Histogram Density Estimators Based Upon a Fuzzy Partition
Kevin Loquin, Olivier Strauss

To cite this version:

HAL Id: lirmm-00366761
https://hal-lirmm.ccsd.cnrs.fr/lirmm-00366761
Submitted on 9 Mar 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Histogram density estimators based upon a fuzzy partition

Kevin Loquin*, Olivier Strauss

Laboratoire d’Informatique, de Robotique et de Microélectronique de Montpellier, Université Montpellier II, 161, rue Ada, 34392 Montpellier Cedex 5, France

Received 8 December 2006; received in revised form 28 August 2007; accepted 9 January 2008

Abstract

This paper presents a density estimator based upon a histogram computed on a fuzzy partition. We prove the consistency of this estimator in the Mean Squared Error (MSE). We give the optimal bin width of the estimator which minimizes the Asymptotic Integrated Mean Squared Error (AIMSE).

© 2008 Elsevier B.V. All rights reserved.

1. Introduction

The histogram is probably the oldest\(^1\) and most widely used density estimator for the presentation and exploration of observed data. A histogram is constructed by partitioning a given reference interval (or domain) \(D\) of \(\mathbb{R}^d\) (we will just focus on the univariate case \(d = 1\)) into \(p\) bins \(A_k\) and counting the number \(A_{ck}\) of observations that belong to each cell \(A_k\).

\[
A_{ck} = \sum_{i=1}^{n} I_{Ak}(X_i).
\]

From these counts, we can derive a consistent estimate \(\hat{f}_{\text{hist}}(x)\) of the underlying probability density function \(f(x)\) at any point \(x\) of \(A_k\), computed by:

\[
\hat{f}_{\text{hist}}(x) = \frac{A_{ck}}{nh}.
\]

However, the histogram estimate has some weaknesses and, particularly, the choice of reference interval and number of cells (or bin width) have a marked effect on the estimated density. In the last 5 years, some authors have suggested that the effect of partitioning arbitrariness can be reduced by replacing the crisp partition by a fuzzy partition (Runkler,

---

\(^*\) Corresponding author.

E-mail addresses: kevin.loquin@lirmm.fr (K. Loquin), olivier.strauss@lirmm.fr (O. Strauss).

\(1\) Probably dating from the works of John Graunt in 1662. See Westergaard (1968).
2. Preliminary concepts

First, the classical subsets of \( \mathbb{R} \) can be generalized in fuzzy subsets of \( \mathbb{R} \).

**Definition 1.** The fuzzy subset \( F \) is defined by its membership function

\[
\mu_F : \mathbb{R} \rightarrow [0, 1]
\]

assigning the value \( \mu_F(x) \in L \) to each element \( x \in \mathbb{R} \) which is the membership degree of \( x \) in \( F \).

The crisp partition, i.e. where \( A_k \) are \( p \) classical subsets of \( \mathbb{R} \), can then be generalized in a fuzzy partition, i.e. where \( A_k \) are \( p \) fuzzy subsets of \( \mathbb{R} \).

**Definition 2.** Let \( D = [a, b] \subset \mathbb{R} \) be the domain of the partition. Let \( m_1 < m_2 < \cdots < m_p \) be \( p \) fixed nodes of \( D \), such that \( m_1 = a \) and \( m_p = b \), with \( p \geq 3 \), and \( \forall k \neq p, m_{k+1} - m_k = h = \text{constant} \). Let \( m_0 := a - h \) and \( m_{p+1} := b + h \). Let \( D' = [m_0, m_{p+1}] \subset \mathbb{R} \) be the extended domain of the partition. We say that the set of the \( p \) fuzzy subsets \( A_1, A_2, \ldots, A_p \), identified with their membership functions \( \mu_{A_1}(x), \mu_{A_2}(x), \ldots, \mu_{A_p}(x) \) defined on \( D' \), form a strong uniform fuzzy partition of the universe, if they fulfil

