

DISTRIBUTIVE NUMBERS: A POST-DEMOGRAPHIC PERSPECTIVE ON PROBABILITY

Adrian Mackenzie

‘We ran the election 66,000 times every night’, said a senior official, describing the computer simulations the campaign ran to figure out Obama’s odds of winning each swing state. ‘And every morning we got the spit-out—here are your chances of winning these states. And that is how we allocated resources’ (Scherer 2012).

IN THE US PRESIDENTIAL ELECTIONS OF NOVEMBER 2012, THE DATA ANALYSIS team supporting the re-election of Barack Obama were said to be running a statistical model of the election 66,000 times every night (Scherer 2012). Their model, relying on polling data, records of past voting behaviour, and many other databases, was guiding tactical decisions about everything from where the presidential candidate would speak, where advertising money would be spent, to the telephone calls that targeted individual citizens (for donations or their vote). Widely reported in television news and internationally in print media (*Time*, *New York Times*, *The Observer*), the outstanding feature of Obama’s re-election seems to me to be the figure of 66,000 nightly model runs. Why so many thousand runs? This question was addressed neither in the media reports nor surprisingly in the online discussion on blogs and other online

forums that followed. A glimmering of an answer appears in more extended accounts of the Obama data analytics efforts (Issenberg 2012) that describe how, in contrast to the much smaller and traditional market research-based targeting of demographic groups used by the Republican campaign for Mitt Romney, the Obama re-election campaign focused on knowing, analysing, and predicting what *individuals* would do in the election. We should note that the Obama data team's efforts are not unique or singular. In very many settings—online gaming, epidemiology, fisheries management, and asthma management (Simpson, Tan, Winn, et al. 2010), similar conjunctions appear. In post-demographic understandings of data, individuals appear not simply as members of a population (although they certainly do that), but themselves as a kind of joint probability distribution at the conjunction of many different numbering practices. If individuals were once collected, grouped, ranked, and trained in populations characterised by disparate attributes (life expectancies, socio-economic variables, educational development, and so on), today we might say that they are distributed across populations of different kinds that intersect through them. Individuals become more like populations or crowds. This chapter seeks to describe, therefore, a shift in what numbers do in their post-demographic modes of existence.

$Pr(A)$: EVENTS AND BELIEFS IN THE WORLD

How can individuals appear as populations? A standard textbook of statistics introduces the idea of probability as event-related numbers in this way:

We will assign a real number $Pr(A)$ to every event A , called the *probability* of A (Wasserman 2003: 3).

Note that this number is 'real' so it can take infinitely many values between 0 and 1. The number concerns 'events', where events are understood as subsets of all the possible outcomes in a given 'sample space' ('the *sample space* Ω is the set of possible outcomes of an experiment. [...] Subsets of Ω are called *Events*'

(Wasserman 2003: 3)). The number assigned to events can be understood in two main ways. Wasserman goes on:

There are many interpretations of $Pr(A)$. The common interpretations are frequencies and degrees of belief. [...] The difference in interpretation will not matter much until we deal with statistical inference. There the differing interpretations lead to two schools of inference: the frequentists and Bayesian schools (Wasserman 2003: 6).

The difference will only matter, suggests Wasserman, in relation to the style of statistical inference. Even apart from these relatively well-known alternative interpretations of probability, the practice of assigning numbers to events in Ω does not, I will suggest, remain stable. If we keep an eye on the machinery that assigns numbers, then we might have a better sense of how events and beliefs themselves might change shape.

Summarising his own account of the emergence of probability, the philosopher and historian Ian Hacking highlights the long-standing interplay of the two common interpretations of probability as frequencies and degrees of belief:

I claimed in *The Emergence of Probability* that our idea of probability is a Janus-faced mid-seventeenth-century mutation in the Renaissance idea of signs. It came into being with a frequency aspect and a degree-of-belief aspect (Hacking 1990: 96).

In this work from 1975, Hacking, writing largely ahead of the marked shifts in probability practice I discuss, claims that there was no probability prior to 1660 (Hacking 1975). As we can verify in the statistics textbooks, there is nothing controversial in Hacking's claim that probability is Janus-faced. Historians of statistics and statisticians themselves regularly describe probability as bifurcated in the same way. Statisticians commonly contrast the frequentist and degree-of-belief, the *aleatory* and the *epistemic*, views of probability. Although the history of statistics shows various distributions and permutations of emphasis on the subjective and objective versions of probability,

statisticians are now relatively happily normalised around a divided view of probability.

Contemporary probability, however, has become entwined with a particular mode of computational machinery—and here we might think of machinery as something like baroque theatre machinery, with its interest in the production of effects and appearances that are never fully naturalised—that deeply convolutes the difference between the epistemic and aleatory faces of probability. Not only is probability a baroque invention, but the fundamental instability that permits recent mutations in probability practice has a distinctively baroque flavour in the way that it combines something happening in the world with something that pertains to subjects. The techniques involved here include the bootstrap (Efron 1975), expectation-maximisation (Dempster, Laird, and Rubin 1977), and Markov-Chain Monte Carlo (Gelfand and Smith 1990). These techniques support increasingly post-demographic treatments of individuals, in which for instance, individuals progressively attract probability distributions, as in Obama's data-intensive re-election campaign.¹ In examining a salient contemporary treatment of probability, my concern is the problem of invention of forms of thought able to critically affirm mutations in probability today. I suggest that these mutations arise in many, perhaps all, contemporary settings where populations, events, individuals, numbers, and calculation are to be found. In such settings, a baroque sense of the enfolding of inside and outside, of belief and events, of approximation and exactitudes offers at least tentative pointers to a different way of describing what is happening as the aleatory and epistemic senses of probability find themselves redistributed.

