

- Ptacek, P. H. & E. K. Sander. 1966. 'Age recognition from voice'. *Journal of speech and hearing Research* 9, 273-7.
- Schötz, S. 2003. 'A first step from analysis to synthesis'. In *Proceedings of the XVth ICPHS*, 2585-8. Barcelona.
- Schötz, S. 2004. 'The role of F_0 and duration in perception of female and male speaker age'. In *Proceedings of Speech prosody 2004*, 379-82. Nara.
- Taylor, P., R. Caley, A. Black & S. King. 1999. *Edinburgh speech tools library. System documentation* Edition 1.2. Website: http://festvox.org/docs/speech_tools-1.2.0/book1.htm.

Susanne Schötz <Susanne.Schotz@ling.lu.se>

Modelling the changing popularity of names

Bengt Sigurd, Mats Eeg-Olofsson and Jørgen Ouren

The popularity of Norwegian first names 1880-2000 can be studied thanks to official Norwegian statistics. The most common curve shows a fast rise and slow fall, which can be approximated by a mathematical gamma frequency function. The curve presumably reflects the development of the parents' enthusiasm.

Introduction

The Norwegian statistical agency (*Statistisk sentralbyrå*) offers a data base on the Internet at www.ssb.no/emner/00/navn/, where the frequencies of several hundred first names from 1880-2000 are shown in diagrams. As has been noted before there is a certain recycling of names in Scandinavia and quite a few names e.g. *Martin*, *Kristian* and *Lars* in Sweden, *Kristine* and *Karoline* in Norway reappear after about 120 years (see publications by the Norwegian statistical agency, and Sigurd & Eeg-Olofsson 2004). Also interesting is the shape of the historical frequency diagrams as most of them display a fast frequency rise followed by a longer slow fall. Such a shape can be approximated and modelled by a frequency function based on the mathematical gamma distribution. We will illustrate typical frequency curves for names and show an approximating curve and gamma frequency function which fits the name *Sverre* well. With somewhat different parameters it should fit several other names and it allows us to predict the development of the popularity of a name.

The study of the developmental frequency patterns of names is interesting since the same patterns are likely to show in other fashion behaviour. Modelling the patterns is not only of linguistic and sociological interest but also of commercial interest as it makes it possible to predict the development of a fashion or the success of a new product.

Types of curves

The name curves generally look like hills with a rise and a fall which can be discerned although the whole hill is not always visible in the Norwegian diagrams which only include frequencies from 1880 to 2000, i.e. 120 years. In

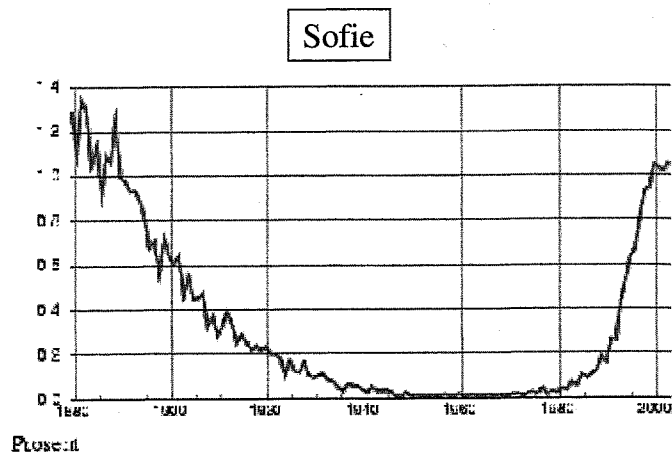


Figure 1. Frequency of children given the name *Sofie* in different years. After *Statistisk sentralbyrå*.

several cases only the fall of a preceding wave or the rise of a second wave or both can be seen, see the diagram *Sofie* (Figure 1).

It is natural to distinguish between symmetric and asymmetric frequency curves. Symmetric names show almost equally long rise and fall parts. This type is illustrated by the name *Åse* which appeared around 1880 and rose to its top after 60 years around 1940 (Figure 2).

