Ann. Inst. Statist. Math.
Vol. 44, No. 4, 721–727 (1992)

ESTIMATING DENSITIES, QUANTILES, QUANTILE DENSITIES AND DENSITY QUANTILES

M. C. JONES

Department of Statistics, The Open University, Milton Keynes MK7 6AA, U.K.

(Received April 24, 1991; revised November 8, 1991)

Abstract. To estimate the quantile density function (the derivative of the quantile function) by kernel means, there are two alternative approaches. One is the derivative of the kernel quantile estimator, the other is essentially the reciprocal of the kernel density estimator. We give ways in which the former method has certain advantages over the latter. Various closely related smoothing issues are also discussed.

Key words and phrases: Kernel smoothing, reciprocal density, sparsity function.

1. Introduction

$\boldsymbol{X} = (X_1, X_2, \dots, X_n)$, a sample of n i.i.d. observations from a continuous distribution, tells us much about the distribution function F via the empirical distribution function $F_n(x) = n^{-1}\sum_{i=1}^n I(X_i \le x)$; here, $I(E) = 1$ if event E occurs and is 0 otherwise. We will often be interested in the density $f = F'$, and, by the introduction of some smoothing, can usefully estimate that quantity from $\boldsymbol{X}$ too e.g. Silverman (1986).

The inverse $Q \equiv F^{-1}$ of F, the quantile function, is sometimes the object of more direct interest than is F itself. The data $\boldsymbol{X}$ relate directly to Q as well simply by taking the left-continuous inverse of F_n, namely the usual empirical quantile function

$$\begin{aligned} Q_n(u) &= \sum_{i=1}^n X_{(i)} I\left(\frac{(i-1)}{n} < u \le \frac{i}{n}\right) \\ &= X_{(1)} + \sum_{i=2}^n (X_{(i)} - X_{(i-1)}) I\left(\frac{(i-1)}{n} < u\right). \end{aligned}$$

Here, $X_{(i)}$, $i = 1, 2, \dots, n$ are the order statistics of the sample and $0 < u < 1$. The derivative of Q ($q \equiv Q'$) is also a function of some interest that may be one's premier target of estimation. Parzen (1979) gives a number of reasons for this; others are interested in q because of its appearance in the asymptotic variance of sample quantiles. We follow Parzen's (1979) terminology in calling q the quantile density function; Tukey (1965) called it the sparsity function. To estimate q, smoothing is again necessary, and it is as natural to base such smoothing on Q_n as it was to estimate f by basing smoothing on F_n, since the same derivative relationship holds in each case.

These first two paragraphs are intended to remind the reader of the twin strands $F, f, f', \ldots$ and $Q, q, q', \ldots$ arising from the inverse relationship $F = Q^{-1}$, $Q = F^{-1}$. The discussion so far has supposed that smoothing best takes place "within strands". But of course there are close relationships "between strands" and in particular

$$q(u) = \frac{1}{f(F^{-1}(u))}, \qquad f(x) = \frac{1}{q(Q^{-1}(x))}. \tag{1.1}$$

So, are we right to approach our smoothing problems in the manner described? For instance, using $\tilde{}$ throughout to denote any suitable smoothed estimate, we were presuming a $\tilde{q}$ of the form $(\tilde{Q})'$ above. But how does this compare with $1/\tilde{f}(\tilde{Q}(u))$? (Actually, $\tilde{Q}$ could be Q_n here.) It is the main purpose of this paper to verify that $(\tilde{Q})'$ does indeed seem preferable in some ways to its competitor. This is something of a relief, since otherwise one might expect the second relationship in (1.1) to offer a more viable alternative to the ubiquitous—and appropriate—$\tilde{f} = (\tilde{F})'$ form of density estimation per se.