EXACT MEANS SIMULATED

The contour plot in Fig. 4.1 was generated by the widely used statistical simulation technique called MCMC (Markov Chain Monte Carlo simulation). MCMC has greatly transformed much statistical practices since the early 1990's. The diagram shows the contours of a distribution (a bivariate normal distribution in

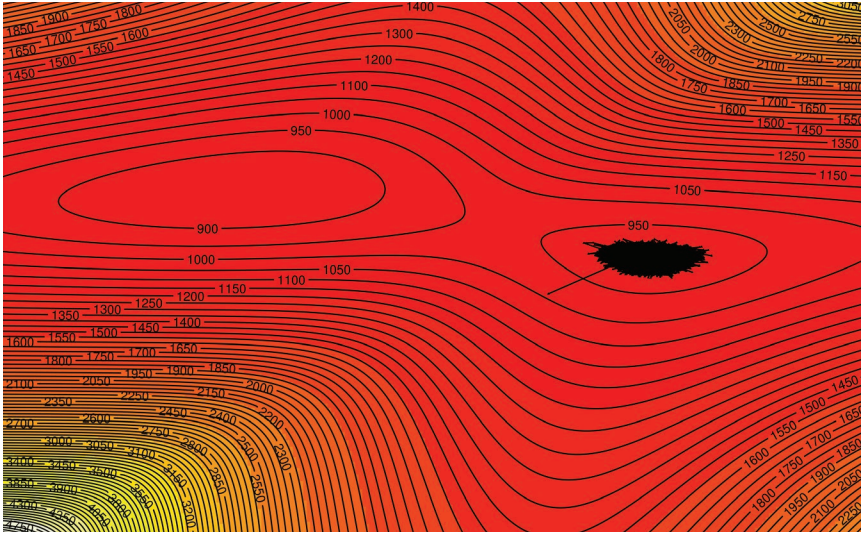


FIG. 4.1 Bivariate Normal Distribution Produced by Gibbs Sampling

this case) generated by MCMC that fits a mixture of two normally-distributed sets of numbers to some data. The topography of this diagram is the product of a simulation of specific kinds of numbers, in this case, the mean values of two normal distributions. The contour lines trace the different values of the means (μ_1, μ_2) of the variables. For the time being, we need to know nothing about what such peaks refer to, apart from the fact they are something to do with probability, with assigning numbers to events. A set of connected points starting on the side of the one of the peaks and clustering on the peak marks the traces of the itinerary of the MCMC algorithm as it explores the topography in search of peaks that represent more likely events or beliefs. Importantly for present purposes, this path comprises 60,000 steps (that is, around the same number mentioned by Obama's data team).

When 'Sampling-Based Approaches to Calculating Marginal Densities', the article that first announced the arrival of MCMC in statistical practice (Robert and Casella 2010: 9) appeared in the *Journal of the American Statistical Association* (Gelfand and Smith 1990), the statisticians Alan Gelfand and Adrian Smith (subsequently Director General of Science and Research in UK government's Department for Innovation, Universities and Skills) stated that

the problem they were addressing was how ‘to obtain numerical estimates of non-analytically available marginal densities of some or all [the collection of random variables] simply by means of simulated samples from available conditional distributions, and without recourse to sophisticated numerical analytic methods’ (1990: 398). Their formulation emphasises the mixture of using things that are accessible—simulated samples—to explore things that are not directly accessible—‘non-analytically available marginal densities [...] of random variables’ (some of this probability terminology will be explored below). For present purposes, the important point is a newly non-analytical probability is in formation here. It lies at some distance from the classical probability calculus first developed in the seventeenth century around games of chance, mortality statistics, and the like.

Note that these statisticians are not announcing the invention of a new technique. They explicitly take up the already existing Gibbs sampler algorithm for image-processing, as described in Geman and Geman (1984), investigate some of its formal properties (convergence), and then set out a number of mainstream statistical problems that could be done differently using MCMC and the Gibbs sampler in particular. They show how MCMC facilitates Bayesian statistical inference—the approach to statistics that shapes basic parameters in the light of previous experience—by reconfiguring six illustrative mainstream statistical examples: multinomial models, hierarchical models, multivariate normal sampling, variance components, and the k-group normal means model. The illustrations in the paper suggest how previously difficult problems of statistical inference can be carried out by sampling simulations. As they state in another paper from the same year, ‘the potential of the methodology is enormous, rendering straightforward the analysis of a number of problems hitherto regarded as intractable’ (Gelfand and Smith 1990: 984).²

Note too that while the MCMC technique has become important in contemporary statistics, and especially in Bayesian statistics (Gelman, Carlin, Stern, et al. 2003), it plays significant roles in applications such as image, speech and audio processing, computer vision, computer graphics, molecular biology and genomics, robotics, decision theory, and information retrieval (Andrieu, De Freitas, Doucet, et al. 2003: 37–38). Usually called an *algorithm*—a series of

precise operations that transform or reshape data—MCMC has been called one of ‘the ten most influential algorithms’ in twentieth-century science and engineering (Andrieu, De Freitas, Doucet, et al. 2003: 5).³ But MCMC is not really an algorithm, or at least, if it is, it is an algorithm subject to substantially different algorithmic implementations (for instance, Metropolis-Hastings and Gibbs Sampler are two popular implementations). In all of these settings, MCMC is a way of simulating a sample of points distributed on a complicated curve or surface (see Fig. 4.1). The MCMC technique addresses the problem of how to explore and map very uneven or folded distributions of numbers. It is a way of navigating areas or volumes whose curves, convolutions, and hidden recesses elude geometrical spaces and perspectival vision. Accounts of MCMC emphasise the ‘high-dimensional’ spaces in which the algorithm works: ‘there are several high-dimensional problems, such as computing the volume of a convex body in d dimensions, for which MCMC simulation is the only known general approach for providing a solution within a reasonable time’ (Andrieu, De Freitas, Doucet, et al. 2003: 5). We might say that MCMC alongside other statistical algorithms such as the bootstrap or EM increasingly facilitates the envisioning of high-dimensional, convoluted data spaces. Simulating the distribution of numbers over folded surfaces, MCMC renders the areas and volumes of folds more amenable to calculation.

