DENSITY ESTIMATION USING INVERSE AND RECIPROCAL INVERSE GAUSSIAN KERNELS

SCAILLET O.* [1]

* HEC Genève and FAME, Université de Genève, Bd Carl Vogt 102, CH - 1211 Genève 4, Suisse. olivier.scaillet@hec.unige.ch

This version: January 2003 (first version : June 2001)

Abstract

This paper introduces two new nonparametric estimators for probability density functions which have support on the non-negative real line. These kernel estimators are based on some inverse Gaussian and reciprocal inverse Gaussian probability density functions used as kernels. We show that they share the same properties as those of gamma kernel estimators: they are free of boundary bias, always non-negative and achieve the optimal rate of convergence for the mean integrated squared error. Monte Carlo results concerning finite sample properties are reported for different distributions and sample sizes.

Key words: Boundary bias, Inverse Gaussian kernel, Reciprocal inverse Gaussian kernel, Gamma kernel, Variable kernel, Density estimation.

JEL Classification: C13, C14. **MSC 2000:** 62G07, 62G08.

---

[1] We thank M. Akritas and the referee for very constructive criticism. We are grateful to I. Gijbels and T. Bouezmarni for many stimulating discussions. We have received fruitful comments from S. X. Chen, A. Chesher, S. Galluccio, M. Hagmann, O. Renault, K. Thompson, and participants at BNP Paribas seminar and the NP conference. The author gratefully acknowledges financial support from the Belgian Program on Interuniversity Poles of Attraction (PAI nb. P4/01) as well as from the Swiss National Science Foundation through the National Center of Competence: Financial Valuation and Risk Management. Part of this research was done when he was visiting THEMA and IRES.

1 Introduction

This paper considers estimation of a probability density function that has bounded support on $[0, \infty)$. Recently CHEN (2000) has proposed a nice way to circumvent the well known boundary bias or edge effect that appears in standard kernel density estimation. Boundary bias is due to weight allocation by the fixed symmetric kernel outside the density support when smoothing is carried out near the boundary. The remedy consists in replacing symmetric kernels by an asymmetric gamma kernel, which never assigns weight outside the support. In addition to nice asymptotic features, CHEN (2000) reports good finite sample performance of this cure through a simulation study.

Here we pursue this idea by proposing two new classes of density estimators. They rely on the use of inverse Gaussian (IG) and reciprocal inverse Gaussian (RIG) probability density functions as kernels in place of the gamma density function. The name 'inverse Gaussian' was introduced by TWEEDIE (1947) who noted the inverse relationship between cumulant generating functions of these distributions and those of Gaussian distributions. They are also known under the name 'Wald' distributions since the same class of distributions was derived by WALD (1947). The IG and RIG kernels have flexible shape and location on the non-negative real line. Their shapes are allowed to vary according to the position of the data points, thus changing the degree of smoothing in a natural way, and their support matches the support of the probability density function under estimation. As gamma kernel estimators, the IG and RIG kernel estimators are free of boundary bias, always non-negative, and achieve the optimal rate of convergence for the mean integrated squared error (MISE) within the class of non-negative kernel density estimators. Furthermore their variance reduces as the position where the smoothing is made moves away from the boundary. In contrast with the gamma kernel estimators, the IG and RIG kernel estimators avoid the presence of the first derivative of the probability density function in their bias. Let us further remark that the Weibull, non-central chi-square, Fisher, lognormal and Pareto distributions are not suitable for use as asymmetric kernels. This can be checked from the arguments needed to show our results (kernel behaviour at boundary, mean and variance of the distribution underlying the kernel specification).

The paper is organised as follows. In Section 2 we introduce the IG and RIG kernel estimators. We compute their bias, variance, optimal MSE and optimal MISE. A comparison is made with gamma kernel estimators. In Section 3 we report Monte Carlo results concerning the finite sample properties of the different asymmetric kernel estimators for various distributions and parameter values. Their performance is compared with the one of smooth optimum boundary and standard symmetric kernel estimators. Section 4 contains some concluding remarks. An appendix gathers technical details.

2 IG and RIG kernel estimators

Let $X_1, \dots X_n$ be a random sample from a distribution with an unknown probability density function f defined on $[0, \infty)$. We assume that f is twice continuously differentiable, and $\int_0^\infty (x^3 f''(x))^2 dx < \infty$.