We will work in terms of kernel smoothing methods (e.g. Silverman (1986)) in particular forms to be introduced as needed in our development. Notationally, we will employ a symmetric probability density function k as kernel ($K' \equiv k$) and use b and h for bandwidths (smoothing parameters). Also, $k_a(\cdot) \equiv a^{-1}k(a^{-1}\cdot)$, $K_a(\cdot) \equiv K(a^{-1}\cdot)$ and $*$ denotes convolution. In Section 2, results on kernel smoothed estimates of Q itself will be briefly reviewed, with a view to appreciating in what sense $\tilde{Q}$ and $(\tilde{F})^{-1}$ are actually equivalent. Such equivalence breaks down as we move on to the mainstream smoothing level of estimating q. The main thrust of the paper, on estimating the density quantile function, is contained in Section 3. Related problems, of estimating densities, reciprocals of densities and density quantiles (reciprocals of quantile densities), are briefly treated in Section 4. Theoretical performance will be measured in squared error terms, as Mean (Integrated) (Weighted) Squared Error (M(I)(W)SE). The usual kind of asymptotic representation of such quantities will be examined (as $n \to \infty$, $b = b(n) \to 0$ such that $nb \to \infty$, and likewise for h). We will not attempt to be very technical in our mathematics (references cited will often do that) but simply assume as much smoothness of f as necessary, mostly the usual assumption of its having two continuous derivatives.

2. Smooth estimation of Q

As the natural kernel estimator of F, $\hat{F}$, simply convolves k_b with F_n (e.g. Azzalini (1981), Reiss (1981)), so we can estimate Q by $\hat{Q} \equiv k_b * Q_n$ (Parzen (1979)) i.e.

$$\hat{Q}(u) = \sum_{i=1}^{n} X_{(i)} \int_{(i-1)/n}^{i/n} k_b(u-y)dy.$$

For comparisons with a number of variations on $\hat{Q}$, see Sheather and Marron (1990). There is a problem with boundary effects that we ignore; boundary kernels or other edge correction methods (e.g. Müller (1991)) should be employed in practice if u is close to 0 or 1. Building on Falk (1984), Sheather and Marron (1990) give the MSE of $\hat{Q}$ as

$$\text{(2.1)} \qquad \text{MSE}\{\hat{Q}(u)\} \simeq \frac{1}{4}b^4\{q'(u)\}^2\sigma_k^4 + \frac{u(1-u)}{n}q^2(u) - \frac{b}{n}q^2(u)\psi(k),$$

where $\sigma_k^2 \equiv \int y^2k(y)dy$ and $\psi(k) \equiv 2\int yk(y)K(y)dy = \int K(y)\{1-K(y)\}dy$. The first term in (2.1) is due to squared bias, the second and third to variance; also, set $b=0$ in (2.1) to obtain the usual asymptotic representation of $\text{MSE}\{Q_n(u)\}$.

Now,

$$\hat{F}(x) = n^{-1}\sum_{i=1}^{n} K_h(x - X_i).$$

We could invert $\hat{F}$ to get an alternative estimate of Q, say $\bar{Q}$, as suggested by Nadaraya (1964). A numerical procedure would be necessary to perform the inversion. Notice that the apparent avoidance of boundary worries (if the support of F is $\boldsymbol{R}$) may be illusory, edge effects being replaced by awkward (if usually implicit) suppositions on tail behaviour. Azzalini (1981) obtains the MSE of $\bar{Q}$:

$$\text{(2.2)} \qquad \text{MSE}\{\bar{Q}(u)\} \simeq \frac{1}{4}h^4\left[\frac{q'(u)}{q^2(u)}\right]^2\sigma_k^4 + \frac{u(1-u)}{n}q^2(u) - \frac{h}{n}q(u)\psi(k).$$

Now, the natural scaling match-up between h and b at a point $x = Q(u)$, if we were to allow h to vary with x and b to depend on u, is via the relationship $b(u) = h(x)f(x)$. And sure enough, if this is so, (2.1) and (2.2) coincide precisely, and there is no (asymptotic) difference in performance between $\hat{Q}$ and $\bar{Q}$ (a phenomenon noticed elsewhere including Lio and Padgett (1991)). One is inclined to view use of constant b in $\hat{Q}$ as a natural choice for "constant kernel" quantile estimation, and hence require h in $\bar{Q}$ to vary with x to match.

As the above concentrates on estimating Q, so there are parallels re estimation of F. In fact, $\hat{Q}^{-1}$ (albeit with a locally varying b) has precisely the same asymptotic properties as does the much more natural $\hat{F}$ with constant h.