What MCMC adds to the world is subtle yet indicative. In their history of the technique, Christian Robert and George Casella, two leading statisticians specialising in MCMC, write that ‘Markov chain Monte Carlo changed our emphasis from “closed form” solutions to algorithms, expanded our impact to solving “real” applied problems and to improving numerical algorithms using statistical ideas, and led us into a world where “exact” now means “simulated”’ (Robert and Casella 2008: 18). This shift from ‘closed form’ solution to algorithms and to a world where ‘exact means simulated’ might be all too easily framed by a post-modern sensibility as another example of the primacy of the simulacra over the original. But here, a baroque sensibility, awake to the convolution of objective-event and subjective-event senses of probability, might allow us to approach MCMC less in terms of a crisis of referentiality and more in terms of the emergence of a new form of distributive number.

How so? The contours of figure 1 define a volume. In its typical usages, the somewhat complicated shape of this volume typically equates to the joint probability of multiple random variables. MCMC, put in terms of the minimal formal textbook definition of probability, is a way of assigning real numbers to events, but according to a mapping shaped by the convoluted volumes created by joint probability distributions. The identification of $Pr(A)$ with a convoluted volume offers great potential to statistics. For instance, political scientists regularly use MCMC in their work because their research terrain—elections, opinions, voting patterns—little resembles the image of events projected by mainstream statistics: independent, identically distributed (‘iid’) events staged in experiments. MCMC allows, as the political scientist Jeff Gill observes, all unknown quantities to be ‘treated probabilistically’ (Gill 2011: 1). We can begin to glimpse why the Obama re-election team might have been running their model 66,000 times each night. In short, MCMC allows, at least in principle, *every* number to be treated as a probability distribution.

1/∞: DISTRIBUTED INDIVIDUALS AS RANDOM VARIABLES

Let us return to the typical problem of the individual voters modelled by the Obama re-election team. Treating every number as a probability distribution involves exteriorising numbers in the service of an interiorising of probability. Techniques of statistical simulation multiply numbers in the world and assign numbers to events, but largely in the service of modifying, limiting, and quantifying uncertainties associated with belief. This folding together of subjective and objective, of epistemic and aleatory senses of probability can be thought as a neo-baroque mode of probability. The baroque sense of probability, especially as articulated by G.W. Leibniz, the ‘first philosopher of probability’ (Hacking 1975: 57), is helpful in holding together these contrapuntal movements. Leibniz’s famously impossible claim that each monad *includes* the whole world is, according to Gilles Deleuze, actually a claim about numbers in variation. Through numbers, understood in a somewhat unorthodox way, monads—the parts of

the world—can include the whole world. Deleuze says ‘for Leibniz, the monad is clearly the most “simple” number, that is, the inverse, reciprocal, harmonic number’ (Deleuze 1993: 129).

Having a world—for the monad is a mode of having a world by including it—as a number entails a very different notion of *having* and a somewhat different notion of number. The symbolic expression of this inclusion is, according to Deleuze: $1/\infty$. The numerator 1 points to the singular individual (remember that for Leibniz, every monad is individual), the denominator, ∞ , suggests a world. The fraction or ratio of 1 to ∞ tends towards a vanishingly small difference (zero), yet one whose division passes through all numbers (the whole world). In what sense is this fraction, in its convergence towards zero, including a world? Deleuze writes that for in the baroque, ‘the painting-window [of Renaissance perspective] is replaced by tabulation, the grid on which lines, numbers and changing characters are inscribed. [...] Leibniz’s monad would be just a such grid’ (1993: 27). This suggests a different notion of the subject, no longer the subject of the world view who sees along straight lines that converge at an infinite distance (the subject as locus of reason, experience, or intentionality), but as ‘the truth of a variation’ (Ibid: 20) played out in numbers and characters tabulated on gridded screens. The monad is a grid of numbers and characters in variation. How could we concretise this? Alongside the individual voters modelled by the Obama re-election team, we might think of border control officers viewing numerical predictions of whether a particular passenger arriving on a flight is likely to present a security risk (Amoore 2009), financial traders viewing changing prices for a currency or financial derivative on their screens (Knorr-Cetina and Bruegger 2002), a genomic researcher deciding whether the alignment scores between two different DNA sequences suggests a phylogenetic relationship, or a player in a large online multi-player game such as World of Warcraft quickly checking the fatigue levels of their character before deciding what to do: these are all typical cases where numbers in long chains of converging variation populate the monadic grid. $\frac{1}{\infty}$ entails a significant shift in the understanding of number. Deleuze writes that ‘the inverse number has special traits: [...] by opposition to the natural number, which is collective, it is individual and distributive’ (1993: 129). If numbers become ‘individual

and distributive’, then the calculations that produce them might be important to map in the specificity of their transformations.

