5-1-2012

The Weighted Hellinger Distance for Kernel Distribution Estimator of Function of Observations

Abdel-Razzaq Mugdadi

*Jordan University of Science and Technology, Irbid, Jordan*

Follow this and additional works at: [http://digitalcommons.wayne.edu/jmasm](http://digitalcommons.wayne.edu/jmasm)

Part of the [Applied Statistics Commons](http://digitalcommons.wayne.edu/jmasm), [Social and Behavioral Sciences Commons](http://digitalcommons.wayne.edu/jmasm), and the [Statistical Theory Commons](http://digitalcommons.wayne.edu/jmasm)

**Recommended Citation**

DOI: 10.22237/jmasm/1335846180
Available at: [http://digitalcommons.wayne.edu/jmasm/vol11/iss1/24](http://digitalcommons.wayne.edu/jmasm/vol11/iss1/24)

This Regular Article is brought to you for free and open access by the Open Access Journals at DigitalCommons@WayneState. It has been accepted for inclusion in Journal of Modern Applied Statistical Methods by an authorized editor of DigitalCommons@WayneState.
The Weighted Hellinger Distance for Kernel Distribution Estimator of Function of Observations

Abdel-Razzaq Mugdadi
Jordan University of Science and Technology, Irbid, Jordan

The asymptotic mean weighted Hellinger distance (AMWHD) is derived for the kernel distribution estimator of a function of observations. In addition, the AMWHD is compared with the asymptotic mean integrated square error (AMISE) of the estimator. A completely data based method is proposed to select the bandwidth in the estimator using the mean weighted Hellinger distance (MWHD).

Key words: Kernel estimation, distribution function estimation, bandwidth, Hellinger distance, mean square error, function of random variables.

Introduction

Given a random sample \( X_1, X_2, \ldots, X_n \) from a distribution \( F(x) \) with unknown density function \( f(x) \), the kernel density estimator (Rosenblatt, 1956) of \( f(x) \) is given by

\[
\hat{f}(x) = \frac{1}{nb} \sum_{i=1}^{n} k \left( \frac{x - X_i}{b} \right),
\]

where \( b \) is the smoothing bandwidth and \( k \) is a symmetric function satisfying \( \int k(x)dx = 1 \).

The kernel distribution function estimator (Nadaraya, 1964) of \( F(x) \) is given by

\[
\hat{F}(x) = \frac{1}{n} \sum_{i=1}^{n} \left( \frac{x - X_i}{b} \right),
\]

where \( K \) is the distribution function of the kernel \( k \), \( K(x) = \int k(u)(du) \), and \( b \) is the bandwidth.

Consider the function \( g(X_1, X_2, \ldots, X_m) \) that depends on \( m \geq 1 \) observations. Assume that \( g \) is a real value and is symmetric in its \( m \) arguments. Frees (1994) proposed an estimate for the density function \( h(t) \) of random variable \( g(X_1, X_2, \ldots, X_m) \) which is given by

\[
\hat{h}(t) = \frac{1}{bh(n\choose m)} \sum_{(n,m)} w \left( t - g(X_{i_1}, \ldots, X_{i_m}) \right),
\]

where \( b \) is the bandwidth, the sum extends over all \( 1 \leq i_1 < i_2 < \ldots < i_m \leq n \), and \( w(.) \) is a kernel function. If \( m = 1 \) and \( g(x) = x \), then the estimator \( \hat{h}(t) \) reduces to the estimator \( \hat{f}(x) \).
The estimator $\hat{h}(t)$ has many applications in real life. For example, in spatial statistics $g$ can be the inter point distance between pairs of objects and in insurance $g$ can be the sum of $m$ claims (Frees, 1994; Ahmad & Fan, 2001; Mugdadi & Ahmad, 2004).

Nadaraya (1964) and Mugdadi and Ghebregiorgis (2005) proposed a kernel distribution estimator of the distribution function of function of observations $H(t)$ as:

$$
\hat{H}(t) = \frac{1}{n} \sum_{(n,m)} W \left( \frac{t - g(X_{i_1}, ..., X_{i_m})}{b} \right),
$$

where $W(x) = \int_{-\infty}^{x} w(t) dt$, $b$ is the bandwidth and the sum extends over all $1 \leq i_1 < i_2 < ... < i_m \leq n$. Theoretical and simulation analyses show that choice of kernel is not crucial for distribution function estimation in the case of independent and identically (i.i.d) random variables; the most important choice is that of bandwidth. A typical way to select the bandwidth is to minimize one error measure, and the most commonly used is the mean integrated square error (MISE) and its asymptotic (AMISE), where

