Chilean Journal of Statistics
Vol. 3, No. 2, September 2012, 113–129

Probabilistic and Inferential Aspects of Skew-Symmetric Models
Special Issue: "IV International Workshop in honour of Adelchi Azzalini's 60th Birthday"

The work of Fernando de Helguero on non-normality arising from selection

Adelchi Azzalini[1,*] and Giuliana Regoli[2]

[1]Dipartimento di Scienze Statistiche, Università di Padova, Italia
[2]Dipartimento di Matematica e Informatica, Università di Perugia, Italia

(Received: 13 October 2011 · Accepted in final form: 9 February 2012)

Abstract

The current literature on so-called 'skew-symmetric distributions' is closely linked to the idea of a selection mechanism operated by some latent variable. We illustrate the pioneering work of Fernando de Helguero who in 1908 put forward a formulation for the genesis of non-normal distributions via a selection mechanism, which perturbs a normal distribution, hence employing a closely connected argument with the one now widely used in this context. Arguably, de Helguero can then be considered the precursor of the current idea of skew-symmetric distributions. Unfortunately, a tragic quirk of fate did not allow him to pursue his project beyond the initial formulation and his work went unnoticed for the rest of the 20th century.

Keywords: Hidden truncation model · History of statistics · Selection models · Selective sampling · Skew-normal distribution · Weighted distributions.

Mathematics Subject Classification: Primary 62E10 · Secondary 62-03.

1. Introduction

There is now a fairly wide literature which adopts the so-called skew-symmetric distribution scheme for the construction of flexible parametric classes of probability distributions. The reader not familiar with this area is referred to Genton (2004) and Azzalini (2005) for introductory accounts, although this literature has expanded considerably since then; see Azzalini (2011) for a concise summary of the key concepts.

An important appeal of the skew-symmetric construction is that it allows various forms of stochastic representation for random variables having such a distribution. This fact implies that early occurrences of the expression of skew-symmetric distributions, in the scalar case, feature in several places in the literature, especially those which include the normal density as its base component; specific instances of this sort will be recalled shortly.

*Corresponding author. Adelchi Azzalini, Dipartimento di Scienze Statistiche, Università di Padova, 35121 Padova, Italia. Email: azzalini@stat.unipd.it

ISSN: 0718-7912 (print)/ISSN: 0718-7920 (online)

http://www.soche.cl/chjs

There is then the natural question: where did the idea of skew-symmetric distributions, or at least of some specific instance of them, first appear in the literature? Questions of this sort seldom have a clear-cut answer and it is well-known that lots of mathematical results are named after someone who did not discover them. However, we believe that, in the present context, some insight can be provided.

Although much of the current literature on skew-symmetric distributions deals with the multivariate case, all our discussion is about univariate distributions, since this is the only relevant case for the question indicated above. To establish notation, consider a univariate skew-symmetric density f written in the form

$$f(x) = 2\, G_0(w(x;\lambda))\, f_0(x), \qquad -\infty < x < \infty, \tag{1}$$

where f_0 is a symmetric density about zero, G_0 is a distribution function with symmetric density about zero, and $w(x;\lambda)$ is an odd function depending on a parameter λ. In practical work, the distribution will usually be shifted by a location parameter and there can be additional parameters which regulate f_0 and G_0.

The very proof that Equation (1) is a proper density integrating to 1 indicates that a stochastic representation of a random variable Z with distribution f is

$$Z \stackrel{d}{=} (X|T < w(X,\lambda)),$$

where X and T are independent random variables with density f_0 and G_0', respectively. A little further elaboration which takes into account the symmetry of X and the oddness of $w(x;\lambda)$ leads to the additional representation

$$Z \stackrel{d}{=} \begin{cases} X, & \text{if } T < w(X,\lambda); \\ -X, & \text{otherwise.} \end{cases} \tag{2}$$

Other stochastic representations exist for important sub-families of the class given in Equation (1), but the above two are the only ones holding for all densities of similar type of Equation (1), at the present state of knowledge. The top branch of Equation (2) suggests considering a more general form of the selection mechanism, that is,

$$Z \stackrel{d}{=} (X|T < \lambda_0 + w(X,\lambda)), \tag{3}$$

where λ_0 is an additional real parameter, which lends the earlier formulation when $\lambda_0 = 0$. The heading 'hidden truncation models' is often used in connection with this construction. Since $\lambda_0 + w(x;\lambda)$ is not an odd function when $\lambda_0 \neq 0$, some of the properties connected to the earlier construction disappear. For instance, we cannot build any longer a two-branch representation like in Equation (2), and in turn this has further implications. The density of Z is now of the form

$$f(x) = k(\lambda_0)\, G_0(\lambda_0 + w(x;\lambda))\, f_0(x), \tag{4}$$

for a suitable normalizing constant $k(\lambda_0)$, which may depend not only on λ_0, but also on the other ingredients of the density, while Equation (1) has a fixed 2 here. The normalizing constant $k(\lambda_0)$ must be computed afresh for each specific choice of the ingredients of Equation (4) and this is feasible only in a limited number of cases; see Arnold and Beaver (2002) for a review which focuses in this direction. The idea of building distributions via a selection mechanism, which operates on a symmetric density, has been brought to a high level of generality by Arellano-Valle et al. (2006) to embrace a very wide set of families.

As predictable, one of the manageable cases of Equation (4) occurs in connection with the normal distribution. Specifically, in Equation (4), taking w linear, $f_0 = \phi$ and $G_0 = \Phi$, the standard normal density and distribution function, respectively, it leads to the so-called extended skew-normal density

$$f(x) = \frac{1}{\Phi\big(\lambda_0/\sqrt{1+\lambda^2}\big)}\, \Phi(\lambda_0 + \lambda\, x)\, \phi(x), \tag{5}$$

with real parameters λ and λ_0. An alternative parameterization in use replaces λ_0 with τ, where $\tau = \lambda_0/\sqrt{1+\lambda^2}$. In practical work, one would introduce also a location and a scale parameter, by considering an affine transformation of the underlying variable, but this is not necessary for our purposes.

