054.

Galton, F. (1863), *Meteorographica*, MacMillan, London.

这位非凡的人在这里发明了气象图, 并通过研究它们发现了北半球的“反气旋”环流模式。

Galton, F. (1889), *Natural Inheritance*, MacMillan, London.

Gentleman, W. M. (1968), 'Matrix multiplication and fast Fourier transformations', *Bell Syst. Tech. J.*, **17**, 1099–1103.

Gillispie, C. C., ed. (1981), *Dictionary of Scientific Biography*, 16 vols., C. Scribner's Sons, New York.

寻找科学家基本信息的首选之处。

Glymour, C. (1980), *Theory and Evidence*, Princeton University Press.

Gnedenko, B. V. & Kolmogorov, A. N. (1954), *Limit Distributions for Sums of Independent Random Variables*, Addison-Wesley, Cambridge, MA.

中译本 Б. В. 哥涅坚科, А. Н. 廓洛莫格若夫. 相互独立随机变数之和的极限分布. 王寿仁, 译. 北京: 科学出版社, 1955.

在第 1 页, 我们发现了一个奇怪的陈述: “事实上, 概率论的所有认识论价值都基于这一点: 大规模随机现象在它们的集体行动中创造了严格的、非随机的规律性。” 有人认为这是为了服务于政治目的。无论如何, 今天概率论最有价值的应用都与不完全信息有关, 而与所谓的“随机现象”无关。随机现象在理论上是未定义的, 而且在自然中是未经确认的。

Goel, P. & Zellner, A. (1986), eds., *Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti*, Elsevier Science Publishers, Amsterdam.

- Gokhale, D. and Kullback, S. (1978), *The Information in Contingency Tables*, Marcel Dekker, New York.
- Goldberg, S. (1983), *Probability in Social Science*, Birkhaeuser, Basel.
- Good, I. J. (1965), *The Estimation of Probabilities*, Research Monographs #30, MIT Press, Cambridge, MA.
- 杰克·古德坚信“物理概率”的存在, 这些概率具有某种独立于人类信息的实在性。因此这本著作具有(对我们来说)不适当的标题。
- Grandy, W. T. & Schick, L. H., eds. (1991), *Maximum Entropy and Bayesian Methods*, Proceedings of the Tenth annual Maximum Entropy workshop, Kluwer Academic Publishers, Holland.
- Grenander, U. & Szegö, G. (1957), *Toeplitz Forms and their Applications*, University of California Press, Berkeley.
- Griffin, D. R. (1958), *Listening in the Dark*, Yale University Press, New Haven.
- 另见 Slaughter, R. H. & Walton, D. W., eds. (1970), *About Bats*, SMU Press, Dallas, Texas.
- Gull, S. F. & Daniell, G. J. (1978), 'Image reconstruction from incomplete and noisy data', *Nature* **272**, 686.
- Gull, S. F. & Daniell, G. J. (1980), 'The maximum entropy algorithm applied to image enhancement', *Proc. IEEE (E)* **5**, 170.
- Gull, S. F. & Skilling, J. (1984), 'The maximum entropy method', in Roberts, J. A., ed., *Indirect Imaging*, Cambridge University Press.
- Hacking, I. (1965), *Logic of Statistical Inference*, Cambridge University Press.
- Hacking, I. (1973), *The Emergence of Probability*, Cambridge University Press.
- Hacking, I. (1984), 'Historical models for justice', *Epistemologia, Special Issue on Probability, Statistics, and Inductive Logic*, **VII**, 191-212.
- Haldane, J. B. S. (1957), 'Karl Pearson, 1857-1957', *Biometrika* **44**, 303-313.
- Hampel, F. R. (1973), 'Robust estimation: a condensed partial survey', *Zeit. Wahrsch, theorie vrw. Beb.* **27**, 87-104.
- Hankins, T. L. (1970), *Jean d'Alembert: Science and the Enlightenment*, Oxford University Press.
- 中译本 托马斯·L·汉金斯. 科学与启蒙运动. 任定成, 张爱珍, 译. 上海: 复旦大学出版社, 2000.
- Heath, D. & Sudderth, W. (1976), 'de Finetti's theorem on exchangeable variables', *Am. Stat.* **30**, 188.
- 一个极其简单的推导.
- Helliwell, R. A. (1965), *Whistlers and Related Ionospheric Phenomena*, Stanford University Press, Palo Alto, CA.

Hellman, M. E. (1979), 'The mathematics of public-key cryptography', *Sci. Am.* **241**, 130-139.

Hewitt, E. & Savage, L. J. (1955), 'Symmetric measures on Cartesian products', *Trans. Am. Math. Soc.* **80**, 470-501.

将德菲内蒂表示定理推广到任意集合。

Hirst, F. W. (1926), *Life and Letters of Thomas Jefferson*, Macmillan, New York.

Hobson, A. & Cheung, B. K. (1973), 'A comparison of the Shannon and Kullback information measures', *J. Stat. Phys.* **7**, 301-310.

Hodges, J. L. & Lehmann, E. L. (1956), 'The efficiency of some nonparametric competitors of the *t*-test', *Ann. Math. Stat.* **27**, 324-335.

Hofstadter, D. R. (1983), 'Computer tournaments of the Prisoner's dilemma suggest how cooperation evolves', *Sci. Am.* **248**, (5), 16-26.

Holbrook, J. A. R. (1981), 'Stochastic independence and space-filling curves', *Am. Math. Monthly* **88**, 426-432.

Jagers, P. (1975), *Branching Processes with Biological Applications*, John Wiley, London.

James, W. & Stein, C. (1961), 'Estimation with quadratic loss', Proc. 4th Berkeley Symp., Univ. Calif. Press, **1**, 361-380.

Jansson, P. A., ed. (1984), *Deconvolution, with Applications in Spectroscopy*, Academic Press, Orlando, FL.

由九位作者写的文章，总结了在引入贝叶斯和最大熵方法之前的最新技术（主要是线性处理方法）。

Jaynes, E. T. (1963), 'Review of Noise and Fluctuations, by D. K. C. MacDonald', *Am. J. Phys.* **31**, 946.

为杰恩斯 (Jaynes, 1976) 所引用，以回应肯奥斯卡·普索恩关于物理学家很少关注噪声的指责。注意到噪声出现在所有物理领域。因此，早在存在统计学家之前，物理学家就在积极研究噪声并且知道正确的处理方法。

Jaynes, E. T. (1973a), 'Survey of the present status of neoclassical radiation theory', in Mandel, L. & Wolf, E., eds., *Proceedings of the 1972 Rochester Conference on Optical Coherence*, Pergamon Press, New York.

Jaynes, E. T. (1973b), 'The well-posed problem', *Found. Phys.* **3**, 477-493.

重印于杰恩斯的文章集 (Jaynes, 1983)。

Jaynes, E. T. (1980a), 'The minimum entropy production principle', *Ann. Rev. Phys. Chem.* **31**, 579-601.

重印于杰恩斯的文章集 (Jaynes, 1983)。

Jaynes, E. T. (1980b), 'What is the question?', in Bernardo, J. M., de Groot, M. H., Lindley, D. V. & Smith, A. F. M., eds., *Bayesian Statistics*, University Press, Valencia, Spain, pp. 618-629.

讨论问题的逻辑, 正如理查德·考克斯 (R. T. Cox, 1978) 指出的, 并且应用于参数估计和假设检验之间的关系. 重印于杰恩斯的文章集 (Jaynes, 1983, 第 382~388 页).

Jaynes, E. T. (1981), 'What is the problem?', in Haykin, S., ed., *Proceedings of the Second SSSP Workshop on Spectrum Analysis*, McMaster University.

以下文章是扩充版本.

Jaynes, E. T. (1982), 'On the rationale of maximum-entropy methods', *Proc. IEEE* **70**, 939-952.

Jaynes, E. T. (1984), 'Prior information and ambiguity in inverse problems', in *SIAM-AMS Proceedings*, Vol. 14, American Mathematical Society, pp. 151-166.

Jaynes, E. T. (1985a) 'Where do we go from here?', in Smith, C. & Grandy, W. T., eds. *Maximum-Entropy and Bayesian Methods in Inverse Problems*, D. Reidel Publishing Co., Dordrecht, pp. 21-58.

Jaynes, E. T. (1985b), 'Entropy and search theory', in Smith, C. & Grandy, W. T., eds. *Maximum-Entropy and Bayesian Methods in Inverse Problems*, D. Reidel Publishing Co., Dordrecht, pp. 443-454.

表明以前在信息论和搜索理论之间寻找联系的失败原因是使用了错误的熵表达式. 事实上, 只要我们在最深层假设空间上定义熵, 两者之间就有非常简单和普遍的联系.

Jaynes, E. T. (1985c) 'Macroscopic prediction', in Haken, H., ed., *Complex Systems - Operational Approaches*, Springer-Verlag, Berlin.

Jaynes, E. T. (1985d), 'Generalized scattering', in Smith, C. & Grandy, W. T., eds., *Maximum-Entropy and Bayesian Methods in Inverse Problems*, D. Reidel Publishing Co., Dordrecht, pp. 377-398.

添加新约束前后的两个最大熵分布的比较中包含的一些有意思的物理预测结论.

Jaynes, E. T. (1986), 'Some applications and extensions of the de Finetti representation theorem', in Goel, P. & Zellner, A., eds., *Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti*, Elsevier Science Publishers, Amsterdam, pp. 31-42.

该定理通常只适用于无限可交换序列, 但是如果舍弃生成函数上的非负性条件, 对于有限序列仍然有效. 这个结论使得它适用于更广泛的问题.

Jaynes, E. T. (1988a), 'The relation of Bayesian and maximum entropy methods', in Erickson, G. J. & Smith, C. R., eds. (1988), *Maximum-Entropy and Bayesian Methods in Science and Engineering*, Vol. 1, pp. 25-29.

Jaynes, E. T. (1988b), 'Detection of extra-solar system planets', in Erickson, G. J. & Smith, C. R., eds. (1988), *Maximum-Entropy and Bayesian Methods in Science and Engineering*, Vol. 1, pp. 147-160.

Jaynes, E. T. (1991), 'Notes on present status and future prospects', in Grandy, W. T. & Schick, L. H., eds. (1991), *Maximum Entropy and Bayesian Methods*, Proceedings of the Tenth annual Maximum Entropy Workshop, Kluwer Academic Publishers, Holland.

对于截止于 1990 年夏天的基本情况的总结.

Jaynes, E. T. (1993), 'A backward look to the future', in Grandy, W. T. & Milonni, P. W., eds., *Physics and Probability: Essays in Honor of Edwin T. Jaynes*, Cambridge University Press, pp. 261-275.

这本是庆祝我 70 岁生日的纪念文集, 其中包含我以前学生和同事的 22 篇文章.

Jefferys, W. H. (1990), 'Bayesian analysis of random event generator data', *J. Sci. Expl.* 4, 153-169.

表明正统显著性检验会严重高估特异功能数据的显著性, 贝叶斯检验由于不依赖于研究者的意图会得出合乎情理的结论.

Jeffreys, H. (1963), 'Review of Savage (1962)', *Technometrics* 5, 407-410.

Jeffreys, Lady Bertha Swirles (1992), 'Harold Jeffreys from 1891 to 1940', *Notes Rec. Roy. Soc. Lond.* 46, 301-308.

对哈罗德·杰弗里斯爵士早年生活简短、有些令人费解的不完整描述, 并附有他 30 多岁的照片. 详细描述了他对植物学的兴趣和早期的荣誉(他于 1910 年进入剑桥圣约翰学院读本科, 同年因一篇关于“岁差和章动”的论文获得亚当斯纪念奖). 但是, 令人惊讶的是, 论文中根本没有提到他在概率论方面的工作! 在 1919~1939 年期间, 这些工作引出了许多文章的发表以及对今天的科学家来说都非常重要的两本书(Jeffreys, 1931, 1939). 杰弗里斯的概率论工作具有根本上的重要性, 并且在他的所有其他工作都已成为历史后仍将具有生命力. 伯莎·杰弗里斯也是一名物理学家, 20 世纪 20 年代后期在哥廷根师从马克斯·博恩, 后来成为剑桥格顿女子学院的院长.

Jerri, A. J. (1977), 'The Shannon sampling theorem - its various extensions and applications', *Proc. IEEE* 65, 1565-1596.

有用公式的教程大合集, 带有 248 篇参考文献.

Johnson, R. W. (1979), 'Axiomatic characterization of the directed divergences and their linear combinations', *IEEE Trans.* IT-7, 641-650.

Kale, B. K. (1970), 'Inadmissibility of the maximum likelihood estimation in the presence of prior information', *Can. Math. Bull.* 13, 391-393.

Kalman, R. E. (1982), 'Identification from real data', in Hazewinkel, M. & Rinnooy Kan, A., eds., *Current Developments in the Interface: Economics, Econometrics, Mathematics*, D. Reidel Publishing Co., Dordrecht-Holland, pp. 161-196.

Kalman, R. E. (1990), *Nine Lectures on Identification*, Lecture Notes on Economics and Mathematical Systems, Springer-Verlag.

Kandel, A. (1986), *Fuzzy Mathematical Techniques with Applications*, Addison-Wesley, Reading, MA.

Kay, S. & Marple, S. L., Jr (1979), 'Source of and remedies for spectral line splitting in autoregressive spectrum analysis', *Proceedings of the 1979 IEEE International Conference on Acoustics, Speech Signal Processing*, October 1978, pp. 469-471.

Kemeny, J. G. & Snell, J. L. (1960), *Finite Markov Chains*, D. van Nostrand Co., Princeton,

NJ.

Kendall, M. G. (1956), 'The beginnings of a probability calculus', *Biometrika* **43**, 1-14; reprinted in Pearson & Kendall (1970).

一项有意思的心理学研究工作. 在试图将概率论早期发展缓慢的原因解释为他人毫无根据的偏见时, 他也无意中暴露了自己毫无根据的偏见, 在我们看来, 这是导致 20 世纪概率论进步缓慢甚至倒退的主要原因.

Khinchin, A. I. (1949), *Mathematical Foundations of Statistical Mechanics*, Dover Publications, Inc., New York.

将计算技巧建立在中心极限定理基础上的尝试, 对于我们当前感兴趣的问题来说不够普遍. 但是对于结构函数和分拆函数之间的拉普拉斯变换关系的处理在今天仍然很有价值, 并且为我们在第二部分中的研究奠定了数学基础.

Kiefer, J. & Wolfowitz, J. (1956), 'Consistency of the maximum likelihood estimation in the presence of infinitely many incidental parameters', *Ann. Math. Stat.* **27**, 887-906.

Kindermann, R. & Snall, J. L. (1980), *Markov Random Fields*, Contemporary Mathematics Vol. 1, AMS, Providence, RI.

Kuhn, T. S. (1962), *The Structure of Scientific Revolutions*, University of Chicago Press; 2nd edn, 1970.

[中译本] 托马斯·库恩. 科学革命的结构 (第四版). 伊安·哈金, 导读. 金吾伦, 胡新和, 译. 北京: 北京大学出版社, 2012.

Kullback, S. (1959), *Information Theory and Statistics*, John Wiley, New York.

这是一部美丽的作品, 从未被正确欣赏过, 因为它比所处的时代先进了 20 年.

Landau, H. J. (1983), 'The inverse problem for the vocal tract and the moment problem', *SIAM J. Math. Anal.* **14**, 1019-1035.

通过与伯格最大熵谱分析密切相关的反射系数技术对语音生成进行建模.

Landau, H. J. (1987), 'Maximum entropy and the moment problem', *Bull. Am. Math. Soc.* **16**, 47-77.

根据多个领域中更一般的问题来解释伯格解. 强烈推荐对其中的数学做更深入的了解.

Legendre, A. M. (1806), 'Nouvelles méthodes pour la détermination des orbits des comètes', Didot, Paris.

Leibniz, G. W. (1968), *General Investigations Concerning the Analysis of Concepts and Truths*, trans. W. H. O'Briant, University of Georgia Press.

Lessard, S., ed. (1989), *Mathematical and Statistical Developments in Evolutionary Theory*, NATO ASI Series Vol. C299, Kluwer Academic Publishers, Holland.

1987 年在加拿大蒙特利尔举行的会议的记录.

Lewis, G. N. (1930) 'The symmetry of time in physics', *Science* **71**, 569.

对熵和信息之间联系的早期认识, 显示出远超 50 年后其他许多人发表文章的理解水平.

Lindley, D. V. (1956), 'On a measure of the information provided by an experiment', *Ann. Math.* **27**, 986-1005.