$$
\text{MISE}(\hat{H}(t)) = E \left[ (\hat{H}(t) - H(t))^2 \right] dt.
$$

Another criterion is the mean Hellinger distance (MHD), where

$$
\text{MHD}(\hat{H}(t)) = E \left[ (\hat{H}^2(t) - H^2(t))^2 \right] dt.
$$

Kanzawaa (1993) discussed the relationship between the asymptotic mean Hellinger distance (AMHD) and the AMISE for $\hat{f}(x)$, Ahmad and Muguadi (2006) examined the relationship between asymptotic mean weighted Hellinger distance (AMWHD) and the AMISE for both $\hat{f}(x)$ and $\hat{F}(x)$, and Mugdadi (2004) studied the AMWHD for $\hat{h}(t)$. This investigation examines the relationship between the AMWHD and the AMISE for $\hat{H}(t)$ and proposes a data method to select the bandwidth for $\hat{H}(t)$ based on the AMWHD $\left(\hat{H}(t)\right)$.

The Asymptotic Mean Weighted Hellinger Distance

One error criterion used to evaluate the estimator is the mean weighted Hellinger distance (MWHD) and its asymptotic (AMWHD), where $\text{MWHD}(\hat{H}(t))$ is defined by:

$$
\text{MWHD}(\hat{H}(t)) = E \left[ (\hat{H}^2(t) - H^2(t))^2 H(t) dt.
$$

It can be argued that

$$
\text{MWHD}(\hat{H}(t)) = \frac{1}{4} \text{MISE}(\hat{H}(t)),
$$

assuming that $\hat{H}(t) = H(t)$ results in:

$$
\text{MISE}(\hat{H}(t)) = \int E[\hat{H}(t) - H(t)]^2 dt
$$

$$
= \int E[\hat{H}^2(t) - H^2(t)]^2 [\hat{H}^2(t) + H^2(t)]^2 dt
$$

$$
= 4 E \int [\hat{H}^2(t) - H^2(t)]^2 H(t) dt
$$

$$
= 4 \text{MWHD}(\hat{H}(t))
$$

Next, the $\text{AMWHD}(\hat{H}(t))$ is derived and compared with the $\text{AMISE}(\hat{H}(t))$. Muguadi and Ghebregiorgis (2005) derived $\text{AMISE}(\hat{H}(t))$ as:

$$
\text{AMISE}(\hat{H}(t)) = \frac{T(H)}{n} - \frac{b}{n} \rho(w) + \frac{b^4}{4} \mu_2(w)^2 R(H^n),
$$

where

\begin{align*}
\rho(w) &= \frac{1}{n} \sum_{i=1}^{n} \frac{w(X_i)}{b^2}, \\
\mu_2(w) &= \frac{1}{n} \sum_{i=1}^{n} \frac{w(X_i)^2}{b^2}, \\
R(H^n) &= \frac{1}{n} \sum_{i=1}^{n} H^n(X_i).
\end{align*}
\[ \rho(w) = 2 \int u W(u) w(u) \, du , \]

\[ T(H) = \int H(t)[1 - H(t)] \, dt \]

and

\[ R(h) = \int h^2(t) \, dt . \]

Theorem

If the fourth derivative of \( H(t) \) exists, then

\[ \text{AMISE} (\hat{H}(t)) = \frac{T(H)}{4} - \frac{b}{4} \rho(w) + \frac{b^4}{16} \mu_2(w)^2 R(H'') . \]

Proof

\[ \text{MWHD} (\hat{H}(t)) = \int \hat{E}H(t)H(t) \, dt - 2 \int \hat{E}H^2(t)H^2(t) \, dt + \int H^2(t) \, dt . \]

Using integration by parts and expanding \( H(t - bu) \) in a 2nd order Taylor’s series about \( t \) results in:

\[
E(\hat{H}(t)) = E \left( \frac{1}{n} \sum_{(n,m)} W \left( \frac{t - g(X_1, X_2, ..., X_m)}{b} \right) \right)
\]

\[
= \int w(u) \left\{ \frac{H(t) - buH(t) + \frac{b^2 u^2 H''(t)}{2}}{b} \right. \\
\quad - \frac{b^3 u^3 H'''(t)}{6} + \frac{b^4 u^4 H^{(4)}(t)}{24} \\
\left. + o(b^4) \right\} \, du
\]

\[
= H(t) + \frac{b^2 H''(t)}{2} \mu_2(w) + \frac{b^4}{24} \mu_4(w) H^{(4)}(t)
\]

therefore,

\[
\int E[\hat{H}(t)H(t)] \, dt 
\]

\[
= \int \left[ H^2(t) \, dt + \frac{b^2 \mu_2(w)}{2} \int H''(t)H(t) \, dt \right] \\
+ \frac{b^4 \mu_4(w)}{24} \int H^{(4)}(t)H(t) \, dt
\]

