Uniform Consistency of Some
Estimates of a Density Function

by

D.S. Moore and E.G. Henrichon

Department of Statistics
Division of Mathematical Sciences
Mimeograph Series No. 168
August, 1968
Uniform Consistency of Some
Estimates of a Density Function

by

D. S. Moore and E. G. Henrichon

Purdue University

1. **Introduction and summary.** Let $X_1, \ldots, X_n$ be independent random
variables identically distributed with absolutely continuous distribution fun-
tion $F$ and density function $f$. Loftsgaarden and Quesenberry [2] propose a
consistent nonparametric point estimator $\hat{f}_n(z)$ of $f(z)$ which is quite easy
to compute in practice. In this note we introduce a step-function approximation
$f^*_n$ to $\hat{f}_n$, and show that both $\hat{f}_n$ and $f^*_n$ converge uniformly (in probability)
to $f$, assuming that $f$ is positive and uniformly continuous in $(\infty, \infty)$. For
more general $f$, uniform convergence over any compact interval where $f$ is
positive and continuous follows. Uniform convergence is useful for estimation
of the mode of $f$, for it follows from our theorem (see [3], section 3) that a
mode of either $\hat{f}_n$ of $f^*_n$ is a consistent estimator of the mode of $f$. The
mode of $f^*_n$ is particularly tractable; it is applied in [1] to some problems
in pattern recognition.

2. **The result.** Choose a non-decreasing sequence of positive integers,
$k(n)$, such that $k(n) \to \infty$ but $k(n) = o(n)$. For any real number $z$, let
$r_{k(n)}(z)$ be the distance from $z$ to the $k(n)$th closest of the observations
$X_1, \ldots, X_n$. Then the univariate form of the Loftsgaarden-Quesenberry estimator
is

$$f_n(z) = \frac{(k(n) - 1)}{n!} \left\{ \frac{1}{2r_{k(n)}(z)} \right\}.$$
We define also the random step-function \( f^*_n \) as follows: let \( X_{i1} \leq X_{2n} \leq \ldots \leq X_{nn} \) be the order statistics from \( X_1, \ldots, X_n \). Then

\[
\begin{align*}
f^*_n(z) &= 0 \quad \text{if } z < X_{i1} \text{ or } z > X_{nn} \\
&= \hat{f}_n(X_{i1}) \quad X_{i1} \leq z < X_{i1+n+1} \quad i = 1, \ldots, n-1 .
\end{align*}
\]

**THEOREM.** If \( f(z) \) is uniformly continuous and positive on \((-\infty, \infty)\) and \((\log n) / k(n) \to 0\), then for every \( \epsilon > 0 \)

\[
(2.1) \quad P \left[ \sup_{-\infty < z < \infty} \left| \hat{f}_n(z) - f(z) \right| > \epsilon \right] \to 0
\]

and

\[
(2.2) \quad P \left[ \sup_{-\infty < z < \infty} \left| f^*_n(z) - f(z) \right| > \epsilon \right] \to 0 .
\]

**Proof.** We will abbreviate (2.1) by \( \hat{f}_n \to f \) (UP) and denote convergence in probability by \( a_n \to a(\mathbb{P}) \). Define

\[
U_k(n)(z) = F(z + r_k(n)(z)) - F(z - r_k(n)(z)).
\]

We show first that

\[
(2.3) \quad \{n/(k(n)-1)\} U_k(n)(z) \to 1(\text{UP}) .
\]

By definition of \( r_k(n)(z) \), the interval \([z - r_k(n)(z), z + r_k(n)(z)]\) contains exactly \( k(n) \) observations, one of which falls at an endpoint of the interval. Suppose the order statistic \( X_{qn} \) is the lower endpoint. Then
\[ \sum_{j=1}^{k(n)-1} \{ F(X_{q+j,n}) - F(X_{q+j-1,n}) \} \leq U_k(n) \]

with the conventions \( F(X_{0,n}) = 0 \) and \( F(X_{n+1,n}) = 1 \). Upper and lower bounds having the same distribution as those in (2.4) exist when \( X_{qn} \) is on upper endpoint. (It is stated in [2] that \( U_k(n) \) has the beta distribution of one of the sums of elementary coverages in (2.4). This is false, since w.p.1 only one endpoint of the interval coincides with an observation; the modifications required to correct the proof of [2] are trivial.)

It is well known that

\[ F(X_{1n}), F(X_{2n}) - F(X_{1n}), \ldots, 1-F(X_{nn}) \]

have the same joint distribution as

\[ Y_1 / S_{n+1}, \ldots, Y_{n+1} / S_{n+1}, \]

where \( Y_1, \ldots, Y_{n+1} \) are independent exponential random variables with mean 1 and \( S_{n+1} = Y_1 + \ldots + Y_{n+1} \). So the upper and lower bounds in (2.4) will converge to 1 (UP) if we can prove that

\[ \max_{0 \leq i \leq n-k(n)+1} \left| \frac{1}{k(n)} \sum_{j=i+1}^{i+k(n)} Y_j / \left( \frac{1}{n} S_{n+1} \right) - 1 \right| \rightarrow 0 \text{ (P)} . \]

