SIMON NEWCOMB – Astronomer and Statistician

Peter Guttorp, Department of Statistics
University of Washington, Seattle, USA

August 4, 1997

Much of the early development in statistics originated in problems of astronomy. The development of least squares by Gauss and Laplace was because of its importance in astronomy. Problems of outliers, and attendant rejection rules, often were introduced in astronomy. The theory of weighted least squares also originated in attempts to assess variability of different data sets measuring the same quantity (Stigler, 1973). One of the leading astronomers in the United States during the late 19th century was Simon Newcomb, whose contributions to probability and statistics must not be overlooked.

Simon Newcomb was born in northern Nova Scotia, Canada, in 1835. His father was a contry school teacher, and the family moved about a lot. His schooling came mainly from reading books that his father obtained for him. At age 16 he needed to start work, and after a stint as an apprentice to a dishonest quack of a medical doctor, he became a school teacher and tutor. In 1857 he was employed as a computer at the American Ephemeris and Nautical Almanac in Cambridge, Massachusetts. The Almanac office duties were limited, and Newcomb managed to combine his work with studies towards a B. Sc. under Benjamin Pierce at Harvard, which degree he obtained in 1858. In 1861 he was appointed professor of mathematics at the Naval Observatory, and in 1877 he also became Superintendent of the Nautical Almanac. He died in 1909, and left behind an impressive legacy of publications (Archibald, 1924 lists 541 published works).

While Newcomb's main contributions were astronomical (he made detailed calculations of the paths of moon and the planets), from a statistical point of view his main contributions were in applying probability theory to data, and in developing what we now call robust statistical methods.

A sequence of Notes on the Theory of Probabilities (Newcomb, 1859-61) were published in a short-lived mathematical problems journal called Mathematics Monthly, to which Newcomb was a frequent contributor. These Notes constitute a brief textbook in elementary probability theory, which looks quite up-to-date even today. Among the most interesting examples he works out are an application of the Poisson process to assess whether or not the Pleiades could result from a random distribution of stars. In fact, he determines the "probability that, if the stars were scattered at random over the heavens, any small sphere selected at random would contain s stars." For a region of size 1 squared degree, and assuming that there are about 1500 stars of fifth and higher magnitude, Newcomb finds the probability of having six stars in this region about 1.28×10^{-7}. While he does not explicitly describe a Poisson process (neither does Clausius, 1858, who arguably is the first scientist to use this process), his calculation makes it clear that "scattered at random" means independently distributed over disjoint sets.

Another interesting application of probability theory to data was Newcomb's discovery of the logarithmic distribution (Newcomb, 1881) as the distribution of leading digits in haphazardly chosen

numbers.

Perhaps the most important contribution of Newcomb in the statistical area was his approach to dealing with outliers in astronomical data. From looking at substantial amounts of data it was clear to him that a normal distribution of errors was inadequate, since the observed tails of the distribution often were fatter than what the normal distribution would require. For example, in looking at a collection of 684 observations of transits of Mercury (Newcomb, 1882) the non-normality could not be explained even by removing some extreme observation. However, it could be the result of combining data with different precision, and the idea of a mixture of normal distributions would therefore seem close at hand. Thus, the contaminated normal distribution was invented.

In a later paper Newcomb (1886) criticized outlier rejection criteria, and developed a new estimation procedure which weighted "more discordant" observations less heavily. The basic idea is to fit a mixture of mean zero normals to the residuals from the sample mean, and compute a posterior mean with respect to a uniform prior for the normal mixture model.

In fact, in later papers (cf. Stigler, 1973) Newcomb criticized outlier rejection techniques for being discontinuous in the data, in essence developing Tukey's sensitivity function, and proposed Huber's M-estimator as a simple robust estimator.

References:

Archibald, R. C. (1924) Simon Newcomb 1835-1909. Bibliography of his life and work. *Mem. Nat. Acad. Sci.* **17:** 19–69.

Campbell, W. W. (1924): Biographical Memoir Simon Newcomb 1835-1909. *Mem. Nat. Acad. Sci.* **17:** 1–18.

Clausius, R. (1858): Über die mittlere Länge der Wege, welche bei Molekularbewegung gasförmigen Körper von den einzelnen Molecülen zurück gelegt werden, nebst einigen anderen Bemerkungen über die mechanishcen Wärmetheorie. *Annalen der Physik* **105:** 239–258.

Newcomb, S. (1859-61): Notes on the Theory of Probabilities. *Math. Mo.* **1:** 136–139, 233–235, 331–355, 349–350; **2:** 134–140, 272–275; **3:** 68, 119–125, 341–349.

Newcomb, S. (1881): Note on the frequency of use of different digits in natural numbers. *Amer. J. Math.* **4:** 39–40.

Newcomb, S. (1882): Discussion and Results of Observations on Transits of Mercury from 1677 to 1881. *Astr. Papers* **1:** 363–487. Published by the U.S. Nautical Almanac Office.

Newcomb, S. (1886): A generalized theory of the combination of observations so as to obtain the best result. *Amer. J. Math.* **8:** 343–366.

Stigler, S. M. (1973) Simon Newcomb, Percy Daniell, and the History of Robust Estimation 1885-1920. *J. Amer. Statist. Assoc.* **68:** 872–879.