The asymmetric names may have a slow increase in the beginning and then a fast fall. This type is not common, but *Øystein* is an example (Figure 3).

The most common type, however, shows a fast rise followed by a slow gradual decrease. Out of the 150 male names displayed on the first page of the web site about 100 can be classified as fast-slow names or gamma names to be defined below. The same is true for the female names.

Gamma names

The frequency pattern of *Sverre* displays a development reflecting a beginning enthusiasm rising fast to a top after ca 30 years followed by a subsequent gradual lack of interest possibly as parents note that there are many children with that name around. The top percentage is generally 0.5–2.5. The decrease of the frequency looks like an exponential curve more or less approaching zero. Data and an approximation are shown in Figure 4.

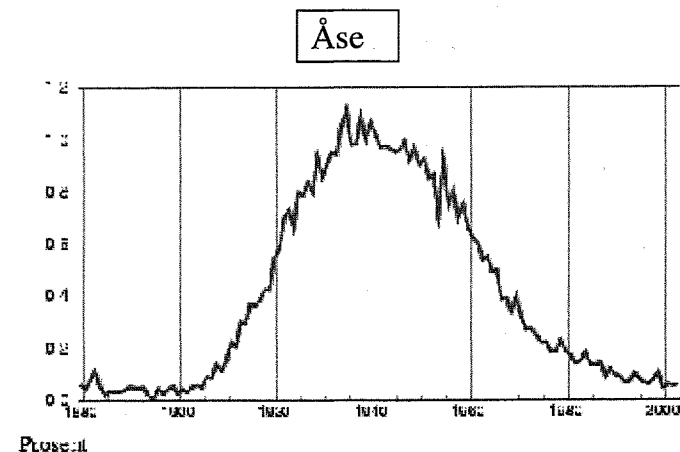


Figure 2. Frequency of children given the name *Åse* different years. After *Statistisk sentralbyrå*.

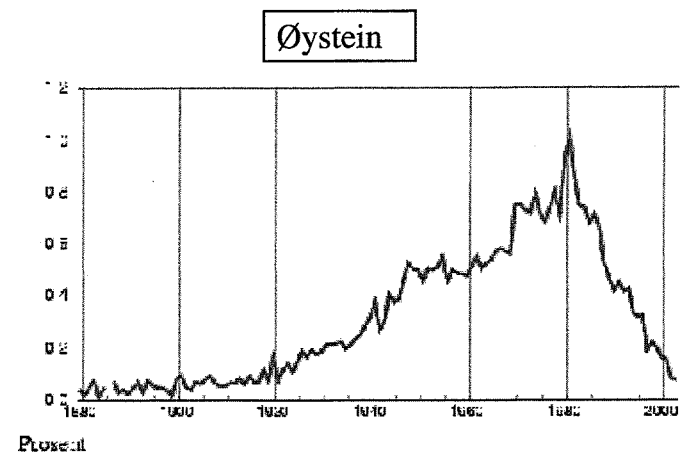


Figure 3. Frequency of children given the name *Øystein* different years. After *Statistisk sentralbyrå*.