Since \( n^{-1} S_{n+1} \rightarrow 1 \) w.p.1 by the law of large numbers, (2.5) will follow if we can show that the sums \( \left( k(n) \right)^{-1} \sum_{i+1}^{i+k(n)} Y_j \) are uniformly near 1 in probability. For any \( \epsilon > 0 \),
\[(2.6) \quad P_n = \Pr \left[ \text{for some } i, \sum_{j=i+1}^{i+k(n)} |Y_j - 1| > k(n) \varepsilon \right]
\leq \sum_{i=1}^{n+1} \Pr \left[ \sum_{j=i+1}^{i+k(n)} (Y_j - 1) > k(n) \varepsilon \right] + \sum_{i=1}^{n+1} \Pr \left[ \sum_{j=i+1}^{i+k(n)} (Y_j - 1) < -k(n) \varepsilon \right].\]

Using the fact that \(\Pr[X > 0] \leq \mathbb{E}[e^{tX}]\) for any random variable \(X\) and \(t > 0\) such that the right side is finite, we obtain
\[
\Pr \left[ \sum_{j=i+1}^{i+k(n)} (Y_j - 1) > k(n) \varepsilon \right] \leq \mathbb{E}[e^{t(\sum_{j=i+1}^{i+k(n)} (Y_j - 1) - k(n) \varepsilon)}] = \left\{ e^{t(1+\varepsilon)/(1-t)} \right\}^{k(n)} \quad 0 < t < 1.
\]

(Recall that a sum of \(k(n) Y_j\)'s has the gamma distribution with parameter \(k(n)\).) Choosing the minimizing value \(t = 1 - (1+\varepsilon)^{-1}\) gives the bound \(\{(1+\varepsilon)e^{-\varepsilon}\}^{k(n)}\). A similar bound holds for each term of the second sum on the right side of (2.6). Therefore, \(P_n \leq (n+1) a(\varepsilon)^{-k(n)}\), where \(a(\varepsilon) > 1\) for \(\varepsilon > 0\). Since \((\log n) / k(n) \to 0\), \(P_n \to 0\) and (2.5) is proved.

It follows from (2.3) that \(U_k(n) \to 0\) (UP) and hence, since \(f\) is everywhere positive, that \(r_k(n) \to 0\) (UP).

To conclude (2.1) we need only (2.3) and the fact that \(U_k(n)/2r_k(n) \to f\) (UP). Since \(f\) is uniformly continuous and \(r_k(n) \to 0\) (UP), this is immediate from the estimate
\[(2.7) \quad \left| \frac{U_k(n)(z)}{2r_k(n)(z)} - f(z) \right| = \left| \frac{1}{2r_k(n)(z)} \int_{z-r}^{z+r} |f(t) - f(z)| \, dt \right|
\leq \max \{ |f(t) - f(z)| : z - r_k(n)(z) \leq t \leq z + r_k(n)(z) \}.
\]
The argument for (2.2) is slightly longer. Let \( i(z) \) be the index such that

\[ X_{i(z), n} \leq z < X_{i(z)+1, n} \]

For any compact interval \( I \), the probability that \( X_{ln} \) and \( X_{nn} \) fall outside \( I \) approaches \( 1 \) as \( n \to \infty \), by positivity of \( f \). Thus \( i(z) \) is defined for all \( z \in I \) with probability approaching \( 1 \) for large \( n \). The Glivenko-Cantelli theorem and uniform continuity of \( f^{-1} \) on \([\alpha, \beta - \alpha] \) for any \( \alpha > 0 \) give that

\[
(2.8) \quad \sup_{z \in I} \left| X_{i(z), n} - z \right| \to 0 \quad (P).
\]

From (2.8) and the fact that \( r_k(n) \to 0 \quad (U_P) \), we can conclude by an estimate analogous to (2.7) that

\[
\sup_{z \in I} \left| \frac{U_k(n)(X_{i(z), n})}{2r_k(n)(X_{i(z), n})} - f(z) \right| \to 0 \quad (P)
\]

and hence, using (2.3), that for any compact interval \( I \) and any \( \epsilon > 0 \),

\[
(2.9) \quad \lim_{n \to \infty} P\left[ \sup_{z \in I} \left| f_n^*(z) - f(z) \right| > \epsilon \right] = 0
\]
If we can establish that for any $\epsilon > 0$ there is a compact interval $I_\epsilon$ such that

$$\lim_{n \to \infty} P\left[ \sup_{z \notin I_\epsilon} |f_n^*(z) - f(z)| > \epsilon \right] = 0,$$

this with (2.9) will imply (2.2).

Since $f(z) \to 0$ as $z \to \pm \infty$, we can choose a compact interval $I^* = [a, b]$ such that $f(z) < \epsilon/2$ outside $I^*$. Then by (2.1), $f_n^*(z) < \epsilon$ for all $z \notin I^*$ with probability approaching 1 as $n \to \infty$. Let $I_\epsilon = [a + c, b + c]$ for some $c > 0$. Then by (2.8) and the fact that $P[X_{in} < a, X_{mn} > b + c] \to 1$, we have that

$$P[X_{i(z), n} \notin I^* \text{ for all } z \notin I_\epsilon \text{ with } X_{in} < z < X_{mn}] \to 1.$$ 

Thus with probability approaching 1, $f_n^*(z)$ is either 0 or $f_n^*(X_{in})$ for some $X_{in} \notin I^*$, for all $z \notin I_\epsilon$. This establishes (2.10).
References