Density given in Equation (5) is an extension of the normal density, whose expression has appeared several times in early literature, as we mentioned near the beginning of this section. The earliest occurrence of this type known to us is Birnbaum (1950), who considered the distribution of the score obtained by a student at some achievement test conditionally on the fact that the score of the same student has exceeded a given threshold at the admission test. Under the assumption of joint normality of the scores at the admission and at the achievement test, it is easy to see that this construction is equivalent to the above one leading to Equation (5).

While the development of Birnbaum (1950) is extremely interesting, his work was targeted to solve a problem in normal distribution theory, not to build a family of distributions alternative to or more general than the normal family to be used for data fitting. There exist other papers which exhibit early occurrences of Equation (5), or of its special case with $\lambda_0 = 0$, building either on the same underlying stochastic representation given in Equation (3) or on some alternative one; see Azzalini (2005, Subsection 2.3) for more information on these occurrences in the literature. The remark made for Birnbaum's work applies equally to these other constructions.

The above constructions fall within the range of weighted distributions examined by Rao (1985) and the related literature. For a random variable X with probability distributions $p(x;\theta)$ which may represent a density or a probability function, depending on the continuous or discrete nature of X, respectively, consider the new distribution

$$p_w(x;\lambda,\theta) = \frac{w(x;\lambda)\, p(x;\theta)}{\mathbb{E}[w(X;\lambda)]}, \tag{6}$$

where $w(x;\lambda)$ is a general non-negative weight function which reflects the effect of the sampling scheme, such that the distribution p_w of the sampled values differs from $p(x;\theta)$. For instance, sampling with probability proportional to size implies the weight function $|x|$. Formulation given in Equation (6) is enormously broad, as it formally allows us to view any possible distribution as a weighted version of any other distribution with the same or a wider support. It is the identification of the weight function, driven by the sampling scheme of the specific problem under consideration, which leads to a fruitful choice of the ingredients of Equation (6).

A sampling scheme considered by Rao (1985) involves two random variables, X and Y, with joint distribution $p(x,y;\theta)$ and a weight function w_2 which is a function of y only. Then, it follows that the marginal distribution of X is of type of Equation (6). Distribution given in Equation (5) falls in this framework, following the route of Birnbaum (1950). For a bivariate normal variable (X,Y) with standardized marginals and correlation ρ, consider the distribution of X conditionally on the fact that $Y > c$, for some threshold c. A simple calculation leads to Equation (5) with $\lambda_0 = -c/\sqrt{1-\rho^2}$ and $\lambda = \rho/\sqrt{1-\rho^2}$.

Notice that here the weight function $w_2(y)$ for $p(x, y; \theta)$ is given by the indicator function of the set (c, ∞), but this gives rise to a weight function of the form $\Phi(\lambda_0 + \lambda\, x)$ in Equation (6). This basic scheme can be extended to the case of a multivariate normal and further on to much more general formulations, which have been discussed in the literature recalled earlier.

In the construction just described and in its various generalizations, the weight function lies between 0 and 1, since it represents a probability. For this case of weight function, we have adopted the term 'selection mechanism' in agreement with the literature recalled at the beginning of this paper. In addition, the term matches well with the idea of 'selective sampling' in use in the distinct but connected literature related to Heckman (1979). Note that some authors use instead the term 'selection' when the weight $w(x; \lambda)$ is the indicator function of a set; however, this situation is more commonly referred to as a 'truncation', as in Rao (1985).

The aim of the present note is to illustrate the proposal presented in 1908 by Fernando de Helguero, a young Italian statistician very active in those years, whose work has been largely ignored for about a century. In two papers whose crucial passages are reported below, he presents a formulation for modelling non-normal frequency distributions, intended as an alternative to Pearson's system of curves, which was predominant at that time. This proposal hinges on the selection mechanism like in Equation (3) applied to a normal population, thus linking this formulation to a possible subject-matter interpretation. For this reason, in spite of the limitations to be described in the course of the paper, it seems to us that de Helguero must be considered the precursor of the present literature on hidden truncation models and related constructions.

2. Life and Scientific Profile

The content of this section is based on the commemorative works of Volterra et al. (1911) and Guerrieri (1972). De Helguero's papers, which will be mentioned shortly, have been reprinted in a book of collected works; see de Helguero (1972).

Fernando de Helguero was born in Pelago near Florence, in Italy, on 1st November 1880 to parents Alberto and Eugenia Bérenger. At the age of nine, fate struck him with the premature death of his mother. After completing primary school in Florence and secondary school in Massa Carrara, he entered the gymnasium of Perugia which he completed in 1899 with distinction. At the University of Rome, he studied mathematics and, in 1903, obtained the licentiateship with the *cum laude* qualification.

Right after completing his mathematics degree, de Helguero started research work under the guidance of Vito Volterra, while working as a secondary school teacher in 1903–1904 to support himself. His interests combined mathematics and biological sciences, and it is then natural that he immediately directed his attention to the recently born discipline of biometry. The connection with Volterra is then not surprising, given that the mathematical work of his mentor was strongly motivated by biological applications. Specifically, the initial research theme of de Helguero was the problem of mixtures of two normal populations, which was of much interest in the statistical literature of those years, with special emphasis on the question of dissection of a mixed population. His work, published in prestigious journals (see de Helguero, 1904b, 1905, 1906a), culminated with his main result in this area, which was an important simplification in the solution of the 9th degree polynomial equation, derived by Pearson using the method of moments, to obtain an estimate of a fitted mixture. The burden required for solving this polynomial equation was at that time a major hindrance to the practical use of Pearson's method of dissection, which has been described by Davenport (1904, p. 40) as "tedious and rarely applicable".

To better pursue his interest for biological sciences, he applied for a grant from the University of Rome, initially awarded for 1904–1905, which allowed him to stop teaching and take courses with the aim of obtaining a degree in natural sciences. The grant was confirmed for 1905–1906 and, while continuing studying natural sciences, he worked as an assistant of Giuseppe Sergi, Professor of Anthropology. In these years, he successfully completed all exams required for the degree in natural sciences, and started working on the final dissertation on biometrical methods applied to anthropology. Later, he managed to complete this dissertation but was unfortunately unable to discuss it in the final exam. This more biology-oriented side of his activity led to another stream of papers; see de Helguero (1904a,c, 1906b, 1907a, 1908a,b).

