# Some Notes on the History of Branching Processes, from my Perspective

## Peter Jagers

## Oberwolfach, January 2009

At a meeting like this, it is far from evident, and indeed, an outsider would never guess that few fields within the mathematical sciences, may be none, cherish their past like branching processes. Ted Harris's classical treatise from 1963 opens by a terse but appetising two-page flashback.  Three years later, David Kendall's elegant overview was published, and like Charles Mode in his monograph (1971), I could borrow from that for the historical sketch opening in my 1975 book, but also add some observations of my own.

At that time we all knew that the French, notably de Candolle and Bienaymé, had considered the nobility and family extinction problem, before Galton publicized it. I also speculated about connections between Bienaymé and the demographer Benoiston de Châteauneuf, who had been studying old French noble families . The plausibility of such contact was corroborated by Chris Heyde and Eugene Seneta in their "I. J. Bienaymé: Statistical Theory Anticipated ", where they also showed that Bienaymé was not only first at formulating the mathematical problem, but indeed knew its solution already in 1845. The original publication has not been found, though, but as pointed out in the recent monograph by Iosifescu *et al.*, Bernard Bru has discovered a proof in a treatise by A. A. Cournot, published only two years after Bienaymé's communication. Though this is not explicitly stated, it seems plausible that Cournot reproduces Bienaymé's argument. (The book by Iosifescu and co-authors also presents an intriguing discussion by Bienaymé arguing that the limited size of mankind (in the mid XIX:th Century!) should show that human mean reproduction must have varied above and below one in historical time.)

There are good reasons for branching processes to keep its heritage alive. Not only is the background in the frequent disappearance of family names, even in growing populations, picturesque and easily understood, it is also something that could not have been explained by prevailing – and long dominating – deterministic population theory. Indeed, it provides convincing arguments for a stochastic population theory, and not only for "small" populations. Further, in spite of its alluring disguise, the family extinction problem concerns an important and basic feature of population development, about frequent extinction and shared ancestry.

It also tells a story of interplay between mathematics, natural science, culture, and society, which we should listen to now that mathematicians tend forget about the social context of their science and even its roots in physical or biological problems. On the other extreme,

post-modern philosophers proclaim that everything is social, as though there were no mathematical or physical truth.

Both positions can be contrasted with Galton's classical formulation in Educational Times 1873:

"PROBLEM 4001: A large nation, of whom we will only concern ourselves with adult males, $N$ in number, and who each bear separate surnames colonise a district. Their law of population is such that, in each generation, $a_0$ per cent of the adult males have no male children who reach adult life; $a_1$ have one such male child; $a_2$ have two; and so on up to $a_5$ who have five.

Find (1) what proportion of their surnames will have become extinct after $r$ generations; and (2) how many instances there will be of the surname being held by $m$ persons."

Rarely does a mathematical problem convey so much of the flavour of its time, colonialism and male supremacy hand in hand, as well as the underlying concern for a diminished fertility of noble families, paving the way for the crowds from the genetically dubious lower classes.

It also exhibits a mathematical theory initiated not by mathematicians but by a broad *savant,* Francis Galton, a polyhistor well versed in mathematics but primarily if anything, a biologist. We see an example falsifying both extremist views on science, that of a pure science, and in particular mathematics, devoid of political meaning and implications; and that degrading science and scientific development into a purely social phenomenon. Indeed, in branching processes, they all meet: pure mathematical development, biology, physics, and demography, and the concoction is spiced to perfection by the social and cultural context in which it is formed.

As is well known, Watson determined the extinction probability as a fixed point of the reproduction generation function. He observed that 1 is always such a fixed point, and from this he and Galton (1874) intriguingly concluded: "All the surnames, therefore, tend to extinction in an indefinite time, and this result might have been anticipated generally, for a surname lost can never be recovered, and there is an additional chance of loss in every successive generation. This result must not be confounded with that of the extinction of the male population; for in every binomial case where $q$ is greater than 2 we have $t_1 + 2t_2 + \&c. + qt_q > 1$, and, therefore an indefinite increase of male population."

