Carta da Presidência

Adriano Polpo
(UFSCar)

Caros leitores, gostaria de dar boas vindas ao Victor Fossaluza (UFSCar) e também agradecê-lo por ter aceito a tarefa de ser o editor deste boletim para o biênio 2013-2014. Ele, em seu primeiro boletim, nos traz um excelente texto de K. V. Mardia e S. B. Cooper sobre o trabalho de Alan Turing. Sendo assim, esperamos muito mais dele para os próximos boletins!

Gostaria também de informá-los que o volume com os anais do EBEB 2012 está disponível em

XI Brazilian Meeting on Bayesian Statistics (2012). Editores: J. M. Stern, M. de S. Lauretto, A. Polpo and M. A. Diniz. *AIP Conference Proceedings*, volume 1490, editora AIP. http://scitation.aip.org/dbt/dbt.jsp?KEY=APCPCS&Volume=1490&Issue=1.

Boa leitura!

Alan Turing and Enigmatic Statistics

Kanti V. Mardia and S. Barry Cooper
School of Mathematics, University of Leeds, Leeds, U.K.
s.b.cooper@leeds.ac.uk

1 Introduction

Enigmatic Alan Turing is known in different ways to different people, like in the story of the elephant and the blind men. Most people have heard of the Turing Test for intelligent machines, but the pure mathematician might be surprised to know that Turing made significant contributions to statistics, while all except the biologists will be surprised to know that Turing's most cited paper deals with the mathematics of emergence of patterns in nature. The 2012 centenary of Alan Turing's birth has seen so many events around the world, with books and papers on his life and work (he even appeared on the cover of Nature), that the Turing legacy is now much better known, at least in academic circles. Below we look briefly at Turing's contribution to statistics, his innovative introduction of Bayesian techniques to cryptography during the 2nd World War – and how the statistics relates to Turing's underlying interest in how the world computes. If the mathematician imagines that Turing knocked off some statistics as a mere ad hoc diversion from the serious business of founding computer science, inventing artificial intelligence and revolutionising developmental biology, she would be missing something basic.

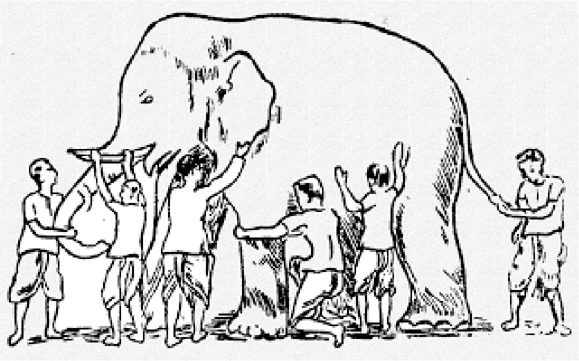

Figure 1: Elephant and the blind men (jainworld.com)

Figure 2: Alan Turing, copyright Beryl Turing

Various important statistical contributions by Turing at Bletchley Park (in 1940-1941) have been recorded by Jack Good (who was main statistical assistant in 1941 to Turing). Good worked with Turing among others in breaking the enigma code; Good (1979) recorded "their" contributions. The article was subsequently elaborated in the commentary to this article by himself in the collected works of Turing (Britton,1992). Recently, a wonderful and readable account has been given in the Book by Mcgrayne (2011) with an up to date history of Bayesian methods. The work has used a combination of several new methods including: 1. Weight of evidence (assigning a tiny non-zero to a rare event which could appear in a larger sample), and 2. Alignment of letters (pairs and triplets of letters in the cipher with substitutions). These are the two main new methods but other methods were Markov Chains, Decision Theory, and Statistical Computing (see, for example, Good, 1992). Since various different statistical methods were used, it will be perhaps right to label these techniques as

"Enigmatic Statistics". The style of developing focussed techniques foreshadowed the style of what is now called Statistical Bioinformatics.

It is now well known that the Enigma was a cryptographic (enciphering) machine used by the German military during WWII. The German navy changed part of the Enigma keys every other day. One of the important cryptanalytic attacks against the naval usage was called Banburismus, a sequential Bayesian procedure (anticipating sequential analysis).

2 Weight of evidence and Empirical Bayes

Suppose that a random sample is drawn from an infinite population of animals of various species, or from a population of words. Let the sample size be N and let n_r distinct species be each represented exactly r times in the sample, so that $\Sigma r n_r = N$, and n_r can be called "the frequency of the frequency r".

