**Notes on Intuitionistic Fuzzy Sets**
**Print ISSN 1310–4926, Online ISSN 2367–8283**
**2025, Volume 31, Number 4, 511–521**
**DOI: 10.7546/nifs.2025.31.4.511-521**

# Modeling uncertainty from bidimensional histograms by intuitionistic fuzzy sets

**Boyan Kolev**

Institute of Mathematics and Informatics, Bulgarian Academy of Sciences
e-mail: bkolev@math.bas.bg

**Received:** 1 November 2025 **Revised:** 20 November 2025
**Accepted:** 24 November 2025 **Online First:** 27 November 2025

**Abstract:** Histograms summarize distributions of numerical data into densities per ranges of values and are typically used for fast approximation of selectivities of range predicates. This paper introduces an approach for intuitionistic fuzzy selectivity estimation from histograms when predicates are expressed in fuzzy terms with membership functions on the numeric values, where the degree of indefiniteness corresponds to the level of uncertainty resulting from the accuracy loss due to approximation. In the case of bidimensional histograms, a method for estimating the joint selectivity of conjunctive predicates over the two dimensions is proposed in a way that the joint selectivity can be considered as a measure of correlation. The approach is validated through the use of SQL queries against a synthetically generated dataset and its corresponding bidimensional histogram.

**Keywords:** Selectivity estimation, Intuitionistic fuzzy sets, Bidimensional histograms, Correlation.

**2020 Mathematics Subject Classification:** 03B52, 62H86, 68P20.

## 1 Introduction

Histogram analysis is widely used for statistical tasks in multiple areas, such as: SQL query optimization [6, 7, 11–13, 16, 17], image processing [14, 15], network traffic analysis [8], biology [5], and many others. Particularly in query optimization, histograms on relational attributes are used for fast selectivity estimation, i.e., approximating the size of the result set returned by a select-from-where query without actually executing the query. For example, the answer to an SQL query[1] like `SELECT count(*) FROM houses WHERE price<1350` can be approximately obtained by querying the histogram of the values in the `price` column. This results in orders of magnitude higher speed compared to querying the dataset instead, however such performance improvement comes at the price of accuracy loss. The returned approximate answer is then used to estimate what fraction of the entire dataset would be filtered by applying the predicate `price<1350`, which is considered as the selectivity of the predicate – a metric used by the optimizer to decide an optimal access method or join order. Now, let us consider that the user wants to apply a vague condition using a fuzzy predicate with membership function on the `price` attribute, e.g. "estimate the fraction of cheap houses", instead of using a strict threshold (this can be achieved with a fuzzy querying system like Intuitionistic Fuzzy PostgreSQL [9, 10]). A histogram analysis of the distribution of price values with respect to such a fuzzy predicate introduces uncertainty (due to the density aggregations per histogram bin) in addition to the membership degree, which we model in intuitionistic fuzzy terms [1, 2] within the scope of this work. Intuitionistic fuzzy histograms have been studied in the literature [3, 4], however the approach presented herewith is different, as it uses a histogram to represent the distribution of elements of two intuitionistic fuzzy sets and a method to quantify the interaction between them.

To illustrate with a more detailed example, let us consider a quite large dataset $D$ of about 1 million records of houses with addresses and prices (Figure 1) and take a look at the unit price (per $m^2$) attribute. The corresponding histogram (Figure 2) splits record counts into equally sized bins (price ranges), each mapped to the number of records with prices within the range. In this example, according to the histogram, if we take an entire bin, we can say there are exactly 7000 houses with prices between 1300 and 1400 EUR/$m^2$ (represented by the entire histogram bin) or approximately 20000 houses with price less than 1350 EUR/$m^2$, summing up the counts in all bins less than 1300 plus an approximation about the count in the bin 1300–1400 where the attribute value is less than 1350. Thus, the selectivity of the predicate "price<1350" would be estimated at about 2%, i.e., the estimated cardinality divided by the total number of rows in the table: $sel_{price<1350}(D) \approx 0.02$. A bidimensional histogram, respectively, shows the joint distribution of two attributes (columns of the relational dataset). Histogram bins then correspond to rectangular ranges of a bidimensional space. In the context of our example, let us consider the distances from the city center of each house (the second column, *km*, on Figure 1) together with prices. To visualize the corresponding bidimensional histogram, a common way is to use a heatmap (Figure 3), where the $x$ axis represents the distance, $y$ axis represents price, and brighter cells correspond to rectangular ranges with higher density. Another way to visualize is with a 3-dimensional graph, where the $z$ axis shows the counts per histogram bin, as in Figure 4. As it can be seen from the graphs, bidimensional histograms might reveal correlations – in this example we observe a negative correlation between price and distance: higher densities are commonly seen for high prices and low distances and vice versa, i.e., "downtown houses are likely expensive". As with unidimensional histograms, fast approximate queries are possible, e.g. for estimating the number of cheap houses close to the city center (or an SQL equivalent).