It is strange that so intelligent a couple as Galton and Watson (the latter turned clergyman but had been second Wrangler at Cambridge, carried on  mathematics and physics as a Rector and even was awarded an honorary D. Sc. by his *Alma Mater*) could have presented, and even believed in this seductive verbiage. Maybe, it is even stranger that it took more than fifty years to rectify it, in particular since Bienaymé had already published a correct statement of the extinction theorem. I always thought the reason simply was that people of the time just did not notice, or bother about, such a mathematical trifle. But according to Heyde and Seneta, "its implications were strongly doubted" already at the time of publication.

And indeed, I checked an (almost) contemporary and non-mathematical criticism quoted by them, by a Swedish historian or political scientist, Pontus Fahlbeck. He was a commoner who married a baroness. Maybe this arose his interest in nobility: he became the author of a monumental two-volume treatise on the Swedish gentry (1898, 1902). There he gives a correct, verbal description of the relation between growth of the whole versus frequent extinction of separate family lines, and writes – somewhat condescendingly or intimately, it may seem: "*Galton*, who with characteristic curiosity considered the question, has tried to investigate to what extent families ... must die out, with the help of a competent person." Fahlbeck then recounts examples considered by Galton, showing that "the tendency is the extinction of all". (The account is not completely lucid.) This is followed by a sequel of questions, and a reassuring answer:

"If this course of events be based on a mathematical law, then it should be as necessary, or not? And what then about our general conclusions, that no necessity forces extinction? Is there not in this a contradiction, which if both arguments are right, as they undoubtedly are, leads to what philosophers call an antinomy? However, mathematical calculations, as applied to human matters, may seem unrelenting but are actually quite innocuous. The necessity lies buried in them like an electrical current in a closed circuit, it cannot get out and has no power over reality." (pp. 133-135, my translation).

As you know, it was another polyhistor, J. B. S. Haldane, chemist, physiologist, geneticist statistician, and prolific political writer in the New Statesman as well as the Daily Worker (He was a notable member of the intellectual British left of the 30's and 40's, beautifully described by Doris Lessing, among others,) who got things basically right, although the really correct formulation was printed slightly later (Steffensen, 1930).

When telling this history and then stating the correct extinction theorem, I sometimes meet the comment at that this is all very amusing, but how does it explain that frequent extinction of family lines occurs even in quickly growing populations? The answer is, of course, that the extinction probability $q$ and mean reproduction $m = f'(1)$ can be large simultaneously, for very convex reproduction generating functions $f$. Indeed, values of 0.75 and 2, respectively, e.g., are obtained for realistic reproductive patterns among human males, or for that sake females, in historic times.

Lecturing in Peking in October 2008, I met with a cute illustration of this, which may well have occurred to some of you. In the *China Daily* I read that Kung Te-chen, who was the 77: th great...grandson of Confucius (Kung Fu-tse) had died on Taiwan at the age of 89. Yes, same surname inherited from father to son for more than 75 generations. Since Confucius (500 B.C.), China's population has undergone a tremendous growth, but as we all know, there are few Chinese family names. Indeed, Wikipedia tells us that three surnames (partly different in different parts of the country) are carried by some 30% of the population. In Korea the situation is even more extreme; half the population has one of the names Kim, Lee, or Park.

Thus, branching processes were born out of a social demographic context. Its first fundamental result, the extinction theorem, has relevance far beyond that, in explaining homogeneity in large populations, as well as (part) of the more than frequent extinction in the course of natural evolution. Indeed, 1991 the palaeontologist David Raup claimed that 99.99% of all species, ever existing on our earth, are extinct now.

When branching processes reappear in scientific literature, between the great wars, the impetus comes from genetics (Fisher and Haldane) and biology more generally. Haldane deduces his approximation for the survival probability, still very important for the consideration of fresh, slightly fitter mutants in a resident population.   In Russia, Kolmogorov coins the term *branching process* itself.

After World War II, the nuclear age arrives. In Stalin's Moscow, Kolmogorov and his disciples, people like the Yaglom twin brothers and B. A. Sevastyanov, try to pursue their research as a purely mathematical undertaking. But of no or little avail. Sevastyanov's thesis was classified, while being written, and since he himself was not deemed reliable he was not allowed to keep it. Every morning a KGB officer opened a safe in the library and handed it out to its author, who continued writing on it until five, when he had to give it back.

Kolmogorov and others, including I believe Sacharov, protested, and finally the ban was lifted (Sevastyanov, 1999). Things had become easier than in the 30's.