Let $K_{IG(m,\lambda)}$ be the density of an $IG(m, \lambda)$ distributed random variable Y defined as:

$$K_{IG(m,\lambda)}(y) = \frac{\sqrt{\lambda}}{\sqrt{2\pi y^3}} \exp\left(-\frac{\lambda}{2m}\left(\frac{y}{m} - 2 + \frac{m}{y}\right)\right), \qquad y > 0.$$

The mean and variance of Y are equal to

$$E[Y] = m, \qquad \mathrm{Var}[Y] = \frac{m^3}{\lambda}.$$

The random variable $Z = 1/Y$ then follows an $RIG(m, \lambda)$ distribution whose density is:

$$K_{RIG(m,\lambda)}(z) = \frac{\sqrt{\lambda}}{\sqrt{2\pi z}} \exp\left(-\frac{\lambda}{2m}\left(mz - 2 + \frac{1}{mz}\right)\right), \qquad z > 0.$$

The mean and variance of Z are equal to

$$E[Z] = \frac{1}{m} + \frac{1}{\lambda}, \qquad \mathrm{Var}[Z] = \frac{1}{\lambda m} + \frac{2}{\lambda^2}.$$

The classes of IG and RIG kernels we consider are:

$$K_{IG(x,\frac{1}{b})}(u) = \frac{1}{\sqrt{2\pi b u^3}} \exp\left(-\frac{1}{2bx}\left(\frac{u}{x} - 2 + \frac{x}{u}\right)\right),$$

and

$$K_{RIG(\frac{1}{x-b},\frac{1}{b})}(u) = \frac{1}{\sqrt{2\pi b u}} \exp\left(-\frac{x-b}{2b}\left(\frac{u}{x-b} - 2 + \frac{x-b}{u}\right)\right),$$

where b is a smoothing parameter satisfying $b + 1/(bn) \to 0$ when n goes to infinity. The estimators of the pdf are

$$\hat{f}_{IG}(x) = n^{-1} \sum_{i=1}^{n} K_{IG(x,\frac{1}{b})}(X_i),$$

and

$$\hat{f}_{RIG}(x) = n^{-1} \sum_{i=1}^{n} K_{RIG(\frac{1}{x-b},\frac{1}{b})}(X_i).$$

These estimators are extremely easy to implement, and very similar to gamma kernel estimators. They are obtained after substitution of the IG and RIG kernels for the gamma kernels used by Chen (2000), namely either:

$$K_{Gam(x/b+1,b)}(u) = \frac{u^{x/b} e^{-u/b}}{b^{x/b+1} \Gamma(x/b+1)}, \qquad u > 0,$$

or

$$K_{Gam(\rho_b(x),b)}(u) = \frac{u^{\rho_b(x)-1} e^{-u/b}}{b^{\rho_b(x)} \Gamma(\rho_b(x))}, \qquad u > 0,$$

with

$$\rho_b(x) = \begin{cases} x/b & \text{if } \ x \geq 2b, \\ \frac{1}{4}(x/b)^2 + 1 & \text{if } \ x \in [0, 2b). \end{cases}$$

Figure 1 plots the shapes of the IG and RIG kernels, together with the shapes of the two gamma kernels for some selected values of x and $b = 0.2$. Let us remark that $K_{IG(x,1/b)}(u)$ tends to zero for all u as x approaches the boundary. This will induce the constraint $\hat{f}_{IG}(0) = 0$, which may be undesirable in some cases. At $x = 0$, $K_{RIG(1/(-b),1/b)}(u)$ tends to zero when u goes to zero, while $K_{Gam(1,b)}(0) = K_{Gam(\rho_b(0),b)}(0) = 1/b$. For $x > 0$, all kernels vanish at $u = 0$. In light of Figure 1, the RIG and second gamma kernels exhibit very similar shapes, except at $x = 0$, whereas the difference between the IG kernel and the first gamma kernel is more marked.

The first proposition is related to the bias of IG and RIG kernel estimators.

Proposition 1 (Bias)

The biases are equal to

$$Bias\{\hat{f}_{IG}(x)\} = \frac{1}{2} x^3 f''(x) b + o(b),$$

and

$$Bias\{\hat{f}_{RIG}(x)\} = \frac{1}{2}xf''(x)b + o(b).$$

The bias is larger, resp. smaller, for the IG kernel estimator for $x > 1$, resp. $x < 1$. The first gamma kernel estimator proposed by CHEN (2000) has a bias equal to $(f'(x) + \frac{1}{2}xf''(x))b$. His second gamma estimator shares the same bias as that of the RIG estimator when $x \geq 2b$, but for $x < 2b$, it involves f' since the bias is then equal to $(\rho_b(x) - x/b)f'(x)b$. Note that the first derivative f' is removed from the bias on the whole support in the IG and RIG kernel estimators which contrasts with the gamma kernel estimators and transformation kernel density estimators based on the logarithmic mapping. IG, RIG and gamma kernel estimators are all free of boundary bias since their bias is $O(b)$ in the interior as well as near the origin. The order of magnitude of the bias does not depend on the location within the density support. Finally, as $\int_0^\infty (x^3 f''(x))^2 dx < \infty$, $x^3 f''(x)$ and $xf''(x)$ converge to zero as $x \to \infty$. So the bias will be smaller as x increases.