Earlier we saw the flat operational definition of probability as assigning real numbers between 0 and 1 to events. By contrast, a random variable ‘is a mapping that assigns a real number to each outcome’ (Wasserman 2003: 19) but this number can vary. If events have probabilities, random variables comprehend a range of outcomes that are mapped to numbers in the form of probability distributions. The practical reality of random variables is variation; variations that usually take the visual forms of the curves of the probability distributions shown in Fig. 4.2. These distributions each have their own history (see Stigler (1986) for a detailed historical account of key developments), but for our purposes the important points are both historical and philosophical. On the one hand, the historical development of probability distributions, particularly the Gaussian or normal distribution, but also lesser known Chi-square, Beta or hypergeometric distributions, displays powerful inversions in which the mapping of numbers to events becomes a mapping of events to numbers. Hacking, for instance, describes

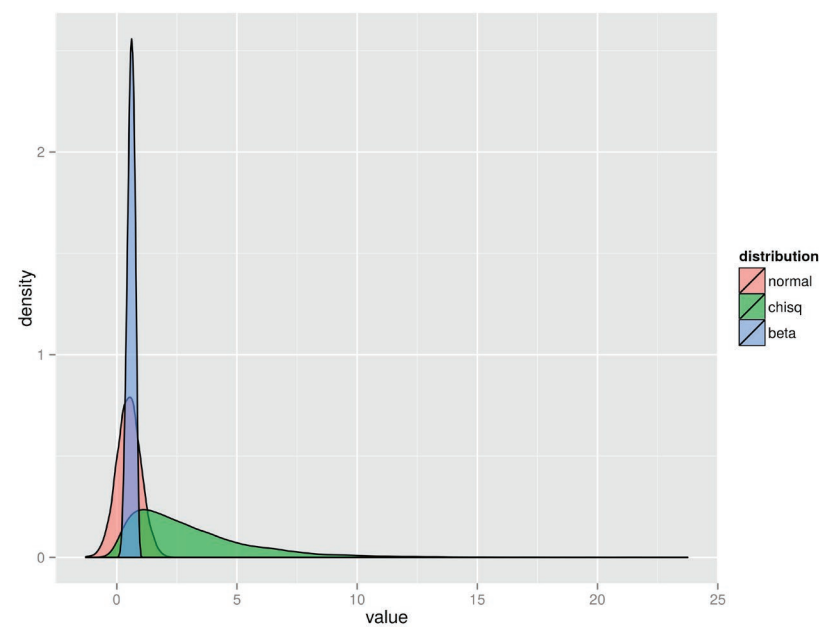


FIG. 4.2 A Variety of Distributions

how the nineteenth-century statistician Adolphe Quetelet began to treat populations. The normal distribution ‘became a reality underneath the phenomena of consciousness’ (Hacking 1990: 205). A whole set of normalisations, often with strongly biopolitical dynamics hinges on this inversion of the relation between numbers and events in nineteenth-century probability practice.

On the other hand, the regularity, symmetry, and above all mathematical expression of these functions in equations, such as the one shown in equation 1, more or less delimited statistical practice. Such expressions offer great tractability since their shape, area, or volume can all be expressed in terms of key tendencies such as μ , the mean and σ , the variance. The eighteenth and nineteenth century development of statistical practice pivots on manipulations that combine or generalise such expressions to an increasing variety of situations. For instance, in figure 2, the normal distribution shown in equation 1 for one variable x becomes a bi-variate normal distribution for two variables x_1 and x_2 . Nevertheless, these equations also limit the range of shapes, areas, and volumes that statistical practice could map onto events. When statisticians speak of ‘fitting a density’ (a probability distribution) to data, they affirm their commitment to the regular forms of probability distributions.

THE ENDLESS FLOW OF RANDOM VARIABLES

Both aspects of this commitment—the curve as underlying reality of events, and the normalised expression of curves in functions whose parameters shape the curve—begin to shift in techniques such as MCMC. In particular, following Deleuze’s discussion of the monad as distributive number, we might say that the probability distributions now function less as the collective form of individuals, and more as the distributive form of individuals across increasingly complex and folded surfaces. We saw above that MCMC inaugurates ‘a world where “exact” now means “simulated”’ (Robert and Casella 2008: 18). This comment links an analytical quality—exactitude—with a calculative, modelling process—simulation. But rather than attesting to the pre-eminence of simulation, we should see techniques such as MCMC as ways of exploring the concavities and convexities,

the surfaces and volumes generated by random variables. Put more statistically, MCMC maps the contoured and folded surfaces that arise as flows of data or random variables come together in one joint probability distribution. These surfaces, generated by the combinations of mathematical functions or probability distributions are not easy to see or explore, except in the exceptional cases where calculus can deliver a deductive analytical ‘closed form’ solution to the problems of integration: finding the area and thereby estimating the distribution function for one variable. By contrast, MCMC effectively simulates some important parts of the surface, and in simulating convoluted volumes, loosens the analytical ties that bind probability to certain well-characterised analytical regular forms such as the normal curve. In this simulation of folded and multiplied probability distributions, the lines between objective and subjective, or aleatory and epistemic probability, begin to shift not towards some total computer simulation of reality but towards a refolding of probability through world and experience. The subjective and the objective undergo an ontological transformation in which calculation lies neither simply on the side of the knowing subject nor inheres in things in the world. These practices perhaps make those boundaries radically convoluted.