In the United States, Ted Harris was employed by the Rand Corporation, an integral part of the military-industrial complex, and his work on electron-photon cascades and Galton-Watson processes with continuous type space (energy) was clearly inspired by nuclear physics. But both he and Sevastyanov saw themselves as mathematicians, though working on a pattern relevant for natural science. Sevastyanov even takes a rather purist stance; I have heard him saying that mathematics is nothing but mathematics, a somewhat unexpected opinion from a mathematician who is neither an algebraist nor a topologist, not even a pure analyst, and who actually after his thesis worked several years in the secret military part of the Steklov Institute, the so called "Box". Maybe it comes naturally to someone who has devoted his life to mathematics in the overly politicised Soviet Union.

With such leaders, it is not surprising that the 50's and 60's was an era of mathematisation. Time structure was added to the simply reproductive branching process in what Bellman and Harris called *age-dependent* processes, depicting populations where individuals could have variable life spans, but split into a random number of children at death, independently of age. Truly age-dependent branching processes were introduced by Sevastyanov, the reproduction probabilities possibly affected by the mother's age at splitting.

The processes thus arising were not Markovian in real time, but could be analytically treated using renewal properties, and the then remarkable renewal theory, which had recently been established by Feller and others. Another development retained the Markov property, but

viewed population evolution as occurring in real time, thus establishing connection to the elementary birth-and-death processes that were flourishing in semi-applied literature.

These approaches however remained in a sort of physical world, far from animal or even plant population dynamics, in the sense that they all considered child-bearing through *splitting* only, like fission, cell division, or molecular replication. Or, for the classical Galton-Watson process, there was the alternative interpretation of disregarding time, and just count generations, as though they did not overlap in real time. The only exception were the models from the-birth-and-death sphere, where exponentially distributed life spans allowed alternative interpretations. That also lead to the first model of populations where individuals could give birth *during* their lives, Kendall's generalised birth-and-death process (1948).

The first monographs, Harris's from 1963 and Sevastyanov's from 1971, as well as Athreya's and Ney's from 1972, however stayed firmly in the tradition of physical splitting. Branching processes remained separated from the deterministic differential equations, matrix, and Lotka-Volterra tradition of population dynamics and mathematical demography. It was the development of a general point process theory that rendered the formulation of general branching processes natural, so as to depict populations where individuals can give birth repeatedly, in streams of events formed by a point process, and possibly even of various types. 1968 time was ripe, and Crump's and Mode's article and mine appeared simultaneously in the winter 68-69. Mine was also part of my Ph.D. thesis, defended in October, fortunately. In those times in Sweden, formal originality was insisted upon, in a somewhat square manner, and in spite of the enormous friendliness of my polite Japanese opponent, Kiyosi Ito himself, I might not have been let through, had the stern local mathematics professors known that some Americans had done the same, sceptical towards probability theory, as they were.  The status of probability within mathematics has certainly changed since then!

The advent of general branching processes meant that branching processes now embraced virtually all mathematical population theory. The dominating mathematical population framework since more than a century was the *stable population theory*, dating back to Quetelet and Lotka. Its real father or forerunner was, however, Euler who deduced its main findings, the exponential increase of population size and how the ensuing stable age distribution is determined by survival and reproduction rates, already around 1750. As I pointed out in my 1975 book, Euler even used rapid population growth as an argument against those incredulous who would not believe that the sons of Adam could have filled the earth during the 5000 years since he was ousted from Eden. Nevertheless, his contributions seem long forgotten in the demographic and mathematical biology communities.

Stable population theory is deterministic but based upon a probabilistic formulation of individual life events. All its findings could now be strictly proven in terms of general branching processes, and basic concepts like average age at childbirth given an interpretation. Furthermore, the stabilisation of population composition could be brought

one step further: stable population theory had only considered the distribution of age in old populations. Age is what could be called an individual property: it is your age and nobody else's. In a population there are however also important relational properties.

My research into this area started in a quaint manner. In my youth, Gothenburg had a well-known doctor caring for the city's alcoholics. Now that he had retired in the late 70's he took up a research idea that he had toyed with for some time. He had made the observation that an astonishing proportion of his patients were first-born.

