

An Application of the Skew-Normal Distribution to the Study of the Defective Hearing Function

*Un'Applicazione della Distribuzione Normale Asimmetrica
allo Studio della Funzione di Perdita Uditiva*

Corrado Crocetta

Dipartimento di Scienze Economiche, Matematiche e Statistiche, Università di Foggia
Via IV Novembre 1, - 71100 Foggia, e-mail: c.crocetta@unifg.it

Nicola Loperfido

Istituto di Scienze Economiche, Università degli Studi di Urbino
Via Saffi, 42 – 61029 (PU) e-mail: nicola@econ.uniurb.it

Riassunto: Nel presente lavoro si dimostra che la media ponderata di 2 statistiche d'ordine corrispondenti ad un'osservazione da una popolazione normale bivariata ha una distribuzione Normale Asimmetrica. Tale risultato trova importanti applicazioni pratiche per lo studio della funzione di perdita uditiva e dell'acuità visiva.

Keywords: Skew-Normal, L-statistics, Exact Distribution, Defective Hearing Function.

1. Introduction

Extreme value theory is mainly a model building tool, but it can also be utilized in statistical evaluations. It concerns the largest and the smallest in a set of random variables. This theory has many interesting application as for example the measure of defective hearing function.

Total deafness segregate the patient and do not allows many important social activity. For this reason the biological damage for total deafness is evaluated 50% whereas the biological damage for deafness in one hear is estimated only 10%.

To evaluate the invalidity percentage I_l doctors usually use a particular L-statistics of the form:

$$I_l = \frac{4 \cdot Y_1 + Y_2}{5} \cdot 0,5 ,$$

where Y_l is the score given the deafness of the best ear and Y_2 is the score given to the worst ear (of the same patient).

This paper provides some useful results in order to estimate the exact distribution of I_l . If we assume that the scores (not ordered) given to deafness of the left and right hear have a Bivariate Normal distribution we shall prove that I_l has a Skew Normal distribution with pdf

$$f(x; \mu, \psi, \lambda) = \frac{2}{\psi} \varphi\left(\frac{x - \mu}{\psi}\right) \Phi\left(\lambda \frac{x - \mu}{\psi}\right) \quad x, \lambda, \mu \in R; \quad \psi \in R^+ , \quad (1)$$

where $\varphi(\cdot)$ and $\Phi(\cdot)$ denote the Standard Normal density function and distribution function, respectively (Azzalini, 1985).

When (1) is the density of a random variable X we write $X \sim SN(\mu, \psi, \lambda)$. The parameter λ controls skewness, which is positive when $\lambda > 0$ and negative when $\lambda < 0$. When $\lambda = 0$ we have $X \sim N(\mu, \psi)$. Despite skewness, these distributions resemble the normal ones in several ways: they are unimodal, their support is the real line and the square of a variable whose distribution is Skew-Normal follows a Chi-Squared distribution. This generalization of the normal law has the advantage of being mathematically tractable and easily interpretable.

2. Main result

The following theorem generalizes results in Roberts (1966) and Loperfido (2002). It shows that any weighted average of the extremes of an exchangeable and bivariate normal random vector is Skew-Normal.

THEOREM: Let X and Y be two random variables whose joint distribution is normal with parameters $E(X)=E(Y)=\mu$; $V(X)=V(Y)=\sigma^2$; $\text{Corr}(X;Y)=\rho$. Let Z the average of the minimum and the maximum between X and Y , with weight α and $1-\alpha$ respectively: $Z=\alpha \cdot \min(X;Y)+(1-\alpha) \cdot \max(X;Y)$ ($0 \leq \alpha \leq 1$). Then Z has a Skew-Normal distribution $SN(\mu; \psi; \lambda)$, where

$$\psi^2 = \frac{\sigma^2}{2} \left[\sqrt{1+\rho} + (1-2\alpha)\sqrt{1+\rho} \right]^2 \quad \lambda = (1-2\alpha) \sqrt{\frac{1-\rho}{1+\rho}}.$$

PROOF:

By definition, $\min(X, Y)$ and $\max(X, Y)$ are the minimum and the maximum between X and Y , respectively. Using properties of minimum and maximum we can write:

$$\min(X, Y) = \left(\frac{X+Y}{2} \right) - \left| \frac{X-Y}{2} \right| \quad \max(X, Y) = \left(\frac{X+Y}{2} \right) + \left| \frac{X-Y}{2} \right|.$$

By definition Z is the weighted mean of $\min(X, Y)$ e $\max(X, Y)$ with weights α and $1-\alpha$ respectively. So

$$Z = \frac{X+Y}{2} + (1-2\alpha) \left| \frac{X-Y}{2} \right|,$$

that gives the following identity:

$$Z = \psi \left(\frac{X+Y-2\mu}{\sigma\sqrt{2+2\rho}} \cdot \frac{\sigma}{\psi} \cdot \sqrt{\frac{1+\rho}{2}} + (1-2\alpha) \cdot \frac{\sigma}{\psi} \cdot \sqrt{\frac{1-\rho}{2}} \cdot \left| \frac{X-Y}{\sigma\sqrt{2-2\rho}} \right| \right) + \mu.$$