Fig. 4.3 shows two plots. The histogram on the left shows the occurrence 10,000 computer generated random numbers between 0 and 1, and as expected, or hoped, they are more or less uniformly distributed between 0 and 1. No single number is much more likely than another. This is simulation of the simplest probability distribution of all, the *uniform* probability distribution in which all events are equally likely. The uniform distribution could be assigned to a random variable. The plot on the right derives from the same set of 10,000 random numbers, but shows a different probability distribution in which events mapped to numbers close to 0 are much more likely than events close to 1. What has happened here? The reshaping of the flow of numbers depends on a very simple multiplication of the simulated uniform distribution by itself:

A real function of a random variable is another random variable. Random variables with a wide variety of distributions can be obtained by transforming a standard uniform random variable $U \approx UNIF(0, 1)$ (Suess and Trumbo 2010: 32).

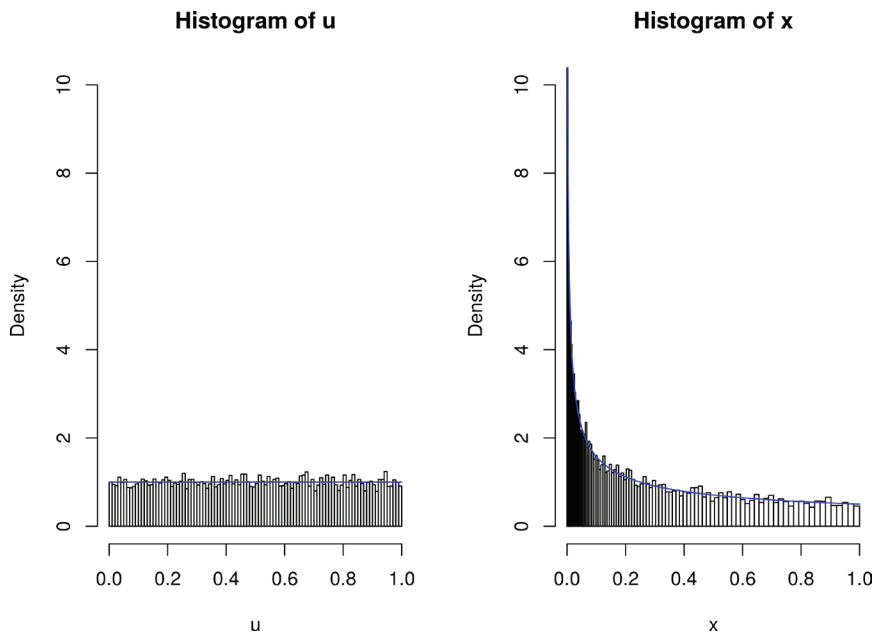


FIG. 4.3 Simulated Distributions

It happens that multiplying the uniform variable by itself (U^2) produces an instance of another random variable, now characterised by the *Beta* distribution, shown on the right of figure 3. While generated by the same set of random numbers, this is now a different random variable. It would be possible to produce that curve of a beta distribution analytically, by plotting points generated by the *Beta* probability density function:

$$f(x; \alpha, \beta) = \text{constant} \cdot x^{\alpha-1} (1-x)^{\beta-1} \text{ where } \alpha = 0.5 \text{ and } \beta = 1 \text{ in equation}$$

But in the case of the plots shown on the right of figure 3, the random variable has been generated from a flow of random numbers. So, from a flow of random numbers, generated by the computer (using a *pseudo-random* number generator algorithm), more random variables result, but with different shapes or probability densities. As Robert and Casella write, ‘the point is that a supply of random variables can be used to generate different distributions’ (2010: 44). Indeed, this is the principle of all Monte Carlo simulations, methods

that ‘rely on the possibility of producing (with a computer) a supposedly endless flow of random variables for well-known or new distributions’ (Ibid: 42). The example shown here is really elementary in terms of the distribution and dimensionality of the random variables involved, yet it illustrates a general practice underpinning the MCMC technique: the reshaping of the ‘supposedly endless flow of random variables’ to produce known or new distributions that map increasingly convoluted volumes and more intricately distributed events.

THE PATH PRECEDES THE TOPOGRAPHY

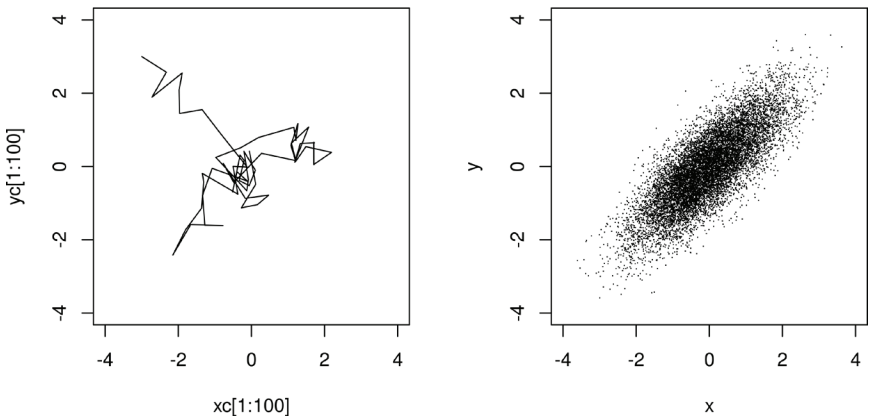


FIG. 4.4 Metropolis-Hastings Algorithm-generated Normal Distribution

In Fig. 4.4, the density of a volume generated by many random numbers (shown on the right as a cloud of points) contrasts with the meandering itinerary of the line on the left. The former plots the now familiar bi-variate normal distribution while the latter shows the path taken by the MCMC algorithm as it *generates* this volume. The path replaces the global analytical solution, the *a priori* analysis or indeed any simply numerical calculation that estimates properties of the volume or surface. Rather, and this difference matters quite a lot, the path constructs the volume as it maps it. The plot on the left precedes the plot on the right, which is effectively a simulated probability distribution derived from the path. Note

too that the path shows only a small selection of the many moves made by the algorithm (approximately 100 of the 40,000 steps).

