Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

Journal of Statistical Planning and Inference 134 (2005) 434–444

journal of statistical planning and inference

www.elsevier.com/locate/jspi

The large-sample distribution of the most fundamental of statistical summaries☆

Arthur Pewsey*

Departamento de Matemáticas, Escuela Politécnica,Universidad de Extremadura, 10071 Cáceres, Spain

Received 4 May 2003; accepted 19 April 2004
Available online 3 August 2004

Abstract

We consider one of the most fundamental of statistical problems, namely that of inference for the mean, standard deviation and coefficients of skewness and kurtosis of an unknown univariate distribution. Assuming the distributional form of the parent population to be unknown, we focus our attention on moment-based inference. As is well-known, the method of moments estimates of the population measures under consideration are the sample mean, standard deviation and coefficients of skewness and kurtosis. Despite being some of the most frequently used of all statistical summaries, it comes as a surprise to find that their full joint distribution has not previously been studied in the literature. We derive a very general theoretical result for the large-sample asymptotic joint distribution of the four estimators and use simulation to explore the validity of the result as a means of approximating the biases, variances and covariances of the estimators for finite sample sizes. The theoretical result is then used to obtain asymptotically distribution-free inferential procedures for the population measures of original interest. Specifically, we propose and investigate the efficacy of bias-corrected and non-bias-corrected methods for point estimation and confidence set construction. We also discuss the relevance of the developed methodology both as an end in itself and as an aid to model formulation.

MSC: 62E20; 62G05; 62G15; 62H10

Keywords: Asymptotic joint distribution; Bias correction; Coefficient of kurtosis; Coefficient of skewness; Large-sample inference; Mean; Skew-normal distribution; Standard deviation

☆ Work supported financially by grant MOV02A044 of the Junta de Extremadura.

* Tel.: +34-927-257221; fax: +34-927-257203.
E-mail address: apewsey@guadiana.unex.es (A. Pewsey).

0378-3758/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2004.04.014

1. Introduction

Suppose we have a sample of data values drawn from some unknown univariate distribution and we wish to conduct inference for the population mean, standard deviation and coefficients of skewness and kurtosis, the summaries chosen as representative measures of the central location, dispersion, asymmetry and peakedness of the parent population. Using established notation, henceforth we will denote the four measures of interest by μ, $\sigma=\mu_2^{1/2}$, $\gamma_1=\beta_1=\mu_3/\mu_2^{3/2}$ and $\beta_2=\mu_4/\mu_2^2$, where $\beta_k=\mu_{k+2}/\mu_2^{(k+2)/2}$ and μ_k denotes the kth central moment about μ. At this early stage in the proceedings we stress that, in considering inference for γ_1, we are not tacitly assuming the parent population to be *symmetric*.

If, initially at least, we are unwilling to assume a specific distributional form for the parent population then, clearly, the problem cannot be tackled using the usual machinery of (parametric) likelihood-based inference. Nevertheless, as long as the appropriate moments can be assumed to exist, we can always use the method of moments as an inferential basis.

Considering the most basic form of inference first, the moment-based point estimates of μ, σ, γ_1 and β_2 are, as is common knowledge, their sample analogues, i.e. $\bar{x}$, $s=m_2^{1/2}$, $g_1=m_3/m_2^{3/2}$ and $b_2=m_4/m_2^2$, where m_k denotes the kth sample moment about $\bar{x}$. However, in order to conduct other forms of inference for the population measures of interest, we require results for the sampling distributions of these four statistics.

Various relevant distributional results appear within the literature. For instance, Kendall and Stuart (1977, p. 258) summarise long-established results for the expectation and variance of the sample mean and standard deviation. Cramér (1946, Chapters 27, 28) considers the large-sample asymptotic normality of well-behaved functions of central moments. His approach to deriving the asymptotic distribution of such functions based on Taylor expansion has become known in the statistical literature as the delta-method. See Mardia (1972, p. 111), Rao (1973, Chapter 6), Serfling (1980, Chapter 3), and Sen and Singer (1993, Section 3.4) for extensions of this technique. Cramér (1946, Chapter 27) obtained expressions for the asymptotic variances of g_1 and the coefficient of kurtosis excess $g_2=b_2-3$ (and hence implicitly that of b_2) in terms of the central moments of the parent population. In their Exercise 3.4.1, Sen and Singer (1993, p. 154) ask their readers to derive the asymptotic variances of g_1^2 and b_2 and their asymptotic covariance.

Nevertheless, although $\bar{x}$, s, g_1 and b_2 are some of the most fundamental and frequently used statistical summaries, it is somewhat surprising to find that general results for their full joint distribution do not appear in the literature. We speculate that this omission is a consequence of the predominant historical use of symmetric distributional forms in the modelling of data, the normal distribution having played a particularly prominent role in that endeavour.

In this article, we obtain a concise theoretical result for the large-sample asymptotic joint distribution of the sample mean, standard deviation and coefficients of skewness and kurtosis. The theorem providing the specific details of the result, together with its proof, is given in Section 2. In Section 3 we present the results of a simulation study designed to explore the validity of the approximation for finite sample sizes. As well as being of interest from a theoretical perspective, the result also provides a basis for conducting asymptotically distribution-free inference for the population measures of interest. In Section 4, we describe