Histogram Density Estimators Based Upon a Fuzzy Partition

Kevin Loquin, Olivier Strauss

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

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 Acc_k of observations that belong to each cell A_k.

$$Acc_k = \sum_{i=1}^{n} \mathbb{1}_{A_k}(X_i). \tag{1}$$

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:

$$\hat{f}_{\text{hist}}(x) = \frac{Acc_k}{nh}. \tag{2}$$

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).

2004; Van Den Berg, 2001; Strauss et al., 2000; Strauss and Comby, 2002). In the present paper, we propose a density estimator based upon a fuzzy partition. We define this kind of partition in Section 2 and present our estimator in Section 3. The MSE consistency is demonstrated in Section 4 and the optimal bandwidth for AIMSE minimization is provided in Section 5.

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

$$\mu_F : \mathbb{R} \to L = [0, 1]$$

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 $\forall k \neq p$, $m_{k+1} - m_k = h = constant$, so, $m_k = a + (k-1)h$. 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) if $x \notin [m_{k-1}, m_{k+1}]$, $\mu_{A_k}(x) = 0$,
(3) $\mu_{A_k}(x)$ monotically increases on $[m_{k-1}, m_k]$ and $\mu_{A_k}(x)$ monotically 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. $\forall x \in D$, $\exists! k_0 \in \{1, \ldots, p-1\}$, *such that* $\forall k \notin \{k_0, k_0+1\}$, $\mu_{A_k}(x) = 0$, *and* $\mu_{A_{k_0}}(x) + \mu_{A_{k_0+1}}(x) = 1$.
2. *for* $k = 1, \ldots, p$, $\int_{m_{k-1}}^{m_{k+1}} \mu_{A_k}(x)\mathrm{d}x = h$.
3. $\exists K_A : [-1, 1] \longrightarrow [0, 1]$ *even, such that,* $\mu_{A_k}(x) = K_A(\frac{x-m_k}{h})\mathbb{1}_{[m_{k-1}, m_{k+1}]}$ *and* $\int_{-1}^{1} K_A(u)\mathrm{d}u = 1$.

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 1
Strong uniform fuzzy partition examples

	Crisp	Cosine	Triangular
$K_A(x) =$	$\mathbb{1}_{[-\frac{1}{2}, \frac{1}{2}]}(x)$	$\frac{1}{2}(\cos(\pi x) + 1)\mathbb{1}_{[-1,1]}(x)$	$(1 - |x|)\mathbb{1}_{[-1,1]}(x)$

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:

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

Besides, we have $\text{MSE}(x) = b^2(x) + \sigma^2(x)$, 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 $u \leftarrow \frac{X_i - m_k}{h}$, point (3) of Proposition 1, and the Taylor expansion of f, we have

$$\mathbb{E}_f[\mu_{A_k}^p(X_i)] = \sum_{m=1}^{r} h^m f^{(m-1)}(m_{ck}) \int_{-1}^{1} K_A^p(u) \left(u - \frac{1}{2}\right)^{m-1} du + \mathcal{O}(h^{r+1}), \qquad (3)$$

$$\mathbb{E}_f[\mu_{A_{k+1}}^p(X_i)] = \sum_{m=1}^{r} h^m f^{(m-1)}(m_{ck}) \int_{-1}^{1} K_A^p(u) \left(u + \frac{1}{2}\right)^{m-1} du + \mathcal{O}(h^{r+1}), \qquad (4)$$

for $f C^r$ 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

$$b_k^2(x) = (\mathbb{E}_f[f_{nh}(x)] - f(x))^2. \qquad (5)$$

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

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

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

Therefore,

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

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

$$f(x) = f(m_{ck}) + (x - m_{ck}) f'(m_{ck}) + \mathcal{O}(h^2). \tag{7}$$

Therefore, from expressions (5)–(7),

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

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

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

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

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

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:

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

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

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

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

$$\mathbb{E}_f(\eta_{i,k})^2 = \mathbb{E}_f[\mu_{A_k}(X_i)^2] - \mathbb{E}_f[\mu_{A_k}(X_i)]^2,$$
$$= h f(m_{c_k}) \int_{-1}^{1} K_A^2(u) \mathrm{d}u - h^2 f^2(m_{c_k}) + \mathcal{O}(h^2),$$

$$\mathbb{E}_f(\eta_{i,k+1})^2 = \mathbb{E}_f[\mu_{A_{k+1}}(X_i)^2] - \mathbb{E}_f[\mu_{A_{k+1}}(X_i)]^2,$$
$$= h f(m_{c_k}) \int_{-1}^{1} K_A^2(u) \mathrm{d}u - h^2 f^2(m_{c_k}) + \mathcal{O}(h^2),$$

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

$$= hf(m_{c_k})\int_0^1 K_A(u)K_A(u-1)\mathrm{d}u - h^2 f^2(m_{c_k}) + \mathcal{O}(h^2).$$

