Journal of Statistical Planning and Inference 193 (2018) 117–135

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

# Nonparametric density estimation for nonnegative data, using symmetrical-based inverse and reciprocal inverse Gaussian kernels through dual transformation

Yoshihide Kakizawa

*Faculty of Economics, Hokkaido University, Nishi 7, Kita 9, Kita-ku, Sapporo 060–0809, Japan*

ARTICLE INFO

*Article history:*
Received 11 February 2017
Received in revised form 13 June 2017
Accepted 22 August 2017
Available online 23 September 2017

*MSC 2010:*
primary 62G07
secondary 62G20

*Keywords:*
Density estimation
Boundary-bias-free
Symmetrical-based inverse Gaussian density
Dual transformation
Log-symmetrical density

ABSTRACT

The classical Birnbaum–Saunders (BS) distribution has recently been generalized in various ways to introduce flexible parametric models for nonnegative data, focusing on the parametric fitting. In this paper, a new symmetrical-based inverse/reciprocal inverse Gaussian density, through dual transformation, is applied to the context of nonparametric density estimation for nonnegative data. The beauty and importance of new density estimator lies in its general formulation via the density generator, including a log-symmetrical kernel density estimator. We provide sufficient conditions under which the proposed estimator has desirable asymptotic properties, and discuss the asymptotic comparison between the proposed estimator and the previous (normal-based) estimator.


## 1. Introduction

The Birnbaum–Saunders (BS) distribution was originally introduced by Birnbaum and Saunders (1969) as a failure time distribution on $[0,\infty) = \mathrm{R}_+$ (say). Its density, with a two-parameter $\alpha, \beta > 0$, is given by $k^{\mathrm{BS}}(s;\alpha,\beta) = \phi(a(s/\beta)/\alpha)A(s/\beta)/(\alpha\beta)$, where $\phi(u) = \exp(-u^2/2)/\sqrt{2\pi}$, $a(t) = t^{1/2} - t^{-1/2}$, and $A(t) = (1/2)(t^{-1/2} + t^{-3/2})$. Díaz García and Leiva (2005) replaced $\phi(u)$ by a symmetrical density $C_g\, g(u^2)$, $u \in \mathrm{R}$, and developed a generalized BS (we call "a symmetrical-based BS") density on $\mathrm{R}_+$, as follows:

$$k_g^{\mathrm{BS}}(s;\alpha,\beta) = C_g\, g\left(\frac{a^2(s/\beta)}{\alpha^2}\right)\frac{A(s/\beta)}{\alpha\beta}, \tag{1}$$

where $g \not\equiv 0$ is a nonnegative function on $\mathrm{R}_+$, called a density generator (see, e.g., Fang et al. (1990, page 35)), such that $1/C_g = \int_{\mathrm{R}} g(u^2)\,du = \int_{\mathrm{R}_+} y^{-1/2} g(y)\,dy$ is well-defined.

Mathematically, the identity $\int_{\mathrm{R}_+} g((s^{1/2}-s^{-1/2})^2/\alpha^2)s^{-1/2}\,ds = \int_{\mathrm{R}_+} g((t^{1/2}-t^{-1/2})^2/\alpha^2)t^{-3/2}\,dt$ $(s = 1/t)$ is helpful to understand the connection to other related densities on $\mathrm{R}_+$, defined by

$$k_g^{\mathrm{IG}}(s;\alpha,\beta) = C_g\, g\left(\frac{a^2(s/\beta)}{\alpha^2}\right)\frac{1}{\alpha\beta}\left(\frac{\beta}{s}\right)^{3/2}, \tag{2}$$

*E-mail address:* kakizawa@econ.hokudai.ac.jp.

http://dx.doi.org/10.1016/j.jspi.2017.08.008
0378-3758/

$$k_g^{\mathrm{RIG}}(s;\alpha,\beta)=C_g\,g\left(\frac{a^2(s/\beta)}{\alpha^2}\right)\frac{1}{\alpha\beta}\left(\frac{\beta}{s}\right)^{1/2}=\frac{s}{\beta}\,k_g^{\mathrm{IG}}(s;\alpha,\beta)\,. \tag{3}$$

Specifically, when $g(y)=\exp(-y/2)=g_{\mathrm{N}}(y)$ (say), these densities are long-standing and popular as the inverse Gaussian (IG) density and its complementary reciprocal (called "a reciprocal IG (RIG)"). Note that the classical IG and RIG densities, due to Tweedie (1957, (1a)–(1d)), were originally parameterized in several ways. Almost half a century later, the distribution having the density (2) was studied by Sanhueza et al. (2008a, Definition 1) as an IG type distribution, with a parameterization $(\alpha,\beta)=(\sqrt{\mu/\lambda},\mu)$, where $\mu,\lambda>0$; in this paper, it is called a symmetrical-based IG distribution, denoted by $\mathrm{IG}_g(\alpha,\beta)$. Similarly, the distribution having the density (3) is called a symmetrical-based RIG distribution, denoted by $\mathrm{RIG}_g(\alpha,\beta)$. In view of the right-hand side of (3), the parameter $\beta>0$ is the mean of the $\mathrm{IG}_g(\alpha,\beta)$ distribution, so that the density is sometimes known as a length-biased version in the literature. We observe that there is a relationship between the symmetrical-based BS and IG distributions, i.e., the density (1) is an equally weighted mixture of the densities (2) and (3); $k_g^{\mathrm{BS}}(s;\alpha,\beta)=(1/2)k_g^{\mathrm{IG}}(s;\alpha,\beta)+(1/2)k_g^{\mathrm{RIG}}(s;\alpha,\beta)$. More generally, a mixture distribution having the density

$$k_g^{\mathrm{MIG}_\epsilon}(s;\alpha,\beta)=(1-\epsilon)k_g^{\mathrm{IG}}(s;\alpha,\beta)+\epsilon k_g^{\mathrm{RIG}}(s;\alpha,\beta)\,, \tag{4}$$

where $\epsilon\in[0,1]$ is a mixing proportion, is called a symmetrical-based mixture IG (MIG) distribution, denoted by $\mathrm{MIG}_{g,\epsilon}(\alpha,\beta)$. See Leiva et al. (2010) with the parameterization $(\alpha,\beta)=(\sqrt{\mu/\lambda},\mu)$, where $\mu,\lambda>0$. The mixture distribution when $g(y)=g_{\mathrm{N}}(y)$ appeared in Jørgensen et al. (1991).