- Lindley, D. V. (1958) 'Fiducial distributions and Bayes' theorem', *J. Roy. Stat. Soc.* **B20**, 102-107.
- Lindley, D. V. (1971) *Bayesian Statistics: A Review*, Society for Industrial and Applied Mathematics, Philadelphia.
- Linnik, Yu. V. (1961), *Die Methode der kleinsten Quadrate in Moderner Darstellung*, Deutscher Verl. der Wiss., Berlin.
- Litterman, R. B. (1985), 'Vector autoregression for macroeconomic forecasting', in Zellner, A. & Goel, P., eds., *Bayesian Inference and Decision Techniques*, North-Holland Publishers, Amsterdam.
- Lukacs, E. (1960), *Characteristic Functions*, Griffin, London.
- Macdonald, P. D. M. (1987), 'Analysis of length-frequency distributions', in Summerfelt, R. C. & Hall, G. E., eds., *Age and Growth of Fish*, Iowa State University Press, pp. 371-384.
- 用于对正态分布和其他分布的混合物进行反卷积操作的计算机程序. "MIX 3.0" 程序可从以下地址获得: Ichthus Data Systems, 59 Arkell St, Hamilton, Ontario, Canada L8S 1N6. 在第 7 章中, 我们注意到这个问题不是很适定的. 以克萨斯承认它 "本质上是困难的", 并且可能无法在用户数据上令人满意地工作. 另见蒂特林顿、史密斯和马科夫的著作 (Titterington, Smith & Makov, 1985).
- Mandel, J. (1964), *The Statistical Analysis of Experimental Data*, Interscience, New York.
- 直接的正统特定方法, 其中之一在杰恩斯的文章 (Jaynes, 1976) 中做了分析.
- Mandelbrot, B. (1977), *Fractals, Chance and Dimension*, W. H. Freeman & Co., San Francisco.
- 中译本 B. 曼德尔布洛特. 分形对象: 形、机遇和维数. 文志英, 苏虹, 译. 北京: 世界图书出版公司, 1999.
- Marple, S. L. (1987), *Digital Spectral Analysis with Applications*, Prentice-Hall, New Jersey.
- Martin, R. D. & Thompson, D. J. (1982), 'Robust-resistant spectrum estimation', *Proc. IEEE* **70**, 1097-1115.
- 显然是在他们的导师约翰·图基的密切指导下写成的, 这延续了他根据直觉而非概率论发明一系列特定工具的实践. 它甚至不承认最大熵或贝叶斯方法的存在. 值得称赞的是, 作者确实通过他们的方法对几个数据集进行了计算机分析——结果对我们来说并不令人兴奋. 获取他们的原始数据并通过像布雷特索斯特 (Bretthorst, 1988) 那样使用概率论的方法来进行分析将会很有趣, 我们认为结果会完全不同.
- Masani, S. M. (1977), 'A paradox in admissibility', *Ann. Stat.* **5**, 544-546.
- Maxwell, J. C. (1850), Letter to Lewis Campbell; reproduced in L. Campbell & W. Garrett, *The Life of James Clerk Maxwell*, Macmillan, 1881.
- McColl, H. (1897) 'The calculus of equivalent statements', *Proc. Lond. Math. Soc.* **28**, 556.
- 对布尔版本的概率论的批评.

- McFadden, D. (1973), 'Conditional logit analysis of qualitative choice behavior', in Zarembka, P., ed., *Frontiers in Econometrics*, Academic Press, New York.
- Mead, L. R. & Papanicolaou, N. (1984), 'Maximum entropy in the problem of moments', *J. Math. Phys.* **25**, 2404–2417.
- Miller, R. G. (1974), 'The jackknife – a review', *Biometrika* **61**, 1–15.
- Mitler, K. S. (1974), *Multivariate Distributions*, John Wiley, New York.
- Molina, E. C. (1931), 'Bayes' theorem, an expository presentation', *Bell Syst. Tech. Publ.* Monograph B-557.
- 与凯恩斯 (Keynes, 1921)、杰弗里斯 (Jeffreys, 1939) 和伍德沃德 (Woodward, 1953) 站在一起, 证明荒野中总是有孤独的声音呼唤一种更明智的推断方法。
- Moore, G. T. & Scully, M. O., eds. (1986), *Frontiers of Nonequilibrium Statistical Physics*, Plenum Press, New York.
- 这里, 几位发言者在贝尔不等式实验的基础上确认了他们的信念, 即“原子不是真实的”, 同时坚持概率是客观真实的信念! 我们认为这是思维投射谬误的一个公开例子, 到了荒谬的地步。
- Munk, W. H. & Snodgrass, F. E. (1957), 'Measurements of Southern Swell at Guadalupe Island', *Deep-Sea Res.* **4**, 272–286.
- 这是图基 (Tukey, 1984) 举出的最伟大的光谱分析例子, 是用其他方法永远无法完成的。反过来, 杰恩斯 (Jaynes, 1987) 用哨声分析回答了这个问题。
- Newton, Sir Isaac (1687) *Philosophia Naturalis Principia Mathematica*, trans. Andrew Motte, 1729; revised and reprinted as *Mathematical Principles of Natural Philosophy*, Florian Cajori, ed., University of California Press (1946).
- 中译本 牛顿. 自然哲学的数学原理. 赵振江, 译. 北京: 商务印书馆, 2020.
- 另见卡乔里的著作 (Cajori, 1928, 1934; 见本参考文献)。
- Neyman, J. & Pearson, E. S. (1933), 'On the problem of the most efficient test of statistical hypotheses', *Phil. Trans. Roy. Soc.* **231**, 289–337.
- Neyman, J. & Pearson, E. S. (1967), *Joint Statistical Papers*, Cambridge University Press.
- 20 世纪 30 年代多篇奈曼-皮尔逊论文的再版, 最初散布在几册不同的期刊上。
- Neyman, J. (1959), 'On the two different aspects of representative method: the method of stratified sampling and the method of purposive selection', *Estadistica* **17**, 587–651.
- Neyman, J. (1962) 'Two breakthroughs in the theory of statistical decision making', *Int. Stat. Rev.* **30**, 11–27.
- 定位和纠正其中的错误是一个很好的家庭作业。
- Neyman, J. (1981), 'Egon S. Pearson (August 11, 1895–June 12, 1980)', *Ann. Stat.* **9**, 1–2.
- Novák, V. (1988), *Fuzzy Sets and their Applications*, A. Hilger, Bristol.
- Nyquist, H. (1924), 'Certain factors affecting telegraph speed', *Bell Syst. Tech. J.* **3**, 324.
- Nyquist, H. (1928), 'Certain topics in telegraph transmission theory', *Trans. AIEE*, **47**, 617–644.

O'Hagan, A. (1977), 'On outlier rejection phenomena in Bayes inference', *J. Roy. Stat. Soc. B* **41**, 358–367.

我们的观点是, 贝叶斯推断没有病态、例外的情况, 特别是没有异常值. 将任何观察值视为“异常值”加以拒绝都违反了理性推断的原则, 这只是表明对问题的表述不正确. 也就是说, 如果能够确定任何观察结果是你指定的模型的异常值, 则该模型无法正确捕获有关生成数据的机制的先验信息. 原则上, 补救措施不是拒绝任何观察值, 而是定义一个更现实的模型(正如我们在讨论稳健性时指出的那样). 然而, 我们承认, 如果严格正确的流程为可疑数据分配了非常低的权重, 那么从数据集中直接删除该值可能是一个合理的近似, 这很容易做到.

Ore, O. (1953), *Cardano, the Gambling Scholar*, Princeton University Press.

Ore, O. (1960), 'Pascal and the invention of probability theory', *Am. Math. Monthly* **67**, 409–419.

Pearson, K. (1892), *The Grammar of Science*, Walter Scott, London.

中译本 卡尔·皮尔逊. 科学的规范. 李醒民, 译. 北京: 商务印书馆, 2012.

1900年、1911年由 A. & C. Black, London 重印, 1937年由 Everyman Press 重印. 科学推理原理的阐述, 主要是因为哈罗德·杰弗里斯深受其影响并高度评价它. 这并不妨碍他指出, 卡尔·皮尔逊在他后来的科学工作中远未应用自己的原则. 有关卡尔·皮尔逊(1857—1936)的传记材料, 见霍尔丹的文章(Haldane, 1957; 见本参考文献).

Pearson, K. (1905), 'The problem of the random walk', *Nature* **72**, 294, 342.

Pearson, K. (1921–1933), *The History of Statistics in the 17'th and 18'th Centuries*, Pearson, E.S., ed., Lectures given at University College, London, Griffin, London (1978).

Penfield, W. (1958), *Proc. Natl Acad. Sci. (USA)* **44**, 59.

脑部手术期间的观察记录, 手术中对大脑特定部位的电刺激导致有意识的患者回忆起各种被遗忘的经历. 这个毫无疑问的真实现象与第18章中的 A_p 分布理论密切相关. 但现在有人指控精神科医生正在让他们的病人——尤其是年幼的孩子——回忆起从未发生过的事情, 并带来灾难性的法律后果. 识别有效和无效回忆的问题似乎进入了一段争议期.

Pierce, J. R. (1980), *Symbols, Signals, and Noise: An Introduction to Information Theory*, Dover Publications, Inc., New York.

对完全的初学者简明地介绍信息论, 但是不涉及当前的重要应用.

Poisson, S. D. (1837), *Recherches sur la Probabilité des Jugements*, Bachelier imprimeur-Libraire, Paris.

泊松分布的首次出现.

Pólya, G. (1921), 'Über eine Aufgabe der Wahrscheinlichkeitsrechnung betreffend die Irrfahrt im Strassennetz', *Math. Ann.* **84**, 149–160.

有时被称为“随机游走”一词的首次出现. 然而, 我们可以指出皮尔逊(Pearson, 1905; 见本参考文献)和瑞利(Rayleigh, 1919; 见本参考文献)都提到过随机游走.

Pólya, G. (1923), 'Herleitung des Gauss'schen Fehlergesetzes aus einer Funktionalgleichung', *Math. Zeit.* **18**, 96–108.

Pontryagin, L. S. (1946), *Topological Groups*, Princeton University Press, Princeton, NJ.

Popper, K. (1958), *The Logic of Scientific Discovery*, Hutchinson & Co., London.

[中译本] 卡尔·波普尔. 科学发现的逻辑. 查汝强, 邱仁宗, 万木春, 译. 北京: 中国美术学院出版社, 2008.

否认归纳的可能性, 理由是每一个科学理论为真的先验概率都为零. 卡尔·波普尔之所以出名, 主要是因为他提出理论可能不会被证明为真、只能证伪的学说. 因此, 一个理论的优点在于它的可证伪性. 第 1 章的三段论表明这其中有明显真理的成分. 爱因斯坦在他的名言中也提到了这一点: “无论多少实验都不能证明我是对的, 但是只要一个实验就可以随时证明我错了.” 然而, 该学说只对断言存在不可观察原因或机制的理论是正确的, 任何断言可观察事实的理论都是其反例.

Popper, K. (1963), *Conjectures and Refutations*, Routledge & Kegan Paul, London.

[中译本] 卡尔·波普尔. 猜想与反驳: 科学知识的增长. 傅季重, 纪树立, 周昌忠, 蒋弋为, 译. 上海: 上海译文出版社, 2015.

Popov, V. N. (1987), *Functional Integrals and Collective Excitations*, Cambridge University Press.

超流体、超导、等离子体动力学、超辐射和相变的应用大纲. 理解这些现象的有用起点, 但是仍然缺乏任何连贯的理论基础——我们认为只有最大熵原理才能提供推理方法.

Prenzel, H. V. (1975), *Dynamic Trendline Charting: How to Spot the Big Stock Moves and Avoid False Signals*, Prentice-Hall, Englewood Cliffs, NJ.

不包含任何概率论或其他数学知识: 只是绘制股票价格的月度变化, 在图上画几条直线, 它们的交点告诉你该做什么以及何时做. 这个系统至少确实可以让人非常清楚地看到四年的美国总统选举周期.

Press, S. J. (1989), *Bayesian Statistics: Principles, Models and Applications*, J. Wiley & Sons, Inc., New York.

[中译本] S. 詹姆士·普雷斯. 贝叶斯统计学: 原理、模型及应用. 廖文, 陈安贵, 等译. 袁卫, 校. 北京: 中国统计出版社, 1992.

包含许多现在可用的贝叶斯计算机程序的列表.

Preston, C. J. (1974), *Gibbs States on Countable Sets*, Cambridge University Press.

在这里有使用“状态”一词来表示概率分布的讨厌做法. 人们无法想象一种更具破坏性的错误、误导性术语.

Priestley, M. B. (1981), *Spectral Analysis and Time Series*, 2 vols., Academic Press, Inc., Orlando, FL; combined paperback edition with corrections (1983).

Puri, M. L., ed. (1975), *Stochastic Processes and Related Topics*, Academic Press, New York.

Quaster, H., ed. (1953), *Information Theory in Biology*, University Illinois Press, Urbana.

Ramsey, F. P. (1931), *The Foundations of Mathematics and Other Logical Essays*, Routledge and Kegan Paul, London.

弗兰克·拉姆齐于 1925 年在剑桥大学获得数学学位, 然后成为国王学院的院士, 其间他与约翰·梅纳德·凯恩斯就经济学理论进行了合作. 如果不是 1930 年就在 26 岁去世, 他

无疑会成为 20 世纪最有影响力的贝叶斯主义者。在这些文章中, 我们可以看到与我们对概率论的阐述非常相似的东西。

Rayleigh, Lord (1919), 'On the problem of random vibrations, and of random flights in one, two or three dimensions', *Edin. & Dublin Phil. Mag. & J. Sci.* **37**, series 6, 321-347.

Reichardt, H. (1960), *C. F. Gauss—Leben und Werk*, Haude & Spener, Berlin.

Reid, C. (1970), *Hilbert*, Springer-Verlag, New York.

中译本 康斯坦丝·瑞德. 希尔伯特: 数学界的亚历山大. 袁向东, 李文林, 译. 上海: 上海科学技术出版社, 2018.

Reid, C. (1959), 'On a new axiomatic theory of probability', *Acta. Math. Acad. Sci. Hung.* **6**, 285-335.

这项工作与我们的工作有几个共同点, 但是阐述方式大不相同。

Rihaczek, A. W. (1981), 'The maximum entropy of radar resolution', *IEEE Trans. Aerospace Electron. Syst.* **AES-17**, 144.

对最大熵方法的另一篇攻击文章, 仍然否认所谓的“超分辨率”的可能性, 尽管多年前约翰·伯格在理论和实践中都已经证明了这一点, 并且到 1981 年为止被许多科学家和工程师例行使用, 如奇尔德斯的文章集 (Childers, 1978; 见本参考文献) 所示。

Rissanen, J. (1983), 'A universal prior for the integers and estimation by minimum description length', *Ann. Stat.* **11**, 416-431.

近几十年来为数不多的新思想之一。我们认为它有着光明的未来, 但是还不能预测它会是什么样子的。

Robbins, H. (1950), 'Asymptotically subminimax solutions of compound statistical decision problems', *Proceedings of the 2nd Berkeley Symposium of Mathematics Statistics and Probability*, University of California Press, pp. 131-148.

斯坦的研究 (Stein, 1956; 见本参考文献) 的先导。

Robbins, H. (1956), 'An empirical Bayes' approach to statistics', *Proceedings of the 3rd Berkeley Symposium on Mathematics, Statistics and Probability I*, University of California Press, pp. 157-164.

Robinson, A. (1966), *Non-standard Analysis*, North-Holland, Amsterdam.

中译本 A. 鲁滨逊. 非标准分析. 申又彬, 王世强, 张锦文, 译. 北京: 科学出版社, 1980. 如何做每一个错误的计算。

Robinson, E. A. (1982), 'A historical perspective of spectrum estimation', *Proc. IEEE* **70**, 855-906.

Robinson, G. K. (1975), 'Some counterexamples to the theory of confidence intervals', *Biometrika* **62**, 155-162.

Rowlinson, J. S. (1970), 'Probability, information and entropy', *Nature* **225**, 1196-1198.