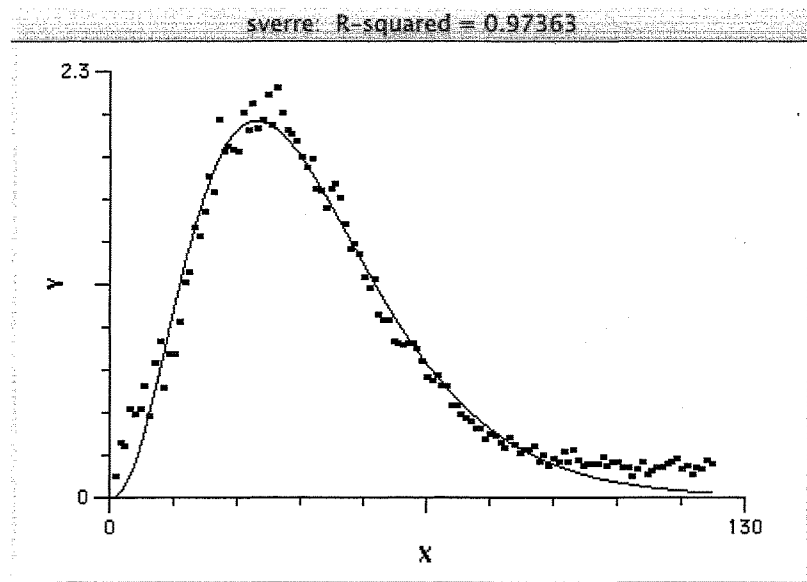


Figure 4. Frequency of children given the name *Sverre* different years and an approximation. Data after *Statistisk sentralbyrå*.

The curve for *Sverre* can be approximated by a frequency curve based on the mathematical gamma distribution, as is shown in the Figure 4. (The gamma distribution is applied to a different kind of data in Sigurd, Eeg-Olofsson & van de Weijer 2004, where the general characteristics of the gamma distribution are discussed in more detail.)

The function for the gamma distribution used here has the form:

$$f(t) = a \cdot t^b \cdot c^t$$

where f is frequency, a is a scaling factor, t is time in years, b is a suitable parameter having to do with the early rise of the curve and c is a parameter < 1 having to do with the exponential decrease of the frequency.

The frequency will rise with the increase of time, but this is counteracted by the later part of the formula where it decreases with time. For the name *Sverre* the parameter values $a = 0.0039568$, $b = 2.60327$ and $c = 0.916608$ give a fairly good fit. If we decrease the exponential value from 0.916608 to e.g. 0.90 the curve will fall faster. The coefficient of determination of the function is as good as $r^2 = 0.97363$.

Some of the data and approximated values for *Sverre* are shown in the following list:

year	data	approx	year	data	approx	year	data	approx
1881	0.118	0.004	1925	1.668	1.582	1965	0.262	0.254
1885	0.454	0.169	1930	1.366	1.347	1970	0.177	0.191
1890	0.840	0.664	1935	0.983	1.117	1975	0.261	0.142
1895	1.162	1.235	1940	0.820	0.906	1980	0.186	0.105
1900	1.730	1.690	1945	0.649	0.722	1985	0.162	0.077
1905	1.880	1.955	1950	0.501	0.567	1990	0.129	0.056
1910	1.989	2.033	1955	0.376	0.439	1995	0.189	0.041
1915	2.076	1.965	1960	0.300	0.336	2000	0.163	0.030
1920	1.781	1.800						

Discussion

The appearance of a name which gets popular generally has to do with the existence of an idol, a royal person, a sports idol, a film star, an artist, but we will not discuss here which persons have started the name waves. Nor will we discuss why certain names show a sharp downfall. But it is clear that the popularity of the name *Adolf* fell by the end of the second world war and the name *Harald* rose when the Norwegian prince was given that name in 1937.

More data are needed in order to support the analysis of name frequencies in other cultures, but it seems highly reasonable to assume that the gamma curve dominates in other cultures where names are given freely as well. Among further research tasks we will mention the study of the frequency development of fashion words and fads. It is clear that new adjectives denoting 'good' or 'bad' are like fashions and appear and fade away in similar ways. In Swedish the expression *alla tiders* 'of all times' was popular in the 1930s but is completely outdated today. Various positive adjectives have entered the English language during the 20th century: *swell*, *cool*, *sexy* and it would be an interesting task to follow their development.

References

- Sigurd, Bengt & Mats Eeg-Olofsson. 2004. 'Modets växlingar hos förnamn i Norden'. *Arkiv för nordisk filologi* 119, 209-25
- Sigurd, Bengt, Mats Eeg-Olofsson & Joost van de Weijer. 2004. 'Word length, sentence length and frequency: Zipf revisited', *Studia linguistica* 58:1, 37-52.

Bengt Sigurd <Bengt.Sigurd@ling.lu.se>

Mats Eeg-Olofsson <Mats.Eeg-Olofsson@ling.lu.se>

Jørgen Ouren, Statistisk sentralbyrå, Oslo, Norway <jou@ssb.no>