In 1906–1907, he returned to teaching, initially at the gymnasium of Asti, but he soon applied for jobs in other places. Of the various positions offered to him, he opted for the one at the Regia Scuola Normale of Messina. While teaching, he continued his own research work, tackling other questions. One of these is about a form of non-linear regression (see de Helguero, 1907b), and another one is a new approach to distribution theory (see de Helguero, 1909a,b). The latter theme is the focus of our interest, and it will be discussed extensively in the subsequent sections.

All this intense and fruitful work legitimated, in spite of his young age, his application for the position of Professor of Statistics at the University of Palermo. Unfortunately, before the evaluation panel of the competition released its report, fate struck him again, at the age of 28. On 28$^{\text{th}}$ December 1908, he was in Messina, working at some new publication although it was a vacation period, when the town was hit by the infamous earthquake that killed some 90,000 to 120,000 people, and Fernando de Helguero was one of them.

3. Perturbation of the Normal Distribution by Selection

3.1 Preliminary remarks

The rest of the present note focuses on two papers in which de Helguero puts forward an innovative formulation to build non-normal distributions. Specifically, these papers are de Helguero (1909a) and de Helguero (1909b), both published posthumously. For simplicity, we shall refer to them as ‘the proceedings paper’ and ‘the journal paper’, respectively. It appears that the author was unable to take care of the final details and proofs reading: both papers have a few obvious misprints and the figures mentioned in the text of the journal paper are missing. These papers are closely connected to each other and they jointly constitute a unique proposal. The journal paper is more extensive and generally more detailed, but the proceedings paper presents in full what de Helguero considered his chief formal development.

The proceedings paper refers to the talk presented at the IV Congress of Mathematicians, held in Rome in April 1908, and this is why we indicated 1908 as the date of de Helguero’s proposal. Browsing the three volumes of the congress proceedings (see Castelnuovo, 1909) is an interesting journey in the mathematical environment of those years, and we digress briefly in this direction.

In a sense, the congress programme was similar to those of the present days, featuring official speeches, committee work and technical talks, with the arrangement of the talks reflecting the *de-facto* hierarchy among the speakers. And there were many important speakers indeed: E. Borel, F. P. Cantelli, C. Carathéodory, G. Darboux, A. R. Forsyth, C. Jordan, A. Liapunoff, H. Minkowski, G. Peano, S. Newcomb, V. Volterra, E. Zermelo, and many others. Some other names were not so known at the time, but they will have become so in a few years, Corrado Gini from Motta di Livenza for one. However, there are also quite striking differences with respect to nowadays conferences. One is the inclusion in the scientific programme of non-technical talks on broad topics, about the long-range perspectives of mathematics and its connections with other disciplines. For instance, Lorentz' talk was on 'Le partage de l'énergie entre la matière pondérable et l'éther', Volterra talked on 'Le Matematiche in Italia nella seconda metà del secolo XIX', Poincaré's talk was about 'L'avenir des mathématiques' (although it was read by Darboux, because Poincaré could not attend the conference for health reasons), Stoermer talked 'Sur les trajectoires des corpuscules électrisés dans le champ d'un aimant élémentaire avec application aux aurores boréales'. Also, the portion of programme allocated to applied areas of mathematics, including statistics, was considerable. Another interesting feature is the range of languages: there were four official languages is use (English, French, German, and Italian), with a prevalence of French in the plenary sessions, but this mix of languages did not appear to prevent people from communicating. Another peculiar aspect is the list of participants itself where, next to the academics, we find representatives of categories, which would not be there in these days. First of all, quite a few participants were secondary school teachers; de Helguero was one of them, many came from Germany, but several from other countries as well. Even more surprising is to find attendants from totally unexpected categories for today standards; here are some examples: l'abbé de Montcheuil, il canonico di Sarzana, Prince Bonaparte from Paris, le Prince de Polignac (avec la Princesse de Polignac et M.$^{\text{lle}}$ A. de Polignac), the actuary G. Lembourg from Bruxelles, Comm. M. Guggenheim and family from Venice, the publishers A. Gauthier-Villars from Paris and U. Hoepli from Milan, Général M. Frolov from Genève, but several other surprising figures could be listed. The overall impression is a sense of cultural openness and interaction with the outside world, which seem to have been lost in the current industry-style research work.

Before proceeding with our main theme, it is convenient to recall briefly a few aspects on the statistical literature of those years. By the end of the 19th century, the normal distribution had already been firmly recognised as the key analytical formulation to approximate empirical frequency distributions of continuous variables, whence its name. At the same time, there was ample evidence that in many cases the normal distribution did not provide an adequate mathematical model for the observed data. Among the various contributions to this question, a special mention is due to the proposals of Gram-Charlier and of Edgeworth, which arises from formal asymptotic expansions of an arbitrary distribution. In practical work, the expansion is truncated to a few terms, leading to an expression which allows for non-normality of the observed data via a modification of the normal distribution based on the observed skewness and kurtosis of the data. In the same period, Karl Pearson published two authoritative papers tackling the same motivating problem from alternative mathematical viewpoints. One paper dealt with distributions arising from the mixture of two normal populations, and specifically with the problems of dissecting the two components; see Pearson (1894). The second one introduced Pearson's system of frequency distributions, later extended to the famous set of 12 families; see Pearson (1895).

Finally, it is worth stressing the importance given at that time to the problem of studying the population distribution. While presently statistics is more often concerned with the problem of finding the relationship between a variable (possibly multivariate) with another one, at Pearson's times the study of the population distribution was regarded as having its own genuine interest.

3.2 Motivation and general framework

In this historical context, de Helguero puts forward his alternative formulation to the problem of abnormal frequency distributions. His programme is set out as follows in the proceedings paper.

Sulla rappresentazione analitica delle curve abnormali

Il compito della statistica nelle sue varie applicazioni alle scienze economiche e biologiche non consiste solo nel determinare la legge di dipendenza dei diversi valori ed esprimerla con pochi numeri, ma anche nel fornire un aiuto allo studioso che vuole cercare le cause della variazione e le loro modificazioni. [. . .]

Invece le curve teoriche studiate dal PEARSON e dall'EDGEWORTH per la perequazione delle statistiche abnormali in materiale omogeneo, mentre dànno con molta approssimazione la legge di variazione (meglio della curva normale perché ne sono delle generalizzazioni), a mio avviso sono difettose in quanto si limitano a dirci che le cause infinitesime elementari della variazione sono *interdipendenti*. Nulla ci fanno sapere sulla legge di dipendenza, quasi nulla sulla relazione colla curva normale che pure deve essere considerata fondamentale.