It can be shown that an estimate of the total probability of unseen species is

$$n_1/N.$$

The work required for obvious reasons calculating the probability that the next word sampled will be one that has not previously been observed. Turing, using what is called an *urn model* in statistics, showed that the expected population frequency of a species represented r times is about $(r+1)n_{r+1}/(Nn_r)$.

The technique is now known as Good–Turing frequency estimation. For a more exact statement, including the need for smoothing the n_r's, and for numerous elaborations and deductions see Good (1953, 1969) and Good and Toulmin (1956). This work is an example of the empirical Bayes method.

In Banks (1996, p.10, col 2)) Good says "For example, I deduced a simple formula for the probability that the next word sampled will be one that has not previously been observed. Makers of dictionaries and teachers of languages ought to know about this work, because it tells you the minimum size of vocabulary required to cover, say, 98% running text."

That is, this work tells you the minimum size of vocabulary (say 98%) that would cover most of the (number of) words most likely to be used; useful for the makers of Dictionaries.

We quote from Robinson (2011): "Suppose a birder spotted 180 different species, many of which were represented by only one bird. Logically, other species must have been missed. A frequentist statistician would count those unseen species as zero, as if they could never be found. Turing, by contrast, assigned them a tiny non-zero probability, thereby factoring in that rare letter groupings might not be present in his current collection of intercepted messages but could appear in a larger sample."

The Bayesian approach to statistics treats unknown parameters as random variables, and prior distributions model information about parameters. In contrast, the classical approach to statistics has no need of prior distributions as it treats unknown parameters as fixed constants. Empirical Bayes is an approach to statistics that lies somewhere between the two. Unknown hyper-parameters in Empirical Bayes are treated as fixed constants (as are the parameters in the classical approach) but in general these are estimated from data unlike in the standard Bayesian approach. For further details on this theme and related theme of odds and probability, weights of evidence, Bayes theorem, some real and insightful examples, we refer to Aitken(1995) and Efron (2010).

3 Alignment of letters

3.1 Description of the coding in Enigma

Before Banburismus could be started on a given day it was necessary to identify which of nine 'bigram' (or 'digraph') tables was in use on that day. In Turing's approach to this identification he had to estimate the probabilities of certain 'trigraphs'. (These trigraphs were used, as described below, for determining the initial wheel settings of messages.) For estimating the probabilities, Turing invented an important special case of the nonparametric (nonhyperparametric) Empirical Bayes method independently of Herbert Robbins. The technique is the surprising form of Empirical Bayes in which a physical prior is assumed to exist but no approximate functional form is assumed for it.

Robinson (2011): "A crucial example of the application of the theorem was Turing's cracking of the German naval cipher Enigma during the Second World War, which played a key part in the Allied victory in 1945. After the war, Turing's wartime assistant, I. J. 'Jack' Good, wrote about Turing's Bayesian technique for finding pairs and triplets of letters in the cipher." Adding: "To avoid censorship under the UK Official Secrets Act, he described it in terms of bird watching."

Good (1992) has given a stage by stage process in coding by Enigma. Let a real message (a triplet sequence) to be coded.

1. The operator would first choose a triplet, say XQV, as a system discriminator, from a table.
2. Next he would set the three wheels at positions G_1, G_2, G_3, which was part of the daily keys
3. At this initial position of the wheels G_1, G_2, G_3, he would encipher his selection M_1, M_2, M_3 (the setting for the real triplet) and obtain the enciphment LRP, say.
4. The six letters XQV and LRP would be further encrypted by the following procedure which does not use the Enigma.
5. First the six letters would be written one under the other at a stagger as

XQV
-LRP

6. Then two letters would be chosen haphazardly to fill a two by four rectangle as (A and L here)

XQVA
LLRP.

7. Then the four vertical pairs XL, QL, VR and AP would be encrypted with the help of a secret printed pairs table, giving, say,

PTOW
XUBN

8. Finally $PTOWXUBN$ would be the first two groups, the "indicator groups", of the enciphered message.
9. There were ten pairing (digraph) tables and which one was to be used would be part of the daily keys.

Each digraph table was reciprocal; for example, if XL became PX, then PX would become XL. This again was helpful both for the encrypter and the cryptanalyst.

