Correlation and regression analysis of the mean and standard deviation of samples of two from a gamma population

Şerafettin Çabukoglu *

School of Business
North Carolina Central University
1801 Fayetteville Street
Durham, North Carolina 27707
U.S.A.

Abstract

It is well-known that a necessary and sufficient condition that a probability distribution be normal is that its mean and standard deviation be independent. Beyond this, little seems to be known about the relationship between the mean $\bar{x}$ and standard deviations s of samples drawn from non normal populations. The exponential distribution and more generally the gamma distribution adequately fit many distributions arising in technological application of statistics, particularly in the fields of reliability studies, life testing, queuing theory, wheatear analysis, and medicine. Inference for parameters of more than two gamma distributions is quite rare in the literature. Tripathi *et al* (1993) proposed a test for parameters of $m \geq 2$ gamma distributions based on a generalized minimum Chi-square procedure. For gamma populations with integer shape parameters $m \geq 2$, this paper derives, among other things, the regression function and the conditional distribution function of s given $X_n F(s \mid \bar{x})$. The latter is particularly important in determining whether the correct gamma distribution (as identified by its shape parameter) is used in connection with its (frequent) application to real world problems in diverse fields. For example, the gamma density function is used both as a p.d.f. and as a Bayesian prior density function in reliability analysis (Mann *et al* (1974), p. 127, p. 379). Mooley (1973) discusses the gamma model for summer monsoon rainfall in millimeters (See Bowman and Shenton (1988), p. 90.). Bordi *et al* (2001) discussed using gamma distribution for drought monitoring in the Mediterranean area. Masyma and Kuroiwa (1951) give data on the sedimentation rate during the period of normal pregnancy, and fit the gamma distribution to the data. (See Bowman and Shenton (1988), p. 90.) Amorosa (1925) used the gamma distribution in analyzing the distribution of income. And of course there are many other applications. But whatever the area or nature of application, the proper use of the gamma distribution depends heavily upon using the correct value of the shape parameter, which can be simply and easily tested by using the conditional distribution

* *E-mail*: `scabukoglu@wpo.nccu.edu`

Journal of Statistics & Management Systems
Vol. 9 (2006), No. 2, pp. 473–484

function of s given $\bar{x}$ to compare the value of s relative to that of $\bar{x}$. Furthermore, when using analytical models involving the mean and standard deviation of gamma distribution. It is necessary to know whether the correlation between them is sufficiently high that it cannot be required. Other interesting results obtained in this paper from the derived distribution $g(\bar{x}, s)$, are:

(1) the regression function is linear.

(2) the scedastic function is quadratic.

(3) for any integer value of the shape parameter m, the coefficient of variation for density function of $\bar{x}$ is $1/(2m)^{1/2}$.

(4) the correlation ratios for the shape parameters $m = 2, 3$ are, respectively, 0.70076 and 0.48564.

The author conjectures that values of the correlation ratios decrease with increasing values of m, but this has not been proven.

Keywords : *Joint distribution, conditional distribution, scedastic function, regression function, correlation ratio.*

1. Introduction

It is well-known that a necessary and sufficient condition for a distribution to be normal is that its mean and standard deviation be independent (Geary (1936)). Beyond this, little seems to be known about their relationship. For example, for what statistical distribution is the relationship linear, and what is the magnitude of the correlation?

The simplest gamma distribution — but nevertheless an important one — is the one for which the shape parameter has the value one. It is usually referred to as the exponential distribution, and is commonly used in life testing and in the reliability analysis of electronic components and systems. It has been discussed in some detail by Çabukoglu (2004). Johnson and Kotz (1970) provide a detailed review of the gamma distribution. Since extension of the analysis to cover the more general complex cases with shape parameter values greater than one unduly lengthens the paper. It was felt advisable to defer their treatment to a later paper. In the literature, inference for parameters of more than two distributions is rare. Bhattacharya (2002) tests general linear combinations for gamma distributions with parameters $m \geq 3$ against inequality restrictions.