Io penso che miglior aiuto per lo studioso potrebbero essere delle equazioni che supponessero una perturbazione della variabilità normale per opera di cause esterne.

On the analytical representation of abnormal curves

The duty of statistics in its various applications to economics and biology does not consist only in identifying the law of dependence of the different values and in expressing it with a few numbers, but also in providing some help to the scholar who wants to search the causes of the variation and their modifications. [. . .]

On the contrary, the theoretical curves studied by PEARSON and by EDGEWORTH for the regularization of abnormal statistics from homogeneous material, while they give with much approximation the law of variation (better than the normal curve because they are generalizations thereof), these are defective in my view because they only limit themselves to tell us that the infinitesimal elementary causes of variation are *interdependent*. They tell us nothing on the law of dependence, almost nothing on the connection with the normal curve, which must still be considered fundamental.

I think that a better help for the scholar could come from some equations which supposed a perturbation of normal variability produced by some external causes.

There are various points to be highlighted here. First of all, it is stated that statistics should not simply produce an adequate numerical fit to the observed data, but also provide an aid to explore the mechanism which generates the data. This important idea features currently in some modern authors, but we are not aware of similar conceptions in those years. From this requirement, it follows that formulations like those of Pearson and Edgeworth are unsatisfactory, because they arise as mere mathematical constructs.

Similarly to a number of his contemporaries, in de Helguero's view, the source of nonnormality is attributed to the lack on independence among the components, which contributes to the genesis of the phenomenon and is expressed by "the infinitesimal elementary causes of variation are *interdependent*". De Helguero's alternative proposal is based on the assumption that the normal distribution remains the one naturally arising, in the sense that the mechanism which underlies the data generation would produce a normal distribution, if some external action did not perturb it, leading to the observed abnormality.

This viewpoint is expressed even more explicitly in the passage of the journal paper reported below. In this passage, he sets out the type of perturbation of normality, which is supposed to operate via a selection mechanism, either removing some individuals lying on one side of the mean or promoting those on the other side.

La presente nota ha lo scopo di fornire a chi si serve della statistica come mezzo di analisi un metodo per interpretare le curve abnormali come deviazioni del tipo normale.

Supponiamo che le cause elementari siano quali le suppone la legge normale: siano cioè infinitesime, in numero infinito, ugualmente tendenti ad elevare od abbassare la media,

The present note aims at providing, to those who make use of statistics as a tool for investigation, a method for interpreting abnormal curves as deviations from the normal one.

Let us suppose that the elementary causes are as the normal law presumes them to be, that is to let them be infinitesimal, infinite in number, equally leading to increase or to deindipendenti; se esse agissero liberamente la seriazione risulterebbe normale. Ma può essere, e ciò spesso deve effettivamente avvenire, che ad esse si aggiungano altre cause *perturbatrici* che tendano ad eliminare gli individui che giacciono da una parte della media o favorire quelli che giacciono dall'altra. La curva sarà abnormale, asimmetrica.

crease the mean, independent. If they acted freely, the distribution would turn out to be normal. But it may happen, and indeed this must often take place, that other *perturbation* causes join in, removing individuals who are on one side of the mean or fostering those who are on the other side. The curve will be abnormal, asymmetrical.

In the journal paper, de Helguero is more specific about his critics to Pearson's system of curves, in the following form.

Esse lasciano molto a desiderare per parecchie ragioni. Sopra tutte grave mi sembra la seguente: non prestano sufficiente aiuto a chi per mezzo della statistica vuole studiare non la manifestazione del fenomeno ma le cause che lo hanno prodotto. Mi spiego con un esempio: quando il biologo ha verificato che la statura dei coscritti dell'esercito americano non segue la legge normale ma piuttosto con grande approssimazione il Tipo IV del Pearson, può studiare con molta esattezza la manifestazione esterna del fatto (calcolare la media, la variabilità, la asimmetria, trovare il valore dei parametri, ecc.), ma riguardo alle cause che lo hanno prodotto può dire solo che esse sono *interdipendenti* e nulla più; la stessa parola userebbe se invece si presentasse un'altra delle curve generalizzate del Pearson.

They leave much to be desired for several reasons. Above all, the following one seems crucial to me: they do not provide adequate help if one intends to examine via statistics not the outcome of the phenomenon but the causes which have produced it. I explain with an example: when the biologist has checked that the height of the conscripts of the American army does not follow the normal law but rather with excellent approximation the Pearson Type IV distribution, he can study with high accuracy the external outcome of the fact (computing the mean, the variability, the skewness, finding the values of the parameters, etc.). But about the causes which have produced it, he can only say that they are *interdependent* and nothing more. The same word would be used if another of Pearson's generalized curves appeared instead.

3.3 Curves perturbed by selection

From the above remarks, the frame of the mathematical development follows quite naturally. We reproduce below the pertaining passage from the journal paper; the proceedings paper is almost identical on this point.

Le curve così ottenute sono *perturbate per selezione*: determiniamone l'equazione.

Sia
$$\frac{c}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{x-b}{\sigma}\right)^2},$$
la variazione normale *ipotetica*, come io la dirò, che risulterebbe senza la causa perturbatrice.

La probabilità che ha un individuo appartenente alla classe x di essere colpito dalla causa perturbatrice deve essere funzione di x, sia $\varphi(x)$. Nella classe x saranno allora colpiti $y\,\varphi(x)$ e rimarranno solo $y - y\,\varphi(x) = y[1-\varphi(x)]$, ossia la curva perturbata avrà l'equazione:
$$\frac{c}{\sigma\sqrt{2\pi}}\,[1-\varphi(x)]\, e^{-\frac{1}{2}\left(\frac{x-b}{\sigma}\right)^2}.$$

The curves so obtained are *perturbed by selection*. Let us determine their equation.

Let
$$\frac{c}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{x-b}{\sigma}\right)^2},$$
be the *hypothetical* normal variation, as I shall call it, which would result without the perturbation cause.