对最大熵原理的攻击表明了对推断性质的一种普遍误解。杰恩斯 (Jaynes, 1978) 对此做出了回答。

- Sampson, A. R. & Smith, R. L. (1984), 'An information theory model for the evaluation of circumstantial evidence,' *IEEE Trans. Systems, Man, and Cybernetics* **15**, 916.
- Sampson, A. R. & Smith, R. L. (1982), 'Assessing risks through the determination of rare event probabilities', *Op. Res.* **30**, 839-866.
- Sanov, I. N. (1961), 'On the probability of large deviations of random variables', IMS and AMS Translations of Probability and Statistics, from *Mat. Sbornik* **42**, 1144.
- Scheffé, H. (1959), *The Analysis of Variance*, John Wiley, New York.
- Schendel, U. (1989) *Sparse Matrices*, J. Wiley & Sons, New York.
- Schlaifer, R. (1959), *Probability and Statistics for Business Decisions: An Introduction to Managerial Economics Under Uncertainty*, McGraw-Hill Book Company, New York.
- 认识到在现实世界决策问题中需要贝叶斯方法的早期著作, 与切尔诺夫和摩西同时期的决策论著作 (Chernoff and Moses, 1959) 形成鲜明的对比.
- Schneider, T. D. (1991), 'Theory of molecular machines', *J. Theor. Biol.* **148**, 83-137.
- 分为两部分, 关注信道容量和能量消耗
- Schnell, E. E. (1960), 'Samuel Pepys, Isaac Newton and probability', *Am. Stat.* **14**, 27-30.
- 我们从中了解到, 帕斯卡和牛顿都有给出正确解但不被相信的经历, 这个问题并不是现代贝叶斯主义者独有的.
- Schrödinger, E. (1945), 'Probability problems in nuclear chemistry', *Proc. Roy. Irish Acad.* **51**.
- Schrödinger, E. (1947), 'The foundation of the theory of probability', *Proc. Roy. Irish Acad. (A)*, **51**, pp. 51-66, 141-146.
- 这篇论文由于使我们能在与我们有同样思考的人员名单中再添加一个杰出的名字, 今天仍然很有价值. 在这里, 薛定谔宣称概率的“频率论”观点不足以满足科学的需要, 并通过在某种程度上与本书第 1 章和第 2 章的精神一致的努力, 试图证明概率观点适用于个别事件而不是事件的“集合”. 他提出了一些巧妙的论证, 但是不知道这些想法已经远远超出了他的工作水平. 他不知道考克斯定理, 而且和当时大多数受过常规训练的科学家一样, 显然从未听说过托马斯·贝叶斯或哈罗德·杰弗里斯. 他既没有给出任何有用的应用, 也没有获得超出杰弗里斯八年前发表的理论结果. 尽管如此, 他的思想在这个问题及其他有争议的问题上都朝着正确的方向发展.
- Shafer, G. (1976), *A Mathematical Theory of Evidence*, Princeton University Press, Princeton, NJ.
- 一个狂热的反贝叶斯主义者试图发展二值概率理论.
- Shafer, G. (1982), 'Lindley's paradox', *J. Am. Stat. Assoc.* **77**, 325-334.
- 显然, 谢弗没有意识到这些内容都在比林德利早了大约 20 年的杰弗里斯的著作 (Jeffreys, 1939, 第 194 页) 中. 但谢弗的其他工作已经表明, 他从未阅读和理解过杰弗里斯的著作.
- Shamir, A (1982), 'A polynomial time algorithm for breaking the basic Merkle-Hellman cryptosystem', in Chaum, D., Rivest, R. L. & Sherman, A. T., eds., *Advances in Cryptology: Proceedings of Crypto 82, 23-25 August 1982*, Plenum Press, New York, pp.

279-288.

Shaw, D. (1976), *Fourier Transform NMR Spectroscopy*, Elsevier, New York.

Sheynin, O. B. (1978), 'S. D. Poisson's work in probability', *Archiv. f. Hist. Exact Sci.* **18**, 245-300.

Sheynin, O. B. (1979), 'C. F. Gauss and the theory of errors', *Archiv. f. Hist. Exact Sci.* **19**, 21-72.

Shiryayev, A. N. (1978), *Optimal Stopping Rules*, Springer-Verlag, New York.

Siegmund, D. (1985), *Sequential Analysis*, Springer-Verlag, Berlin.

没有提及贝叶斯定理或可选停止!

Simmons, G. J. (1979), 'Cryptography, the mathematics of secure communication', *The Math. Intelligencer* **1**, 233-246.

Sinai, J. G. (1982), *Rigorous Results in the Theory of Phase Transitions*, Akadémiai Kiado, Budapest.

Skilling, J., ed. (1989), *Maximum Entropy and Bayesian Methods*, Proceedings of the Eighth Maximum Entropy Workshop, Cambridge, UK, August 1988, Kluwer Academic Publishers, Dordrecht, Holland.

Smith, C. R. & Grandy, W. T., eds. (1985), *Maximum-Entropy and Bayesian Methods in Inverse Problems*, D. Reidel Publishing Co., Dordrecht, Holland.

Smith, C. R. & Erickson, G. J., eds. (1987), *Maximum-Entropy and Bayesian Spectral Analysis and Estimation Problems*, D. Reidel Publishing Co., Dordrecht, Holland.

Smith, D. E. (1959), *A Source Book in Mathematics*, McGraw-Hill Book Co., New York.
包含费马-帕斯卡通信.

Smith, W. B. (1905), 'Meaning of the Epithet Nazorean', *The Monist* **15**, 25-95.

得出的结论是, 在尼西亚会议之前, “拿撒勒” 不是地理名称, 它有另外的含义.

Sonett, C. P. (1982), 'Sunspot time series: spectrum from square law modulation of the Hale cycle', *Geophys. Res. Lett.* **9**, 1313-1316.

Spinoza, B. (1663), 'Renati des Cartes Principiorum philosophiae pars I, & II, more geometrico demonstratae,' *Ethics*, part 2, Prop. XLIV: 'De natura Rationis no est res, ut contingentes; sed, ut necessarias, contemplari.'

翻译过来, 这个命题是: “把事物看成偶然的, 不是理性的本性; 相反, 事物应该被认为是必然的.”

Spitzer, F. (1964), *Principles of Random Walk*, van Nostrand, New York.

历史背景及当前现状.

Stein, C. (1945), 'A two sample test for a linear hypothesis whose power is independent of the variance', *Ann. Math. Stat.* **16**, 243-258.

Stein, C. (1956), 'Inadmissibility of the usual estimator for the mean of a multivariate normal distribution', *Proceedings of the 3rd Berkeley Symposium*, vol. 1, pp. 197-206, University of California Press.

首次宣称“斯坦收缩”现象.

Stein, C. (1959), 'An example of wide discrepancy between fiducial and confidence intervals', *Ann. Math. Stat.* **30**, 877–880.

Stein, C. (1964), 'Inadmissibility of the usual estimate for the variance of a normal distribution with unknown mean', *Ann. Inst. Stat. Math.* **16**, 155–160.

斯坦的不可容许性发现虽然令接受过常规训练的统计学家感到震惊, 但是对贝叶斯主义者来说一点儿也不感到吃惊或不安. 它只是说明了我们已经清楚的事情: 忽略所有先验信息的可容许性准则在实际问题中具有潜在危险. 在这里, 该准则可以将实际上的最优估计量作为“不可容许的”加以拒绝, 正如第 13 章中简要说明的那样.

Stigler, S. M. (1974a), 'Cauchy and the Witch of Agnesi', *Biometrika* **61**, 375–380.

Stigler, S. M. (1974b), 'Gergonne's 1815 paper on the design and analysis of polynomial regression experiments', *Historia Math.* **1**, 431–477.

Stigler, S. M. (1982a), 'Poisson on the Poisson distribution', *Stat. & Prob. Lett.* **1**, 33–35.

Stigler, S. M. (1982b), 'Thomas Bayes's Bayesian inference', *J. Roy. Stat. Soc.* **A145**, 250–258.

Stone, M. & Springer-Verlag, B. G. F. (1965), 'A paradox involving quasi prior distributions', *Biometrika* **52**, 623–627.

Stromberg, K. (1979), 'The Banach-Tarski paradox', *Am. Math. Monthly* **86**, 151–160.

关于全等集合.

Student (1908), 'The probable error of a mean', *Biometrika* **6**, 1–24.

Takeuchi, K., Yanai, H. & Mukherjee, B. N. (1982), *The Foundations of Multivariate Analysis*, J. Wiley & Sons, New York.

Taylor, R. L., Daffer, P. Z. & Patterson, R. F. (1985), *Limit Theorems for Sums of Exchangeable Random Variables*, Rowman & Allanheld Publishers.

介绍欧几里得空间和巴拿赫空间中离散时间可交换序列的极限定理的已知版本.

Thomas, M. U. (1979), 'A generalized maximum entropy principle', *Operations Res.* **27**, 1188–1195.

Tikhonov, A. N. & Arsenin, V. Y. (1977), *Solutions of Ill-posed Problems*, Halsted Press, New York.

中译本 A. H. 吉洪诺夫, B. Я. 阿尔先宁. 不适定问题的解法. 王秉忱, 译. 陈恕行, 校. 北京: 地质出版社, 1979.

特定数学方法的集合, 作者不断尝试对没有逆运算的运算符取逆. 他从没有意识到这些是推断问题, 而不是取逆问题.

Todhunter, I. (1865), *A History of the Mathematical Theory of Probability*, Macmillan, London; reprinted 1949, 1965, by Chelsea Press, New York.

Todhunter, I. (1873), *A History of the Mathematical Theories of Attraction and the Figure of the Earth*, 2 vols., Macmillan, London; reprinted 1962 by Dover Press, New York.

Toraldo di Francia, G. (1955), 'Resolving power and information', *J. Opt. Soc. Am.* **45**, 497-501.

首次认识到广义逆问题本质的文章, 作者试图使用信息论, 但没有看到(贝叶斯)概率论是解决问题的合适工具.

Train, K. (1986), *Qualitative Choice Analysis Theory, Econometrics, and an Application to Automobile Demand*, MIT Press, Cambridge, MA.

Truesdell, C. (1987), *Great Scientists of Old as Heretics in 'The Scientific Method'*, University Press of Virginia.

历史记录表明, 数学物理中一些最伟大的进展是在很少甚至根本没有实验基础的情况下取得的, 这似乎是对通常宣称的“科学方法”的挑战. 这只是表明构建创造性假设对于根据给定假设进行推断的首要重要性. 不幸的是, 今天我们虽然有完善且非常成功的推断理论, 但是根本没有关于最优假设构建的正式理论, 而且最近成功的例子也很少.

Tukey, J. W. (1960), 'A survey of sampling from contaminated distributions', in Olkin, I., ed., *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*, Stanford University Press, California, pp. 448-485.

Tukey, J. W. (1962), 'The future of data analysis', *Ann. Math. Stat.* **33**, 1-67.

对于所有试图预言将来自己的主观偏见会成真的人来说, 这是一个强有力的教训.

Tukey, J. W. (1978), 'Granger on seasonality', in Zellner, A., ed., *Seasonal Analysis of Time Series*, US Dept of Commerce, Washington.

将贝叶斯推断的本质视为“羞怯地隐藏在形式方法背后”偷偷干不光彩事情的有趣观点.

Tukey, J. W. (1984), 'Styles of spectrum analysis', *Scripps Inst. Oceanography Ref. Series* 84-85, March, pp. 100-103.

对所有理论原理的强烈攻击, 包括自回归模型、最大熵和贝叶斯方法等. 出现在第 103 页上的“最大熵的倡导者”正是埃德温·杰恩斯.

Tukey, J. W., Bloomfield, P. Brillinger, D. & Cleveland, W. S. (1980), *The Practice of Spectrum Analysis*, University Associates, Princeton, New Jersey.

1980 年 12 月的课程笔记.

Tukey, J. W. & Brillinger, D. (1982), 'Spectrum estimation and system identification relying on a Fourier transform', unpublished.

这篇罕见的文章是作为 1982 年 9 月关于频谱分析的 IEEE 特刊的特邀论文而写的, 但是它的长度(不完整版本为 112 页)使得它出现在特刊里. 因为它是一份重要的历史文献, 我们希望它能在别处出版. 图基的文章(Tukey, 1984; 见本参考文献)包含其中的一部分内容.

Valery-Radot, R. (1923), *The Life of Pasteur*, Doubleday, Page & Co., Garden City, New York.

中译本 R. 瓦莱里-拉多. 微生物学的奠基人巴斯德. 陶亢德, 董元骥, 译. 北京: 科学出版社, 1985.

Van Campenhout, J. M. & Cover, T. M. (1981), 'Maximum entropy and conditional probability', *IEEE Trans. Info. Theor.* **IT-27**, 483-489.

对物理学家自 1928 年以来称为“统计力学的达尔文-福勒方法”的概括总结。

van den Bos, A. (1971), 'Alternative interpretation of maximum entropy spectral analysis', *IEEE Trans. Info. Theor.* **IT-17**, 493-494.

重印于奇尔德斯的文章集 (Childers, 1978; 见本参考文献)。表达了对最大熵频谱分析的几种担忧, 杰恩斯的文章 (Jaynes, 1982; 见本参考文献) 对此进行了回答。

Varian, H. (1978), *Microeconomic Analysis*, Norton & Co., New York.

中译本 哈尔·R. 范里安. 微观经济分析 (第三版). 王文举, 译. 北京: 中国人民大学出版社, 2015.

Vasicek, O. (1980), 'A conditional law of large numbers', *Ann. Prob.* **8**, 142-147.

Wald, A. (1941), *Notes on the Theory of Statistical Estimation and of Testing Hypotheses*, Mimeographed, Columbia University.

此时, 沃尔德试图使得他的学生确信, 贝叶斯方法是完全错误的, 无法处理推断问题。九年后, 他自己的研究使他得出了相反的观点。

Wald, A. (1942), *On the Principles of Statistical Inference*, Notre Dame University Press.

Wald, A. (1943), 'Sequential analysis of statistical data: theory', Restricted report dated September 1943.

Waldmeier, M. (1961), *The Sunspot Activity in the Years 1610-1960*, Schulthes, Zürich.

也许是所有数据集中被分析得最多的。

Walley, P. (1991), *Statistical Reasoning with Imprecise Probabilities*, Chapman & Hall, London.

由于担心非正常先验, 他引入了先验“近乎无知类”(NIC)的概念。从那时起, 许多人试图精确定义可用先验的 NIC。我们认为任何导致正常后验分布的先验都是可用的并且可能有用, 注意到这一点可以简化这一概念。显然, 一个给定的非正常先验是否能够做到这一点, 不仅取决于先验的任何性质, 还取决于先验和似然函数的联合行为——也就是说, 通过先验、模型和数据。还需要更多吗?

Watson, J. D. (1968), *The Double Helix*, Signet Books, New York.

中译本 詹姆斯·沃森. 双螺旋: 发现 DNA 结构的故事. 刘望夷, 译. 上海: 上海译文出版社, 2016.

描述导致发现 DNA 结构过程的著名图书。它之所以成为畅销书, 是因为激起了没有任何科学知识的人的强烈好评, 他们很高兴有人认为象牙塔里的科学家有与他们一样肮脏的动机。这不是具有技术知识的科学家的看法, 其中一位私下对本书作者说: “看起来最糟糕的人是沃森本人。”但这是古老的历史, 对于今天的我们来说, 有趣的问题是: 如果应用于 X 射线衍射数据的贝叶斯推断原理在 1950 年被开发并简化为计算机程序, 这一发现过程是否会明显加快? 我们怀疑, 如果通过计算机程序 [如适配这一问题的布雷特索斯特 (Bretthorst, 1988) 的程序] 分析乍看起来非常令人困惑的罗莎琳德·富兰克林的第一张“A 结构”照片, 是否会立即指出双螺旋是最可能的结构 (或者至少说“螺旋”存在具有

可能性, 并且事后可以被肉眼识别到)。从广义上讲, 这一问题与雷达目标识别非常相似。关于 DNA 发现故事的另一个版本, 以及对事件过程的一些不同的回忆, 见克里克的著作 (Crick, 1988; 见引用文献)。

Wax, N., ed. (1954), *Noise & Stochastic Processes*, Dover Publications, Inc., New York.

Wehrl, A. (1978), 'General properties of entropy', *Rev. Mod. Phys.* **50**, 220-260.

Whittle, P. (1954), Comments on periodograms, Appendix to H. Wold (1954; this bibliography), pp. 200-227.

Whittle, P. (1957), 'Curve and periodogram smoothing', *J. Roy. Stat. Soc. B* **19**, 38-47.

Whittle, P. (1958), 'On the smoothing of probability density functions', *J. Roy. Stat. Soc. B* **20**, 334-343.

Wigner, E. P. (1967), *Symmetries and Reflections*, Indiana University Press, Bloomington.

从概率论的角度来看, 这里重印的最有趣的文章是 #15, “自我复制单元存在的概率”。通过计算从 (一个生物 + 环境) 的初始状态到 (两个相同生物 + 相容环境) 的最终状态的量子力学转换概率, 他得出结论: 因为要满足的方程数量大于未知数的数量, 所以复制的概率为 0。由于事实是复制存在, 如果这个论证正确, 只能说明量子理论是错误的。

Wilbraham, H. (1854), 'On the theory of chances developed in Professor Boole's "Laws of Thought"', *Phil. Mag. Series 4*, **7**, (48), 465-476.

对布尔版本概率论的批评。

Williams, P. M. (1980), 'Bayesian conditionalisation and the principle of minimum information', *Brit. J. Phil. Sci.* **31**, 131-144.

Wilson, A. G. (1970), *Entropy and Urban Modeling*, Pion Limited, London.