Let us now examine the variance for $x \in (0, \infty)$ and take a strictly positive constant κ.

Proposition 2 (Variance)
i) For $x/b \to \infty$ (interior x), $x > 0$, the variances are equal to

$$Var[\hat{f}_{IG}(x)] = \frac{1}{2\sqrt{\pi}} n^{-1} b^{-1/2} x^{-3/2} f(x) + o(n^{-1} b^{-1/2}),$$

and

$$Var[\hat{f}_{RIG}(x)] = \frac{1}{2\sqrt{\pi}} n^{-1} b^{-1/2} x^{-1/2} f(x) + o(n^{-1} b^{-1/2}).$$

ii) For $x/b \to \kappa$ (boundary x), $x > 0$, the variances are equal to

$$Var[\hat{f}_{IG}(x)] = \frac{1}{2\sqrt{\pi}} n^{-1} b^{-2} \kappa^{-3/2} f(x) + o(n^{-1} b^{-2}),$$

and

$$Var[\hat{f}_{RIG}(x)] = \frac{1}{2\sqrt{\pi}} n^{-1} b^{-1} (\kappa^{-1/2} + \frac{7}{16}\kappa^{-3/2}) f(x) + o(n^{-1} b^{-1}),$$

For interior x the variance is smaller, resp. larger, for the IG, resp. RIG, kernel than for the RIG, resp. IG, kernel when $x > 1$, resp. $x < 1$. The variance expression in the

RIG case is equal to the approximation obtained for both gamma kernel estimators under $x/b \to \infty$. As pointed out by CHEN (2000), a unique feature of such estimators is that the variance coefficient decreases as x increases. This compensates for a potential larger bias when compared to kernels with compact support. Besides it has an advantage in estimating densities that have sparse areas because more data points can be pooled to smooth in areas with fewer observations. This can be viewed as a kind of robustness property of asymmetric kernels.

For boundary x the variance of the RIG estimator shares the same order as the one of the gamma kernel estimators. The following ranking can be obtained from a numerical comparison of the different multipliers of $n^{-1}b^{-1}f(x)$: $V_{Gam2} \geq V_{RIG} > V_{Gam1}$ for $\kappa \in [1.2272, \infty)$, and $V_{RIG} > V_{Gam2} > V_{Gam1}$ otherwise. The variance of the IG estimator is of a higher order. Let us remark that comparison with other techniques for boundary correction such as use of smooth optimum boundary kernels is difficult to make. Indeed such a comparison becomes bandwidth dependent since boundary x is there defined as $x/h \to q$. We need then to put $h = \sqrt{b}$ and $q = \kappa\sqrt{b}$ to ensure an amount of smoothing in the same scale.

Note also that the variance of the IG estimator is zero when $x = 0$ since $\hat{f}_{IG}(0) = 0$ by construction. The variance of the RIG kernel at $x = 0$ is finite if $E[(K_{RIG(\frac{1}{-b},\frac{1}{b})}(X_i))^2] < \infty$. Unfortunately the trick used in the proof of Proposition 2 cannot be applied here since $K_{RIG(\frac{1}{-b},\frac{2}{b})}(u))$ does not correspond to a properly defined density of an RIG distributed random variable (its variance is zero). The variance of both gamma kernel estimators at $x = 0$ is $n^{-1}b^{-1}f(0)/2 + o(n^{-1}b^{-1})$.