1. \( \mu_{A_k}(m_k) = 1 \),
2. \( \mu_{A_k}(x) = 0 \) if \( x \notin [m_{k-1}, m_k] \),
3. \( \mu_{A_k}(x) \) monotonically increases on \( [m_{k-1}, m_k] \) and \( \mu_{A_k}(x) \) monotonically decreases on \( [m_k, m_{k+1}] \),
4. \( \forall x \in D', \exists k \), such that \( \mu_{A_k}(x) > 0 \),
5. \( \forall x \in D, \sum_{k=1}^{p} \mu_{A_k}(x) = 1 \) (strength condition),
6. \( \forall x \in [0, h], \mu_{A_k}(m_k - x) = \mu_{A_k}(m_k + x) \),
7. \( \forall x \in [m_k, m_{k+1}], \mu_{A_k}(x) = \mu_{A_{k-1}}(x - h) \) and \( \mu_{A_{k+1}}(x) = \mu_{A_k}(x - h) \).

**Proposition 1.** Let \( (A_k)_{k=1, \ldots, p} \) be a strong uniform fuzzy partition on \( D \), extended on \( D' \), then

1. \( \forall x \in D, \exists k \in \{1, \ldots, p - 1\} \), such that \( \forall k \notin \{0, k_0 + 1\}, \mu_{A_k}(x) = 0 \), and \( \mu_{A_{k_0}}(x) + \mu_{A_{k_0+1}}(x) = 1 \).
2. \( \forall k \in \{1, \ldots, p\}, \int_{m_{k-1}}^{m_{k+1}} \mu_{A_k}(x)dx = h \).
3. \( \exists K_A : [-1, 1] \rightarrow [0, 1] \) even, such that, \( \mu_{A_k}(x) = K_A\left(\frac{x - m_k}{h}\right) \mathbb{1}_{[m_{k-1}, m_{k+1}]} + \int_{-1}^{1} K_A(u)du = 1 \).

Table 1 provides illustrative examples of membership functions of strong uniform fuzzy partitions that are obtainable by point (3) of Proposition 1. The first column contains the crisp partition, the second one, the cosine partition and the last column, a triangular fuzzy partition, formed by fuzzy triangular numbers.

<table>
<thead>
<tr>
<th></th>
<th>Crisp</th>
<th>Cosine</th>
<th>Triangular</th>
</tr>
</thead>
<tbody>
<tr>
<td>( K_A(x) = )</td>
<td>( \mathbb{1}_{[-\frac{1}{2}, \frac{1}{2}]}(x) )</td>
<td>( \frac{1}{2} (\cos(\pi x) + 1) \mathbb{1}_{[-1, 1]}(x) )</td>
<td>( (1 -</td>
</tr>
</tbody>
</table>

3. Fuzzy-partition-based histogram density estimator

The accumulated value \( Acc_k \) is the key feature of the histogram technique. It represents the number of observations in complete agreement with the label represented by the restriction of the real line to the interval (or bin) \( A_k \). As the partitioning is highly arbitrary, the histogram technique is known to be very sensitive to the choice of both reference interval and number of cells (or bin width). As mentioned before, the effect of this arbitrariness can be reduced by replacing the crisp partition by a fuzzy partition of the real line. The paper Loquin and Strauss (2006) underlines...
this property of the fuzzy histogram. Note that this approach is quite similar to the binned kernel density estimator approach (Hardle and Scott, 1992; Scott and Sheather, 1985).

Let \((A_k)_{k=1,...,p}\) be a strong uniform fuzzy partition of \(D\), then the natural extension of expression (1) induces a distributed vote:

\[
Acc_k = \sum_{i=1}^{n} \mu_{A_k}(X_i).
\]

As far as the density estimator (2) is concerned, since \(A_k\) is a fuzzy subset, expression (2) no longer holds for any \(x \in A_k\), but normalized accumulators \(\frac{Acc_k}{nh}\) now have degrees of truth inherited from the fuzzy nature of \(A_k\). Then, the \(\frac{Acc_k}{nh}\) value is more true at \(m_k\) than at any other point of \(D\). We propose to once again use the concept of strong uniform fuzzy partition of \(p\) fuzzy subsets to provide an interpolation of the \(p\) points \((m_k, \frac{Acc_k}{nh})\) on \(D\).