Wold, H. (1954), *Stationary Time Series*, Almquist and Wiksell, Stockholm.

Yockey, H. P. (1992), *Information Theory in Molecular Biology*, Cambridge University Press.

Zabell, S. L. (1982), 'W. E. Johnson's sufficientness postulate', *Ann. Stat.* **10**, 1091-1099.

在杰恩斯的论文 (Jaynes, 1986b) 中有讨论。

Zabell, S. L. (1988), 'Buffon, Price, and Laplace: scientific attribution in the 18'th century', *Arch. Hist. Exact Sci.* **39**, 173-181.

Zellner, A. (1984), *Basic Issues in Econometrics*, University of Chicago Press.

近期讨论和说明科学推断重要原则的 17 篇文章的重印合集。与之前的引用文献一样, 这对更广泛的读者来说是有价值的, 而不是仅对标题中所提到领域中的人有价值。这些问题和例子是在经济学的背景下陈述的, 但是这些原则本身具有普遍的有效性和重要性。在我们看来, 它们对物理学、生物学、医学和环境政策的重要性甚至比对经济学更高。请务必阅读标题为“因果关系和计量经济学”的第 1.4 章。确定是否存在因果作用的问题对于物理学至关重要, 人们可能期望物理学家能对其进行最佳分析。然而, 泽尔纳在这里给出了比物理学文献或任何其他“硬”科学都要复杂得多的处理方法。他提出了与我们在本书中强调的相同的观点, 并通过令人信服的例子说明了为什么先验信息在任何判断中都是绝对必要的。

Zubarev, D. N. (1974), *Nonequilibrium Statistical Thermodynamics*, Plenum Publishing Corp., New York.

中译本 Д. Н. 祖巴列夫. 非平衡统计热力学. 李沅柏, 郑哲洙, 译. 北京: 高等教育出版社, 1982.

一部了不起的作品. 作者开发了几乎所有的最大熵分拆函数算法作为特定工具, 但是随后拒绝了给出其理由以及解释其为什么工作的最大熵原理! 结果, 他只愿意将形式方法用于它能够解决的一小部分问题, 从而几乎失去了该方法的所有真正价值. 这个例子很好地证明了, 通过对概率进行正统观念化, 即使有所有必要的数学知识, 有用的应用程序可以如何变得无力.

译后记

“你可知道我前四分之三的努力完全是为了最后四分之一。”

——托马斯·哈代，《德伯家的苔丝》

一位物理学家的概率观

20 世纪 70 年代，一位物理学家参加了主要由计算机与航空航天行业的统计学家参与的可靠性与质量控制研讨会。在会上，他做了题为“置信区间与贝叶斯区间”的主题演讲。他举了几个简单、常见的显著性检验问题，并对使用正统的置信区间方法和贝叶斯方法的解答进行对比，得出的结论是：对于每一个问题，正统的置信区间方法给出的答案显然违背普通人的直觉，而贝叶斯方法得出的结论则合情合理。这让台下炸开了锅。“下去吧！”很多人喊道，“这完全是胡说八道！像置信区间这种理论牢靠且有效的方法怎么可能有这种表现？你简直是在诽谤伟人！奈曼绝对不可能提出在如此简单的问题上失效的理论！你如果连简单的算术都不会，就没有必要跑到这里来做这样的演讲！”

于是，演讲者被轰下了台。在气氛稍微缓和之后，他再次上台，向大家一步一步地展示得出结论的数学计算过程。在场的人都斜视着他，企图第一个找到他计算过程中的错误。整个检查过程持续了整整 4 小时，但是没有人能发现其中有任何错误，因为问题和计算都很简单，他得出的结果显然是对的。于是，很多人的反应变成了：“我的天啊！为什么原来没有人告诉我这个呢？我的教授和教科书从来没有提到这一点！看来我得回去重新检查一下许多人所做的工作了！”

由于得不到普遍接受正统统计学教育的杂志编辑的认可，这个会议报告的内容直到十年后才得以正式发表，而这位演讲者正是本书的作者，物理学家埃德温·汤普森·杰恩斯。他既没有获得过诺贝尔物理学奖，甚至也不是美国科学院院士。如果不做考证，我们大多数人可能也不知道他的博士论文导师是著名物理学家尤金·维格纳（1963 年诺贝尔物理学奖得主）。他生前的主要身份只是华盛顿大学圣路易斯分校和斯坦福大学教授，所以他在 20 世纪 70 年代上台演讲时也不可能因为带着名人或头衔光环而受到任何优待。他在物理学上的主要贡献是 1957 年提出热动力学的最大熵原理，以及 1963 年提出量子光学的杰恩斯-卡明斯（Jaynes-Cummings）模型。但是正如作者所言，他对于理论物理学问题的兴趣只是短暂的

细枝末节,最长久的兴趣还是在概率论上,并且在该领域中进行了长达 40 年的持续探索与思考.这种持续探索与思考的结晶就是这本遗著《概率论沉思录》(英文原名是 *Probability Theory: The Logic of Science*,直译为《概率论:科学的逻辑》),其主要思想是将概率论视为传统亚里士多德逻辑的扩展.在这种思想框架下,布尔逻辑只是概率逻辑的一种特殊情形.众所周知,传统数学是以演绎逻辑为基础的,而概率论却可以作为科学推断(归纳逻辑 + 演绎逻辑)的理论基础.这样,作为扩展逻辑的概率论就是可以融合归纳推理与演绎推理的统一理论.

有人可能会问:且慢,概率论不是一门数学学科吗?它与科学有何相关?的确,概率论的公理化是 20 世纪数学方面最重要的进展之一.现代概率论的诞生以柯尔莫哥洛夫 1933 年的奠基性著作《概率论基础》(*Foundations of the Theory of Probability*)的发表为标志.同时,现代统计学也在 20 世纪中如火如荼地发展,其代表人物是皮尔逊、费希尔、奈曼等.概率统计可以说是机器学习、人工智能最重要的基础之一.甚至可以说,概率论以及统计学对现代社会的影响无论是在思想层次还是现实层次,绝不亚于 20 世纪中出现的物理学的相对论与量子力学.

希尔伯特在 1900 年提出了 23 个待解决的数学问题,其中的第 6 个问题是用数学的公理化方法推演出全部物理原理,包括概率和力学.概率论的公理化问题在 1933 年由苏联数学家柯尔莫哥洛夫解决.从此,概率论成为一门数学学科.但是到目前为止,物理学包含力学的公理化工作并未完成,而且很多人相信这可能永远无法完成,即使完成,也没有什么实际意义.毕竟物理学与数学本质上还是存在很大差别的:物理学家使用数学作为工具,却又清楚地知道物理理论的正确与否不在于逻辑上是否自洽或成立,而在于是否与实际相符.

值得注意的是,概率论公理化之前经过了三个多世纪的发展,如本书作者所说,主要是以数学物理学家(丹尼尔·伯努利、拉普拉斯、高斯、勒让德、泊松等)为主体进行的.物理学家(哈罗德·杰弗里斯是个例外)基本上没有参与 20 世纪初概率论及现代统计学的蓬勃发展进程:皮尔逊、费希尔有生物学背景,奈曼主要是数学家.在作者看来,这主要是由于当时的物理学家都忙于相对论与量子力学的研究,等到他们感到在这两个领域已经没有太多重要的事情可做,将目光转向统计学时,却发觉统计学已经被正统统计学家所主宰,但是内部仍然派别林立,对很多问题没有统一的认识.这在物理学家看来是非常难以接受的:如果一门所谓科学的理论在内部都缺乏统一性,我们又该如何相信它的真理性呢?

哈罗德·杰弗里斯在 1939 年完成的《概率论》(*Theory of Probability*)是贝叶斯概率论的经典之作.当时,贝叶斯思想还少有人知,遑论得到欣赏和认同,而正统统计如日中天,该书出版后在很多年之内在主流学术界影响甚微,甚至是受

强烈批评的对象。“贝叶斯主义者”最初几乎是正统统计学家用来嘲讽另一派统计同行的侮辱性用词。正统统计学家对于贝叶斯主义者最主要的批评是他们使用的“先验”具有主观性，缺乏基本的“科学客观性”。作为杰弗里斯思想继承者的杰恩斯，对杰弗里斯的辩护和费希尔的批评散见于本书很多章节中。贝叶斯统计阵营内也有几个不同的流派（其主要差别见后文），杰恩斯可以说是客观贝叶斯派的代表人物。在客观贝叶斯主义者看来，“先验”主要与信息相关，不同人掌握的信息可能不同，这很正常。这些不同的“先验信息”是客观存在的，没有任何主观的成份。而且先验信息无处不在，这对于物理学家来说几乎是显而易见的。费希尔有句名言是“让数据自己说话（Let the data speak for themselves）”，但是数据不会自己说话。在面对数据时，物理学家掌握的先验信息至少还有已知的物理定律，他们会同时利用已知理论模型和数据进行分析和推断，而不是仅仅依靠数据。杰恩斯认为，有生物学背景的统计学家之所以不认同先验，是因为他们主要应用统计方法于生物统计问题上，而生物学中本来就没有什么说得出口的理论。

有时被当成数学物理学家的杰恩斯当然对于数学家没有什么成见，正如书中所说明的，对于柯尔莫哥洛夫的概率论公理系统，作者其实更多的是认同而非反对。但是他认为，如果采取本书的框架，概率论的应用范围可以比柯尔莫哥洛夫的概率论更加广泛。现代社会中的很多人对数学家有着莫名的崇拜，普遍认为数学家是这个世界上最聪明的人。但是，物理学家一般不属于对数学家顶礼膜拜的群体，因为有追求的物理系学生所学的数学教材通常与数学系学生所学的教材是同等难度的，而且他们也会深刻理解科学与数学之间的区别。因此，物理学家一般不会仰视数学家，而只会平视他们。杰恩斯不认为使用集合论和测度论来重构概率论有多大的实际意义。他对于基于测度论的概率论的批评主要是由于现代测度论引入了实无穷的概念和理论，不谨慎的使用会带来很多悖论。引入测度论的确增加了概率论的“数学严格性”，但是这不能增进我们对于概率本身的理解，也通常无助于实际问题的解决。我希望读者能通过本书掌握概率论在数学之外的另一种视角：物理学或者说是科学的视角。按照这一视角，现在一般被称为“高等概率论”的内容严格来说应该称为“数学概率论”或者“基于测度论的概率论”才合适。我之所以强调这一点，是因为发觉有些朋友说想要深入掌握概率论而准备学习“高等概率论”。如果理解“高等概率论”其实就是“基于测度论的概率论”，那么在读完初等概率论教材后，为了深入理解概率论的思想以及实际科学推断的需要，其实更应该花时间研读这本《概率论沉思录》。

虽然贝叶斯统计现在没有像 20 世纪那样受到普遍歧视，但是国内的绝大多数数理统计学教材主要还是在传统频率派统计的框架下介绍统计学，最多在简单

介绍一下贝叶斯统计后增加“贝叶斯派和传统派的争论仍将长期存在”“先验分布的客观性常引起争议”“实际贝叶斯方法还有很多困难”等评价，甚至还会加上一句“贝叶斯统计大体仍处于弱势地位”。但是实际情况并非如此。贝叶斯统计经过几代贝叶斯主义者艰苦不懈的努力，到目前为止已经为自己争夺了数理统计学的至少半壁江山。国际流行的概率论教材《概率导论》^①（迪米特里·伯特瑟卡斯、约翰·齐齐克利斯著）中有两章介绍统计推断：一章是贝叶斯统计推断，另一章是经典统计推断。该书第1章则一开始就指出，作为信念程度的概率解释有时与频率解释同样必要且有用。著名的机器学习教材 MLaPP（全名 *Machine Learning: A Probabilistic Perspective*，凯文·墨菲著）同样用贝叶斯统计和频率派统计两章来介绍统计推断，甚至是以贝叶斯视角来概括整个机器学习，只是为了避免意识形态的争论，而选择了“概率视角”（probabilistic perspective）而非“贝叶斯视角”（bayesian perspective）的用词。另一本经典书籍 PRML（全名 *Pattern Recognition and Machine Learning*，克里斯托弗·毕晓普著）也被誉为贝叶斯机器学习的圣经，因为它不仅向我们展示了一切都可以用贝叶斯解释的信仰，并且对于几乎所有重要经典机器学习算法都描述了对应贝叶斯版本（贝叶斯线性回归、贝叶斯逻辑回归、贝叶斯神经网络等）。可见在国际主流概率统计与机器学习界，贝叶斯统计学并不是还处在“弱势地位”这么简单，而是大家越来越意识到其重要性，并且会作为与频率派统计并列甚至更重要的地位来做介绍。杰弗里斯的《概率论》被称为现代贝叶斯统计的奠基之作，而这本《概率论沉思录》则是对杰弗里斯概率论的直接继承和发展。希望深入理解贝叶斯统计的读者自然不应该忽视这两本经典图书。

我在本译后记中将补充一些自己对与本书相关问题的说明与理解，希望能在翻译之外提供一些额外的价值。^②

概率论公理与可列可加性

现代公理化的概率论教科书一般通过如下方式引入概率[比如已经更新到第9版的谢尔登·罗斯的《概率论基础教程》(*A First Course in Probability*)]. 首先引入以下三个概率公理.

(1) 公理 1 (非负性): $0 \leq P(E) \leq 1$.

(2) 公理 2 (规范性): $P(S) = 1$.

(3) 公理 3 (可列可加性): $P(\bigcup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} P(E_i)$, 所有命题 E_i 两两相斥.

^① 《概率导论》(第2版·修订版)》中文版已由人民邮电出版社出版(ituring.cn/book/3018)。——编者注

^② 本书的其它阅读辅助材料可参见 github.com/HairenLiao/jaynes-probability-cn.

这样,有限可加性就变成了一个定理(令后面无限项为空集,每一项概率为0即可).绝大多数教科书理所当然地这样做,并且不做任何解释,仿佛这一切都是“显然”的!我不知道一般读者是否觉得这些公理都很容易理解.这也让人想起 R. 柯朗和 H. 罗宾在《什么是数学》(*What Is Mathematics*)中对序列极限概念的评价:“这是序列极限概念的一个抽象描述.初次遇到它时暂时不理解是不足为怪的.遗憾的是某些课本的作者故弄玄虚,他们不做充分的准备,而只是把这个定义直接向读者列出,好像做些解释就有损于数学家的身份似的.”概率论中对于可列可加性(又称完全可加性)的不加解释类似于微积分教材中对于极限概念的不加解释.可列可加性这一公理其实并非不言自明,初次接触该公理有疑惑才是正常的表现.

在这方面描述得比较清晰的是钟开莱教授的《初等概率论》(*Elementary Probability Theory*),其中定义概率测度如下.

样本空间 S 上的概率测度是 S 子集上满足以下三个公理的函数.

- (i) 对于任何集合 $A \subset S$, 函数值非负: $P(A) \geq 0$.
- (ii) 对于任意两个不相交集合 A 和 B , 其并集 $A \cup B$ 的函数值等于 A 和 B 分别的函数值之和.

$$P(A \cup B) = P(A) + P(B), \text{ 如果 } A \cap B = \emptyset.$$

- (iii) S (作为子集) 的函数值等于 1.

$$P(S) = 1.$$

接着,他将有限可加性看作公理 (ii) 的一个自然扩展:

- (iv) 对于任何有限数量的不相交集合 A_1, \dots, A_n , 我们有

$$P(A_1 \cup \dots \cup A_n) = P(A_1) + \dots + P(A_n). \quad (1)$$

所以有限可加性可以代替 (ii) 成为三大公理之一. 然后,他将可列可加性看作公理 (ii) 或 (iv) 的一个扩展:

- (ii*) 对于可数的无限个不相交集合 A_k ($k = 1, 2, \dots$), 我们有

$$P\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} P(A_k). \quad (2)$$

这样,公理 (iv) 可以看作公理 (ii*) 的一个特殊情况. 但钟开莱教授强调的是:可列可加性公式 (2) 不能通过对有限可加性公式 (1) 取 $n \rightarrow +\infty$ 推导得到,并在后面做了详细的解释. 要点是从有限可加性到可列可加性的推导过程中需要允许极限与概率可交换,但是这种可交换性并不是无条件成立的!

如果觉得钟开莱教授的描述不够权威,那么我们可以看看公理化概率论的代表人物柯尔莫哥洛夫在《概率论基础》中的描述.

设 E 是元素的集合,我们称之为基本事件, S 是 E 的子集的集合. S 的元素称为随机事件.

- I. S 是集合的域.
- II. S 包含集合 E .
- III. 对 S 中的每个集合 A 分配一个非负实数 $P(A)$. 这个数 $P(A)$ 称为事件 A 的概率.
- IV. $P(E) = 1$.
- V. 如果 A 和 B 没有共同元素,那么 $P(A \cup B) = P(A) + P(B)$.