In the interior, the optimal mean squared errors based on

$$b^*_{IG} = \left(\frac{1}{2\sqrt{\pi}}\frac{f(x)}{f''(x)^2}\right)^{2/5} x^{-3}n^{-2/5},$$

$$b^*_{RIG} = \left(\frac{1}{2\sqrt{\pi}}\frac{f(x)}{f''(x)^2}\right)^{2/5} x^{-1}n^{-2/5},$$

are given by:

$$MSE^*_{IG} = \frac{5}{4}\left(\frac{1}{2\sqrt{\pi}}f(x)\right)^{4/5} f''(x)^{2/5}n^{-4/5},$$

$$MSE^*_{RIG} = \frac{5}{4}\left(\frac{1}{2\sqrt{\pi}}f(x)\right)^{4/5} f''(x)^{2/5}n^{-4/5}.$$

Notice that both MSE^*_{IG} and MSE^*_{RIG} only depend on $f(x)$ and not on x itself. They are the same as MSE^*_{Gam2} of the second gamma kernel estimator when $x \geq 2b$. They are also equal to MSE^*_{Gau} of the standard density estimator relying on the Gaussian kernel. In the interior their efficiency is thus equal to the efficiency of the Gaussian kernel, namely .951. The price to pay to avoid boundary bias is here a little suboptimality with respect to the Epanechnikov kernel. The optimal MSE^*_{Gam1} for the first gamma kernel estimator is different and can not be compared directly since it depends on the first derivative of the density as well as x itself.

The increase in the variance near the boundary can be shown as in CHEN (2000) to have a negligible impact on the integrated variance. Regarding global properties the optimal bandwidths and mean integrated squared errors are thus:

$$b^{**}_{IG} = \frac{\left(\frac{1}{2\sqrt{\pi}}\int_0^\infty x^{-3/2}f(x)dx\right)^{2/5}}{\left(\int_0^\infty (x^3 f''(x))^2 dx\right)^{2/5}} n^{-2/5},$$

$$b^{**}_{RIG} = \frac{\left(\frac{1}{2\sqrt{\pi}}\int_0^\infty x^{-1/2}f(x)dx\right)^{2/5}}{\left(\int_0^\infty (x f''(x))^2 dx\right)^{2/5}} n^{-2/5},$$

and

$$MISE^{**}_{IG} = \frac{5}{4}\left(\frac{1}{2\sqrt{\pi}}\int_0^\infty x^{-3/2}f(x)dx\right)^{4/5}\left(\int_0^\infty (x^3 f''(x))^2 dx\right)^{1/5} n^{-4/5},$$

$$MISE^{**}_{RIG} = \frac{5}{4}\left(\frac{1}{2\sqrt{\pi}}\int_0^\infty x^{-1/2}f(x)dx\right)^{4/5}\left(\int_0^\infty (x f''(x))^2 dx\right)^{1/5} n^{-4/5}.$$

This means that both estimators achieve the optimal rate of convergence for the MISE within the class of non-negative kernels (class of second order kernel functions). Observe also that the order $O(n^{-2/5})$ of the optimal bandwidths is the same as that for non-negative kernels when expressing the amount of smoothing in the same scale ($b = h^2$). The optimal $MISE^{**}_{RIG}$ is equal to the optimal $MISE^{**}_{Gam2}$ for the second gamma kernel estimator. Besides CHEN (2000) has shown that $MISE^{**}_{Gam1} \geq MISE^{**}_{Gam2}$ if both $\int(f')^2$ and $\int(f'')^2$ are finite. The comparison with $MISE^{**}_{IG}$ is unclear. Further theoretical properties are reported in BOUEZMARNI and SCAILLET (2002). They concern the weak and strong uniform consistency as well as L_1 consistency of asymmetric kernel density estimators.

Finally let us remark that bandwidth selection for symmetric kernels is frequently based in practice on the rule of thumb proposed by Silverman (1986). This rule is in fact optimal according to the MISE for the normal probability density function. An analogous rule may be suggested for the lognormal probability density function in the IG and RIG cases. Indeed when $\log X$ follows a normal distribution with parameters μ and σ^2 we have:

$$\begin{aligned}
\int_0^\infty x^{-1/2} f(x)dx &= \exp(\frac{1}{8}(\sigma^2 - 4\mu)), \\
\int_0^\infty x^{-3/2} f(x)dx &= \exp(\frac{1}{8}(9\sigma^2 - 12\mu)), \\
\int_0^\infty (xf''(x))^2 dx &= \frac{12 + 4\sigma^2 + \sigma^4}{32\sqrt{\pi}\sigma^5} \exp(\frac{1}{4}(9\sigma^2 - 12\mu)), \\
\int_0^\infty (x^3 f''(x))^2 dx &= \frac{12 + 68\sigma^2 + 225\sigma^4}{32\sqrt{\pi}\sigma^5} \exp(\frac{1}{4}(\sigma^2 + 4\mu)).
\end{aligned}$$