3. Smooth estimation of q

As $q = Q'$, the natural direct kernel estimate of q is $\hat{q} \equiv (\hat{Q})'$:

$$\begin{aligned}\hat{q}(u) &= \sum_{i=2}^{n} X_{(i)}\left\{k_b\left(u - \frac{(i-1)}{n}\right) - k_b\left(u - \frac{i}{n}\right)\right\}\\ &= \sum_{i=2}^{n}(X_{(i)} - X_{(i-1)})k_b\left(u - \frac{(i-1)}{n}\right) - X_{(n)}k_b(u-1) + X_{(1)}k_b(u).\end{aligned}$$

Various authors (e.g. Parzen (1979), Falk (1986), Welsh (1988), Babu and Rao (1990)) have considered this appealing kernel quantile density estimator which runs a kernel smoother through consecutive order statistic spacings located on the grid i/n, $i = 1, 2, \ldots, n$.

MSE properties of $\hat{q}$ have been derived by Falk (1986), among others. A particularly attractive feature of this direct kernel estimator is that to $O(n^{-1/2})$ its bias looks like $k_b * q$, so its leading bias term depends on unknowns simply through q''. Specifically,

$$\text{MSE}\{\hat{q}(u)\} \simeq \frac{1}{4}b^4\{q''(u)\}^2\sigma_k^4 + \frac{1}{nb}q^2(u)\kappa \tag{3.1}$$

where $\kappa = \int k^2(y)dy$. This expression tells us that rates are just the same as for kernel density estimation per se, shows how the coefficients of b and n depend on q, and might be used as a basis for bandwidth selection, and so on.

Interest here, however, is centred on how this MSE compares with that of the alternative estimator $\bar{q}(u) \equiv 1/\hat{f}(Q_n(u))$, where $\hat{f} \equiv (\hat{F})'$ is the usual kernel density estimator. This (obvious) alternative has also appeared in the literature, if less frequently than has $\hat{q}$ (e.g. Sheather (1987), Hall *et al.* (1989)). Here, we have taken Q_n as estimator of Q inside $\hat{f}(\cdot)$ for convenience: actually any $\tilde{Q}$ which differs from Q by order $n^{-1/2}$ (or even something slightly bigger) would do, a smoothed quantile estimator with appropriate degree of smoothing perhaps being preferable globally to avoid piecewise constancy of the estimate.

Writing $1/\hat{f}(x) = 1/[f(x) + \{\hat{f}(x) - f(x)\}] \simeq \{1/f(x)\} \times [1 - \{\hat{f}(x) - f(x)\}/f(x)]$, and carefully translating from x to $u = F(x)$, we get

$$\text{MSE}\{\bar{q}(u)\} = \frac{1}{4}h^4\left[\frac{q(u)q''(u) - 3\{q'(u)\}^2}{q^3(u)}\right]^2\sigma_k^4 + \frac{1}{nh}q^3(u)\kappa.$$

Once again, taking (pointwise) $h = h(x) = bq(u)$, we can make this match up better with expression (3.1). Explicitly,

$$\text{MSE}\{\bar{q}(u)\} = \frac{1}{4}b^4\left[q''(u) - 3\frac{\{q'(u)\}^2}{q(u)}\right]^2\sigma_k^4 + \frac{1}{nb}q^2(u)\kappa. \tag{3.2}$$

The variance contributions to (3.1) and (3.2) are now identical. There remains, however, a persistent difference in the biases. While that of $\hat{q}$ is $q''(u)$, the bias of $\bar{q}$ is actually $[3\{q'(u)\}^2/q(u)] - q''(u)$. Simplicity and interpretability may favour the former, but at any point u it is unclear which of $\hat{q}$ and $\bar{q}$ is better. Indeed, the squared bias of $\hat{q}$ is only less than or equal to that of $\bar{q}$ if $q(u)q''(u) \le 1.5\{q'(u)\}^2$ and $q'(u) \neq 0$. (The main condition is equivalent to $f(x)f''(x) \ge 1.5\{f'(x)\}^2$.) Very roughly, it seems that in general $\bar{q}$ might do the better job towards the centre of the distribution and $\hat{q}$ would tend to be preferable more in the shoulders and towards the tails. At stationary points the two have the same behaviour to this degree of approximation.