 Studying the literature, he found that not only Gothenburg alcoholics, but also poets, statesmen, and people suffering from various mental disorders had been found often to be first-born. Galton had even claimed that the firstborns were the motor of history, since they were more often "men of note". The retired doctor realised that the apparent overrepresentation of firstborns could be an artefact, and performed a primitive but adequate simulation experiment. He bought a huge and long paper table cover drawing the family trees of an invented but realistic population on a long white paper table cloth, and started to draw an invented but realistic set or parallel family trees, evolving along the roll. Towards the end he then sampled individuals at random, or at least haphazardly. Many were firstborns. Now he wanted to discuss with me.

I found the probability of being first-born in an old single-type supercritical general branching process. It is $E[e^{-\alpha\tau}]$, where $\alpha$ is the Malthusian parameter and $\tau$ mother's age at her first bearing. Since the Malthusian parameter equals ln 2 divided by the doubling time and the latter is usually larger than age at first bearing, we see that the probability of being first-born tends to be larger than 0.5, even in populations with large broods or families (Jagers, 1982). (In an old critical population the probability of being first-born is larger than one over the expected sibship size, due to Jensen's inequality.)

The important is, however, that being first-born is not a property of your own life and birth-time. It concerns your relation to your sibship. Thus, this simple observation led on to an investigation of how the whole pedigree, family structure, and type distribution in multi-type populations stabilise during exponential growth. A strict framework for general branching processes in abstract type spaces was formulated, related to branching tree ideas due to Neveu and Chauvin (1986). In these, type distribution and ancestry, and hence mutational history could be traced backward in a Markov renewal structure. Our group published a whole sequel of papers on these topics during the 80's and 90's, and indeed a final (?) attempt to popularise the admittedly heavy theory by restriction to discrete time quite recently (Jagers and Sagitov, 2008). Stable pedigrees and backward times was virgin land, with one exception only, the investigations by Joffe and Waugh into kin numbers in Galton-Watson processes (1981 and 1982).

In the mean time, deterministic population dynamics had advanced through work by eminent mathematicians like Odo Diekmann and Mats Gyllenberg, inspired by the biologist

Hans Metz. They had realised that the differential equations formulations they had been brought up with were becoming a straitjacket, and turned to semigroups of positive operators, yielding a theory corresponding to the Markov renewal theory of expectations of multitype general branching. However, they took a further step, considering the feedback loop individual -> population -> environment -> individual. Through this theory, *structured population dynamics,* they were able to analyse the fascinating new ideas that Metz and his followers had advanced to explain evolution, under the name of *adaptive dynamics*.

This was a new challenge to branching processes, and is being met in a series of path-breaking papers by Sylvie Méleard and her co-workers. We have also tried to formulate models investigating the consistency of adaptive dynamics, and in particular the problem of *sympatric speciation*, i.e. how successive small mutations can lead to new species, and their coexistence – but with less success so far.

The general problem of interaction in population dynamics is elusive. On one hand, the very concept of population builds upon individuals in some sense being the agents, those changing the population by their actions. The branching process idea is to make this vague idea of "individual initiative" precise by sharpening it into the requirement of stochastic independence between individuals. This is proper as an idealisation, but obviously takes us far from reality. In special cases this can be remedied, as in the models considered by Méléard and Champagnat and Lambert, or in the population size dependence studied by Kersting, Klebaner, and others, which allows an understanding of the linear growth occurring in the famous polymerase chain reactions, PCR, (cf. Haccou et al). But a real liberation from independence, replacing it by exchangeability in some form, e.g., remains out of reach.

This overview has been rather centred on my own interests, branching processes as a form of theoretical biology. It has centred on the supercitical case, which was the main interest of my own expansive youth. My recent papers, quite suitably deal with the path and time to extinction (2007). However, most of the revival branching processes and related areas experienced in the 90's, and which continues to this day has a different character. Mainly it is purely mathematical; partly it is inspired by computer algorithms. The whole area of superprocesses and measure-valued Markov branching processes, seems to belong to the former realm, whereas random trees though certainly a pure mathematical area also has drawn upon both phylogenetics and computer science.  But these are areas where others have much more insight than I, and I leave it to you to comment upon the impressive growth of these fields during the past three or so decades.

References

Bienaymé, I. J. (1845), De la loi de multiplication et de la durée des familles. *Soc. Philomath. Paris Extraits, Sér. 5,* 37–39.

Bru, B. (1991), A la récherche de la demonstration perdue de Bienaymé. *Math. Inform. Sci. Humaines* **29,** 5-17.