The probability that an individual belonging to class x is hit by the perturbation cause must be a function of x, say $\varphi(x)$. In class x there will be $y\,\varphi(x)$ affected [individuals] and only $y - y\,\varphi(x) = y[1-\varphi(x)]$ will remain, that is the perturbed curve will have the equation:
$$\frac{c}{\sigma\sqrt{2\pi}}\,[1-\varphi(x)]\, e^{-\frac{1}{2}\left(\frac{x-b}{\sigma}\right)^2}.$$

Rimane da stabilire la natura della funzione $\varphi(x)$. Essa rappresenta una probabilità, onde il suo valore deve essere compreso fra 0 ed 1. In questo scritto io suppongo la legge di selezione lineare: nulla vieta di fare delle ipotesi diverse (supponendola per es. esponenziale si giunge pure ad un tipo di equazione molto semplice), ma a questa ipotesi qui mi limito perchè a me sembra la più semplice ed importante.

Sia dunque $\varphi(x) = A(x-b) + B$, dove b è la media della variazione ipotetica; essa acquista il valore 0 per $x = b - B/A$ che dovrà perciò trovarsi fuori del campo di variazione se la legge è di semplice selezione; esso acquista il valore 1 per $x = b + (1-B)/A$ che rappresenterà il limite della variazione poichè per tale valore e per i succesivi la selezione colpisce tutti gli individui di ogni classe.

Sostituendo e ponendo $y_0 = c(1-B)$, $\alpha = -\sigma A/(1-B)$, si ha l'equazione

$$y = \frac{y_0}{\sigma\sqrt{2\pi}} \left(1 + \frac{\alpha(x-b)}{\sigma}\right) e^{-\frac{1}{2}\left(\frac{x-b}{\sigma}\right)^2}.$$

It remains to establish the nature of the function $\varphi(x)$. This represents a probability, hence its value must lie between 0 and 1. In this writing, I assume a linear selection law: nothing prevents us from making different assumptions (supposing, e.g., it to be exponential one arrives at a very simple form of equation), but here I restrict myself to this assumption because it seems to me the simplest and most important.

Let then be $\varphi(x) = A(x-b) + B$, where b is the mean of the hypothetical variation. This takes the value 0 when $x = b - B/A$, which therefore must lie outside the range of variation if the law is one of simple selection. It takes the value 1 when $x = b + (1-B)/A$, which represents the bound of variation because for this value and for those beyond it the selection hits all individuals of all classes.

On substituting and setting $y_0 = c(1-B)$, $\alpha = -\sigma A/(1-B)$, one gets the equation

$$y = \frac{y_0}{\sigma\sqrt{2\pi}} \left(1 + \frac{\alpha(x-b)}{\sigma}\right) e^{-\frac{1}{2}\left(\frac{x-b}{\sigma}\right)^2}.$$

The last expression is of the currently employed type as in Equation (4), once we notice that the second factor of the product is proportional to the distribution function of a uniform random variable. There is the irrelevant formal difference that, following a standard practice of that time, de Helguero worked with an unnormalized frequency distribution integrating to the sample size instead of one. Also, the above constructive argument coincides with the selection mechanism currently considered in connection with Equations (1) and (4). These facts support the view of de Helguero as the precursor of the current literature linked to Equations (1) and (4).

The following step is to find the four coefficients, y_0, b, σ and α, which represent the normalizing constant, the mean and the standard deviation of the hypothetical normal distribution and the coefficient of perturbation, respectively. As already mentioned, here normalization is to be intended as equalizing the integral of the curve to the number of observations. The process involves computing the moments from order 0 (which he calls "the area") to 3, equating the theoretical moments to the observed ones, and solving the equations with respect to the desired coefficients.

In order to compute the required moments, de Helguero considers two procedures: one which he denotes approximate, described in the journal paper, and another one he denotes exact, described in the proceedings paper, which we shall examine in more detail in the sequel. However, in both variants, he proceeds in a somewhat different way from the original plan, since he takes into consideration only the constraint $1 - \varphi(x) > 0$, ignoring the other one, $1 - \varphi(x) < 1$. As a consequence, he effectively works with the distribution

$$y = \begin{cases} 0, & \text{if } x \leq x_1; \\ \dfrac{y_0}{\sigma\sqrt{2\pi}} \left(1 + \dfrac{\alpha(x-b)}{\sigma}\right) \exp\left(-\dfrac{1}{2}\left(\dfrac{x-b}{\sigma}\right)^2\right), & \text{if } x_1 \leq x; \end{cases} \tag{7}$$

when $\alpha > 0$. The case $\alpha < 0$ is similar, except that the support set is $(-\infty, x_1]$. Notice that this revision of the initial model implies dropping one of the coefficients involved.

The omission of the constraint $1-\varphi(x) < 1$ is not discussed explicitly, but the motivation behind it seems to be found in a qualitative argument presented on p. 245 of the journal paper. Its essence is that the perturbation mechanism may not only decrease the number of individuals, but also increase it. For example, he says, let us consider the distribution of the wages of a set of workers, which in the first instance is taken to be of normal type. If the economic conditions of the country worsen, a certain number of them will move elsewhere searching for better conditions, and these will more likely be the workers with lower wages, hence producing a thinning in the lower tail of the distribution. Let us suppose instead that the owner of the factory needs more workers and he wants them to be skilled and industrious. Then, they are likely to have wages in the upper tail of the distribution, which will become thicker, and it is plausible that the increase of individuals at a given level x, although proportional to the number of existing individuals, has a proportionality constant which increases linearly with x. Hence, de Helguero concludes that the same mathematical formulation introduced earlier is capable of describing also the new one, and in fact it can handle even the coexistence of both types of perturbation.

Clearly, there is a zero lower limit for the thinning process of a tail, but there is no corresponding limit for its thickening. In addition, the fact of working with absolute frequency distributions seems to us to facilitate this type of reasoning. Finally, distribution given in Equation (7) has a similar support to the one of the Pearson type IV distribution, which de Helguero repeatedly mentioned as frequently arising in biological applications.