满足公理 I~V 的集合系统 S 和 $P(A)$ 的明确赋值一起称为概率域.

可见,他已经开始在有限可加性的基础上定义概率.而可列可加性是在第 2 章“无限概率域”中通过一个额外的公理 VI (连续性公理) 引入的.

VI. 对于 S 中事件的递减序列 $A_1 \supset A_2 \supset \cdots \supset A_n \supset \cdots$, 其中 $\bigcap_{n=1}^{\infty} A_n = \emptyset$, 以下等式成立:

$$\lim_{n \rightarrow +\infty} P(A_n) = 0.$$

对于这一补充公理,柯尔莫哥洛夫说明道:

对于无限域,连续性公理 VI 被证明是独立于公理 I~V 的.由于新公理只对无限概率域至关重要,因此几乎不可能如 1.2 节阐述公理 I~V 那样阐明其经验意义.因为,在描述任何可观察的随机过程时,我们只能得到有限的概率域.无限的概率域只能作为真实随机过程的理想化模型出现.我们可以任意地将自己限定在那些满足公理 VI 的模型上.我们发现这在各种不同的研究中是方便有效的.

可见他对于公理 VI 的非经验性有非常清晰的认识.公理 VI 只是作为一种方便的假设引入的,并没有经验意义.

关于这一点,日本数学家伊藤清在其《伊藤清概率论》^①中也有描述,他虽然在完全可加性的基础上定义概率,但是也明确说明:“仅依靠形式的推理是不能导出完全可加性的,将概率的完全可加性作为基础来假设,是数学上的理想化模式,你渐渐地便能理解这种理想化不是与实际相悖的,而是与其一致的.”

① 《伊藤清概率论(修订版)》中文版已由人民邮电出版社出版(ituring.cn/book/2835). ——编者注

本书在三大合情条件的基础上推演出整个概率理论. 在这种概率理论中, 数学化公理基本都变成了推论或者方便的约定. 比如按照本书的推导过程, 规范性公理 $P(S) = 1$ 就是一种方便的约定. 若不做这种约定, 概率论的加法与乘法规则的形式就会变得更加复杂, 但是无关本质. 非负性公理 $0 \leq P(E) \leq 1$ 也可以替换为 $1 \leq P(E) < +\infty$, 以此构建另一种形式的概率理论. 本书只是在有限可加性的基础上定义与使用概率, 并且认为这已经可以满足实际应用的需求. 作者也不是完全排斥可列可加性, 只有当存在明确定义且行为良好的极限时为了方便才使用可列可加性. 值得说明的是, 柯尔莫哥洛夫概率系统是自治却不完备的. 自治性在于存在现实的对象, 它们的确满足这些公理, 说明这些公理之间是和谐的. 不完备性在于即使对于同一集合而言, 其中对象的概率也可以通过不同方式来满足概率论公理. 这种不完备性正是概率论能够应用于物理世界中各种不同随机事件的基础.

也论无穷大

无穷大有潜无穷与实无穷之分. 无穷的观念自古就有, 但是大家(包括亚里士多德、高斯、马克思等)普遍接受的是潜无穷, 而不是实无穷. 近代著名数学家康托尔和他的门徒建立了实无穷的数学理论. 康托尔创造的实无穷得到了一批著名数学家(如戴德金、魏尔斯特拉斯和希尔伯特等)的认同和赞许. 希尔伯特甚至有一句相关的名言: “没有人能够把我们从康托尔创造的乐园中驱赶出去!” 这种理论虽然形式上很迷人, 却从其诞生之日起, 就受到一批同样重量级的数学家(如克罗内克、庞加莱等)的猛烈批评和攻击. 将“实无穷”引入数学仿佛将“上帝”引入数学, 从而导致了数学“确定性的丧失”.

康托尔创造的无穷世界的确足够美好和理想. 比较典型的是“希尔伯特无穷旅店”的例子: 一家无穷旅店里的每个房间里已经住满了人, 这时又来了一个百人旅行团要住店. 店主不会像一般旅店店主那样说“本店已住满, 请往别家”, 而是让原来住第 1 间房的顾客挪到第 101 间, 原来住第 2 间房的顾客挪到第 102 间……这样就能腾出 100 间房让新来的旅行团成员都住下. 无论再来多少旅行团、来多少人, 店主都可以通过类似的方法解决问题. 多么神奇和美妙!

实无穷在物理学中更无地位. 物理学家基本只与有限、实在的经验世界打交道. 唯一相关的问题是宇宙的有限无限问题. 宇宙无限指的是宇宙在空间上是实无穷的. 按照目前标准宇宙模型的一般解释, 现代宇宙学并没有断言宇宙是有限的还是无限的. 只是不少人, 包括爱因斯坦和霍金等都倾向于相信宇宙是有限的. 这是为什么呢? 首先是由于实无穷大的观念不可理解. 如上所述, 实无穷即使在

数学上也是一个很有争议的概念。其次是由于物理学的方法论。物理学是一门实验科学，实际上也是一门经验科学。但是我们的经验没有也不可能接触到实无穷大。即使宇宙真的是实无穷的，目前也没有验证这一事实的可靠方法（即宇宙无限的命题如果为真，那么既不能被证否，也不能被证实；但是宇宙有限如果被证实，则可以否定宇宙无限的说法）。如果连宇宙都极有可能是有限的，那么真实世界上何来实无穷的集合呢？虚构的世界再美好，也不能代替我们在真实世界中的生存；虚拟世界中解决问题的方式也不能随便拿来解决实际问题。

本书只在传统潜无穷的意义上使用无穷大，并且通过很多示例表明不考虑极限过程而直接跳到实无穷概念的危险性。

随机变量的迷雾

在概率论中，“随机变量”概念的引入是很重要的一件事情。不知道读者在初学概率论时是否有过疑惑：什么是随机变量？为什么要引入随机变量？有了样本空间、 σ 代数以及概率测度，难道不是一切都定义清楚了吗？一般的概率论图书对此通常也是直接给出定义，一带而过。概率论大师钟开莱教授没有轻易放过这个重要问题。在《初等概率论》第 4 章中，他花了两节（4.1 节“什么是随机变量？”和 4.2 节“随机变量从何而来？”）的篇幅来对此进行解释。但是这种追求明晰的做法似乎引起了某些读者的不满，觉得作者解释以后反而让自己困惑了。的确，通过阅读这两节，我们原来觉得清晰的样本空间概念可能变得模糊，因为在一些实际问题中，样本空间可能很难描述，有时使用样本图更好，有时则直接使用随机变量更好。

事实上，随机变量的直接定义是样本空间上的实值函数，其引入是为了能够使用数学分析的方法研究概率论。随机变量化也是一个一切皆“量化”的过程，有了这种数值化过程，就可以使用强大的数据分析工具来统一处理了。不过我们也要注意：这种数值化的过程也可能使一些实际问题中的物理量的意义消失。下面，我将使用机器学习中的典型点击概率预估问题来说明这一点。

点击概率预估是计算广告学中的一个基本问题，它指的是预估给定用户 (U) 在一定场景 (C) 下点击某个广告 (A) 的概率，可以用 $P(Y|X = (A, U, C))$ 来表示。一般用 $Y = 1$ 表示曝光后点击，用 $Y = 0$ 表示曝光后不点击。类似的条件概率估计问题在机器学习中的重要性也是不言而喻的。该问题中的样本空间是什么？从数学上讲是乘积空间 $A \times U \times C$ ，也就是三元组（广告，用户，场景）的集合，这没有什么问题。核心是，以上每个方面都可能需要从多个维度来描述，广告至少可以分解为（广告创意，广告主，...），用户至少可以分解为（性别，身高，体重，居住城市，...），

场景至少可以分解为 (手机型号, 广告位所在 APP, 广告位 ID, ...), 所以可以使用 $A = (A_1, A_2, \dots, A_m)$, $U = (U_1, U_2, \dots, U_p)$, $C = (C_1, C_2, \dots, C_q)$ 来分别表示. 这样 X 可以展开成 n 维变量空间 $(A_1, A_2, \dots, A_m, U_1, U_2, \dots, U_p, C_1, C_2, \dots, C_q) = (X_1, X_2, \dots, X_n)$, 其中 $n = m + p + q$. 随机变量化就是将每一维都数值化, 比如用 $U_1 = 1$ 表示性别男, 用 $U_1 = 0$ 表示性别女; 将每一手机型号映射成一个整数 ID; 身高、体重本身就是数值变量, 直接列出其值即可. 一切似乎都很自然, 这样我们就拥有了一个 n 维随机变量 X , 然后预估概率 $P(Y|X)$. 但是随机变量的强大之处在于: 如果我们有两个随机变量 X_1 和 X_2 , 那么 $X_1 + X_2$ 、 $X_1 - X_2$ 、 $X_1 X_2$ 、 X_1/X_2 ($X_2 \neq 0$) 以及任何函数 $f(X_1, X_2)$ 都是随机变量. 概率论图书中关于随机变量的章节可能主要是介绍如何计算各种随机变量函数的分布. 发现其中的问题了吗? 在将以上现实问题概率模型化以后, 类似 (性别 + 身高 + 体重 + 手机型号) 等都是合法的随机变量了! 但是这有什么物理意义吗? 有人肯定会说: 这不是数学本身的问题, 数学模型化本来就是抽象化的过程, 一定会抛弃现实物理上的一些所谓意义. 在数学上可以进行函数操作不代表有物理意义, 最终是否及如何使用还在于实际应用者. 不同物理量的运算也未必没有意义, 比如 Swolf 就是一个衡量游泳效率的合成指标, 由一定距离内的游泳速度与划水次数相加得到. 我用这个例子只是想表明, 实际问题的随机变量化实际上可能做了很大的抽象, 特别是当对分类变量与数值变量都同样进行数值化时, 但在此基础上的概率论本身允许的各种数学运算未必是有物理意义的. 当然, 现代机器学习模型对于具有 R 个不同值的分类变量的处理未必是直接将其转化为一个维度上的 R 个不同数值, 而可能是通过 One-Hot 编码的方式将其转化 R 维向量, 但是这种处理方式可能产生数据稀疏和同一维度数据的借鉴性丧失等问题. 总之, 实际数据随机变量化的方式及如何在建模时保持物理意义非常值得注意. 我希望读者在读完此段后不会认为我将随机变量的概念解释得更加令人费解.

波利亚的合情推理

本书的三大思想来源是波利亚的合情推理、考克斯定理以及杰弗里斯概率论. 波利亚对于合情推理模式的描述体现在其 1954 年出版的《数学与猜想: 合情推理模式》(*Mathematics and Plausible Reasoning: Patterns of Plausible Inferences*) 一书中. 该书对于类比和归纳推理的定性规则有更多的举例和说明. 但是作者在试图将定性形式转化为定量形式、使用概率论来定量描述归纳推理逻辑时, 认为遇到了不可克服的困难: 作者试图估计牛顿定理可靠的概率, 但是在正确预测了一些罕见的事实后, 作者认为其可靠性应该至少提高上万倍! 这样, 在未做预测

时, 牛顿定理可靠的概率就不能超过万分之一. 作者认为这是不可接受的, 因此不可能发展出定量理论. 杰恩斯在本书中对于产生该困难的原因进行了解释: 作者是在做模型比较, 而在做这种比较时明确指明备择模型是至关重要的, 作者实际上计算的是几率而非概率. 在贝叶斯理论中, 谈论一个模型或假设成立的绝对概率是没有意义的, 只有条件概率, 而没有无条件的概率. 将这个疑难解决之后, 波利亚的合情推理定性理论就可以向定量的概率论发展. 值得说明的是, 传统哲学认为推理主要分为演绎推理与归纳推理, 这里的归纳推理严格来说应该改为合情推理才合适. 归纳推理和类比推理在数学发现中都起着很大作用, 但是这两种推理都只是合情推理的特殊情况. 归纳推理是一种从个别到一般的推理, 这种“个别”一般有多个不同实例, 但是合情推理却可以从单一实例中根据合情性推断很多事实: 比如考古学家从考古挖掘发现的某个朝代的一件文物通过合情推理对这一朝代诸多原来尚不确定的问题的答案做出推断.

杰弗里斯概率论

杰弗里斯是著名地球物理学家, 在地球物理学和地震学等方面都做出了突出贡献. 同时, 他也以作为学术权威顽固地攻击魏格纳的大陆漂移(板块构造)假说而著名. 不过, 这也不能掩盖他在统计推断和概率论方面的突出贡献. 他在 20 世纪 30 年代写的《科学推断》和《概率论》至今还在发挥着影响. 他的一位同事曾经感慨: “我写过五本书, 但是没有一本还在印刷. 杰弗里斯的书(指其《概率论》)则在出版 80 年之后还在重印.”

杰弗里斯在《概率论》前言中指出, 该书旨在发展一种根据观测数据进行推断的自洽且实用的方法, 实际上是将概率论作为归纳推理的形式法则来建构整个理论. 他首先提出了建立该理论的一些指导规则.

- (1) 所有假设都必须明确表述, 结论必须从假设中得出.
- (2) 理论必须是自洽的, 也就是说, 从假设和任何一组观测数据中都不能得出矛盾的结论.
- (3) 给出的任何规则都必须适用于实践. 除非被定义的事物在出现时能根据定义被识别出来, 否则定义就是无用的. 某个事物的存在或某个量的估计不应涉及不可能实施的实验.
- (4) 理论必须明确说明它所做的推论有可能是错误的. 定律可能包含可调整的参数, 这些参数可能被错误地估计, 或者定律本身可能在事后被发现需要修改. 事实上, 为了考虑新的信息(相对论和量子论就是明显的例

子), 科学定律经常需要修改, 因此没有确凿的理由认为目前的任何定律都是最终的。但是, 我们确实在同样的意义上接受归纳推理: 我们有一定的信心, 相信它在任何特定情况下都是正确的, 尽管这种信心并没有逻辑上的确定性。

- (5) 理论不得先验地否定任何经验命题: 任何精确表述的经验命题都必须在给定适量相关证据的条件下, 能够在上一条规则的意义上正式予以接受。
- (6) 公设的数量应当尽量减少。
- (7) 虽然我们不认为人类大脑是完美的推理器, 但是必须承认它是有用且唯一可用的推理器。理论不必详细体现实际的思维过程, 但应与之大体一致。
- (8) 鉴于归纳法的复杂性较大, 我们不能指望它可以比演绎法发展得更详尽。因此, 如果对这里所发展的理论的类似反对意见会使普遍接受的纯数学的一部分失效, 我们将认为这种反对意见没有分量。

根据以上规则, 杰弗里斯指出: “这些规则排除了任何用无限可能性的观察集合来定义概率的尝试, 因为我们实际上不可能进行无限次的观察。根据规则 (3), 维恩概率极限、费希尔的无限总体假设、吉布斯的无限集合对我们来说都是无用的。”

在这八条指导规则的基础上, 杰弗里斯提出以下三条约定和七条公理, 并由此推演出所有概率论定理。

约定 1 我们将给定数据中较大的数值分配给可能性较大的命题 (因此, 将相等的数值分配给可能性相等的命题)。

约定 2 如果给定 p , q 和 q' 是相斥的, 那么根据数据 p 分配给 “ q 或 q' ” 的数值就是分配给 q 和 q' 的数值之和。

约定 3 如果 p 蕴涵 q , 则 $P(q|p) = 1$ 。

公理 1 (概率可比较性公理)。给定 p , q 比 r 的可能性大、相等或小的这三个选项中不可能有两个为真。

公理 2 (概率可传递性公理)。如果 p, q, r, s 是四个命题, 给定 p , q 的可能性大于 r , 且 r 的可能性大于 s , 那么给定 p , q 的可能性大于 s 。

公理 3 (与演绎逻辑相容性)。从命题 p 推导出的所有命题在数据 p 上的概率相等, 与命题 p 不一致的所有命题在数据 p 上的概率相等。

公理 4 (加法公理)。如果给定 p , q 和 q' 不同时为真, r 和 r' 不同时

为真, q 和 r 的可能性相同, q' 和 r' 的可能性相同, 那么给定 p , “ q 或 q' ” 和 “ r 或 r' ” 的可能性相同.

公理 5 (与实数对应性公理). 按“更有可能”关系排序的给定数据的概率的集合, 可以按递增顺序与一个实数的集合一一对应.

公理 6 (蕴涵推理公理). 如果 pq 蕴涵 r , 那么 $P(qr|p) = P(q|p)$.

公理 7 (乘法公理). 对于任何命题 p, q, r , 有 $P(qr|p) = P(q|p)P(r|qp)/P(q|qp)$.

考虑到篇幅, 这里只列出杰弗里斯概率论的基本框架, 尽管其公理体系看似比柯尔莫哥洛夫体系复杂得多, 但是可以看出他从一开始就将自己的理论与根据数据进行科学推断以及现实应用联系起来. 强烈建议想深入理解贝叶斯概率和统计理论的读者阅读杰弗里斯的《概率论》全书.