[1] In this example, we consider a dataset of houses with the unit price per $m^2$ in the `price` attribute.

This paper presents a methodology for intuitionistic fuzzy representation of the distribution of dataset entries with respect to vague terms that derives from a bidimensional histogram over attributes of the dataset. The methodology is described in detail in Section 2 and consists of the following steps: (a) defining membership functions to evaluate fuzzy predicates corresponding to the vague terms; (b) evaluating conjunctions on two predicates and estimating the joint selectivity; (c) intuitionistic fuzzy estimation of the joint selectivity when derived from a bidimensional histogram; (d) discussion on how the resulting joint selectivity is related to correlation.

| address | km | EUR | m² | price |
|---|---|---|---|---|
| 13 Apple Str. | 11.8 | 210000 | 100 | 2100 |
| 92 Green Str. | 2.3 | 168000 | 48 | 3500 |
| 2 Elephant Str. | 16.1 | 78000 | 65 | 1200 |
| 57 Wisdom Str. | 9.3 | 378000 | 135 | 2800 |
| 18 Central Str. | 1.1 | 347100 | 89 | 3900 |
| ... | ... | ... | ... | ... |

Figure 1. Sample of a large dataset of house records with address, distance from city center, and price per $m^2$

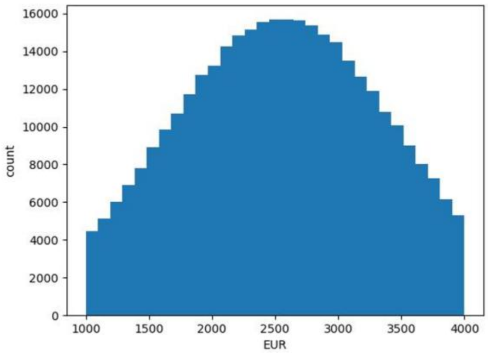


Figure 2. Histogram of the distribution of unit prices – densities are assigned to equally sized histogram bins (price ranges)

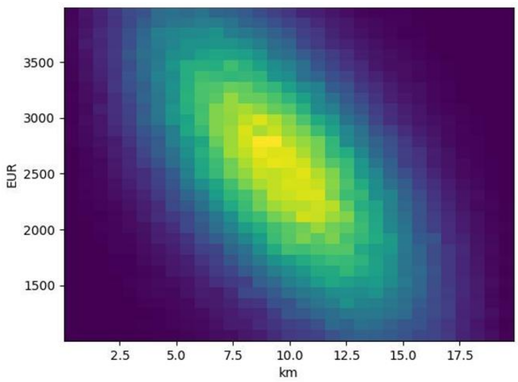


Figure 3. Bidimensional histogram as a heatmap – brighter cells correspond to rectangular ranges with higher densities

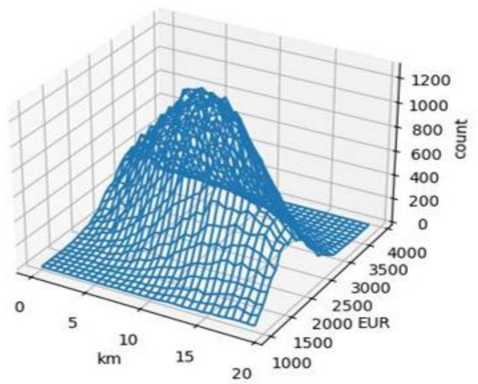


Figure 4. Bidimensional histogram as a 3D graph – densities are shown in the *z*-axis

## 2 Intuitionistic fuzzy selectivity estimations

This section describes the approach in detail with examples in the context of the house records dataset, for more comprehensive illustration. Let us consider the houses dataset along with a bidimensional histogram on the distance and price attributes and some vague terms to express fuzzy predicates, in particular: (a) "house is cheap" or "house is expensive" that can be evaluated through a membership function that transforms the house price to a degree of truth and (b) "house is near the city center" or "house is far from the center" that can be evaluated through a membership function that takes into account the distance attribute. We define membership functions, which are further used for computing the selectivity.