What form of rule regulates the itinerary of this path? ‘Consider the Markov chain defined by $X^{t+1} = \sigma X^t + \epsilon(t)$ where $\epsilon(t) \in (0, 1)$ ’, write Robert and Casella (2010: 169). The Markov chain—the first MC in MCMC—knows nothing of the normal distribution, yet simulates it by using a flow of random numbers to construct random variables, and then using another stream of random numbers to nudge that random variable into a particular shape. The idea of using a ‘random walk’ to explore the folds of a volume dates back to the work of physicists Nicholas Metropolis and Stanislaw Ulam in the late 1940s’ modelling particles in nuclear reactions (Metropolis 1949).⁴ Purely randomly sampled points are just as likely to lie in low probability regions (valleys and plains) as in the high probability peaks. Metropolis proposed a move which becomes the *modus operandi* of subsequent MCMC work (and hence justifies the high citation count): ‘we place the N particles in any configuration [...] then we move each of particles in succession’ (Ibid: 1088). As well as generating a sample of random numbers that represent particles in a system, they submit each simulated particle to a test. Physically, the image here is that they displace each particle by a small random amount sideways. Having moved the particle/variable, they calculate the resulting slight change in the overall system state, and then decide whether that particular move puts the system in a more or less probable state. If that state is more likely, there is some probability that the move is allowed; otherwise the particle goes back to where it was. Having carried out this process of small moves for all the particles, they can calculate the overall system state or property. The process of randomly displacing the particles by a small amount, and always moving to the more probable states, effectively maps the possibly bumpy terrain of the joint probability density. In many minute moves, the simulation begins to steer the randomly generated values points towards the peaks that represent interesting high-valued features on the surface.

In contemporary MCMC, the folds and contours mapped by the Markov Chains are no longer particles in physical systems but random variables with irregular probability distributions. But the connection between iteration and itinerary holds firm. By generating many itineraries, a topography begins to take

shape and appear. The computationally intensive character of MCMC arises from the iteration needed to construct many random walks across an uneven surface in order to ensure that all of its interesting features have been visited. As we saw in Fig. 4.4, the surface appears by virtue of the Markov chain paths that traverse it.

CONCLUSION

What in the redistribution of events and beliefs in the world, the random variables as distributive individuals, or the paths that precede the terrain they traverse helps us make sense of what was happening: each night in the Obama election team; each time players are matched in Microsoft's Xbox-live player matching system; in the epidemiological models of public health authorities forecasting influenza prevalence (Birrell, Ketsetzis, Gay, et al. 2011); or for that matter, in the topic models that have recently attracted the interest of humanities and social science researchers sifting through large numbers of documents (Mohr and Bogdanov 2013)? I have been suggesting that a redistribution of number is occurring in all these settings, and is perhaps generalised across them. In this redistribution, probabilities no longer simply normalise individuals and groups in partitioned spaces and ranked orders (Foucault 1977) as they might have in a nineteenth-twentieth-century statistical treatments of populations. What might be surfacing in somewhat opaque and densely convoluted forms such as MCMC is a post-demographic rendering of a world in which individuals become something like joint distributions. It is likely that these joint distributions, and their effects on the chances of donating or voting were the target of the Obama data analytics team's night modelling efforts.

This is not to say that a world is clearly and distinctly expressed in these techniques. Against the common tendency to see probability as split between two main interpretations, the aleatory and the epistemic, the frequencies-of-events versus the degrees-of-belief, we see their convoluted embrace in techniques such as MCMC. In this setting, neither the objectivist (frequentists) nor subjectivist (Bayesian) interpretations of probability work well. For the objectivist

interpretations of probability, MCMC presents the difficulty that all parts of the statistical model potentially become random variables or probability distributions, including the parameters of the statistical model itself. For the subjective interpretations, while MCMC means that all parameters can become random variables, these variables only become available for belief via the long chains of numbers that arise in the computations, gradually converging towards the central tendencies or significant features we see in the contour plots. From the post-demographic perspective, both interpretations miss the redistribution of probability as randomly generated but topographically smooth surfaces whose many dimensions support complicated conjunctions of events.

What we might instead see in MCMC and similar techniques is a redistribution of chance, a refiguration of the chance tamed during the last few centuries in the development of concepts of probability and then the techniques of statistics with their reliance on controlled randomness. In these techniques, randomness is again redistributed in the world. This happens materially in the sense that computational machinery generates long converging series of random numbers in order to map the curved topography of the joint probability distributions. But it also happens more generally as a staging of events. Many of Gilles Deleuze's articulations of a baroque sensibility take the form of curves. He describes, for instance, the world as 'the infinite curve that touches at an infinity of points an infinity of curves, the curve with a unique variable, the convergent series of all series' (Deleuze 1993: 24). In Deleuze's account, curves act as causes: 'the presence of a curved element acts as a cause' (1993: 17). This claim begins to make more sense as we see the curved surfaces of joint probability distributions acting as the operational or control points in so many practical settings (asthma studies, multi-player game coordination, epidemiological modelling, spam filtering, and so on). The particles, maps, images and populations figure in a baroque sensibility as curves that fold between outside and inside, creating partitions, relative interiorities and exteriorities.