丹尼斯·林德利的概率统计思想

丹尼斯·林德利 (1923—2013) 是英国统计学家、决策理论家以及著名的贝叶斯统计的倡导者. 本书第 18 章开篇引用的就是他的话. 他给世界留下的遗产除了林德利佯谬 (Lindley's paradox)、林德利过程 (Lindley processes) 以及克伦威尔规则 (Cromwell's rule) 等, 还包括 2000 年发表在 *Journal of the Royal Statistical Society: Series D (The Statistician)* 上的“统计哲学” (“The philosophy of statistics”) 一文以及 Wiley 出版社 2006 初版、2014 年修订再版的《理解不确定性》 (*Understanding Uncertainty*) 一书. 这些都是值得阅读的贝叶斯思想方面的重要文献. 这里仅列出其中一些重要的观点.

- (1) 统计学是对不确定性的研究, 统计学家是处理不确定性的专家.
- (2) 对于不确定性的度量必须遵循概率原则, 概率是不确定性的唯一合适的表达.
- (3) 概率表示“信念度” (degree of belief), 它表达的是人与世界的一种关系. 一方面, 它不能离开人; 另一方面, 它也离不开外部世界.
- (4) 概率总是有条件的, 没有无条件的概率, 因为一个人对一件事情的不确定性总是依赖于其知识水平.
- (5) 概率必须遵守以下规则.
 - (a) 规则 1 (凸性). 对于所有 A 和 B , 有 $0 \leq p(A|B) \leq 1$ 和 $p(A|A) = 1$.
 - (b) 规则 2 (加法规则). 如果 A 和 B 是互斥的, 那么给定 C , 有

$$p(A \cup B|C) = p(A|C) + p(B|C).$$

(c) 规则 3 (乘法规则). 对于所有 A, B, C , 有

$$p(AB|C) = p(A|BC)p(B|C).$$

(d) 规则 4 (可聚集性). 如果 $\{B_n\}$ 是 C 的一个划分 (可能是无限划分), 对于所有 n 都有 $p(A|B_nC) = k$, 那么 $p(A|C) = k$.

(6) 科学的统一性在于其方法, 而非其内容. 科学方法在人文方面的应用甚少.

另外, 林德利也曾为 1974 年德菲内蒂出版的两卷本《概率论》(主观贝叶斯概率论的代表著作) 作序, 并在其中预言贝叶斯主义会在 50 年后被广泛接受. 德菲内蒂这两卷本《概率论》的主题, 按照他的说法, 是“概率并不存在”(Probability does not exist). 这当然只是故作惊人之语, 德菲内蒂真正想表明的是, 概率不像距离一样是世界本身的纯客观性质, 而是依赖于主体, 因此概率必然是主观的. 值得注意的是, 林德利认为自己从频率主义者开始, 过渡到客观贝叶斯主义者, 最终成为跟德菲内蒂一样的主观贝叶斯主义者. 他跟本书作者杰恩斯所持的客观贝叶斯主义立场还是有很大差别的.

频率派、客观贝叶斯派、主观贝叶斯派, 究竟谁正确?

到目前为止, 我们看到至少有三种针对概率的观点: 传统频率派, 以本书作者杰恩斯为代表的客观贝叶斯派, 以及以德菲内蒂为代表的主观贝叶斯派. 这里将概率的古典定义和统计定义都归到频率派里, 因为两者都是用两个数的比值 (频率) 来定义概率的. 贝叶斯派其实还包含经验贝叶斯派等, 这里姑且不论. 有一个基本的问题: 如果将概率论视为数学或者科学理论, 那么究竟哪一种观点才是正确的呢?

这个问题的争论在最近的差不多一百年间持续进行, 有些人甚至认为事关“唯心”与“唯物”之争. 比较典型的是曾经影响一代人的莫斯科大学教授格涅坚科在《概率论教程》(丁寿田译, 1956 年人民教育出版社第 1 版) 中所持的观点. 在该书第 1 章第 2 节 (对概率定义的种种见解) 中, 他对“把数学概率当作认识主体的‘信念程度’”这一“主观唯心论”定义进行了严厉的批评, 并且认为作为对个体事件的信念程度的“概率”没有意义, 因此绝不加以研究. 不过值得一提的是, 这本书在 1998 年出版了英文第 6 版, 虽然格涅坚科在其中仍然不认同主观概率定义, 但是意识形态意味已经大大减少. 总之, 频率派对于贝叶斯主义的主要批评是其作为信念度的概率缺乏基本的“科学客观性”.

再来说说主观贝叶斯派和客观贝叶斯派的区别. 两者都将概率看作个人的“信念度”(degree of belief), 但是更确切地说, 客观贝叶斯主义认为概率是个人的

“合理信念度”(degree of reasonable belief). 何谓“合理”? 就是确定其值的过程要满足本书第 1 章提到的用实数表示、与常识定性对应以及一致性这三大合情条件, 最终其推理规则满足概率论的基本加法规则和乘法规则. 所以本书设计了一个理想的“合情推理机器人”来做这种合情推理, 这种机器人在给定先验信息和数据的情况下做出的概率推断是唯一的. 只要给定同样的先验信息和数据, 我们每个人都应该做出与这个合情推理机器人同样的概率推断, 所以这种推断过程是完全“客观的”, 没有任何主观成分. 同时, 客观贝叶斯派认为先验概率分布也不是纯个人的选择, 而是完全由证据确定的. 事实上, 客观贝叶斯主义认为将先验信息唯一地转化为先验概率分布的问题是概率论的重要问题之一. 本书作者认为该问题还没有被完全解决, 本书中探讨的最大熵、变换群等方法都是重要的工具. 而主观贝叶斯派认为先验概率在很大程度上取决于个人, 也允许有不同. 这样, 在给定新数据后, 每个人得到的后验概率也会有不同, 不同的人得到的概率必然是不可比较的. 杰恩斯对于德菲内蒂主观贝叶斯派的第一个批评(见附录 A.2)就是这种纯主观性——这似乎不是在做科学研究, 而是属于心理学的研究范围. 另外, 德菲内蒂的主观贝叶斯理论将概率论建立在需要满足“连贯性”(coherence)的基础上, 而杰恩斯认为概率推理要满足考克斯意义上的“一致性”(consistency)才是最为重要的. 虽然有可能进行连贯但是不一致的推理, 但满足一致性的推理则一定会满足连贯性. 由于概率估计的唯一性, 客观贝叶斯方法估计的概率一般需要做校准(calibration)工作, 因为估计的概率最终还是需要或者说可以用数据来验证的. 比如上面提到的计算广告学中典型的点击概率预估问题, 模型直接给出的概率预估结果未必满足校准性的要求, 一般需要进一步的概率校准工作才是真正客观贝叶斯意义上的概率.

那么, 这三种观点究竟哪种正确? 也许还需要再等几十年, 大家才会对该问题的答案形成基本共识. 我在这里也只能简单谈谈自己目前对这一问题的思考和体悟. 贝叶斯派经过与频率派的百年抗争才争取到现在至少半壁江山的地位, 还是能说明其强大生命力的. 或然事件与不确定性在人类生活中普遍存在. 类似“某件事发生的可能性有多大”的问题在人类思想中是自然且很早就存在的. 这里所谓的“可能性”就蕴含着朴素的概率思想. 当然, 这里的“某件事”的类型有很多种: 可能是重复性随机事件, 比如抛硬币、扔骰子等; 也可能是在集体现象中体现的随机性, 比如放射性核衰变(每个原子核发生衰变的概率很小, 但是有大量几乎等价的原子核, 所以衰变呈现泊松分布的统计规律); 更为普遍的则是所谓的“个体事件”, 即非重复性事件. 明天是否会下雨? 下个月某个城市的房价是否会上涨? 基普乔格如果参加 2024 年奥运会, 继续获得马拉松金牌的概率有多大? 商

朝灭亡的时间具体是哪一年? (考古学家可能需要根据考古证据进行推断。) 对于这些事件, 无论我们做出怎样的概率判断, 都不太可能通过结果或实验去验证具体概率值的正确与否。但是若说对于这些问题, 概率论坚决不能应用和讨论, 那么概率论的应用范围也就太狭隘了。甚至可以说, 人在社会生活中面临最多的是各种“个体事件”, 而重复性随机事件和集体现象只是其中的特殊情况。频率派将概率论限制在重复性随机事件和集体现象的讨论和研究上显然有些画地为牢的意味。贝叶斯派则完全没有这种限制, 无论是主观贝叶斯派还是客观贝叶斯派都是如此。本书将概率论视为扩展逻辑的理论, 即合情推理或归纳推理的形式化理论, 还有哪种事件不能被其研究呢? 频率派完全根据数据进行的推断 ($p(\theta|D)$) 可以看作贝叶斯推断 ($p(\theta|D, I)$) 在没有什么先验信息时的特例。

这当然不是说频率派概率论完全没有独立的意义。一方面, 这种理论对于重复性随机事件和集体现象的研究具有很大的指导意义, 得出的结论也很客观、可靠。另一方面, 我们对于概率的赋值, 无论是主观贝叶斯派还是客观贝叶斯派, 都是在与频率派一致的意义上理解的。假如某人预测基普乔格在 2024 年奥运会马拉松比赛中获得金牌的概率是 0.99, 尽管 2024 年奥运会马拉松比赛只有一次, 但我们知道他说这句话的意义是: 假定比赛可以重复进行 100 次, 基普乔格会赢 99 次。客观贝叶斯派的概率验证和校准都需要根据这一原则。值得说明的是, 对于某些事件, 概率的预测虽然严格来说是个体事件的概率预测, 但是由于会做多次预测, 因此大致可以作为集体现象进行验证。比如天气预报和上面提到的计算广告学中的点击概率的预估: 虽然每个地方、每天的天气都是唯一的, 不同用户、场景、广告的组合都是独特的, 单次预测的准确与否很难验证, 但是可以通过某个集合上预估的平均概率与真实统计频率的比较来确定模型预测是否存在系统偏差。

那么又当如何看待客观贝叶斯派与主观贝叶斯派之间的分歧呢? 这似乎可以联系到中国思想史上的孟子荀子之辩, 或者朱子王阳明之辩。哲学上认为孟子、王阳明是主观唯心主义哲学家, 荀子、朱子是客观唯心主义哲学家, 大致可以将孟子、王阳明视为主观贝叶斯主义者, 将荀子、朱子视为客观贝叶斯主义者。对于相同的问题, 他们有不同的看法, 那么究竟谁对谁错呢?

客观贝叶斯派将不同人概率评估的不同归结为掌握信息的不同, 这种信息可能是数据, 也可能是先验信息。本书作者在 5.3 节讨论了意见分歧与趋同的问题以及在客观贝叶斯框架下如何解释它们。但是我们看到, 在对一些问题的分歧似乎不完全能用掌握的信息不同来解释。也许可以在客观贝叶斯推断 ($p(\theta|D, I)$) 的基础上再加上一层代表文化价值观或公理系统的先验的推断 ($p(\theta|D, I, V)$, 其中 V 代表文化价值观或者公理系统等)。如果底层的文化价值观或者公理系统不

同, 那么即使接受同样的信息, 对于世界的判断也会不同. 如此看来, 主观贝叶斯派似乎更好些, 因为这一流派的理论可以讨论一切事件和问题. 的确, 我们也不可能阻止任何人对任何事件发生的“概率”进行评估. 但是科学意义上的研究和讨论似乎需要终止于客观贝叶斯主义哲学, 因为纯主观的看法虽然并非一定没有意义, 但是恐怕很难讨论出个是非曲直出来. 我们也可以说, 主观贝叶斯派和客观贝叶斯派的侧重点不太一样. 人文方面的理论是偏主观的, 而科学的理论则是偏客观的.

概率论与因果推断

这里再来简单说说朱迪亚·珀尔 (Judea Pearl) 的因果推断理论. 因果关系的数学化当然是科学界划时代的大事, 珀尔也因此获得了 2011 年的图灵奖. 不过需要注意的, 珀尔的因果推断工作是在其概率论工作的基础上完成的, 他意识到基于规则系统的人工智能的局限, 于是在 20 世纪 80 年代提出了基于概率推理的贝叶斯网络理论, 用概率知识表达认知推理, 其成果体现在 1988 年出版的《智能系统中的概率推理》(*Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*) 一书中. 这是采用概率方法来研究人工智能的里程碑式著作. 珀尔在此基础上又花了十几年进一步发展了因果推断理论.

珀尔认为因果关系的学习者必须熟练掌握至少三个层级的认知能力: 观察能力 (seeing)、行动能力 (doing) 与想象能力 (imaging), 分别对应着相关、干预和反事实. 第一层级是观察能力, 是指发现环境中的规律的能力, 要求我们基于被动观察做出预测. 这一层级的基础工具是概率统计, 特别是条件概率计算与变量之间的相关性分析. 这些年来广受推崇的深度学习算法 (深度神经网络) 仍然处在因果关系之梯的最低层. 第二层级是行动能力, 涉及预测对环境进行干预后的结果, 并根据预测结果选择行动方案. 第三层级是想象能力, 涉及根据观察到的实际数据进行反事实推理, 从而推断一个无法被观察到的反事实世界里将会发生的事情. 这其中离不开基础性的解释因果过程的理论模型.

对于频率派和贝叶斯派之争, 珀尔无疑是坚定地站在了贝叶斯一方. 他认为正统统计学对于因果关系的漠视和无能的深层原因在于它对于客观数据优先于主观认识的过分坚持. 无论是贝叶斯网络理论还是因果推断理论都离不开“主观”构建的贝叶斯网络和因果图.

因果推断相对概率论当然有新的内容: 比如 do 算子, 反事实推理等. 学术界对于“因果关系不能简化为概率”的认识来之不易. 束缚在概率论的框架里, 试图通过概率公式 $P(Y|X) > P(Y)$ 来推断因果关系实际上是不成立的, 因为 X 与

Y 之间可能存在共同的因, 即混杂因子 (confounder). 只有通过 do 算子, 看到 $P(Y|\text{do}(X)) > P(Y)$, 我们才可以说 X 导致了 Y . 也就是说, $P(Y|\text{do}(X))$ 一般不等于 $P(Y|X)$, 这是因果推断理论的基本认识.

但是我们也不应该觉得有了因果推断理论, 概率论就过时了. 作为扩展逻辑的概率推断实际上是比较因果推断更一般、通用的推理方法. 杰恩斯作为物理学家, 自然不会否定因果关系的重要性. 他在本书中也多次谈到因果关系, 同时也多次警告不要将因果依赖性与逻辑依赖性混淆. 因果关系是有时间箭头的, 只能从因到果, 而且其间必须有真实的物理作用; 而逻辑推断则既可以从因到果, 也可以从果到因, 即可以从后来发生的事情去推断前面发生的事情, 不管其中有无因果关系.