Global considerations can be thought of as helping us determine which estimator has the "better" collection of points at which it is the preferable one. Simply integrating MSE's will (usually) lead to integrals that are not finite, so we employ a MIWSE of the form $\text{MIWSE}(\tilde{q}) = E\int q^w(u)\{\tilde{q}(u) - q(u)\}^2du$. Choice of power w

has a nonnegligible effect on our answers. We intuitively like the idea of choosing w so that the MIWSE is "on a par" with the (unweighted) MISE of an estimator $\tilde{f}(x)$ of the density $f(x)$; $w = -3$ achieves this.

Practical difficulties with implementing the local variation of h with f notwithstanding, the integrated weighted squared bias (IWSB) of either $\tilde{q}$ above has the form $b^4\tilde{R}(q)\sigma_k^4/4$ where $\tilde{R}(q)$ is that functional of q appropriate to $\tilde{q}$. In fact,

$$\hat{R}(q) = \int q^w(u)\{q''(u)\}^2 du,$$

while

$$\bar{R}(q) = \int q^{w-2}(u)[q(u)q''(u) - 3\{q'(u)\}^2]^2 du.$$

Write

$$I(q) = \int q^{w-1}(u)\{q'(u)\}^2 q''(u)du \quad \text{and} \quad J(q) = \int q^{w-2}(u)\{q'(u)\}^4 du.$$

Integration by parts yields that, for well-behaved q,

$$I(q) = -\frac{1}{3}(w-1)J(q),$$

so that

$$\bar{R}(q) = \hat{R}(q) + (2w+7)J(q).$$

Since $J(q) > 0$, it follows that IWSB($\hat{q}$) $<$ IWSB($\bar{q}$) provided $w > -7/2$. This range of w for which $\hat{q}$ "wins" in MIWSE terms includes, of course, our favoured $w = -3$. The direct estimate $\hat{q}$ therefore tends to dominate the indirect estimate $\bar{q}$ more often than the reverse in sufficient of the areas that most weight is given to by the usual kind of MISE (in f estimation terms). $\bar{q}$ appears to take over when less notice is taken of performance in the tails of the distribution, in agreement with our comment about the pointwise comparison. $\hat{q}$ clearly wins when the unweighted case $w = 0$ is meaningful.

To close this section, it may be useful to remind the reader that another popular method of quantile density function estimation, namely the Siddiqui-Bloch-Gastwirth estimator (e.g. Hall and Sheather (1988), for definition and references) is precisely a version of $\hat{q}$ (as noted by various earlier authors).

4. Related problems

While little in our comparisons is totally clear cut, we have demonstrated ways in which "direct" kernel estimators of q may be thought preferable. Another is their simpler extension to derivative estimation.

We very briefly address estimation of f in a similar light. This is, of course, the standard problem of nonparametric density estimation (e.g. Silverman (1986)) which has the usual kernel density estimator $\hat{f}$ as a standard solution. But as q

might be estimated directly or via f, so f might be estimated directly, or (perversely?) via q i.e. define $\bar{f} \equiv 1/\hat{q}(F_n(x))$ (or else involving other appropriate $\tilde{F}$).

The MSE of $\hat{f}$ is very well known; it is

$$\text{MSE}\{\hat{f}(x)\} \simeq \frac{1}{4}h^4\{f''(x)\}^2\sigma_k^4 + \frac{1}{nh}f(x)\kappa. \tag{4.1}$$

The equivalent formula for $\bar{f}$ at $x = Q(u)$ using $b = b(u) = hf(x)$ turns out to be

$$\text{MSE}\{\bar{f}(x)\} \simeq \frac{1}{4}h^4\left[f''(x) - 3\frac{\{f'(x)\}^2}{f(x)}\right]^2\sigma_k^4 + \frac{1}{nh}f(x)\kappa. \tag{4.2}$$