Recalling the definition of ψ and λ we obtain:

$$Z = \psi \left(\frac{X + Y - 2\mu}{\sigma\sqrt{2+2\rho}} \cdot \frac{1}{\sqrt{1+\lambda^2}} + \frac{\lambda}{\sqrt{1+\lambda^2}} \cdot \left| \frac{X - Y}{\sigma\sqrt{2-2\rho}} \right| \right) + \mu .$$

By hypothesis the joint distribution of X and Y is bivariate normal with parameters $E(X)=E(Y)=\mu$; $V(X)=V(Y)=\sigma^2$; $Corr(X;Y)=\rho$. Using the properties of the multivariate normal we have

$$\begin{pmatrix} \frac{X + Y - 2\mu}{\sqrt{2\sigma^2 + 2\sigma^2 \cdot \rho}} \\ \frac{X - Y}{\sqrt{2\sigma^2 - 2\sigma^2 \cdot \rho}} \end{pmatrix} \sim N \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right] .$$

Azzalini (1986) and Henze (1986) show that the distribution of $\psi \left(U\sqrt{1-\delta^2} + \delta|W| \right) + \xi$ is Skew Normal $SN \left(\xi, \psi, \delta/\sqrt{1-\delta^2} \right)$ if

$$\begin{pmatrix} U \\ W \end{pmatrix} \sim N \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right] \quad \begin{matrix} -1 < \delta < 1 \\ \omega > 0 \end{matrix} .$$

We can then write $Z \sim SN(\mu, \psi, \lambda)$ and complete the proof.

This result is easily applicable to find the sampling distribution of I_l . If we take $\alpha = 4/5$ and denote by $Y_1 = \min(X, Y)$ and $Y_2 = \max(X, Y)$ respectively the minimum and the maximum of the deafness score given to a patient, we find that the exact distribution of I_l is Skew Normal.

The same result can be applied to a similar measure used in ophthalmology. To measure the individual's total vision impairment I_2 ophthalmologist use the L- statistic

$$I_2 = \frac{3 \cdot Y_1' + Y_2'}{4} ,$$

where Y_1' is the score of the eye with the best acuity and Y_2' is the one with the worst acuity.

When measuring visual acuity, attention focuses on the extremes of two normally distributed measures, together with one or more covariates. Olkin and Viana (1995) examine the correlation structure between worst acuity, best acuity and a covariate. They assume that acuities in eyes of the same individual are identically distributed and that their joint distribution is bivariate normal. Under these assumptions (Viana, 1998) best (worst) acuity is skew-normal (Roberts, 1966; Loperfido, 2002).

This result adds a new important plug in the puzzle of the exact distributions of linear combination of order statistics. A major limitation of the L-statistics for small samples is the intractability of appropriate sampling distributions.

The above theorem, that covers also samples from normal populations, represents a progress in that direction.

3. Numerical example

The following statistics are based on a sample of $N=10$ males working in the same factory that have been submitted to the hearing test using an increasing signal for five frequencies: 500, 1000, 2000, 3000 and 4000 Hz. The frequencies have a different importance in evaluating the hearing damage: 25%, 30%, 35%, 8% and 2%, respectively.

The invalidity percentage has been calculated using the index I_I described in the introduction. The index I_I is commonly used by the insurance companies to evaluate the invalidity percentage and to refund the biological damage.

Using the data collected we estimated the defective hearing function for the factory examined. Applying the result given in the previous paragraph we show that said function is Skew-Normal with mean $\mu=2,24$, scale parameter $\psi=7,42$ and shape parameter $\lambda=0,04$.

This result allows comparison between risks of biological damage of different type of activity and to estimate, in a more precise way, future claim of damages for different kind of factories. Using this theoretical result insurance companies will be able to foreseen the risk and the stock necessary to refund people that have been damaged.

The knowledge of the exact sampling distribution allow the use of all the statistical tools available for estimation of parameters, testing hypothesis, etc.

References

- Azzalini A. (1985) A class of distributions which includes the normal ones, *Scand. J. Statist.*, 12, 171-178.
- Azzalini A. (1986) Further results on a class of distributions which includes the normal ones, *Statistica XLVI*, 199-208.
- Bargagnana M., and Others (2000) Guida orientativa per la valutazione del danno biologico permanente, *Società Italiana di Medicina Legale e delle Assicurazioni*, 57, 62.
- Henze, N. (1986) A probabilistic representation of the 'skew-normal' distribution, *Scand. J. Statist.*, 13, 271-275.
- Loperfido N. (2002) Statistical implications of selectively reported inferential results. *Statist. & Prob. Lett.*, 56, 13-22.
- Viana M. (1998) Linear combinations of ordered symmetric observations with application to visual acuity, *Handbook of Statistics*, N. Balakrishnan & C. R. Rao (Eds.), Elsevier Science, 17, 513-524.
- Olkin I. and Viana M. (1995) Correlation analysis of extreme observations from a multivariate normal distribution, *J. Am. Statist. Assoc.*, 90, 1373-1379.
- Roberts C. (1966) A correlation model useful in the study of twins, *J. Am. Statist. Assoc.*, 61, 1184-1190.