本书的影响

能评估一本专业图书价值的未必是这本书的知名度或者销量, 主要还得看它的专业影响力, 特别是对领域内思考最深入的专家是否有影响. 我在写下这段文字时查到《概率论沉思录》英文版在 Google Scholar 上的引用量是 8953, 对照起来, 德菲内蒂的两卷本《概率论》的引用量是 4538, 而威廉·费勒的《概率论及其应用》^①的引用量是 61972. 鉴于费勒的书第 1 卷初版出版于 1950 年, 最终的第三版也早已在 1970 年出版, 杰恩斯的这本《概率论沉思录》到目前为止的引用量还是非常高的. 资深计算机技术和概率统计专家凯文·范霍恩 (Kevin S. Van Horn) 认为本书是 20 世纪最重要的概率论著作之一, 并在其个人主页上为本书英文版写了详尽的非正式勘误表, 并提供了一些很有价值的评论和补充材料, 非常值得在阅读本书时参考. 机器学习专家凯文·墨菲在著名的机器学习教材 MLaPP 第 2 章“概率论”中推荐了三本优秀教材, 首先就是这本《概率论沉思录》, 并且在多处引用了本书的观点和结论. 伊恩·古德费洛、约书亚·本吉奥和亚伦·库维尔所著号称 AI 圣经的《深度学习》第 3 章“概率论与信息论”推荐扩展阅读的概率图书也是这本《概率论沉思录》, 此外还提到了珀尔的《智能系统的概率推理》. 1974 年菲尔茨奖得主大卫·曼福德 (David Mumford) 曾在 2000 年发表过一篇文章 “The Dawning of the Age of Stochasticity”, 主要思想是: 亚里士多德的逻辑学两千多年来一直统治着西方知识分子的思维, 所有精确理论、科学模型, 甚至思维过程本身的模型, 原则上都受逻辑的约束; 但是现在, 概率论和统计推断已经成为科学模型, 尤其是思维过程模型的更好的基础, 也是

^① 《概率论及其应用 (卷 1·第 3 版)》和《概率论及其应用 (卷 2·第 2 版)》中文版已由人民邮电出版社出版 (ituring.cn/book/2793 和 ituring.cn/book/2790). ——编者注

理论数学的重要组成部分，甚至是数学本身的基础。这种视角的转变将在下一个世纪对几乎所有数学理论产生影响。曼福德在这篇文章中也多次提到杰恩斯的思想，并加以引用与讨论。值得说明的是，曼福德引用的这本书还是杰恩斯放在网上的版本。杰恩斯于 1998 年去世，这本书的英文版 5 年后才由他的学生及同事拉里·布雷特霍斯特（G. Larry Bretthorst）正式编辑出版。

概率观世界

概率论在这个时代的重要性无论再怎么强调都不为过。现在所谓的人工智能革命都可以说是概率革命的延续。前面已经提到，20 世纪概率论的数学化及概率统计后续发展的现实重要性绝不亚于相对论与量子力学的出现。按照本书的阐述，概率论将归纳推理形式化和定量化，使得我们拥有了一个强大而重要的推理工具。概率论作为数学是独特的，因为它虽然也像其他数学理论一样通过演绎推理来发展，却可以被看作在描述归纳推理的过程。作为其应用的统计学一般会被认为是自然科学而非纯数学（至少对于应用统计而言是如此）。因此可以说，这本《概率论沉思录》既是一本数学书，也是一本科学哲学书，还可以被看作一本逻辑学书，甚至一本生活智慧书（可以看作《易经》的现代科学版：两者都教我们如何面对这个充满不确定性、易变的世界。古人凡事不决做占卜，我们凡事不决算概率）。一般的科学或数学书对于个人的日常生活其实是没有多大实际作用的，更多的是作为科学研究或技术应用的辅助。概率论却不相同，它可以帮助我们更好地认识这个世界并且更好地生活。本书对保险的原理有很好的解释，对于如何认识特异功能也有讨论。如果理解其中的思想，相信读者对该不该买保险、该不该炒股、如何看待超自然现象等都会有自己更好的判断。

从概率的视角看世界也会产生不同的认知。比如，从概率论的角度来说，我们只能以概率的方式认识世界，这种认识不可避免地带有不确定性。所谓科学理论，只是解释世界的一种模型而已。公认的科学理论只是目前科学共同体认为正确概率最高的模型。模型总是有局限性和适用范围的，往往只是近似描述了真实世界某个方面的特征。非科学的模型未必就没有真理存在。从某种程度上来说，我们现在掌握的科学知识是由基本确定的知识、不太确定的知识以及高级迷信组成的。以物理学为例，牛顿力学已经成为基础科学，确定性很大；但是现代宇宙学的不确定性却很大，主流的宇宙大爆炸理论未必可靠，而其中所谓的平行宇宙理论在我看来是一种高级迷信（我不想为此做太多的解释和辩论，可以说这是我的主观贝叶斯主义判断）。

翻译因缘

查看购买记录,我应该是 2009 年 8 月 15 日在海淀图书城的九章书店(很遗憾,该书店现已关闭)看到了人民邮电出版社出版的本书英文影印版,看了一下前言觉得备受吸引,于是就买了下来。这也是我沉迷物理大约十年左右,却最终转行做数据挖掘工作的第二年。其后断断续续看了前言和前几章,虽然还是觉得很吸引自己,但也没有连续看完。正如我买的大多数新书一样,这本书在我的书架上又沉寂了大概四五年。直到 2015 年左右,我在一家 DSP 广告公司做算法优化工作时,才下定决心将其看完。在 DSP 广告算法优化过程中,CTR/CVR 预估是最基础的任务之一,核心就是一种概率预估。记得在将近两年时间里,几乎每个周六,我都会花一上午左右的时间来看这本书。北宋程颐谈到读《论语》时说:“有读了全然无事者,有读了后其中得一两句喜者,有读了之后好之者,有读了后直不知手之舞之足之蹈之者。”在读这本书时,我的确好几次体会到这种近乎手舞足蹈的感觉。什么是好书呢?就是当思考某个问题却又百思不得其解时,你发现书的作者在讨论同样的问题,并将你带到更深处;或者当你觉得已经对某个问题理解得非常清晰,没有任何疑义时,作者的讨论让你完全颠覆了原来的认知。比如我在做转化率预估时,犹疑是直接做 CTCVR(曝光转化率)预估还是分别预估 CTR(点击率)和 CVR(点击转化率)再相乘。此时我恰巧读到本书 5.6 节中对 $(C \rightarrow B \rightarrow A)$ 和 $(C \rightarrow A)$ 两种推断模型的比较:这不就是曝光点击转化和曝光转化吗?而在读到第 21 章对稳健性数据分析的讨论时,我完全颠覆了具有稳健性总是更好的认知。可以说,阅读这本书改变了我对概率论、统计学、数据挖掘、机器学习、人工智能甚至整个世界的认知,也激发了我对概率论和贝叶斯统计的热情。在随后几年里,我将本来准备深入学习机器学习的时间几乎全部用来重新深入学习概率论了。

2019 年,由于知道我有购买英文原版书的习惯,在人民邮电出版社工作的朋友问我有没有人工智能方面值得引进的好书推荐。那时我对珀尔的因果推断理论热情较高,除林德利的《理解不确定性》外,主要推荐了珀尔的三本书^①,另外问了他杰恩斯的这本《概率论沉思录》是否有人翻译。他问了出版社后说还没有,因为这本书实在是太厚了,似乎没有人愿意翻译。我本来对于翻译科技书这种事是没有什么热情的,觉得人生精力有限,能够阅读英文文献本应该是现代科研人

^① 这三本书中的《为什么:关于因果关系的新科学》与《因果论:模型、推理和推断》目前已经有中译本。但是《智能系统中的概率推理》似乎还没有受到应有的重视。这三本书中最容易读的是《为什么》,因为它几乎可以说是一本关于因果论的科普书。但是《因果论》作为专业书就比较难啃了。在我看来,先读完《智能系统中的概率推理》再去研读《因果论》可能才是循序渐进的方式。

员的基本素养，花费那么多精力来翻译一本书还不如去读几本新书。后来，我觉得要求所有人能够自由阅读英文原著可能有点苛求于人，再者这本书的英文影印版在国内已经不能再版，国内读者除高价购买英文原版书（原来的影印版二手书也被炒到了三四百元一本）外，基本没有了获取纸质书的渠道。因此，我觉得这种经典书籍的翻译还是值得的，不完全是为人，也是为己——翻译的过程肯定也是译者多次研读而有更多收获的过程。除此之外也有点宿命的感觉：自己原来一直痴迷物理并学习了那么多年，却最终没有以物理研究为业，总有点“无颜面对江东父老”之感，所以对于物理学家杰恩斯的这本遗著，我的喜爱之情难以言表，而且自视为作者的知音。也许我是命中注定要来翻译本书吧，这也算是对自己多年的物理学学习与研究有所交代。于是，我主动请缨来翻译本书，出版社也给予了支持，图灵公司当时的总经理武卫东先生还特意通过微信与我确认了翻译意向。翻译过程虽然辛苦，但也的确充满了愉悦，因为这种经典书籍几乎每读一遍都会有新的收获。

致谢

首先感谢本书特约编辑江志强老师、黄志斌老师，图灵公司的编辑杨琳老师、张子尧老师和岳新欣老师。虽然我作为译者需要为译文的总体质量负责，但是他们的工作也使得其中的错漏大大减少。另外，带评注的引用文献和参考文献也是本书的重要组成部分，译文也在翻译和编辑时列出，提供了额外的价值。

在翻译过程中，我曾经就某些词的翻译与经济金融领域的学者刘庆先生、陈义博士以及冯佳奇先生进行了讨论，在此也向他们表示感谢。本书中 *desiderata* 一词的翻译曾经多次修改，比如“理想条件”“必备条件”“预期条件”等，都不太满意，最后参考了刘庆先生的建议定为“合情条件”。我的妻子王军丽女士作为曾经的专业校对就我翻译的初稿做了基础的校对工作。感谢朋友贺瑞君先生为本书的翻译出版牵线搭桥。另外，我也想在此感谢初中的英语老师郑咏申老师。初中毕业时她送我的那套纯英文《新概念英语》一直陪伴着我后面的英语学习。正是这套书，让我从高中阶段就养成了阅读英文原版书的习惯。

由于译者的学识和水平有限，译文难免有不妥之处，欢迎广大读者批评指正。

廖海仁

2023年10月31日于江西宝峰

人名索引

A

阿巴思诺特, John Arbuthnot, 129, 130, 645
阿贝尔, Niels Henrik Abel, 26, 353, 354, 421
阿达马, Jacques Salomon Hadamard, 524
阿基米德, Archimedes, 42, 626, 646
阿喀琉斯, Achilles, 119
阿米蒂奇, Peter Armitage, 157
埃尔米特, Charles Hermite, 222-223, 629, 673
埃里亚斯, Peter Elias, 408
埃利斯, Robert Leslie Ellis, 296, 649, 652, 673
埃姆斯, 小, Adelbert Ames Jr., 125
爱德华兹, Anthony William Fairbank Edwards, xv, 237, 315, 658, 682
爱德华兹, Harold. M. Edwards, 625, 626
爱因斯坦, Albert Einstein, 125, 128, 132, 139, 190, 193, 215, 218, 221, 269, 338, 345, 416, 465, 472, 574, 587, 646, 647, 692, 707
安德鲁斯, D. F. Andrews, 224
安妮女王, Queen Anne, 645
安斯科姆, Francis John Anscombe, 287, 645
奥采尔, János D. Aczél, 26, 31, 629
奥尔金, Ingram Olkin, 243
奥卡姆的威廉, William of Ockham, 563,

564, 566-569, 571-574

奥斯勒, Sir William Osler, 516, 646

B

巴尔, Avron Barr, 677
巴克, B. Buck, xi
巴拿赫, Stefan Banach, 5, 696
巴纳德, George Alfred Barnard, 187, 195, 198, 203, 236, 238, 646
巴斯德, Louis Pasteur, 293, 471, 473, 574
柏辽兹, Hector Louis Berlioz, 8, 11
保罗, Paul, 252
鲍尔弗, Arthur Balfour, 462
贝多芬, Ludwig van Beethoven, 8, 11, 59
贝尔, Eric Temple Bell, 625, 638, 646, 649, 661, 667, 690
贝尔曼, R. Bellman, 372, 542
贝尔纳多, José M. Bernardo, 114, 683
贝塞尔, Friedrich Wilhelm Bessel, 673
贝特朗, Joseph Louis François Bertrand, 184, 185, 358, 360, 364, 365, 454, 647, 648
贝特曼, Frederick Bateman, 114, 308, 665
贝叶斯, Thomas Bayes, ix, x, 2, 107, 111, 118, 119, 121, 122, 124-126, 128, 131-133, 135, 138, 140, 141, 156, 196, 230, 232, 233, 241, 247, 271, 275, 276, 280, 281, 286, 292, 293, 302, 318, 321, 342, 346, 355, 358, 360, 368, 377, 380, 383,

- 385, 389, 395, 401, 403, 405,
427, 429, 439, 447, 450, 455,
458, 463, 465, 467-470, 472,
474, 476, 477, 482, 486, 492,
494, 498, 499, 501, 503-506,
508, 509, 514-516, 518-521,
525, 527, 528, 534, 541,
554-556, 561, 563-565,
567-569, 576, 579, 580, 582,
584-586, 612, 614, 618, 619,
635, 650, 651, 654, 656-658,
661, 662, 665, 666, 671,
673-675, 679, 680, 682, 683,
685, 687, 689, 691, 693-697,
701-704, 710, 712-716, 718
- 本福德, Frank Albert Benford, 669, 677
- 本吉奥, Yoshua Bengio, 717
- 本森, P. George Benson, 213, 665
- 彼得, Peter, 252
- 比奥, Jean-Baptiste Biot, 118
- 比恩, William Bennett Bean, 516
- 比耶尔斯, Ambrose Bierce, 368
- 毕达哥拉斯, Pythagoras of Samos, 480
- 毕晓普, Christopher M. Bishop, 704
- 波利亚, George Pólya, vi-viii, xi, xv, 5,
35, 127, 130, 131, 138, 228,
253, 293, 460, 646, 650, 668,
675, 709, 710
- 波普尔, Karl Raimund Popper, 58, 261,
292, 293, 307, 467, 657, 661,
668, 676, 692
- 玻尔, Niels Henrik David Bohr, 58, 308,
345, 680
- 玻尔兹曼, Ludwig Eduard Boltzmann,
196, 215, 227, 296, 410, 574
- 玻色, Satyendranath Bose, 269, 416
- 伯恩鲍姆, Allan Birnbaum, 237, 238
- 伯恩斯, Arthur Burns, 473
- 伯格, John Parker Burg, 688, 693
- 伯杰, James O. Berger, 237, 377, 379,
385, 390
- 伯努利, Daniel Bernoulli, vii, 130, 131,
192, 197, 220, 296, 302,
369-371, 374, 376, 387, 458,
702
- 伯努利, James Bernoulli, vi, 40, 41, 49,
56, 64, 69, 70, 78, 107, 113,
116, 136, 145, 154, 178, 192,
267, 275, 276, 278, 296, 461
- 伯特瑟卡斯, Dimitri P. Bertsekas, 704
- 博恩, Max Born, 687
- 博尔特凯维奇, Ladislaus von
Bortkiewicz, 167
- 博克斯, George Edward Pelham Box,
286, 295-297, 648
- 博克斯, Joan Fisher Box, 461, 648
- 博雷尔, Félix Edouard Justin Émile
Borel, 105, 359, 364, 435, 455,
588, 613, 637, 648, 663, 678
- 博林, Edwin Garrigues Boring, 114
- 布尔, George Boole, 8, 9, 47, 60, 61, 63,
235, 296, 646, 649, 666, 699
- 布尔巴基, Nicolas Bourbaki, ix, 421,
634, 653, 655, 663, 667
- 布丰, Georges Louis Leclerc Comte de
Buffon, 359
- 布拉赫, Tycho Brahe, 192
- 布莱克曼, Ralph Beebe Blackman, 473,
490, 647
- 布莱克韦尔, David Harold Blackwell,
233, 330
- 布朗, Robert Hanbury Brown, 195, 679
- 布劳威尔, Luitzen Egbertus Jan
Brouwer, 634
- 布勒丹, Jean-Denis Bredin, 121
- 布雷斯韦尔, Ronald N. Bracewell, 679

布雷特索斯特, G. Larry Bretthorst, v,
xi, xv, 118, 196, 197, 225, 441,
473, 516, 661, 689, 698, 718
布卢姆菲尔德, Peter Bloomfield, 487,
488, 490, 675
布罗德, Charlie Dunbar Broad, 146
布罗斯, Irwin D. J. Bross, 679

C

赤池弘次, Hirotugu Akaike, 250

D

达尔穆瓦, Georges Darmon, 484
达尔文, Charles Galton Darwin, 410,
590, 650, 653, 655, 672, 698
达尔文, Charles Robert Darwin, viii,
125, 126, 216-218, 257, 296,
525, 574
达朗贝尔, Jean Le Rond d'Alembert, 421
戴德金, Julius Wilhelm Richard
Dedekind, 707
戴森, Freeman John Dyson, 210, 524,
628, 665, 681
戴维, Philip A. Dawid, 438
戴维斯, Frances A. Davis, 285, 492
道金斯, Richard Dawkins, 218, 650
德菲内蒂, Bruno de Finetti, viii, 35,
252, 263, 298, 316, 346, 395,
422, 435, 548-550, 581, 583,
612, 614, 615, 651, 685, 713,
714, 717
德格罗特, Morris H. ("Morrie") de
Groot, 203, 233, 390, 651
德雷福斯, Alfred Dreyfus, 121
德摩根, Augustus de Morgan, 187, 195,
198, 202, 203, 205, 296, 645,
652, 655, 657, 662
德维纳茨, Allen Devinatz, 11

狄拉克, Paul Adrien Maurice Dirac, 210,
224, 628
狄利克雷, Johann Peter Gustav Lejeune
Dirichlet, 119, 421, 627, 660
笛卡儿, René Descartes, 269
迪皮尤, David J. Depew, 218
棣莫弗, Abraham de Moivre, 107, 178,
191, 227, 273, 275
蒂奇马什, Edward Charles Titchmarsh,
627, 629, 631, 665, 673
蒂特林顿, Donald Michael Titterton,
224, 689
刁锦寰, George C. Tiao, 286
杜哈明, Jean Marie Constant Duhamel,
628