To compute the moment ν_n of order n of Equation (7), after translating it to $b = 0$, de Helguero assumes without loss of generality that $\alpha > 0$. Consider the integral

$$I_n = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\frac{\sigma}{\alpha}}^{\infty} x^n \, e^{-\frac{x^2}{2\sigma^2}} \, \mathrm{d}x,$$

such that I_0 is the standard normal distribution function evaluated at $1/\alpha$ and

$$I_1 = \sigma \, z\left(\frac{1}{\alpha}\right),$$

where z denotes the standard normal density. From the recurrence relationship

$$I_n = \sigma \left(-\frac{\sigma}{\alpha}\right)^{n-1} z\left(\frac{1}{\alpha}\right) + (n-1)\sigma^2 \, I_{n-2},$$

he obtains that

$$\nu_n = y_0 \left(I_n + \frac{\alpha}{\sigma} I_{n+1}\right), \qquad n = 0, \ldots, 3.$$

From here, the expressions of the moments of the normalized distribution up to order 3 are obtained. Specifically, after re-shifting the distribution back to location b, one arrives at the normalizing factor

$$y_0 = \nu_0 \, \frac{1}{I_0 + \alpha \, z(\alpha^{-1})},$$

where the area ν_0 can be set equal to 1 to adopt the convention in use nowadays, and

$$b = \mu_1 - \sigma\, H^{-1}\,,$$
$$\sigma^2 = \mu_2\, \frac{H^2}{2\,H^2 - \alpha^{-1}H - 1}\,,$$
$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{\left(\alpha^{-1}FH^2 + 3FH - 2\right)^2}{\left(2\,H^2 - \alpha^{-1}H - 1\right)^3},$$

where μ_1 is the mean value of the observed distribution, μ_2 and μ_3 are its central moments of order 2 and 3, respectively, and

$$F = \frac{z(\alpha^{-1})}{I_0}, \qquad H = \frac{1}{\alpha} + F\,.$$

To obtain estimates of b, σ and α, de Helguero replaces μ_1, μ_2 and μ_3 by their sample counterparts and solve the above equations. Since these equations are highly non-linear, their solution is accomplished with the aid of suitable tables provided in the paper. These tables relate α with the Pearson index of skewness

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3},$$

across the range $\alpha = 0.30(0.01)1.00$ and $1/\alpha = 1.00(0.01)0.00$. The tables provide also, for each given α, the log-transformed values of the non-linear functions of α appearing as the right-most term of the above expressions of σ^2 and b.

All subsequent steps after the drop of the constraint $1 - \varphi(x) < 1$ are coherent with this revised model. Hence, Equation (7) is properly normalized and its moments are given correctly. Consequently, the estimation procedure based on the method of moments delivers consistent estimates.

The procedure which we have summarized is the one that de Helguero denotes as exact, while in the approximate procedure, described in the journal paper, the integrals I_n are computed by adopting the second branch of Equation (7) over the whole range of x. This approximation leads to simpler expressions for ν_n, but their use is recommended only for small values of α.

Both procedures are illustrated with a range of real data, with more space allocated to this numerical work in the journal paper. Most of these examples have previously been examined by Pearson (1895). An additional dataset refers to the distribution of the wages of Belgian mine workers in 1896 and 1900. This last example is discussed somewhat more extensively than the others, and the comparison of the parameters at the two time points suggests an interesting conjecture on the underlying mechanism of the evolution of the wages.

3.4 Retaining the original constraints

We examine what would have been obtained retaining both constraints $0 < \varphi(x) < 1$. Recall that in the de Helguero's construction the parameters A and B of $\varphi(x)$ are such that the intersection points of $\varphi(x)$ with 0 and 1 fall outside the range of variation of the data, which implies that $0 < B < 1$.

On setting

$$y_0 = c(1-B), \qquad \alpha = -\sigma\,\frac{A}{1-B}, \qquad \beta = -\sigma\,\frac{A}{B},$$

the points where $\varphi(x)$ takes on the value 0 and 1 can be written as

$$x_0 = b - \frac{B}{A} = b + \frac{\sigma}{\beta}, \qquad x_1 = b + \frac{1-B}{A} = b - \frac{\sigma}{\alpha}\,,$$

respectively. Here β represents an additional parameter. This is required since φ was originally written as a two-parameter function. Hence, clearly, it cannot be re-written as a function of α only.

In the case $\alpha > 0$, we have $x_1 < x_0$ and the density is

$$y = \begin{cases} 0, & \text{if } x \le x_1; \\ \dfrac{\beta}{\alpha+\beta}\,\dfrac{c}{\sigma\sqrt{2\pi}}\left(1+\dfrac{\alpha(x-b)}{\sigma}\right)\exp\left(-\dfrac{1}{2}\left(\dfrac{x-b}{\sigma}\right)^2\right), & \text{if } x_1 \le x \le x_0; \\ \dfrac{c}{\sigma\sqrt{2\pi}}\exp\left(-\dfrac{1}{2}\left(\dfrac{x-b}{\sigma}\right)^2\right), & \text{if } x \ge x_0; \end{cases} \tag{8}$$

where we have taken $\varphi(x) = 0$ for $x > x_0$ by continuity and monotonicity. If $\alpha < 0$, then $x_0 < x_1$ and all inequalities of Equation (8) must be reversed. On defining

$$I_n(\xi) = \int_\xi^\infty x^n\,\frac{\exp(-\frac{1}{2}(x/\sigma)^2)}{\sigma\sqrt{2\pi}}\,\mathrm{d}x,$$

and denoting now by ν_n the unnormalized nth order moment of the density given in Equation (8) shifted to $b = 0$, one arrives at

$$\begin{aligned}\nu_0 &= c\left\{\frac{\beta}{\alpha+\beta}\left[I_0(x_1) - I_0(x_0) + \alpha\sigma^{-1}\left(I_1(x_1) - I_1(x_0)\right)\right] + I_0(x_0)\right\} \\ &= \frac{c}{\alpha+\beta}\{\alpha\Phi(-\beta^{-1}) + \beta\Phi(\alpha^{-1}) + \alpha\beta[z(\alpha^{-1}) - z(\beta^{-1})]\},\end{aligned}$$

and similar expressions can be obtained for the subsequent terms ν_n.

In the present days, we want a density normalized to one, so we set $\nu_0 = 1$ and get the corresponding value of c as a function of α and β. In the special case $\alpha = \beta$, we obtain

$$\frac{\beta}{\alpha+\beta} = \frac{1}{2}, \qquad \nu_0 = \frac{c}{2},$$

so that $c = 2$ when $\nu_0 = 1$. This lends a density like in Equation (1) where, up to the b shift, the normal density in Equation (8) is multiplied by the distribution function of a random variable uniform on the interval $(-\sigma/\alpha, \sigma/\alpha)$.