It should be remarked that, in this paper, the term "symmetrical-based IG" (rather than the term "generalized IG (GIG)") is used for the density (2), because, historically, the density on $\mathrm{R}_+$, with a three-parameter $\sigma,\omega>0$ and $\nu\in\mathrm{R}$, defined by

$$k_{\nu,\omega,\sigma}(s)=\frac{s^{\nu-1}}{2\sigma^{\nu}K_\nu(\omega)}\exp\left\{-\frac{\omega}{2}\left(\frac{s}{\sigma}+\frac{\sigma}{s}\right)\right\} \tag{5}$$

(see, e.g., Jørgensen (1982, page 1) with parameterization $(\sigma,\omega)=(\sqrt{\chi/\psi},\sqrt{\chi\psi})$, where $\chi,\psi>0$), is popularly known as a GIG density. Here, $K_\nu(\cdot)=K_{-\nu}(\cdot)$ is the modified Bessel function of the third kind and with index $\nu$. Such a GIG family of the densities contains the classical IG and RIG densities as special cases $\nu=-1/2$ and $1/2$, respectively; note that $K_{1/2}(\omega)=\{\pi/(2\omega)\}^{1/2}\mathrm{e}^{-\omega}$.

Many authors have introduced the above-mentioned models (1)–(4) to study the density/cumulative distribution/hazard functions, moments, transformations, generation of random numbers, and so on, mainly focusing on the parametric likelihood inference for nonnegative data. In this paper, using such flexible densities as kernels, we are concerned with nonparametric estimation of the probability density that has the support $\mathrm{R}_+$. In that case, the standard kernel density estimator (Rosenblatt, 1956) is, in general, inconsistent near the boundary, due to the so-called boundary bias (see, e.g., Wand and Jones (1995, Subsection 2.11)). Consequently, various remedies for removing the boundary bias have been suggested. Jones (1993) gave an extensive review of the boundary corrections (renormalization, reflection, and generalized jackknifing) until 1993. See Zhang et al. (1999) for more advanced reflection techniques in the 1990s. On the other hand, over the last decade, there has been a growing interest in the use of asymmetric kernel (AK) whose support matches the support of the density to be estimated. Chen (1999, 2000) did pioneering studies of nonparametric density estimation using gamma densities (or beta densities) as kernels, for nonnegative data (or the data from the unit interval). Scaillet (2004) proposed using IG and RIG densities, rather than gamma density. See also Jin and Kawczak (2003) for other applications of BS and log-normal (LN) densities. Igarashi and Kakizawa (2014) and Igarashi (2016) reformulated Jin and Kawczak's (2003) and Scaillet's (2004) estimators, by applying a weighted LN density and a mixture of modified Bessel (MMB) densities. It should be remarked that Igarashi and Kakizawa (2014) renamed the GIG density (5) as a modified Bessel density, noting that its normalizing constant involves only $K_\nu(\cdot)$; the modified Bessel function of the third kind.

In this paper, we further develop a general framework for boundary-bias-free AK density estimation. The contribution of this paper is fivefold: First, the densities (1)–(4) are further extended so that the resulting four-parameter density nests the log-symmetrical (LS) density as a special case. Here, the LS density, being a natural extension of the LN density $k^{\mathrm{LN}}(s;\mu,\sigma)=\exp\{-(\log s-\mu)^2/(2\sigma^2)\}/(\sqrt{2\pi\sigma^2}s)$ in a familiar form, is the univariate analogue of (multivariate) log-elliptical density (see Fang et al. (1990, Section 2.8)). Second, the present paper allows a unified treatment via density generator $g$. We provide a set of conditions of $g$ under which a family of the proposed estimators has desirable asymptotic properties. Third, a subfamily of the LS kernel density estimators is newly proposed, as an alternative to the LN kernel density estimator (Igarashi, 2016). Fourth, some existing estimators had an unrealistic problem of "a zero value at the origin", which is not suitable for estimating the density $f(0)>0$. We resolve such a bad definition of the previous estimators (Jin and Kawczak, 2003; Marchant et al., 2013). Fifth, the asymptotic comparison between the proposed estimator and the previous (normal-based) estimator is discussed in terms of the mean integrated squared error (MISE).

The rest of the paper is organized as follows. In Section 2, we introduce an additional parameter $q\in\mathrm{R}$ to create a new distribution with a four-parameter $(\alpha,\beta,\epsilon,q)$ and a density generator $g$, focusing on a broad applicability in the nonparametric density estimation for nonnegative data. In Section 3, we propose a new symmetrical-based $\mathrm{MIG}^{\langle q\rangle}$ kernel density estimator. Section 4 presents some asymptotic properties of the proposed estimator. In Section 5, we discuss the asymptotic comparison between new estimator and the previous (normal-based) estimator. Section 6 contains a simulation study to demonstrate its finite sample performance. Proofs of the results in Section 4 are given in the Appendix.