**Definition 3.** The fuzzy histogram density estimator defined on \(D\) is given by

\[
f_{nh}(x) = \frac{1}{nh} \sum_{k=1}^{p} Acc_k \mu_{B_k}(x),
\]

where \((B_k)_{k=1,...,p}\) is a strong uniform fuzzy partition defined on \(D\).

It can easily be shown that \(f_{nh} \geq 0\), and \(\int f_{nh}(x)\,dx = 1\) and that \(f_{nh}\) goes through the \(p\) points \((m_k, \frac{Acc_k}{nh})\).

4. MSE consistency

One way to evaluate \(f_{nh}\) is via some measure of its local difference from \(f\). One of the most common measures is the Mean Squared Error:

\[
MSE(x) \triangleq \mathbb{E}_{f}[\{f_{nh}(x) - f(x)\}]^2.
\]

Besides, we have \(MSE(x) = b^2(x) + \sigma^2(x)\), where \(b^2(x)\) is the squared bias and \(\sigma^2(x)\) is the variance of the estimator \(f_{nh}\) in \(x\). We will bound \(b^2(x)\) and \(\sigma^2(x)\) under the following assumptions

**Assumption 1.** \(f\) is a \(C^2\) function with bounded derivatives,

**Assumption 2.** \(K_A\) and \(K_B\), as defined by point (3) of Proposition 1, are square integrable.

Prior to bounding \(b^2(x)\) and \(\sigma^2(x)\), useful equalities are provided to easily approximate them. Let \(m_{ck}\) be the center of \([m_k, m_{k+1}]\). Thanks to a change of variable \(u \leftarrow \frac{x - m_{ck}}{h}\), point (3) of Proposition 1, and the Taylor expansion of \(f\), we have

\[
\mathbb{E}_{f}[\mu_{A_k}(X_i)] = \sum_{m=1}^{r} h^m f^{(m-1)}(m_{ck}) \int_{-1}^{1} K^p_A(u) \left(u - \frac{1}{2}\right)^{m-1} du + \mathcal{O}(h^{r+1}), \tag{3}
\]

\[
\mathbb{E}_{f}[\mu_{A_k+1}(X_i)] = \sum_{m=1}^{r} h^m f^{(m-1)}(m_{ck}) \int_{-1}^{1} K^p_A(u) \left(u + \frac{1}{2}\right)^{m-1} du + \mathcal{O}(h^{r+1}), \tag{4}
\]

for \(f \in C^r\) with bounded derivatives.

**Squared bias bounding**

For \(x \in [m_k, m_{k+1}]\), the fuzzy histogram density estimator is given by

\[
f_{nh}(x) = f_{nh}(m_k) \mu_{B_k}(x) + f_{nh}(m_{k+1}) \mu_{B_{k+1}}(x),
\]

with \(f_{nh}(m_k) = \frac{Acc_k}{nh}\) and \(f_{nh}(m_{k+1}) = \frac{Acc_{k+1}}{nh}\). The squared bias is given by

\[
b^2_k(x) = (\mathbb{E}_{f}[f_{nh}(x)] - f(x))^2. \tag{5}
\]
By implementing expressions (3) and (4) with \( p = 1 \) and \( r = 2 \) (see Assumption 1), we get

\[
\mathbb{E}_f[f_{nh}(m_k)] = \sum_{i=1}^{n} \frac{\mathbb{E}_f[\mu_{A_k}(X_i)]}{nh} = f(m_c) - \frac{h}{2} f'(m_c) + O(h^2),
\]

\[
\mathbb{E}_f[f_{nh}(m_{k+1})] = \sum_{i=1}^{n} \frac{\mathbb{E}_f[\mu_{A_{k+1}}(X_i)]}{nh} = f(m_c) + \frac{h}{2} f'(m_c) + O(h^2).
\]