Which leads to

$$\sigma_k^2(x) = \frac{f(m_{c_k})}{nh}\left[\mu_{B_k}^2(x)\int_{-1}^1 K_A^2(u)\mathrm{d}u + 2\mu_{B_k}(x)\mu_{B_{k+1}}(x)\int_0^1 K_A(u)K_A(u-1)\mathrm{d}u\right.$$

$$\left. + \mu_{B_{k+1}}^2(x)\int_{-1}^1 K_A^2(u)\mathrm{d}u\right] - \frac{f^2(m_{c_k})}{n} + \mathcal{O}\left(\frac{1}{n}\right).$$

According to Cauchy–Schwartz, $\int_0^1 K_A(u)K_A(u-1)\mathrm{d}u \le \int_{-1}^1 K_A^2(u)\mathrm{d}u$. Therefore

$$\sigma_k^2(x) \le \frac{f(m_{c_k})}{nh}\int_{-1}^1 K_A^2(u)\mathrm{d}u.$$

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

$$\sigma^2(x) \le \frac{\max_{k=1,\dots,p-1} f(m_{c_k})}{nh}\int_{-1}^1 K_A^2(u)\mathrm{d}u. \tag{9}$$

Conclusion

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

Theorem 2. *Under Assumptions 1 and 2,*

$$\mathrm{MSE}(x) \le \max_{k=1,\dots,p-1} f'(m_{ck})^2 h^2 + \frac{\max_{k=1,\dots,p-1} f(m_{c_k})}{nh}\int_{-1}^1 K_A^2(u)\mathrm{d}u,$$

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

$$h \longrightarrow 0 \quad and \quad nh \longrightarrow +\infty \Longrightarrow MSE(x) \longrightarrow 0.$$

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 $\mathrm{IMSE} = \int_D \mathrm{MSE}(x)\mathrm{d}x$. We proved that this upper bound, called AIMSE for Asymptotic Integrated Mean Squared Error, is given by:

$$\mathrm{AIMSE} = h^2 C_1 \int_D f'(x)^2\mathrm{d}x + \frac{C_2}{nh}, \tag{10}$$

with $C_1 = \int_0^1 K_B^2(u)\mathrm{d}u\int_{-1}^1 K_A^2(u)\mathrm{d}u + 2\int_0^1 K_B(u)K_B(u-1)\mathrm{d}u\int_0^1 K_A(u)K_A(u-1)\mathrm{d}u + \int_{-1}^0 K_B^2(u)\mathrm{d}u\int_{-1}^1 K_A^2(u)\mathrm{d}u$ and $C_2 = \frac{1}{4}\int_0^1 (K_B(y-1)-K_B(y))^2\mathrm{d}y + 2\int_0^1 K_B(y)y\mathrm{d}y - \frac{5}{12}$, hence

$$h^* = \left(\frac{C_2}{2nC_1\int_D f'(x)^2\mathrm{d}x}\right)^{\frac{1}{3}} \tag{11}$$

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

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 $K_A(x) = K_B(x) = \frac{1}{2}\mathbb{1}_{[-1,1]}(x)$, expression (10) becomes

$$\mathrm{AIMSE} = \frac{h^2}{12}\int_D f'(x)^2\mathrm{d}x + \frac{1}{nh}.$$

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

Hardle, W., Scott, D., 1992. Smoothing in low and high dimensions by weighted averaging using rounded points. Computational Statistics 7, 97–128.

Loquin, K., Strauss, O., 2006. Fuzzy histograms and density estimation. SMPS 2006. Soft Methods in Probability and Statistics 45–52.

Runkler, T.A., 2004. Fuzzy histograms and fuzzy chi-squared tests for independence. In: IEEE International Conference on Fuzzy Systems, vol. 3. pp. 1361–1366.

Scott, D., Sheather, S., 1985. Kernel density estimation with binned data. Communications in Statistics. Theory and Methods 14, 1353–1359.

Scott, D.W., 1979. On optimal and data-based histograms. Biometrika 66 (3), 605–610.

Scott, D.W., 1992. Multivariate Density Estimation. Wiley Interscience.

Strauss, O., Comby, F., 2002. Estimation modale par histogramme quasicontinu. LFA'02 Rencontres Francophones sur la Logique Floue et ses Applications. Montpellier, pp. 35–42.

Strauss, O., Comby, F., Aldon, M., 2000. Rough histograms for robust statistics. In: ICPR'2000 15th International Conference on Pattern Recognition. Barcelona, Catalonia, Spain, pp. 3–8.

Van Den Berg, J., 2001. Probabilistic and statistical fuzzy set foundations of competitive exception learning. In: IEEE International Fuzzy Systems Conference. pp. 1035–1038.

Waltman, L., Kaymak, U., Van Den Berg, J., 2005. Fuzzy histograms: A statistical analysis. In: EUSFLAT-LFA 2005 Joint 4th EUSFLAT 11th LFA Conference. Barcelona, pp. 605–610.

Westergaard, H., 1968. Contributions to the History of Statistics. Agathon Press.