If \( Z \) is a random variable with a standard normal distribution, then

\[
\hat{H}(t) = E(\hat{H}(t)) + Z \sqrt{\text{Var}(\hat{H}(t))} .
\]

Mugdadi and Ghebregiorgis (2005) derived \( \text{Var}(\hat{H}(t)) \), this is given by:

\[
\text{Var}(\hat{H}(t)) = 
\left[ O \left( \frac{1}{n} \right) + O \left( \frac{b^2}{n} \right) + \left( \frac{n}{m} \right)^{-1} H(t)(1 - H(t)) \right] \\
+ \left( \frac{n}{m} \right)^{-1} o(b) - \left( \frac{n}{m} \right)^{-1} bH'(t) \rho(w)
\]

Thus,

\[
\hat{H}(t) 
\]

\[ 
= H(t) + \frac{b^2 H''(t)}{2} \mu_2(w) + \frac{b^4 H^{(4)}(t)}{24} \mu_4(w) \\
+ Z \left[ O \left( \frac{1}{n} \right) + O \left( \frac{b^2}{n} \right) + \left( \frac{n}{m} \right)^{-1} H(t)(1 - H(t)) \right] \\
+ \left( \frac{n}{m} \right)^{-1} o(b) - \left( \frac{n}{m} \right)^{-1} bH'(t) \rho(w)
\]

\[ 
= H(t) \\
+ Z \left[ \frac{1}{2H(t)} \frac{H'(t)\rho(w)}{H(t)} + \frac{1}{24H(t)} \left( \frac{1}{n} \right) + O(b^2) \right]
\]
therefore,

\[
\hat{H}_t^2(t) = \left[ 1 + \frac{b^2 H''(t) \mu_z(w)}{4H(t)} + \frac{b^4 \mu_2(w) H^{(4)}(t)}{48H(t)} \right] + \frac{Z}{2} \left[ \frac{1-H(t)}{n} \frac{H(t)}{m} - \frac{bH'(t)\rho(w)}{n} \right]^{\frac{1}{2}} \\
+ O\left( \frac{1}{n} \right) + O(b^2) \\
- \frac{b^4}{32} \left( \frac{H''(t)}{H(t)} \right)^2 (\mu_z(w))^2 \\
- \frac{Z^2}{8} \left[ \frac{1-H(t)}{n} \frac{H(t)}{m} - \frac{bH'(t)\rho(w)}{n} \right] \\
+ O\left( \frac{1}{n} \right) + O(b^2)
\]

and

\[
E(\hat{H}_t^2(t)) = \left[ 1 + \frac{b^2 H''(t) \mu_z(w)}{4H(t)} + \frac{b^4 \mu_2(w) H^{(4)}(t)}{48H(t)} \right] \\
- \frac{b^4}{32} \left( \frac{H''(t)}{H(t)} \right)^2 (\mu_z(w))^2 \\
- \frac{1}{8} \left[ \frac{1-H(t)}{n} \frac{H(t)}{m} - \frac{bH'(t)\rho(w)}{n} \right] \\
+ O\left( \frac{1}{n} \right) + O(b^2)
\]

thus

\[
\int [EH_t^2(t)\hat{H}_t^2(t)] dt \approx
\left[ \int H^2(t) dt + \frac{b^2}{4} \mu_2(w) \int H''(t)H(t) dt \right. \\
+ \frac{b^4}{48} \mu_4(w) \int H^{(4)}(t)H(t) dt \right. \\
- \frac{b^4}{32} \mu_2^2(w) \int (H''(t))^2 dt \left. \\
- \frac{1}{8} \left( \frac{n}{m} \right) \int H(t)(1-H(t)) dt \right. \\
+ \frac{b}{8} \rho(w) + O\left( \frac{1}{n} \right) + O(b^2)
\]

therefore,

\[
AMWHD(\hat{H}(t)) = \frac{b^4}{4} \mu_2^2(w)R(H'') \\
+ \frac{1}{4(n)} \int (H(t) \times [1-H(t)] dt - \frac{b}{4(n)} \rho(w)
\]

Under these conditions, the following corollaries can be proven.