3.2 The significance of alignment of letters

From the description above, it is clear that alignments of letters was a critical step in breaking the code. The idea is somewhat similar to alignment of DNA and protein sequences (see, for example, Durbin (1995). Indeed,the DNA connection has been mentioned in various writings the following: Robinson (2011 "Turing, by contrast, assigned them a tiny non-zero probability, thereby factoring in that rare letter groupings might not be present in his current collection of intercepted messages but could appear in a larger sample. The same technique was later adopted in **DNA sequencing** and by artificial-intelligence analysts." We have inserted BOLD lettering for "sequencing".

Good (1992, p.214) says "The game of Banburismus involved putting together large numbers of pieces of probabilistic information somewhat like the reconstruction of DNA sequences."

4 Turing's Statistics in Context

This work of decoding is a very early successful story of interdisciplinary research. Perhaps this is somewhat different from the early days of the subject, when the statisticians such as Galton and R.A.Fisher played a leading role in interdisciplinary research. It seems that with the floods of large-scale data, computer scientists and statisticians with computing skills have a major part in creating impact." Mardia and Gilks (2005) have named this approach Holistic Approach which is more and more now required.

4.1 Statistics and levels of abstraction

As Turing's mentor Professor Max Newman (1955) observes, Turing was far from being a detached theoretician, describing Turing as "at heart more of an applied than a pure mathematician". Turing's engagement with how the world 'computes' is visceral, with David Leavitt (2006) portraying Turing as *identifying* with his computing machine abstractions.

His respect for the complexity of the world as information emerges in his interest in type theory, which seeks to clarify the way in which mathematical objects can occupy different 'levels of abstraction' — for instance, describing integers as of type 0, reals as of type 1, sets of reals (or geometrical shapes) as type 2, and so on. Clarity about type was exploited by Bertrand Russell to rescue early 20th century mathematics from the paradoxes. While in the real world, statistics provides a fundamentally important route to reducing scientific entities of apparent higher type to data which we can handle computationally. As we know, every computer today is an embodiment of Turing's universal computing abstraction, and these cannot comfortably cope with data above the level of type 1.

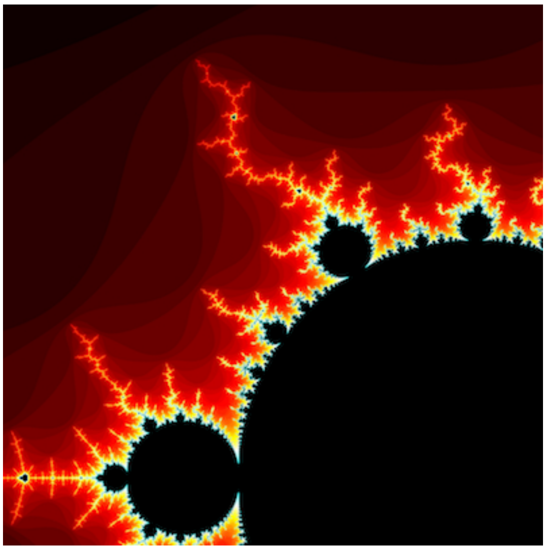

Figure 3: Mandelbrot set, courtesy of Niall Douglas

A typical example of type reduction, with consequent computational accessibility, is provided by the Mandelbrot set. As the set of complex numbers the mathematics provides (a type 2 object) the question of its computability is an open question. But a little mathematics gives us a digital approximation representable digitally on our computer screen — hence the plethora of beautiful images on the web, and that accompanying this article. Of course, the sampling process is not interesting enough to be termed "statistics", any more than the image of our family via a digital camera is. But it does exhibit a level of computability that any sampled reality will have. As we saw from the elephant and the blind men, in general there is an art to sampling and interpretation, that on the one hand reduces complex information to useful data, and on the other delivers a recognisable approximation to truth. For higher type information with very complex structure — for instance chaotic or turbulent contexts, such as weather, or economics, or coded messages, in which emergent non-local phenomena are the objects of interest — the reduction may be fraught with difficulties. The statistics is a challenge and an art, and this is what so engaged the creativity and genius of those working at Bletchley Park in the early 1940s.

4.2 Scaling the informational hierarchy

So what is the link between Turing's most abstract mathematics, his 1939 paper written in Princeton under the guidance of Alonzo Church, and his hands-on practical involvement with real-world complexity of information?