This leads to:

$$\begin{aligned}
b^{**}_{IG} &= \left(\frac{16\sigma^5 \exp(\frac{1}{8}(7\sigma^2 - 20\mu))}{12 + 68\sigma^2 + 225\sigma^4} \right)^{2/5} n^{-2/5}, \\
b^{**}_{RIG} &= \left(\frac{16\sigma^5 \exp(\frac{1}{8}(-17\sigma^2 + 20\mu))}{12 + 4\sigma^2 + \sigma^4} \right)^{2/5} n^{-2/5}.
\end{aligned}$$

In applied work the unknown parameters μ and σ^2 may be estimated by the empirical mean and empirical variance computed on the logarithm of the data. However this rule of thumb tends to provide bandwidths values which are very small. Monte Carlo experiments (not reported here) show that this leads to unsatifactory finite sample properties. Hence we do not advocate the use of this type of rule of thumb as quick bandwidth selection device for asymmetric kernel estimators.

Explicit but lengthy expressions can also be computed for Weibull and Gamma distributions. These expressions show that $MISE^{**}_{IG}$, resp. $MISE^{**}_{RIG}$, is well-defined if $\gamma \geq 3/2$, resp. $\gamma \geq 4/3$, for the $W(\gamma, \lambda)$ case, and $MISE^{**}_{IG}$, resp. $MISE^{**}_{RIG}$, is well-defined if $\gamma \geq 3/2$, resp. $\gamma \geq 3/2$ for the $Gam(\gamma, \lambda)$ case.

3 Monte Carlo results

In this section we wish to investigate the finite sample properties of the four asymmetric kernel estimators. We compare their performance with the one of two standard symmetric kernel estimators, namely the Epanechnikov and Gaussian kernel estimators. We also consider the smooth boundary optimum boundary kernel estimator obtained by modification of the Epanechnikov kernel in MÜLLER (1991). The experiments are based on 1000 random samples of length $n = 3^5 = 243$, $n = 486$, and $n = 972$. For each simulated sample and each estimator considered, integrated squared errors (ISE) were computed from a grid of bandwith values proportional to $n^{-2/5}$ for the asymmetric kernels and proportional to $n^{-1/5}$ for the other kernels. Numerical integration was performed by a Gauss Legendre quadrature with 96 knots. Minimum average integrated squared errors are reported in Table 1 for various distributions, namely Gamma, Weibull and lognormal distributions, and various parameter values.

Table 1: **Average ISE**

$n=243$	IG	RIG	Gam1	Gam2	Epa	Gau	Bou
$Gam(1.5,1)$	.03858	.00492	.00478	.00475	.00630	.00622	.01078
$Gam(2,1)$	.01350	.00296	.00360	.00316	.00391	.00392	.00889
$Gam(3,1)$	.00833	.00202	.00226	.00200	.00253	.00252	.00737
$W(1.5,1)$	.04715	.00631	.00713	.00651	.00726	.00726	.01076
$W(2,1)$	.02562	.00644	.00751	.00653	.00642	.00662	.00915
$W(3,1)$	.02064	.00944	.01368	.01194	.00798	.00835	.01050
$LN(0,1)$	.02040	.00611	.02157	.02107	.01031	.01019	.00814
$LN(1,1)$	.00740	.00224	.00403	.00344	.00378	.00373	.00357
$LN(2,1)$	.00238	.00081	.00098	.00084	.00134	.00133	.00198

$n=486$	IG	RIG	Gam1	Gam2	Epa	Gau	Bou
$Gam(1.5,1)$	.02165	.00286	.00294	.00280	.00400	.00394	.00618
$Gam(2,1)$	.00795	.00172	.00210	.00181	.00239	.00240	.00535
$Gam(3,1)$	.00504	.00123	.00137	.00122	.00155	.00156	.00522
$W(1.5,1)$	.04021	.00375	.00438	.00389	.00463	.00462	.00622
$W(2,1)$	.01479	.00387	.00494	.00398	.00395	.00407	.00570
$W(3,1)$	.01258	.00574	.01186	.00977	.00485	.00507	.00758
$LN(0,1)$	.01206	.00364	.02095	.02025	.00648	.00643	.00548
$LN(1,1)$	.00424	.00134	.00350	.00290	.00237	.00236	.00226
$LN(2,1)$	.00139	.00049	.00065	.00050	.00084	.00084	.00121