Where are we in the folded volumes that result from this distributive treatment of numbers? Sensations of change, movement, texture, and increasingly of something happening are attributable to distributive numbers. These machineries stage new convergences between numbers coming from the world, numbers

coming from belief or subjects, and numbers that lie somewhere between the world and knowing subject. I suggested above that we might need to reconceptualise individuals less as the product of biopolitical normalisation and more as a mode of including the world. To the extent that we monadically include the world in such stagings, to the extent that we become the most simple, individual distributive numbers, $1/\infty$ numbers that can only be integrated in simulated surfaces and volumes, then events or what happens are assigned according to the distributive numbers and their curves. What would it mean to be aware of those curves, to have a sense of the joint probability distributions that subtly shape the public health initiatives, the phone calls, or advertisements we receive from a marketing drive, or the price of a product? If normalisation and its statistical techniques sought to strategically manage human multiplicities, to what end do the redistributive numbers we have been discussing tend? The task here, it seems to me, is to identify in the joint probability distributions what is put together, and how assigning numbers to events changes in the light of this joining or concatenating of curves on folded surfaces. There is a kind of generativity here, since the demographic categories and rankings shift and blur on a more differentiated yet integrated or connective surface.

NOTES

1 This chapter will not trace the complicated historical emergence of probability and its development in various statistical approaches to knowing, deciding, classifying, normalising, governing, breeding, predicting, and modelling. Historians of statistics have documented this in great detail and tracked how statistics is implicated in power-knowledge in various settings (Stigler (1986); Hacking (1990); Daston (1994); Porter (1996)).

2 A rapid convergence on MCMC follows from the 1990's onwards. Gibbs samplers appear in desktop computer software such as the widely used WinBUGS ('Windows Bayes Using Gibbs Sampler') written by statisticians at Cambridge University in the early 1990's (Lunn, Thomas, Best, et al. 2000), and MCMC quickly moves into the different disciplines and applications found today.

3 In making sense of the change described by Robert and Casella, scientific histories of the technique are useful. The brief version of the history of MCMC might run as follows: physicists working on nuclear weapons at Los Alamos in the 1940's (Metropolis 1949)

first devised ways of working with high-dimensional spaces in statistical mechanical approaches to physical processes such as crystallisation and nuclear fission and fusion. Their approach to statistical mechanics was later generalised by statisticians (Hastings 1970). It was taken up by ecologists working on spatial interactions in plant communities during the 1970's (Besag 1974), revamped by computer scientists working on blurred image reconstruction (Geman and Geman 1984), and then subsequently seized on again by statisticians in the early 1990's (Gelfand and Smith 1990). In the 1990's, it became clear that the algorithm could make Bayesian inference—a general style of statistical reasoning that differs substantially from mainstream statistics in its treatment of probability (Mcgrayne 2011)—practically usable in many situations. A vast, still continuing, expansion of Bayesian statistics ensued, nearly all of which relied on MCMC in some form or other (Thompson Reuters Web of Knowledge shows 6 publications on MCMC in 1990, but over 1000 *each year* for the last five years in areas ranging from agricultural economics to zoology, from wind-power capacity prediction to modelling the decline of lesser sand eels in the North Sea; similarly NCBI Pubmed lists close to 4000 MCMC-related publications since 1990 in biomedical and life sciences, ranging from classification of new-born babies' EEGs to within-farm transmission of foot and mouth disease; searches on 'Bayesian' yield many more results).

4 It is hardly surprising that scientists working at the epicentre of the 'closed world' (Edwards 1996) of post-WWII nuclear weapons research should develop such a technique. In 1953, Metropolis, the Rosenbluths, and the Tellers were calculating 'the properties of any substance which may be considered as composing of interacting individual molecules' (Metropolis, Rosenbluth, Rosenbluth, et al. 1953: 1087) (for instance, the flux of neutrons in a hydrogen bomb detonation). In their short but still widely cited paper, they describe how they used computer simulation to deal with the number of possible interactions in a substance, and to thereby come up with a statistical description of the properties of the substance. Their model system consists of a square containing only a few hundred particles. These particles are at various distances from each other and exert forces (electric, magnetic, and so on) on each other dependent on the distance. In order to estimate the probability that the substance will be in any particular state (fissioning, vibrating, crystallising, cooling down, etc.) they needed to integrate over the many dimensional space comprising all the distance and forces between the particles (this space is a typical multivariate joint distribution). As they write, 'it is evidently impossible to carry out a several hundred dimensional integral by the usual numerical methods, so we resort to the Monte Carlo method' (Ibid: 1088), a method that Nicholas Metropolis and Stanislaw Ulam had already described in an earlier paper (Metropolis 1949). Here the problem is that the turbulent randomness of events in a square containing a few hundred particles thwarts calculations of the physical properties of the substance. They substitute for that non-integrable turbulent randomness a controlled flow of random variables generated by a computer. While still somewhat random (i.e. pseudo-random), these Monte Carlo variables taken together approximate to the integral of the many dimensional space.