F

法恩, Terrence L. Fine, 611, 616
法诺, R. M. Fano, 599
范伯格, Stephen E. Feinberg, 461
范丹齐格, David van Dantzig, 670, 673
范霍恩, Kevin S. Van Horn, 717
范里安, Hal R. Varian, 698
范米特, David Van Meter, 396, 408
菲茨, Gary Fitts, 410, 411, 416
菲利克斯, L. Félix, 638, 663
菲耶勒, E. C. Fieller, 678, 683
斐兹杰惹, George Francis FitzGerald,
587
费根鲍姆, Edward A. Feigenbaum, 677
费勒, William Srecko Feller, 66, 84, 89,
187, 195, 198, 199, 203, 252,
253, 275, 302, 303, 309, 374,
433-435, 460, 473, 490, 491,
622, 623, 717
费马, Pierre de Fermat, 296, 695
费希尔, Sir Ronald Aylmer Fisher, 44,
141, 159, 165, 184-186, 199,

- 217, 230-235, 237-242, 263,
275, 285, 296, 297, 375, 382,
384, 452, 459-467, 470, 473,
477, 480, 481, 487, 489-492,
498, 514, 515, 532, 566, 648,
653, 654, 658, 661, 682, 702,
703, 711
- 费希纳, Gustav Theodor Fechner, 88,
91, 682
- 冯·博尔特凯维奇, Ladislaus von
Bortkiewicz, 167
- 冯·亥姆霍兹, Hermann Ludwig
Ferdinand von Helmholtz, 458,
471, 615
- 冯·林德曼, Carl Louis Ferdinand von
Lindemann, 136, 625
- 冯·米泽斯, Richard von Mises, 275,
359, 365, 385, 460
- 冯·诺伊曼, John von Neumann, 6, 21,
587, 674
- 弗拉姆斯蒂德, John Flamsteed, 288
- 弗雷德霍姆, Erik Ivar Fredholm, 450
- 弗雷歇, René Maurice Fréchet, 484
- 弗雷泽, Donald Alexander Stuart
Fraser, 237, 351, 439, 443, 444,
446, 447, 449
- 福尔摩斯, Sherlock Holmes, 59
- 福克, Adriaan Daniël Fokker, 195, 680
- 福克斯, John Leroy Folks, 237, 515
- 福勒, Ralph Howard Fowler, 410, 590,
698
- 傅里叶, Jean Baptiste Joseph Fourier,
118, 208, 209, 221-224, 487,
488, 550, 626, 627, 629, 630,
640, 642, 643, 665, 672
- 富兰克林, Rosalind Franklin, 618, 698
- G**
- 盖尔曼, Murray Gell-Mann, 634
- 盖塞尔, Seymour Geisser, 447, 451, 464,
466
- 高尔顿, Sir Francis Galton, 205,
213-218, 221, 224-226, 294,
459, 667
- 高斯, Johann Carl Friedrich Gauss, ix,
42, 53, 80, 103, 104, 107, 162,
165, 178, 187, 190, 191, 193,
195, 198, 199, 202, 203, 207,
216, 221-223, 226, 227, 229,
232, 234, 268, 296, 340, 403,
406, 420, 421, 424, 434, 435,
454, 458, 462, 469, 473, 479,
489, 495, 509, 515, 555, 556,
558, 562, 571, 578, 585, 623,
624, 626, 643, 647, 651, 661,
671, 702, 707
- 哥白尼, Nicolaus Copernicus, 574
- 哥德尔, Kurt Gödel, 6, 11, 15, 43-45,
79, 326, 392, 393, 654
- 戈埃尔, P. K. Goel, 203
- 戈德曼, S. Goldman, 675, 682
- 戈德斯坦, Herbert Goldstein, 299, 656
- 戈赛特, William Sealey Gossett, 462
- 格宾斯, Keith E. Gubbins, 196
- 格兰特, John Graunt, 287, 657
- 格雷, Christopher G. Gray, 196
- 格利穆尔, Clark N. Glymour, 93
- 格林, George Green, 628
- 格林伍德, Major Greenwood, 290, 657,
658
- 格罗瑟, Morton Grosser, 127
- 格涅坚科, Boris Vladimirovich
Gnedenko, 713
- 古德, Irving John(Jack) Good, 89, 110,
113, 119, 136, 298, 656, 658,
684
- 古德费洛, Ian Goodfellow, 717

古尔德, Stephen Jay Gould, 185, 218,
294, 657

H

哈代, Godfrey Harold Hardy, 628, 629,
632, 665

哈代, Thomas Hardy, 701

哈蒂根, John A. Hartigan, 351

哈尔, Alfréd Haar, 47, 351, 658

哈雷, Edmund Halley, 220, 221,
287-291, 500, 657

哈密顿, A. G. Hamilton, 11, 15

哈密顿, William Rowan Hamilton, 524

哈特利, Ralph Vinton Lyon Hartley,
587, 588

海明威, Ernest Hemingway, 596

海森伯, Werner Karl Heisenberg, 345

亥姆霍兹, Hermann Ludwig Ferdinand
von Helmholtz, 458, 471, 615

汉金斯, Thomas L. Hankins, 684

汉塞尔, Charles Edward Mark Hansel,
119

汉森, William W. Hansen, 183, 408

豪森, Cowlin Howson, 120, 658

豪斯多夫, Felix Hausdorff, 633, 635, 678

赫尔维茨, Adolf Hurwitz, 47

赫奇斯, Larry Vernon Hedges, 243

赫斯, Victor Hess, 472

赫希, Warren M. Hirsch, 374

赫歇尔, Sir Frederick William Herschel,
127

赫歇尔, Sir John Frederick William
Herschel, 189, 190, 193, 494,
647

赫胥黎, Thomas Huxley, 308

赫兹, Heinrich Rudolf Hertz, 196

亨佩尔, Carl Gustav "Peter" Hempel,
136, 454, 456, 656

胡贝尔, Peter Jost Huber, 578, 655

怀特, Anthony J. Whyte, 132

怀特海, Alfred North Whitehead, 111,
125

惠更斯, Christiaan Huygens, 134, 398

惠特克, Edmund Taylor Whittaker, 421

霍尔丹, John Burdon Sanderson
Haldane, 156, 538, 691

霍金, Stephen William Hawking, 707

霍兰, J. D. Holland, 646

霍普夫, Eberhard Frederick Ferdinand
Hopf, 673

霍伊特, William G. Hoyt, 132

J

基普乔格, Eliud Kipchoge, 714, 715

吉布斯, Josiah Willard Gibbs, 218, 264,
273, 274, 279, 296, 312, 313,
339, 345, 374, 416, 419, 587,
635, 656, 664, 669, 675, 680,
711

伽利略, Galileo Galilei, 58, 281, 574,
633, 654, 661

加德纳, Martin Gardner, 186, 655, 659

加勒, Johann Gottfried Galle, 127

杰恩斯, Edwin Thompson Jaynes, v, ix,
9, 110, 126, 209, 218, 251, 274,
275, 281, 308, 326, 345, 348,
411, 439, 441, 445, 450-452,
470, 475, 480, 486, 496, 508,
512, 514, 515, 621, 635,
648-650, 655, 659-661, 663,
669-671, 673, 679, 680, 685,
686, 689, 690, 693, 697-699,
701, 703, 710, 713, 714,
717-720

杰斐逊, Thomas Jefferson, 346, 661

杰弗里, Richard Carl Jeffrey, 132

杰弗里斯, Lady Bertha Swirles Jeffreys, 462, 687

杰弗里斯, Sir Harold Jeffreys, iv, vi-viii, x, xiv, xv, 2, 35, 44, 128, 132, 146, 156, 168, 171-173, 176, 194, 199, 230, 235, 255, 263, 275, 293, 297, 305, 306, 317, 345, 350, 353-355, 357, 367, 393, 407, 419, 441-444, 446, 447, 449-452, 457, 460-467, 472, 484, 490, 495, 515, 528, 530, 532, 566, 574, 632, 635, 653, 654, 657, 658, 661, 662, 675, 676, 680, 687, 690, 691, 694, 702-704, 709-712

杰文斯, William Stanley Jevons, 296, 662

居维叶, Georges Cuvier, 528

K

卡登, Joseph B. Kadane, 422

卡尔达诺, Girolamo Cardano, 296

卡尔纳普, Rudolf Carnap, 263, 355, 534, 536-538

卡克, Mark Kac, 623

卡拉巴, R. Kalaba, 372, 542

卡莱尔, C. J. Carlile, 224, 225

卡罗尔, Lewis Carroll, 632

卡明斯, Fred Cummings, 701

卡奈曼, Daniel Kahneman, 121, 124

卡文迪许, Henry Cavendish, 464

卡西莫多, Quasimodo, 394

开尔文勋爵, William Thomson (Lord Kelvin), 40, 196

开普勒, Johannes Kepler, 471

凯恩斯, John Maynard Keynes, 9, 16, 35, 39, 194, 199, 215, 219, 461, 528, 548, 663, 690, 692

凯撒, Gaius Julius Caesar, 119

凯斯, John Caius, 462

凯特尔, Lambert Adolphe Jacques Quetelet, 221, 224, 225

坎伯兰, William G. Cumberland, 497

坎特里尔, Albert Hadley Cantril, 125

康德, Immanuel Kant, 83

康菲尔德, Jerome Cornfield, 447, 451

康托尔, Georg Ferdinand Ludwig Philipp Cantor, ix, 626, 707

考克斯, David Roxbee Cox, 680

考克斯, Richard Threlkeld Cox, vi-viii, x, xv, 16, 26, 31, 35, 43, 45, 63, 93, 102, 105, 111, 126, 136, 238, 249, 282, 305, 387, 395, 454, 456, 460, 476, 527, 580, 614, 616, 617, 622, 629, 650, 663, 686, 694, 709, 714

柯尔莫哥洛夫, Andrei Nikolaevich Kolmogorov, viii, 47-48, 105, 435, 455, 609-613, 615, 669, 678, 680, 702, 703, 706, 707, 712

柯朗, Richard Courant, 705

柯匹, Irving Marmer Copi, 15, 649

柯西, Augustin-Louis Cauchy, 200, 206, 211, 234, 421, 469, 470, 515, 516, 533, 562, 624

科恩, Carl Cohen, 649

科恩, Paul R. Cohen, 677

克拉默, Carl Harald Cramér, 58, 165, 276, 281, 284, 298, 309, 382, 460, 478, 484, 486, 526, 543, 630, 635, 650

克莱因, Morris Kline, ix, xiv, 420, 625, 634, 635, 638, 653, 663, 667

克莱因, Norbert Klein, 310

克雷格, John Craig, 82, 110

克里克, Francis Crick, 132, 618, 650, 699

克里西, M. A. Creasy, 678, 680, 683
 克伦威尔, Oliver Cromwell, 288, 712
 克罗, Edwin L. Crow, 285, 492
 克罗内克, Leopold Kronecker, ix, 144,
 625, 628, 667, 707
 克洛珀, C. J. Clopper, 543
 肯德尔, Sir Maurice George Kendall, 84,
 359, 375, 376, 460, 484, 515,
 645-647
 肯纳德, Earle Hesse Kennard, 227
 肯普索恩, Oscar Kempthorne, 237, 515,
 685
 库尔贝克, Solomon Kullback, 374
 库尔茨, Paul Kurtz, 119
 库尔诺, August Antoin Cournot, 130,
 296, 652, 673
 库普曼, Bernard Osgood Koopman, 233,
 486, 667
 库维尔, Aaron Courville, 717

L

拉奥, Calyampudi Radhakrishna Rao,
 233, 382, 484, 612
 拉达尼, Shaul P. Ladany, 392
 拉东, Johann Karl Augus Radon, 588,
 637
 拉格朗日, Joseph-Louis Lagrange, 220,
 311, 313, 322, 329, 334, 412,
 486, 590, 601
 拉马克, Jean-Baptiste Lamarck, 217
 拉马努金, Srinivasa Aiyangar
 Ramanujan, 392
 拉姆齐, Frank Plumpton Ramsey, 692
 拉普拉斯, Pierre-Simon Laplace, vi, vii,
 8, 23, 41, 42, 80, 106, 107, 118,
 129-131, 136, 141, 143, 146,
 156, 162, 164, 165, 178, 191,
 193, 198, 199, 207, 216, 220,
 221, 227, 229, 234, 263-265,
 273, 275, 295-297, 302, 305,
 310, 316, 318, 347, 350, 355,
 358, 370, 374, 376, 387, 389,
 407, 419, 445, 458, 460-462,
 469, 471, 528-535, 538-542,
 544-548, 550, 552, 554, 558,
 574, 581, 583, 618, 619, 646,
 649, 652, 662, 666, 668, 673,
 688, 702
 拉瓦锡, Antoine-Laurent de Lavoisier,
 646
 莱布勒, Richard Arthur "Dick" Leibler,
 374
 莱布尼茨, Gottfried Wilhelm Leibniz,
 625
 莱恩, David Lane, 464
 莱曼, Erich Lehmann, 477
 莱特希尔, Sir Michael James Lighthill,
 210, 628, 665, 672
 莱文森, Norman Levinson, 675
 赖米, R. A. Raimi, 677
 兰伯特, Johann Heinrich Lambert, 220
 兰登, Vernon D. Landon, 193, 406
 兰开斯特, Henry Oliver Lancaster, 286
 勒贝格, Henri Léon Lebesgue, 42, 106,
 224, 588, 623, 629, 630
 勒让德, Adrien-Marie Legendre, 162,
 226, 296, 334, 335, 339, 458,
 550, 702
 勒韦里耶, Urbain Jean Joseph Le
 Verrier, 127, 130, 131
 雷法, Howard Raiffa, 377, 378
 雷舍尔, Nichola Rescher, 4
 黎曼, Georg Friedrich Bernhard
 Riemann, 224, 351
 李, Y. W. Lee, 675
 里德, Constance Reid, 464
 利特尔, Roderick J. A. Little, 499, 665

林德利, Dennis Victor Lindley, 114, 447,
518, 525, 680, 694, 712, 713,
719
林德曼, Carl Louis Ferdinand von
Lindemann, 136, 625
林肯, Abraham Lincoln, 619
林奇, Dorothy Maud Wrinch, 146, 675
刘维尔, Joseph Liouville, 218, 312
卢卡西维茨, Jan Lukasiewicz, 3, 665
卢斯, Robert Duncan Luce, 377, 378
卢斯特德, Lee Lusted, 516, 646, 665
鲁宾, Donald Bruce Rubin, 499, 665
伦珀, Gerhard Rempe, 310
伦琴, Wilhelm Konrad Röntgen, 6
罗, David E. Rowe, 638
罗宾, Herbert Ellis Robbins, 705
罗伯逊, Bernard Robertson, 138
罗德里格斯, Carlos Rodriguez, xv, 222
罗森克兰茨, Roger D. Rosenkrantz, 35,
565
罗思曼, T. Rothman, 655
罗斯, Edward John Routh, 299
罗斯, Sheldon M. Ross, 704
罗素, Bertrand Arthur William Russell,
111, 525, 633
罗亚尔, Richard M. Royall, 497
洛厄尔, Percival Lowell, 132
洛夫克拉夫特, Howard Phillips
Lovecraft, 186
洛伦兹, Hendrik Antoon Lorentz, 190,
587
吕埃勒, David Pierre Ruelle, 112, 670

M

马尔可夫, Andrei Andreyevich Markov,
74, 76-78, 603, 605
马赫, Ernst Waldfried Josef Wenzel
Mach, 587
马科夫, U. E. Makov, 224, 689
马克思, Karl Heinrich Marx, 707
马克斯菲尔德, Margaret W. Maxfield,
285, 492
马克威克, Betty Markwick, 664
马洛斯, C. L. Mallows, 224
马乔尔, Robert E. Machol, 392
麦考利, V. A. Macaulay, xi
麦克多诺, James T. McDonough, 596
麦克拉夫, James T. McClave, 213, 665
麦克利里, John McCleary, 638
麦克斯韦, James Clerk Maxwell, 2, 6,
189, 190, 195, 227, 296, 366,
462, 465, 614, 647, 658, 669
麦克塔格特, John Ellis McTaggart, 632
曼福德, David Bryant Mumford, 717,
718
梅根塔勒, Ottmar Mergenthaler, 595
孟德尔, Gregor Johann Mendel, 132,
574, 653
米德尔顿, David Middleton, 396, 408,
477, 631
米勒, David W. Miller, 292
米切纳, James Michener, 596
密斯纳, Charles W. Misner, 139
米泽斯, Richard von Mises, 275, 359,
365, 385, 460
密立根, Robert Andrews Millikan, 552
摩根斯坦, Oskar Morgenstern, 674
摩西, Lincoln E. Moses, 348, 371, 377,
389-391, 410, 480, 694
墨菲, Edward Aloysius Murphy Jr., 192
墨菲, Kevin P. Murphy, 704, 717
莫兰, Patrick Alfred Pierce Moran, 359
莫里森, Donald G. Morrison, 392
莫里斯, C. Morris, 666
莫利纳, Edward Charles Dixon Molina,
646