## 2.1 Membership functions

For computing the degree of truth of a fuzzy predicate (the corresponding membership function) that is to be evaluated on a particular attribute *x* of the dataset (in our example, *x* can refer to *km* or *price*), let us first use the following substitutions:

- *N* is the size of the dataset;
- $M_x$ is the mean of the values in *x*, i.e.:

$$M_x = \frac{1}{N}\sum_{i=1}^{N} x_i ; \tag{1}$$

- $S_x$ is the standard deviation of the values in *x*, i.e.:

$$S_x = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(x_i - M_x\right)^2} ; \tag{2}$$

- $Z_x$ is a normalizer function that computes the z-score of a value of *x* (all z-scores have a mean of 0 and standard deviation of 1), i.e.:

$$Z_x\left(x_i\right) = \frac{x_i - M_x}{S_x} ; \tag{3}$$

- $F_x$ is a scaling factor, which is a positive number such that it is guaranteed that, for each value of *x*, $\frac{Z_x(x_i)}{F_x} \in [-1, 1]$. An appropriate choice for $F_x$ would be the maximal absolute *z*-score:

$$F_x = \max_i \left|Z_x\left(x_i\right)\right| . \tag{4}$$

Then, the membership function $\mu_x$ is a simple linear transformation of *x* through its *z*-score scaled to the interval [0, 1] by the following formula:

$$\mu_x\left(x_i\right) = \frac{1}{2}\left(1 \pm \frac{Z_x\left(x_i\right)}{F_x}\right) . \tag{5}$$

Here, the *z*-score is negated in case of fuzzy predicates that require that the membership degree decreases monotonically when the value of *x* increases. For example, if *x* is distance, then positive sign is taken when evaluating the predicate "far" and negative for the predicate "near". Similarly, for evaluating "expensive" and "cheap" on the price attribute.

The computation of the summary metrics *N*, $M_x$, $S_x$, and $F_x$ might be an expensive operation if done on the entire dataset, but once they are computed initially, they can be incrementally updated as data gets continuously appended.

## 2.2 Joint selectivity and correlation

For the discussion that follows in this subsection, let us substitute with $x$ and $y$ the vectors of values of the *km* and *price* attributes respectively; thus, $x_i$ and $y_i$ are the values of *km* and *price* of the *i*-th dataset entry. Let $\mu_x$ and $\mu_y$ be the functions over elements of $x$ and $y$ that assign memberships of a house entry to the fuzzy sets that represent "houses near city center" and "cheap houses", respectively, i.e.:

$$\mu_x(x_i) = \frac{1}{2}\left(1 - \frac{Z_x(x_i)}{F_x}\right); \quad \mu_y(y_i) = \frac{1}{2}\left(1 - \frac{Z_y(y_i)}{F_y}\right) \tag{6}$$

Summing up the values of $\mu_x$ and $\mu_y$ over all records in the dataset $D$ gives the cardinalities of the corresponding finite fuzzy sets, which, divided by the cardinality $N$ of the universe $D$, results in the selectivities $sel_{near}$ and $sel_{cheap}$ of the fuzzy predicates "near" and "cheap", as defined in (7) below. Note that at this point, the discussion still concerns the entire dataset $D$ and not yet the histogram, i.e., we do not yet introduce intuitionistic fuzziness as there is no uncertainty from accuracy loss.

$$sel_{near}(D) = \frac{1}{N}\sum_{i=1}^{N}\mu_x(x_i); \quad sel_{cheap}(D) = \frac{1}{N}\sum_{i=1}^{N}\mu_y(y_i) \tag{7}$$