Back in the late 1930s, Turing was puzzled by the fact that Kurt Gödel's Incompleteness Theorem told us that even restricting our attention to the basic theory of the natural numbers — just a part of what we can abstract from the real world — we discover that truth soon passes out of our control. Given any useable theory containing basic arithmetic (one where we can computably recognise the axioms and rules of deduction) one can easily write down a true statement not provable in it. Of course, thought Turing, this means one has an inductive way of computably expanding the theory, so potentially defeating Gödel's theorem. Turing succeeded in carrying out a transfinite induction which did indeed take us into realms unknown. The process was refined and equipped with more power in later years (by Sol Feferman, and Michael Rathjen) to take us to even dizzier proof theoretic heights.

However, a key element in the inductive process was the choice of computable 'fundamental sequences' to take us through limit points of the computable ordinals used to notate the tower of theories. The mathematical difficulties in keeping control of the process are essentially those in evidence in complex real world situations only handleable via statistical sampling. In tune with the subtleties of the statistical route to truth, the fundamental sequence (the logician's counterpart of the statisticians sampling procedure) gives a computable (but by no means computably choosable) route up the informational mountain. And in the mathematics, one needs an *oracle* providing more than computably derivable information to identify the route. On a real mountain, one may need individual brilliance to get to the top — though once a route is identified it is computable, can be shared, and others can subsequently follow it. Turing (1939) contains the famous quotation:

> Mathematical reasoning may be regarded . . . as the exercise of a combination of . . . intuition and ingenuity. . . . In pre-Gödel times it was thought by some that all the intuitive judgements of mathematics could be replaced by a finite number of . . . rules. The necessity for intuition would then be entirely eliminated. In our discussions, however, we have gone to the opposite extreme and eliminated not intuition but ingenuity, and this in spite of the fact that our aim has been in much the same direction.

The mathematician interprets this as an explanation of the mismatch between the subjectively experienced creative process leading to a new theorem, and the axiomatic proof she shares with her colleagues and students. For the statistician, there is a very similar message. The approach to the informational mountain may be full of uncertainty and theoretically rich devices: but success can be shared with others.

For the Enigma operator, the success of the coding process depended on having a computable route up the informational mountain that was incomputable to the less well-equipped observer from afar. What the decoders at Bletchley Park depended on was the human brain having hidden type-traversing resources, at times in the form of statistical wizardry.

5 Morphogenesis, Statistics and Alan Turing's AI

Alan Turing's final years at Manchester were very much taken up with approaches to higher type computability. By then stored program computers, as anticipated by the 1936 Universal Turing Machine, were in operation. This was already giving rise to a powerful paradigm of algorithmic ubiquity and a new digital age. A whole range of ways of bridging the gap between what the by then actual computers could handle, and the informational complexity of social and natural formation, was inhabiting Turing's thoughts in the years running up to June 7, 1954.

The work on morphogenesis (the emergence of form in nature) brought a reassertion of old certainties from logic in an unexpected way. Turing was able to point to the *definability* of a range of natural formations via descriptions (differential equations) based on computable causal relations from the underlying chemistry. This gave explicit descriptions which led to computable solutions and computer generated simulations. Although the mathematics pointed to the likelihood of more complicated instances – possibly differential equations with incomputable solutions – the theory did point to type reduction via approximations built on explicit descriptions.

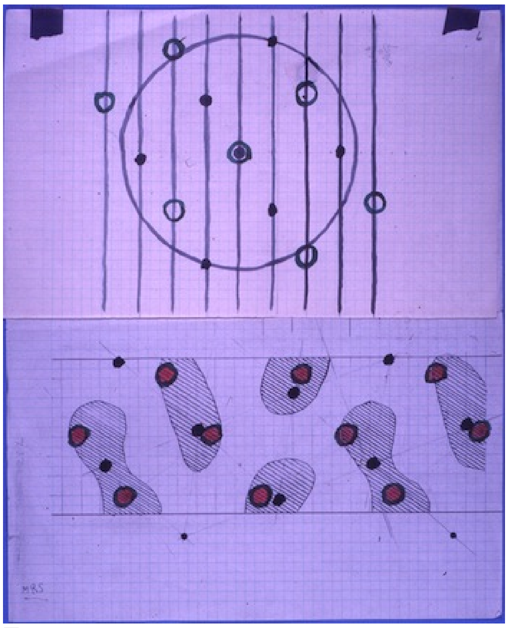

Figure 4: Colored diagram showing patterns of dappling and calculations, made by Turing in connection with work on morphogenesis. Courtesy of P. N. Furbank

