Generalized Skew Normal Model

Ramesh C. Gupta
Department of Mathematics and Statistics,
University of Maine.

Rameshwar D. Gupta
Department of Computer Science and Applied Statistics,
University of New Brunswick, Canada.

Abstract

The skew normal distribution proposed by Azzalini (1985) can be a suitable model for the analysis of data exhibiting a unimodal density function having some skewness present, a structure often occurring in data analysis. In this paper, we study a generalization of the basic Azzalini model proposed by Balakrishnan, as a discussant of Arnold and Beaver (2002). The basic structural properties of the model including the reliability properties are presented. Estimation and testing of hypothesis of the skew parameter are discussed. Some comparisons of the models in terms of mean, variance and skewness are provided. Two data sets are analyzed.

Key Words: Normal distribution, skewness, log concave, stochastic ordering, maximum likelihood estimator.


1 Introduction

This paper is concerned with a generalization of a skew normal distribution. The probability density function of the skew normal distribution, proposed by Azzalini (1985) is given by

$$\phi (z; \lambda) = 2\phi (z) \Phi (\lambda z), \quad -\infty < z < \infty,$$  \hspace{1cm} (1.1)

where $\phi (z)$ and $\Phi (z)$ denote the $N(0, 1)$ density and distribution function, respectively. The parameter $\lambda$ varies in $(-\infty, \infty)$ and regulates the skewness and $\lambda = 0$ corresponds to the standard normal case. For $\lambda = 1$, it represents the distribution of the maximum of two independent standard

\*Correspondence to: Department of Mathematics and Statistics University of Maine Orono, ME 04469-5752. USA. E-mail: rcgupta@maine.edu
Received: February 2003; Accepted: July 2003
normal variables. In general, it is true that the maximum of two exchangeable random variables having bivariate normal distribution is a skew normal random variable. More specifically, for a standard bivariate normal with correlation coefficient $\rho$, the maximum has a skew normal distribution (1.1) with parameter $\sqrt{(1 - \rho)/(1 + \rho)}$. The density given by (1.1) enjoys a number of formal properties which resemble those of the normal distribution, for example if $Z$ has a pdf (1.1), then $Z^2$ has a chi-square distribution with one degree of freedom. From a practical point of view, the density (1.1) can be suitable for the analysis of data exhibiting a unimodal density but with some skewness present, a structure often occurring in data analysis. Some examples of unimodal densities with some skewness present include gamma, log gamma and log normal, see Klawless (1982). A motivation of the above model has been elegantly exhibited by Arnold et al. (1993) as follows. Suppose students admitted to a college are screened on their SAT scores and their progress is monitored with respect to their grade point average (GPA). Let $(X, Y)$ denote their (GPA, SAT). Assuming that $(X, Y)$ follow a bivariate normal distribution and assuming that only those students are admitted to college whose SAT scores are more than average, the distribution of $X$ follows a distribution whose standardized version is given by (1.1). In fact Arnold and Beaver (2000a, b) have given a general method of constructing a skewed distribution.

A multivariate version of (1.1) has been recently studied by Azzalini and Dalla Valle (1996) and Azzalini and Capitanio (1999). This distribution represents a mathematically tractable extension of the multivariate normal density with the addition of parameters to regulate skewness. These authors demonstrate that the multivariate skew-normal distribution has a reasonable flexibility in real data fitting, while it maintains some convenient formal properties of the normal density.

Recently, an excellent survey on skewed multivariate models has been presented by Arnold and Beaver (2002). This survey article contains comments by several discussants. One of the discussants, Balakrishnan proposed a generalization of Azzalini’s model as follows:

A random variable $Z$ is said to have a generalized skew normal distribution, $\text{GSN}(\lambda)$ if its probability density function is given by

$$
\phi_n(z; \lambda) = \frac{[\Phi(\lambda z)]^n \phi(z)}{C_n(\lambda)},
$$

(1.2)