Chauvin, B. (1986) Arbres et processus de Bellman-Harris. *Ann. Ins. H. Poincaré.* **22,** 2, 209-232.

Champagnat, N. (2006), A microscopic interpretation for adaptive dynamics trait substitution sequence models. *Stoch. Proc. Appl.* **116**, 1127-1160.

Crump, K. S. and Mode, C. J. (1968 and 1969), A general age-dependent branching process I and II. *J. Math. Anal. Appl.* **24**, 494-508, and **25**, 8-17.

Cournot, A. A. (1847), *De l'origine et des limites de la correspondence entre l'algèbre at la géométrie.* Hachette, Paris.

Dieckmann, U. and Doebeli, M. (1999) On the origin of species by sympatric speciation. *Nature* **400**, 354-357.

Dieckmann, U. and Law, R. (1996), The dynamical theory of coevolution. *J. Math. Biol.* **34**, 579-612.

Fahlbeck, P. E. (1898, 1902), Sveriges adel, statistisk undersökning öfver de å riddarhuset introducerade ätterna, I and II. Also available in German.

Galton, F., and Watson, H. W. (1874), On the probability of extinction of families. *J. Anthropol. Inst.* **4**, 138-144.

Haccou, P., Jagers, P., and Vatutin, V. A. (2005), *Branching Processes: Variation, Growth, and Extinction of Populations*. Cambridge. U. Press, Cambridge.

Harris, T. E. (1963), *The Theory of Branching Processes.* Springer-Verlag, Berlin etc. (Reprinted as a Dover Book 1989).

Heyde, C. C., and Seneta, E. (1977), *I. J. Bienaymé Statistical Theory Anticipated*. Springer-Verlag, Berlin etc.

Iosifescu, M., Limnios, N., and Oprisan, G. (2007), Modèles stochastiques. Hermes Lavoisier, Paris.

Jagers, P. (1969) A general stochastic model for population development. *Skand. Aktuarietidskr.* **52**, 84-103.

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

Jagers, P. (1982), How probable is it to be firstborn? and other branching process applications to kinship problems. *Math. Biosci.* **59**, 1-15.

Jagers, P. (1989) General branching processes as Markov fields. *Stoch. Proc. Appl.* **32,** 183-212.

Jagers, P. and Nerman, O. (1996), The asymptotic composition of supercritical, multi-type branching populations. Séminaire de probabilités de Strasbourg, **30**, *Springer Lecture Notes in Mathematics,* **1626,** 40-54.

Jagers, P., Klebaner, F., and Sagitov, S. (2007), On the path to extinction. *Proc. Nat. Acad. Sci.* 104, 6107-6111.

Jagers, P., Klebaner, F., and Sagitov, S. (2007), Markovian paths to extinction. *Adv . Appl. Probab.* 39, 569-587.

Jagers, P. and Sagitov, S. (2008), General branching processes in discrete time as random trees. *Bernoulli* 14:4, 949-962.

Joffe, A. and Waugh, W. A. O'N. (1982), Exact distributions of kin numbers in a Galton-Watson process. *Journ. Appl. Probab.* **19,** 767-775.

Kendall, D. G. (1948), On the generalized 'birth-and-death' process. *Ann. Math. Statist.* **19**, 1-15.

Kendall, D. G. (1966), Branching processes since 1873. *J. London Math. Soc* **41**, 385-486.

Metz, J. A. J., Geritz, S. A. H., Meszéna, G., Jacobs, F. J. A., and and van Heerwaarden, J.S. (1996). Adaptive dynamics, a geometrical study of nearly faithful reproduction. In: *Stochastic and Spatial Structures of Dynamical Systems.* North-Holland, Amsterdam.

Mode, C. J. (1971), *Multitype Branching Processes.* Elsevier, New York.

Nerman*,* O. and Jagers, P. (1984). The stable doubly infinite. pedigree process of supercritical branching populations. *Z. Wahrsch. Verw. Gebiete* **65**, 445-460.

Neveu, J. (1984) Arbres et processus de Galton-Watson. . *Ann. Ins. H. Poincaré*. **22,** 2, 199-207.

Raup, D. M. (1991), *Extinction*. Norton*,* New York.

Sevastyanov, B. A. (1999) Lecture at acceptance of Honorary Doctorate, Chalmers University of Technology.