Corollary 1

\[
AMWHD(\hat{H}(t)) \approx \frac{AMISE(\hat{H}(t))}{4}
\]

Similar to Powell and Stocker (1996), the optimal bandwidth to minimize the AMWHD(\hat{H}(t)) is shown in corollary 2.
Corollary 2

\[ b_{opt} = \left( \frac{\mu}{\mu_\rho^2} \right)^{\frac{1}{3}} \left( \frac{n}{m} \right)^{-\frac{1}{3}} \]

Bandwidth Selection

The choice of bandwidth is very important in the Kernel density estimator as well as in the Kernel distribution estimator. One of the simplest methods to select bandwidth is based on equation (2.8). Assume that the data is from a normal distribution with mean equal to zero and variance \( \sigma^2 \). If \( s^2 \) is the variance of the data, \( g(X_{i_1}, \ldots, X_{i_m}) \) for all \( 1 \leq i_1 < \ldots < i_m \leq n \), then \( R(h') = \frac{1}{4\pi s^3} \), thus

\[ b_{opt} = \left( \frac{\rho(w)4\pi}{\mu_\rho^2} \right)^{\frac{1}{3}} \left( \frac{n}{m} \right)^{-\frac{1}{3}} s \]

(3.1)

This shows that \( b_{opt} \) depends only on the standard deviation and on the Kernel function.

A completely different data based method is proposed to select the bandwidth for a Kernel distribution estimator of the function of observations. The method is based on minimizing the \( MWHD(\hat{H}(t)) \). The \( MWHD(\hat{H}(t)) \) is defined as:

\[ MWHD(\hat{H}(t)) = E\left[ \frac{1}{H^2(t)} - \frac{1}{H^2(t)} H(t)dt \right] \]

Minimizing \( MWHD(\hat{H}(t)) \) is therefore equivalent to minimizing \( MWHD1(\hat{H}(t)) \), where,

\[ MWHD1(\hat{H}(t)) = \int (E\hat{H}(t)H(t)dt - 2\int (E\hat{H}^2(t))H(t) \frac{1}{2} dt) \]

Thus, \( MWHD1(\hat{H}(t)) \) can be estimated as follows. Let \( m_{(i)} \) be a fixed choice of \( m \) variables and let \( I(A) \) be the indicator function. Also, define

\[ H_{n,m_{(i)}}(t) = \frac{1}{\binom{n}{m}} \sum_{(n,m),m=m_{(i)}} I(t-g(X_{i_1}, \ldots, X_{i_m})) \]

and

\[ \hat{H}_{-m_{(i)}}(t) = \frac{1}{\binom{n}{m}} \sum_{(n,m),m=m_{(i)}} W\left( \frac{t-g(X_{i_1}, \ldots, X_{i_m})}{b} \right) \]

which is the distribution estimated based on a sample with \( m_{(i)} \) deleted. Thus, \( MWHD1(\hat{H}(t)) \) is estimated by

\[ MWHD.EST = \frac{1}{\binom{n}{m}} \sum_{1 \leq i_1 < \ldots < i_m < n} \left[ \hat{H}_{-m_{(i)}}(g(X_{i_1}, \ldots, X_{i_m})) \times H_{n,m_{(i)}}(g(X_{i_1}, \ldots, X_{i_m})) \right] \]

As noted, there are many applications for \( \hat{H}(t) \). One example, introduced by Free (1993), regards an insurance claims problem. Table 1 shows total hospital charges (in dollars) in one Wisconsin (USA) hospital for females aged 30-49 in the year 1989.

Consider the case \( m = 2 \) and the function \( g(X_1, X_2) = X_1 + X_2 \). By minimizing \( MWHD.EST \), the bandwidth is determined to be 0.437. Figure 1 shows the kernel distribution function for \( g(X_1, X_2) \) using data in Table 1. It is clear that the kernel
estimate is smooth and \( \hat{H}(t) \) is 0 when \( t \leq 0 \) because the sum of the charges should be positive; also, \( \lim_{t \to -\infty} \hat{H}(t) = 1 \).

Table 1: Total 1989 Hospital Charges (USD) for Females Aged 30-49

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2337</td>
<td>1765</td>
<td>1802</td>
</tr>
<tr>
<td>2179</td>
<td>2467</td>
<td>2011</td>
</tr>
<tr>
<td>2348</td>
<td>3609</td>
<td>2270</td>
</tr>
<tr>
<td>4765</td>
<td>2141</td>
<td>3425</td>
</tr>
<tr>
<td>3041</td>
<td>1850</td>
<td>3558</td>
</tr>
<tr>
<td>2088</td>
<td>3191</td>
<td>2315</td>
</tr>
<tr>
<td>2872</td>
<td>3020</td>
<td>1642</td>
</tr>
<tr>
<td>1924</td>
<td>2473</td>
<td>5878</td>
</tr>
<tr>
<td>2294</td>
<td>1898</td>
<td>2101</td>
</tr>
<tr>
<td>2182</td>
<td>7787</td>
<td>2242</td>
</tr>
<tr>
<td>2138</td>
<td>6169</td>
<td>5746</td>
</tr>
</tbody>
</table>

References