His lifelong preoccupation with human thought processes took Turing in a very different direction. The famous Mind paper adheres to the faith in the key role of the by then commercially produced digital computer. But the key role of the human judges in the implementation of the Turing Test for machine intelligence is very significant. This, and the radio broadcasts and the more popular talks and writings, show an acceptance of complementary roles for humans and machines. The role of mistakes, uncertainty, interaction, 'common sense' and, of course, the lessons of 1939 ('intuition') and Bletchley Park (Bayesian methods) — all point to a world in which logic and statistical methods work together, in complementary ways.

Since Turing's passing in 1954, the history of artificial intelligence has tended to confirm this picture. There is more and more a sense that the computer and the human mind work in rather different ways. Today we are more and more aware of the power and limitations of our computational techniques. At the time, the main aim of Turing's 1936 paper was to demonstrate the inadequacy of algorithms and related forms of reasoning. There is a growing sense that human thinking has a much in common with statistical processes as logical ones. This is good news for both humans and statisticians! Maybe digital computers will not supersede brains. But all those mistakes and the 'weather in the brain'? The brain shares with statistics an ability to handle large and complex assemblies of information. It is the algorithmic backbone that is provided by ever growing computer power.

6 References

Aitken, C. G. G. (1995). *Statistics and the Evaluation of Evidence for Forensic Scientists.* John Wiley and Sons.

Banks, D. L. (1996). A Conversation with I. J. Good. *Statistical Science.* **11**, , 1–19.

Cooper, S.B. (2012). Turing's Titanic machine? *Communications of the ACM.* **55** (3), 74–83.

Durbin, R., Eddy, S., Krogh, A. and Mitchison, G. (1998). *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids.* Cambridge University Press.

Efron, B. (2010). *Large-Scale Inference Empirical Bayes Methods for Estimation, Testing, and Prediction.* Cambridge University Press.

Good, I. J. (1950). *Probability and the Weighing of Evidence.* London: Griffin.

Good, I. J. (1953). The population frequencies of species and the estimation of population parameters. *Biometrika* **40**, 237-64.

Good, I. J. (1979). Turing's statistical work in World War II, Studies in the history of probability and statistics. XXXVII. *Biometrika* **66**, 393-396.

Good, I. J. (1988). The Interface between Statistics and Philosophy of Science. *Statistical Science*, **3**, , 386–412.

Good, I. J. (1992). Introductory remarks for the article in Biometrika 66 (1979), 'A. M. Turing's Statistical Work in World War II'. In: *Collected works of A. M. Turing: Pure Mathematics* (ed. J.L. Britton). North-Holland, Amsterdam, London.

Good, I. J. (2000). Turing's anticipation of empirical Bayes in connection with the cryptanalysis of the naval Enigma. *J. Statist. Comput. Simul.* ,**66**, 101–111

Good, I. J. and Toulmin, G. H. (1956). The number of new species, and the increase in population coverage, when a sample is increased. *Biometrika*, **43**, 45-63.

Leavitt, D. (2006). *The Man Who Knew Too Much: Alan Turing and the Invention of the Computer.* W.W. Norton.

Mardia, K.V. and Gilks, W.R. (2005) Meeting the Statistical Needs of 21st-century Science. *Significance,* **2**, 162–165.

McGrayne, S. B. (2011). *The Theory that would not die* . Yale University Press. New Haven and London.

Newman, M.H.A. (1955). Alan Mathison Turing. In *Biographical Memoirs of the Fellows of the Royal Society*, **1**, pp. 253–263.

Robinson, A. (2011). Known unknowns. *Nature* , **475**, 450–451.

Simpson, E. (2010) Bayes at Bletchley Park. *Significance,* **7**, 76–80.

Turing, A.M. (1939), Systems of logic based on ordinals, *Proc. London Math. Soc.* (2) **45**, 161–228. Reprinted in A. M. Turing, *Collected Works: Mathematical Logic* (eds. R.O. Gandy, C.E.M. Yates). North-Holland, Amsterdam, London, pp. 81-148.