$n = 972$	IG	RIG	Gam1	Gam2	Epa	Gau	Bou
$Gam(1.5, 1)$	.01579	.00171	.00201	.00177	.00257	.00254	.00393
$Gam(2, 1)$	.00463	.00105	.00133	.00109	.00152	.00152	.00376
$Gam(3, 1)$	.00299	.00073	.00084	.00073	.00094	.00095	.00416
$W(1.5, 1)$	.01913	.00227	.00287	.00236	.00297	.00296	.00388
$W(2, 1)$	.00855	.00236	.00365	.00273	.00245	.00252	.00393
$W(3, 1)$	.00764	.00354	.01064	.00899	.00296	.00310	.00593
$LN(0, 1)$	.00661	.00216	.02143	.02067	.00403	.00402	.00423
$LN(1, 1)$	.00241	.00079	.00335	.00277	.00148	.00147	.00159
$LN(2, 1)$	.00080	.00029	.00048	.00034	.00053	.00052	.00083

We may observe that the RIG kernel estimator and the second gamma kernel estimator have similar performance with a small advantage of the former over the latter (except for gamma densities). This was expected due to their close shapes. The second gamma kernel estimator performs better than the first. This was already observed by Chen (2000) on Gamma distributed data. The IG kernel estimator is almost always dominated by the others. As also expected, both symmetric kernel estimators perform well for distributions exhibiting low probability mass near the boundary ($W(3, 1)$). The smooth optimum boundary kernel does not seem to make a good job in general. Besides use of smooth optimum boundary kernels based on polynomials with finite support may lead to negative density estimates. This feature may be another reason to prefer asymmetric kernel estimators. The following table gives the number of cases where at least one knot in the Gaussian quadrature with a negative density has been detected on the one thousand simulated samples.

Table 2: Number of cases with negative density

$n = 243$	$Gam(1.5, 1)$	$Gam(2, 1)$	$Gam(3, 1)$
	780	644	422
	$W(1.5, 1)$	$W(2, 1)$	$W(3, 1)$
	784	653	119
	$LN(0, 1)$	$LN(1, 1)$	$LN(2, 1)$
	550	678	519
$n = 486$	$Gam(1.5, 1)$	$Gam(2, 1)$	$Gam(3, 1)$
	771	764	617
	$W(1.5, 1)$	$W(2, 1)$	$W(3, 1)$
	785	701	119
	$LN(0, 1)$	$LN(1, 1)$	$LN(2, 1)$
	763	672	607
$n = 972$	$Gam(1.5, 1)$	$Gam(2, 1)$	$Gam(3, 1)$
	773	854	803
	$W(1.5, 1)$	$W(2, 1)$	$W(3, 1)$
	789	685	106
	$LN(0, 1)$	$LN(1, 1)$	$LN(2, 1)$
	772	763	679

4 Concluding remarks

We have proposed two new kernel estimators for probability density functions defined on $[0, \infty)$, namely IG and RIG kernel estimators. Such densities are encountered in a wide variety of applications in the biological sciences and economics. The estimators have good finite sample properties, and should therefore be useful in applied work involving nonparametric techniques, see e.g. Härdle and Linton (1994), Pagan and Ullah (1999), Hall (2001) for some examples. Obiously the new classes of kernels considered here can also be exploited in regression curve estimation or hazard rate estimation (see Fernandes and Grammig (2000) for a convincing use in goodness-of-fit testing procedures for duration models).

APPENDIX

Proof of Proposition 1

1) Bias of the IG kernel estimator

We have:

$$E[\hat{f}_{IG}(x)] = \int_0^\infty K_{IG(x,\frac{1}{b})}(y)f(y)dy = E[f(\xi_x)],$$

where ξ_x follows an $IG(x, \frac{1}{b})$ distribution. From the expressions for the mean and variance of an IG distributed random variable, we deduce $\mu_x = E[\xi_x] = x$ and $V_x = \text{Var}[\xi_x] = x^3 b$. Then we get by Taylor expansion:

$$\begin{aligned} E[f(\xi_x)] &= f(\mu_x) + \frac{1}{2} f''(x)V_x + o(b) \\ &= f(x) + \frac{1}{2} x^3 f''(x)b + o(b), \end{aligned}$$

which gives the first statement.