BIBLIOGRAPHY

- Amoore, L., 'Lines of Sight: On the Visualization of Unknown Futures', *Citizenship Studies*, 13.1 (2009), 17–30 <<http://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=36353240&site=ehost-live>> [accessed 19 May 2010]
- Andrieu C., N. De Freitas, A. Doucet, et al., 'An Introduction to MCMC for Machine Learning', *Machine Learning*, 50.1 (2003), 5–43
- Besag J., 'Spatial Interaction and the Statistical Analysis of Lattice Systems', *Journal of the Royal Statistical Society. Series B (Methodological)* (1974), 192–236
- Birrell P.J., G. Ketsetzis, N. J. Gay, et al., 'Bayesian Modeling to Unmask and Predict Influenza a/H1N1pdm Dynamics in London', *Proceedings of the National Academy of Sciences*, 108.45 (2011), 18238–18243 <<http://www.pnas.org/content/108/45/18238.abstract>> [accessed 14 November 2011]
- Daston L., 'How Probabilities came to be Objective and Subjective', *Historia Mathematica*, 21.3 (1994), 330–344
- Deleuze, G., *The Fold: Leibniz and the Baroque* (London: Athlone Press, 1993)
- Dempster A. P., N. M. Laird, and D. B. Rubin, 'Maximum Likelihood from Incomplete Data via the EM Algorithm', *Journal of the Royal Statistical Society. Series B (Methodological)*, 39.1 (1977), 1–38
- Edwards, P. N., *The Closed World: Computers and the Politics of Discourse in Cold War* (Inside Technology Series) (Cambridge, MA, and London: MIT Press, 1996)
- Efron, B., 'The Efficiency of Logistic Regression Compared to Normal Discriminant Analysis', *Journal of the American Statistical Association*, 70.352 (1975), 892–898 <<http://www.jstor.org.ezproxy.lancs.ac.uk/stable/2285453>> [accessed 24 January 2015]
- Foucault, M., *Discipline and Punish: The Birth of the Prison* (New York: Vintage, 1977)
- Gelfand, A. E., and A. F. M. Smith, 'Sampling-based Approaches to Calculating Marginal Densities', *Journal of the American Statistical Association*, 85.410 (1990), 398–409 <<http://www.jstor.org.ezproxy.lancs.ac.uk/stable/2289776>> [accessed 14 January 2011]
- Gelman, A., J. B. Carlin, H. S. Stern, et al., *Bayesian Data Analysis, Second Edition* (Chapman and Hall/CRC, 2003)
- Geman, S., and D. Geman, 'Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images', *Pattern Analysis and Machine Intelligence, IEEE Transactions*, 6 (1984), 721–741
- Gill, J., *Introduction to the Virtual Issue: Bayesian Methods in Political Science, Political Analysis*, 2 (2011) <http://www.oxfordjournals.org/our_journals/polana/virtualissue2.html> [accessed 9 January 2013]
- Hacking, I., *The Emergence of Probability* (Cambridge and New York: Cambridge University Press, 1975)
- *The Taming of Chance* (Cambridge: Cambridge University Press, 1990)

- Hastings, W., 'Monte Carlo Sampling Methods using Markov Chains and their Applications', *Biometrika*, 57.1 (1970), 97–109
- Issenberg, S., 'The Definitive Story of how President Obama Mined Voter Data to Win a Second Term', *MIT Technology Review*, (2012) <<http://www.technologyreview.com/featuredstory/509026/how-obamas-team-used-big-data-to-rally-voters/>> [accessed 9 January 2013]
- Knorr-Cetina, K., and U. Bruegger, "'Traders'" Engagement with Markets: A Postsocial Relationship', *Theory, Culture & Society*, 19.5.6 (2002), 161–185
- Lunn, D., A. Thomas, N. Best, et al., 'WinBUGS—a Bayesian Modelling Framework: Concepts, Structure, and Extensibility', *Statistics and Computing*, 10.4 (2000), 325–337
- Mcgrayne, S. B., *The Theory that Would not Die: How Bayes' Rule Cracked the Enigma Code, Hunted down Russian Submarines, and Emerged Triumphant from Two Centuries of Controversy* (New Haven, CT: Yale University Press, 2011)
- Metropolis, N., 'The Monte Carlo Method', *Journal of the American Statistical Association*, 44 (1949), 335–341
- Metropolis, N., Rosenbluth, A.W. Rosenbluth, M. N., et al., 'Equation of State Calculations by Fast Computing Machines', *The Journal of Chemical Physics*, 21 (1953) (1953) <<http://link.aip.org/link/%3FJCPSA6/21/1087/1>> [accessed 10 October 2012]
- Mohr, J. W., and P. Bogdanov, 'Introduction—Topic Models: What they Are and why they Matter', *Poetics*, 41.6 (2013), 545–569 <<http://www.sciencedirect.com/science/article/pii/S0304422X13000685>> [accessed 16 December 2013]
- Porter, T. M., *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life* (Princeton: Princeton University Press, 1996)
- Robert, C., and G. Casella, 'A History of Markov Chain Monte Carlo-subjective Recollections from Incomplete Data', *arXiv:0808.2902* (2008) <<http://arxiv.org/abs/0808.2902>> [accessed 1 January 2011]
- Robert, C. P., and G. Casella, *Introducing Monte Carlo Methods with r* (New York: Springer, 2010)
- Scherer, M., 'How Obama's Data Crunchers Helped him Win', *CNN.com* (2012) <<http://www.cnn.com/2012/11/07/tech/web/obama-campaign-tech-team/index.html>> [accessed 20 November 2012]
- Simpson, A., V. Y. F. Tan, J. Winn, et al., 'Beyond Atopy: Multiple Patterns of Sensitization in Relation to Asthma in a Birth Cohort Study', *American Journal of Respiratory and Critical Care Medicine*, 181.11 (2010), 1200–1206 <<http://www.ncbi.nlm.nih.gov/pubmed/20167852>> [accessed 16 May 2011]
- Stigler, S. M., *The History of Statistics: The Measurement of Uncertainty before 1900* (Cambridge, MA: Harvard University Press, 1986)
- Suess, E. A., and B. E. Trumbo, *Introduction to Probability Simulation and Gibbs Sampling with r. 1st edn.* (New York: Springer, 2010)
- Wasserman, L., *All of Statistics: A Concise Course in Statistical Inference* (New York: Springer, 2003)