The variances of $\hat{f}$ and $\bar{f}$ are the same, the biases differ in ways entirely parallel to the discussion in q-space. Interestingly, this means that there are places (actually, likely to be in the centre of the distribution, specifically where $f(x)f''(x) > 1.5\{f'(x)\}^2$, $f'(x) \neq 0$, again) where the reciprocal of the direct estimator of the quantile density function outperforms $\hat{f}$. That said, the analogous global MSE manipulations do turn out to favour $\hat{f}$. Weighting integrated MSE's by $f^w(x)$, one finds that $\hat{f}$ is superior to $\bar{f}$ in MIWSE terms provided $w > -7/2$ once more. But now the natural choice of w is most likely $w = 0$, so we are further into the region of superiority of the direct estimate, and there is also a greater difference between the two. Overall, there is little here to force us to rethink our use of $\hat{f}$ as the basic kernel density estimator.

Estimating the density quantile function, $f(Q(u))$, rather than $q(u)$ or $f(x)$, is more akin to the latter because of the ability to estimate Q to order $n^{-1/2}$, an amount negligible compared with smoothing rates. Thus, this author sees the more appropriate estimator of $f(Q(u))$ as being $\hat{f}$ at an appropriate $\tilde{Q}(u)$ rather than the reciprocal of $\hat{q}(u)$. Likewise, $1/f(x)$ might be estimated via direct estimation of q rather than via f.

Finally, our theoretical interest has been couched in MSE terms. However, tailoring loss functions to end points implies possibly alternative comparisons. For instance, Hall and Sheather (1988) show that, for certain confidence interval and hypothesis testing problems concerning sample quantiles, instead of MSE's squared bias and variance, the important aspects of $\tilde{q}$'s behaviour are determined by bias itself and variance. It therefore turns out that much of what we have done here remains pertinent; a twist is that even when $q'(u) = 0$, biases differ in sign between $\hat{q}$ and $\bar{q}$.

Acknowledgements

The referees are thanked for comments which led to presentational improvements.

References

Azzalini, A. (1981). A note on the estimation of a distribution function and quantiles by a kernel method. *Biometrika*, **68**, 326–328.

Babu, G. J. and Rao, C. R. (1990). Estimation of the reciprocal of the density quantile function at a point, *J. Multivariate Anal.*, **33**, 106–124.

Falk, M. (1984). Relative deficiency of kernel type estimators of quantiles, *Ann. Statist.*, **12**, 261–268.

Falk, M. (1986). On the estimation of the quantile density function, *Statist. Probab. Lett.*, **4**, 69–73.

Hall, P. and Sheather, S. J. (1988). On the distribution of a studentized quantile, *J. Roy. Statist. Soc. Ser. B*, **50**, 381–391.

Hall, P., DiCiccio, T. J. and Romano, J. P. (1989). On smoothing and the bootstrap, *Ann. Statist.*, **17**, 692–704.

Lio, Y. L. and Padgett, W. J. (1991). A note on the asymptotically optimal bandwidth for Nadaraya's quantile estimator, *Statist. Probab. Lett.*, **11**, 243–249.

Müller, H.-G. (1991). Smooth optimum kernel estimators near endpoints, *Biometrika*, **78**, 521–530.

Nadaraya, E. A. (1964). Some new estimates for distribution functions, *Theory Probab. Appl.*, **9**, 497–500.

Parzen, E. (1979). Nonparametric statistical modelling (with comments), *J. Amer. Statist. Assoc.*, **74**, 105–131.

Reiss, R.-D. (1981). Nonparametric estimation of smooth distribution functions, *Scand. J. Statist.*, **8**, 116–119.

Sheather, S. J. (1987). Assessing the accuracy of the sample median: estimated standard errors versus interpolated confidence intervals, *Statistical Data Analysis Based on the L_1-norm and Related Methods* (ed. Y. Dodge), 203–215, North-Holland, Amsterdam.

Sheather, S. J. and Marron, J. S. (1990). Kernel quantile estimators, *J. Amer. Statist. Assoc.*, **85**, 410–416.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London.

Tukey, J. W. (1965). Which part of the sample contains the most information?, *Proc. Nat. Acad. Sci. U.S.A.*, **53**, 127–134.

Welsh, A. H. (1988). Asymptotically efficient estimation of the sparsity function at a point, *Statist. Probab. Lett.*, **6**, 427–432.