2) Bias of the RIG kernel estimator

Along the same lines we have:

$$E[\hat{f}_{RIG}(x)] = \int_0^\infty K_{RIG(\frac{1}{x-b},\frac{1}{b})}(y)f(y)dy = E[f(\xi_x)],$$

where ξ_x follows an $RIG(\frac{1}{x-b}, \frac{1}{b})$ distribution. By Taylor expansion and using $\mu_x = E[\xi_x] = x$ and $V_x = \text{Var}[\xi_x] = xb + b^2$, we deduce:

$$\begin{aligned} E[f(\xi_x)] &= f(\mu_x) + \frac{1}{2} f''(x)V_x + o(b) \\ &= f(x) + \frac{1}{2} x f''(x)b + o(b), \end{aligned}$$

which ends the proof.

Proof of Proposition 2

1) Variance of the IG kernel estimator

The variance is equal to:

$$\begin{aligned}\mathrm{Var}[\hat{f}_{IG}(x)] &= n^{-1}\mathrm{Var}[K_{IG(x,\frac{1}{b})}(X_i)] \\ &= n^{-1}E[(K_{IG(x,\frac{1}{b})}(X_i))^2] + O(n^{-1}).\end{aligned}$$

Let η_x be an $IG(x, \frac{2}{b})$ distributed random variable. Hence $\mu_x = E[\eta_x] = x$ and $V_x = \mathrm{Var}[\eta_x] = x^3 b/2$. We have

$$E[(K_{IG(x,\frac{1}{b})}(X_i))^2] = B_b E[\eta_x^{-3/2} f(\eta_x)],$$

where $B_b = (4\pi b)^{-1/2}$. By Taylor expansion we get:

$$\begin{aligned}E[\eta_x^{-3/2} f(\eta_x)] &= \mu_x^{-3/2} f(\mu_x) + \frac{1}{2}\left(\frac{15}{4}x^{-7/2} f(x) - 3x^{-5/2} f'(x) + x^{-3/2} f''(x)\right) V_x + o(b) \\ &= x^{-3/2} f(x) + \frac{1}{4}\left(\frac{15}{4}x^{-1/2} f(x) - 3x^{1/2} f'(x) + x^{3/2} f''(x)\right) b + o(b), \\ &= x^{-3/2} f(x) + O(b),\end{aligned}$$

which leads to the first result.

2) Variance of the RIG kernel estimator

We have:

$$\mathrm{Var}[\hat{f}_{RIG}(x)] = n^{-1}E[(K_{RIG(\frac{1}{x-b},\frac{1}{b})}(X_i))^2] + O(n^{-1}).$$

Let η_x be an $RIG(\frac{1}{x-b}, \frac{2}{b})$, so that

$$E[(K_{\frac{1}{x-b},\frac{1}{b}}(X_i))^2] = B_b E[\eta_x^{-1/2} f(\eta_x)],$$

where $B_b = (4\pi b)^{-1/2}$. Since $\mu_x = E[\eta_x] = x - b/2$ and $V_x = \mathrm{Var}[\eta_x] = xb/2$, we obtain by Taylor expansion :