Therefore,

\[
\mathbb{E}_f[f_{nh}(x)] = f(m_c) + \frac{h}{2} f'(m_c)(\mu_{B_{k+1}}(x) - \mu_{B_k}(x)) + O(h^2). \tag{6}
\]

Now, for \( x \in [m_k, m_{k+1}] \), \( f(x) \) can be approximated by

\[
f(x) = f(m_c) + (x - m_c) f'(m_c) + O(h^2). \tag{7}
\]

Therefore, from expressions (5)–(7),

\[
b_k^2(x) = f'(m_c)^2 \left[ \frac{h}{2}(\mu_{B_{k+1}}(x) - \mu_{B_k}(x)) + (m_c - x) \right]^2 + O(h^4).
\]

For \( x \in D \), the squared bias is bounded by

\[
b^2(x) \leq \max_{k=1,\ldots,p-1} f'(m_c)^2 \left[ \frac{h}{2}(\mu_{B_{k+1}}(x) - \mu_{B_k}(x) + 1) \right]^2, \quad \text{because } m_c - x \leq \frac{h}{2}
\]

\[
b^2(x) \leq \max_{k=1,\ldots,p-1} f'(m_c)^2 \left[ h\mu_{B_{k+1}}(x) \right]^2, \quad \text{because } \mu_{B_k}(x) + \mu_{B_{k+1}}(x) = 1
\]

\[
b^2(x) \leq \max_{k=1,\ldots,p-1} f'(m_c)^2 h^2. \tag{8}
\]

**Variance bounding**

The variance of the estimator (4), for \( x \in [m_k, m_{k+1}] \), is given by \( \sigma_k^2(x) = \mathbb{E}_f[f_{nh}(x) - \mathbb{E}_f[f_{nh}(x)]^2] \).

Let us define \( \eta_{i,k} = \mu_{A_k}(X_i) - \mathbb{E}_f[\mu_{A_k}(X_i)] \). We can easily obtain:

\[
\sigma_k^2(x) = \frac{1}{(nh)^2} \mathbb{E}_f \left( \mu_{B_k}(x) \sum_{i=1}^{n} \eta_{i,k} + \mu_{B_{k+1}}(x) \sum_{i=1}^{n} \eta_{i,k+1} \right)^2
\]

\[
= \frac{1}{(nh)^2} \mu_{B_k}(x) \sum_{i=1}^{n} \mathbb{E}_f(\eta_{i,k})^2 + \frac{1}{(nh)^2} 2\mu_{B_k}(x)\mu_{B_{k+1}}(x) \sum_{i=1}^{n} \mathbb{E}_f(\eta_{i,k}\eta_{i,k+1})
\]

\[
+ \frac{1}{(nh)^2} \mu_{B_{k+1}}(x) \sum_{i=1}^{n} \mathbb{E}_f(\eta_{i,k+1})^2.
\]

By implementing expressions (3) and (4) with \( p = 2 \) and \( r = 1 \) (see Assumption 1), we get

\[
\mathbb{E}_f(\eta_{i,k})^2 = \mathbb{E}_f[\mu_{A_k}(X_i)^2] - \mathbb{E}_f[\mu_{A_k}(X_i)]^2,
\]

\[
= hf(m_c) \int_{-1}^{1} K^2_A(u) du - h^2 f^2(m_c) + O(h^2),
\]

\[
\mathbb{E}_f(\eta_{i,k+1})^2 = \mathbb{E}_f[\mu_{A_{k+1}}(X_i)^2] - \mathbb{E}_f[\mu_{A_{k+1}}(X_i)]^2,
\]

\[
= hf(m_c) \int_{-1}^{1} K^2_A(u) du - h^2 f^2(m_c) + O(h^2),
\]
$$\mathbb{E}_f(\eta_{i,k+1}) = \mathbb{E}_f[\mu_{A_k}(X_i)] - \mathbb{E}_f[\mu_{A_{k+1}}(X_i)] = hf(m_{ck}) \int_0^1 K_A(u)K_A(u-1)du - h^2 f^2(m_{ck}) + O(h^2).$$