It is easy to reparameterize Equation (8), when $b = 0$ and $\sigma = 1$, into the form of Equation (4), where f_0 is the $\mathrm{N}(0,1)$ density, G_0 is the distribution function of a $\mathrm{U}(-1/2,\,1/2)$ variate and the parameters in the argument of G_0 in Equation (4) are given by

$$\lambda_0 = \frac{\beta-\alpha}{2\,(\alpha+\beta)}, \qquad \lambda = \frac{\alpha\,\beta}{\alpha+\beta}\,.$$

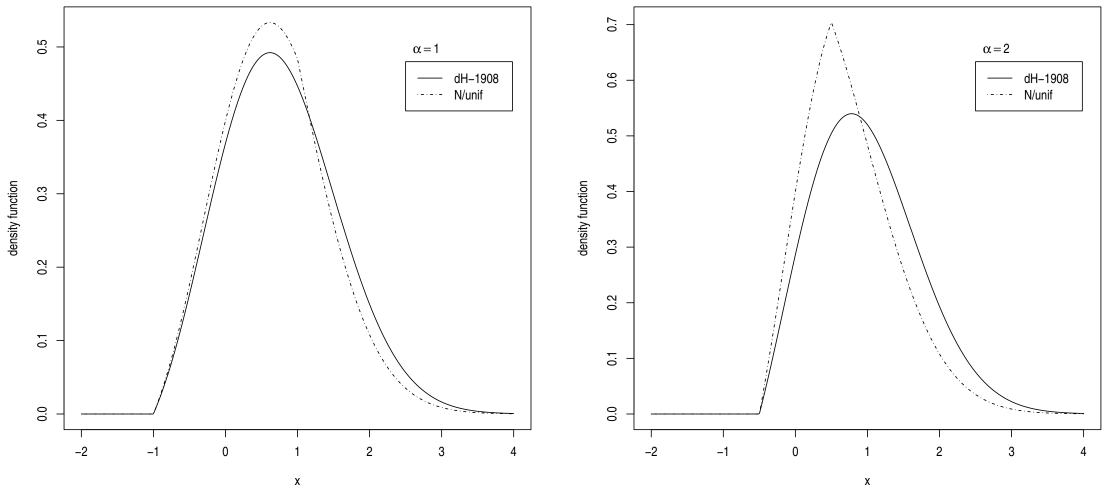

Figure 1. The continuous line represents de Helguero's curve given in Equation (7), the dashed line is the density given in Equation (8) in the symmetric interval case with $\alpha = \beta$, with $\sigma = 1$ in both cases. Left panel: $\alpha = 1$. Right panel: $\alpha = 2$.

Figure 1 displays the graphical appearance of Equation (7) and of the symmetric interval case of Equation (8), when $\alpha = \beta = 1$ and $\alpha = \beta = 2$, and in both cases $\sigma = 1$. For $\alpha = 1$, the curves are quite similar, while for $\alpha = 2$ there is a marked difference, with Equation (7) exhibiting a smooth behaviour over the whole support, while Equation (8) is noticeably spiky at right end-point of the interval $(-\sigma/\alpha, \sigma/\alpha)$.

4. Discussion

Let us recapitulate the main steps of de Helguero's formulation. There is the qualitative requirement that a theoretical frequency curve should ideally not only be able to fit adequately the observed frequency distribution, but it should also provide some help to understand the phenomenon which regulates the data generation. It is assumed that the mechanism underlying the data would produce a normal distribution if some external agents did not perturb it, producing non-normality of the actually observed distribution. The initial formulation of de Helguero postulated a selection mechanism where the probability of censoring an observation x depends on the value of x via a function $\varphi(x)$. The same function $\varphi(x)$ is later viewed as a way for representing not only a censoring mechanism, but also an increase in the number of individuals with values x. The relationship between x and $\varphi(x)$ is examined in detail in the linear case $\varphi(x) = A(x - b) + B$, but this choice is made for simplicity and other type of functions can be considered, notably an exponential function. After a distribution of this kind, and hence a specific function $\varphi(x)$, has been fit to the data, the interpretation of the estimated $\varphi(x)$ is driven by subject-matter considerations, possibly with more than one tentative interpretation.

From the point of view of the current literature on skew-symmetric distributions, it is clearly the initial formulation of de Helguero that is the interesting point. Even, if he later combines it with another mechanism leading him on a slightly different route, the key fact remains that the initial idea lucidly encapsulates the driving concept of the current skew-symmetric construction. Another point to be stressed is his pursue of a formulation capable of offering a subject-matter interpretation. Note also that in the second stage of the formulation, where the requirement that $\varphi(x)$ is a probability is dropped, Equation (7) can be viewed as a distribution like in Equation (6) with weight function w not constrained in $(0, 1)$.

There is ample evidence that, if de Helguero's life had not been interrupted so prematurely, he would have made a major contribution to the development of statistics, both for the discipline as a whole and for its specific development in Italy. The loss of his potential contribution to the discipline was immediately perceived by his contemporaries, and it is witnessed by the commemorative work of Volterra et al. (1911), both in the very fact that this initiative was taken as well as in the highly appreciative statements made in there. Of these, we only mention that Vito Volterra reported on a letter written to him by Karl Pearson when he heard of de Helguero's death. In this letter, Pearson expressed his high consideration for de Helguero, who had been among the first people in Italy to recognise the importance of the new statistical methods, in which he had seen wide opportunities for a mathematician. Also, Pearson manifested the convincement that, if de Helguero had survived, he would have made important contributions to the progress of biometry, the discipline to which he was so dedicated and whose principles he mastered perfectly.