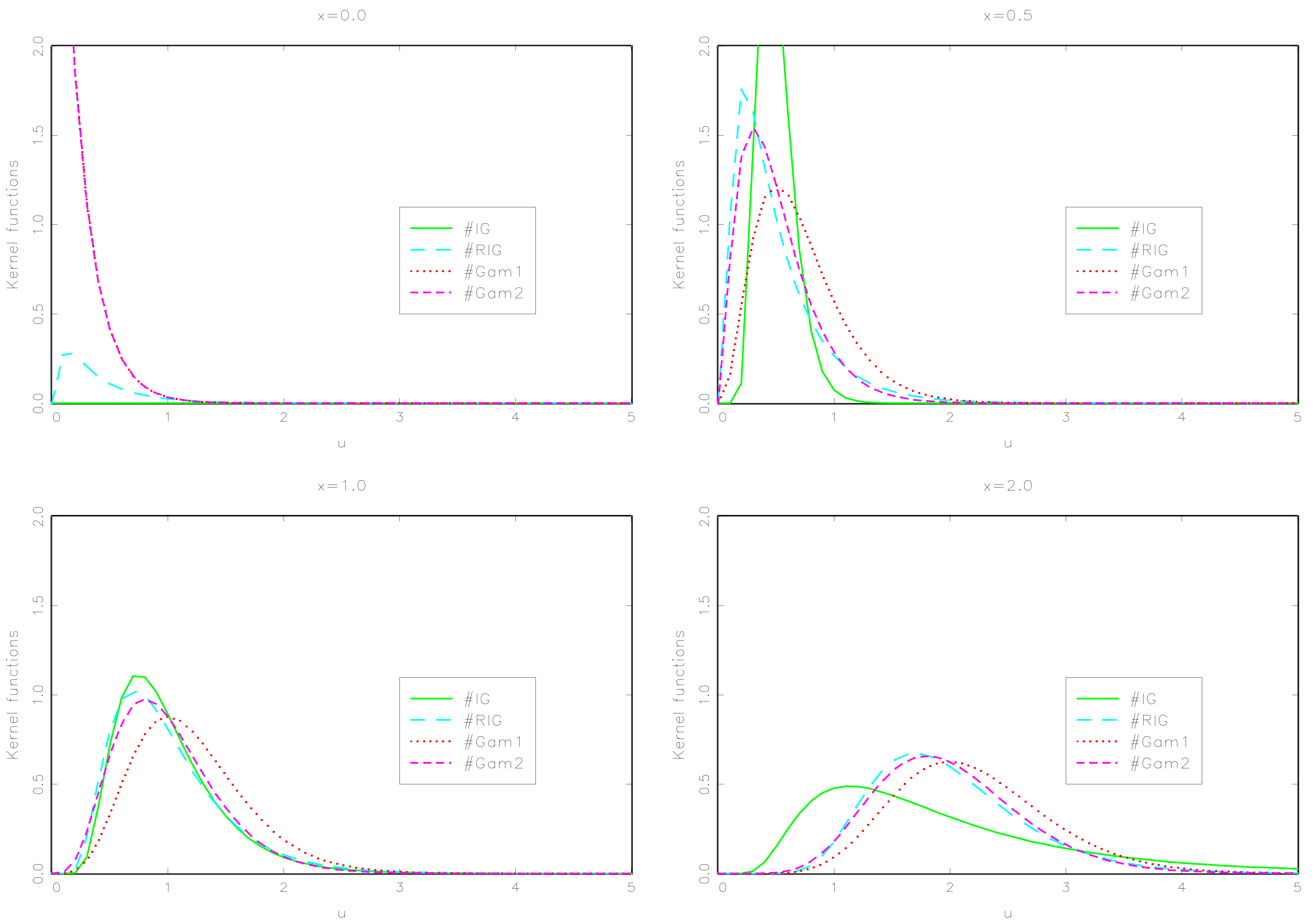
$$E[\eta_x^{-1/2} f(\eta_x)] = \mu_x^{-1/2} f(\mu_x) + \frac{1}{2}\left(\frac{3}{4}x^{-5/2} f(x) - x^{-3/2} f'(x) + x^{-1/2} f''(x)\right) V_x + o(b)$$

$$
\begin{aligned}
&= x^{-1/2}f(x) + \frac{1}{4}\left(\frac{7}{4}x^{-3/2}f(x) - 3x^{-1/2}f'(x) + x^{1/2}f''(x)\right)b + o(b), \\
&= x^{-1/2}f(x) + O(b),
\end{aligned}
$$

which gives the second result.

REFERENCES :

- Bouezmarni, T. and O. Scaillet (2002): "Consistency of Asymmetric Kernel Density Estimators and Smoothed Histograms with Application to Income Data", UCL DP.
- Chen, S. X. (2000): "Probability Density Function Estimation Using Gamma Kernels", *Ann. Inst. Stat. Math.*, **52**, 471-480.
- Chen, S. X. (2001): "Local Linear Smoothers Using Asymmetric Kernels", *Ann. Inst. Stat. Math.*, **52**, 312-323.
- Fernandes, M. and J. Grammig (2000): "Nonparametric Specification Tests for Conditional Duration Models", Fondation Getulio Vargas, mimeo.
- Hall, P. (2001): "Biometrika Centenary: Nonparametrics", *Biometrika*, **88**, 143-165.
- Härdle, W. and O. Linton (1994): "Applied Nonparametric Methods", in *Handbook of Econometrics*, **IV**, Eds Engle R. and McFadden D., North-Holland, Amsterdam.
- Müller, H. (1991): "Smooth Optimum Kernel Estimators near Endpoints", *Biometrika*, **78**, 521-520.
- Pagan, A. and A. Ullah (1999): *Nonparametric Econometrics*, Cambridge University Press, Cambridge.
- Silverman, B. (1986): *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London.
- Tweedie, M. (1947): "Functions of a Statistical Variate with Given Means with Special Reference to Laplacian Distributions", *Proceedings of the Cambridge Philosophical Society*, **43**, 41-49.
- Wald, A. (1947): *Sequential Analysis*, Wiley, New York.


x=0.0
x=0.5
x=1.0
x=2.0
Kernel functions
u
#IG
#RIG
#Gam1
#Gam2