Which leads to

$$\sigma^2_k(x) = \frac{f(m_{ck})}{nh} \left[ \mu^2_B(x) \int_{-1}^1 K^2_A(u)du + 2\mu_B(x)\mu_{B_{k+1}}(x) \int_{-1}^1 K_A(u)K_A(u-1)du \right. $$

$$+ \mu^2_{B_{k+1}}(x) \int_{-1}^1 K^2_A(u)du \left. \right] - \frac{f^2(m_{ck})}{n} + O\left(\frac{1}{n}\right).$$

According to Cauchy–Schwartz, $f_0^1 K_A(u)K_A(u-1)du \leq f_0^1 K^2_A(u)du$. Therefore

$$\sigma^2_k(x) \leq -\frac{f(m_{ck})}{nh} \int_{-1}^1 K^2_A(u)du.$$ 

For $x \in D$, the variance is bounded by

$$\sigma^2(x) \leq \frac{\max_{k=1,\ldots,p-1} f(m_{ck})}{nh} \int_{-1}^1 K^2_A(u)du.$$  

(9)

**Conclusion**

The following theorem summarizes inequalities (8) and (9).

**Theorem 2.** Under Assumptions 1 and 2,

$$\text{MSE}(x) \leq \max_{k=1,\ldots,p-1} f'(m_{ck})h^2 + \frac{\max_{k=1,\ldots,p-1} f(m_{ck})}{nh} \int_{-1}^1 K^2_A(u)du,$$

hence, $\forall x \in D$, $f_{nh}(x)$ is consistent in the MSE, i.e.

$$h \rightarrow 0 \quad \text{and} \quad nh \rightarrow +\infty \Rightarrow \text{MSE}(x) \rightarrow 0.$$

**5. IMSE consistency and optimization**

Optimization of our estimator consists of finding the bandwidth $h^*$ which minimizes the upper bound of the IMSE, given by $\text{IMSE} = \int_D \text{MSE}(x)dx$. We proved that this upper bound, called AIMSE for Asymptotic Integrated Mean Squared Error, is given by:

$$\text{AIMSE} = h^2C_1 \int_D f'(x)^2 dx + C_2,$$

where $C_1 = \int_0^1 K^2_B(u)du \int_{-1}^1 K^2_A(u)du + 2 \int_0^1 K_B(u)K_B(u-1)du \int_0^1 K_A(u)K_A(u-1)du + \int_{-1}^0 K^2_B(u)du \int_{-1}^1 K^2_A(u)du$ and

$$C_2 = \frac{1}{4} \int_0^1 (K_B(y-1) - K_B(y))^2 dy + 2 \int_0^1 K_B(y)y dy - \frac{5}{12},$$

hence

$$h^* = \left(\frac{C_2}{2nC_1 \int_D f'(x)^2 dx}\right)^{\frac{1}{3}}$$

(11)

which leads to $\text{AIMSE} = O(n^{-2/3}).$

This AIMSE result (10), which you can find in a different form in Waltman et al. (2005), is a generalization of the AIMSE of the classical histogram, which can be found in Scott (1992, 1979). Indeed, with $K_A(x) = K_B(x) = \frac{1}{2}[1,1)(x)$, expression (10) becomes

$$\text{AIMSE} = \frac{h^2}{12} \int_D f'(x)^2 dx + \frac{1}{nh}.$$"
6. Conclusion

Here we have presented a histogram density estimator based upon a fuzzy partition. We have proved its consistency in MSE and obtained its optimal bin width for AIMSE minimization. The main advantage of this tool is that it improves the robustness of the histogram density estimator with respect to the partitioning arbitrariness. Illustrations of this fuzzy histogram density estimator property are provided in Loquin and Strauss (2006).

References