In retrospect, the impact of de Helguero's death emerges today even deeper than it appeared at his times. From the specific viewpoint of the theme which we have discussed in this note, his approach was not followed up by anyone, and in fact this formulation passed unnoticed for the rest of the century, with the only exception of Mengarini (1909), as far as we know. However, the loss was much greater from the broad perspective of the development of statistics and biometry in Italy. De Helguero appeared to be in fairly regular contact with Karl Pearson and he shared conceptual framework, research themes and methodology with the British statistical environment. This was quite different from the Italian one, definitely focused on applications in social sciences and generally reluctant to adopt the new ideas coming from Pearson and colleagues, both on the side of methodology and on the application side of natural sciences. It is therefore no surprise that Benini in his contribution to Volterra et al. (1911) acknowledged that it would have been very doubtful that de Helguero could have obtained the chair of statistics in Palermo he had applied for. In spite of the unfavourable academic environment, de Helguero was pursuing his mission with energetic commitment and there is little doubt that in the long run he could have managed to narrow the gap between the Italian and British statistical environments. In other words, he had started acting like a bridge between the two cultures, and we can trust that he would have continued in the future, if he had been given the chance. With his death, this connecting bridge collapsed and there was no other Italian statistician playing a similar role. In the subsequent decades, Italian statistics strengthened its peculiar characteristics, marking its distance from the approach stemming from the Anglo-Saxon and other cooperating schools, as proudly maintained by Gini (1926, 1965). Holding these peculiar characteristics meant a long separation period of Italian statistics from the mainstream evolution of the discipline.

Acknowledgements

We are grateful to three referees for comments which led to a better presentation of the material.

References

Arellano-Valle, R.B., Branco, M.D., Genton, M.G., 2006. A unified view on skewed distributions arising from selections. Canadian Journal of Statistics, 34, 581–601.

Arnold, B.C., Beaver, R.J., 2002. Skewed multivariate models related to hidden truncation and/or selective reporting (with discussion). Test, 11, 7–54.

Azzalini, A., 2005. The skew-normal distribution and related multivariate families (with discussion). Scandinavian Journal of Statistics, 32, 159–188.

Azzalini, A., 2011. Skew-symmetric families of distributions. In Lovric, M., (ed.). International Encyclopedia of Statistical Sciences. Volume 19. Springer, New York, pp. 1344–1346.

Birnbaum, Z.W., 1950. Effect of linear truncation on a multinormal population. The Annals of Mathematical Statistics, 21, 272–279.

Castelnuovo, G., (ed.) 1909. Atti del IV Congresso Internazionale dei Matematici. 6-11 Aprile 1908, Roma. Tipografia della R. Accademia dei Lincei. Available at `http://www.mathunion.org/ICM`.

Davenport, C.B., 1904. Statistical Methods: With Special Reference to Biological Variation (second, revised printing edition). John Wiley & Sons, New York.

de Helguero, F., 1904a. Grandezza e forma degli organismi in somatometria. Atti della Società Romana di Antropologia, XI, 17–26.

de Helguero, F., 1904b. Sui massimi delle curve dimorfiche. Biometrika, 3, 84–98.

de Helguero, F., 1904c. Variazioni del numero di fiori ligulari del Bellis perennis. Bollettino dell'Orto Botanico di Napoli, 2, 133–144.

de Helguero, F., 1905. Per la risoluzione delle curve dimorfiche (followed by a corrigendum). Biometrika, 4, 230–231.

de Helguero, F., 1906a. Per la risoluzione delle curve dimorfiche. Reale Accademia dei Lincei – Memorie della Classe di Scienze Fisiche, Matematiche e Naturali, 5, 164–203.

de Helguero, F., 1906b. Variazione ed omotiposi nelle infiorescenze di Cichorium Intybus L. Biometrika, 5, 184–189.

de Helguero, F., 1907a. Il valore delle differenze sessuali dal punto di vista biometrico. Atti della Società Romana di Antropologia, XIII, 87–96.

de Helguero, F., 1907b. Sulla determinazione dei parametri di alcune funzioni per mezzo dei dati sperimentali. Nuovo Cimento, 5, 1–7.

de Helguero, F., 1908a. Dell'influenza del gozzo sulle statistiche della statura. Atti della Società Romana di Antropologia, XIV, 61–69.

de Helguero, F., 1908b. Interpretazione fillotassica dei poligoni fitostatistici. In Borzì, A., (ed.). Contribuzioni alla Biologia Vegetale. Volume IV, parte II. Palermo, pp. 169–191.

de Helguero, F., 1909a. Sulla rappresentazione analitica delle curve abnormali. In Castelnuovo, G. (ed.). Atti del IV Congresso Internazionale dei Matematici (Roma, 6–11 Aprile 1908), volume III, sezione III-B, pp. 288–299. Reale Accademia dei Lincei. Available at `http://www.mathunion.org/ICM/ICM1908.3/Main/icm1908.3.0288.0299.ocr.pdf`.

de Helguero, F., 1909b. Sulla rappresentazione analitica delle curve statistiche. Giornale degli Economisti, 38, 241–265.

de Helguero, F., 1972. Scritti vari. Padova: Cedam. Collected papers published by the "Istituto di Studi Economici, Finanziari e Statistici dell'Università di Roma". Volume 12.

Genton, M.G., (ed.) 2004. Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality. Chapman & Hall/CRC, Boca Raton, FL.

Gini, C., 1926. The contributions of Italy to modern statistical methods. Journal of The Royal Statistical Society, 89, 703–724.

Gini, C., 1965. On the characteristics of Italian statistics. Journal of The Royal Statistical Society Series A - Statistics In Society, 128, 89–109.

Guerrieri, G., 1972. Fernando de Helguero: la figura e l'opera scientifica. Giornale degli Economisti e Annali di Economia, 11/12, 806–834.

Heckman, J.J., 1979. Sample selection bias as a specification error. Econometrica, 47, 153–161.

Mengarini, P., 1909. Sulla distribuzione per salario dei minatori di carbone del Belgio (1896–1900). Giornale degli Economisti, 39, 183–196.

Pearson, K., 1894. Contributions to the mathematical theory of evolution. Philosophical Transactions of the Royal Society A: Mathematical, Physical & Engineering Sciences, 185, 71–110.

Pearson, K., 1895. Contributions to the mathematical theory of evolution. II. Skew variation in homogeneous material. Philosophical Transactions of the Royal Society A: Mathematical, Physical & Engineering Sciences, 186, 343–414.

Rao, C.R., 1985. Weighted distributions arising out of methods of ascertainment: what population does a sample represent? In Atkinson, A.C., Fienberg, S.E., (eds.). A Celebration of Statistics. Chapter 24. Springer-Verlag, New York, pp. 543–569.

Volterra, V., de Helguero, A., Gallardo, A., Reina, V., Benini, R., Sergi, G., 1911. In memoria di Fernando de Helguero. Tipografia Nazionale G. Bertero & C., Roma.