Now, let us consider the composite fuzzy predicate "cheap houses near the city center", defined by conjunction of the "near" and "cheap" predicates and evaluated by applying the "product" T-norm to compute degrees of membership. The selectivity of this predicate on the dataset $D$ we refer to as "joint selectivity", as it takes into account the interaction between the estimations of the terms "near" and "cheap":

$$sel_{near\&cheap}(D) = \frac{1}{N}\sum_{i=1}^{N}\mu_x(x_i)\,\mu_y(y_i). \tag{8}$$

Expanding the expression (8) using the formulas (6) results in:

$$\begin{aligned} sel_{near\&cheap}(D) &= \frac{1}{N}\sum_{i=1}^{N}\frac{1}{4}\left(1 - \frac{Z_x(x_i)}{F_x}\right)\left(1 - \frac{Z_y(y_i)}{F_y}\right) \\ &= \frac{1}{4} + \frac{1}{4F_xF_y}\frac{1}{N}\sum_{i=1}^{N}Z_x(x_i)Z_y(y_i) - \frac{1}{4F_x}\frac{1}{N}\sum_{i=1}^{N}Z_x(x_i) - \frac{1}{4F_y}\frac{1}{N}\sum_{i=1}^{N}Z_y(y_i) \end{aligned}. \tag{9}$$

The last two terms of (9) are equal to zero, since the sum of all z-scores equals to zero, as per (3). The other sum in the expansion (9) contains the scalar product of the z-scores of $x$ and $y$, which, divided by $N$, is equal to the Pearson correlation coefficient of $x$ and $y$. This results in the following linear dependency between the correlation corr($x, y$) and the fuzzy predicate selectivity:

$$\mathrm{corr}(x,y) = F_xF_y\left(4sel_{near\&cheap}(D) - 1\right). \tag{10}$$

In fact, *corr(x,y)*, as defined in (10), is equal either to the Pearson correlation, which is the case in our example (where $\mu_x$ and $\mu_y$ are both monotonically decreasing), or to its negation, as it would be if the two participating fuzzy terms required membership functions of different monotonicity as per (5). Either way, it is a measure of how correlated the two terms are, in particular, how likely it is to select houses that are both cheap and near downtown. As we will see in the validation section, for the dataset we use, corr($x$, $y$) ≈ –0.5, i.e., negative, meaning that it is not very likely to find cheap downtown houses. However, if selectivity were evaluated for the condition "near and expensive", it would result in a positive corr($x$, $y$) ≈ 0.5.

## 2.3 Intuitionistic fuzzy selectivity from histogram

The relational view of a bidimensional histogram $H$ built on the dataset $D$ (e.g., as in the first three columns of Figure 5) represents non-overlapping rectangular ranges of the distance-price space, each mapped to the number of house entries within the range. A histogram size is typically orders of magnitude less than the size of the entire dataset. For example, if the histogram splits data in 30 bins per dimension, it would contain only 900 records. This compact representation allows for fast selectivity estimation, but it comes at the price of minor loss of information. For this reason, we use intuitionistic fuzzy estimations (IFE) that assign degrees of indefiniteness corresponding to the accuracy loss due to the range widths at each dimension. For the definitions that follow, let us recall that $x$ and $y$ correspond to the attributes *km* and *price* respectively and assume that [ $h_x^{min}$ , $h_x^{max}$ ] and [ $h_y^{min}$ , $h_y^{max}$ ] are the ranges on $x$ and $y$ of the histogram entry $h$, $h_n$ is the number of records (density) within the rectangular range represented by $h$, and $\langle\mu_x, \nu_x\rangle$ and $\langle\mu_y, \nu_y\rangle$ are the pairs of membership and non-membership functions of the predicates "near" and "cheap" over a histogram record. In fact, when $\mu$ is decreasing (which is the case for both $\mu_x$ and $\mu_y$ in our example), then $\nu$ is increasing and corresponds to the opposite term, in particular, $\nu_x$ and $\nu_y$ are equivalent to the membership functions of the predicates "far" and "expensive", respectively. Therefore, the definitions of $\mu_x$ and $\mu_y$ both use the upper bounds of the ranges, since they are both decreasing, while $\nu_x$ and $\nu_y$ apply to the lower bounds, as follows:

$$\mu_x(h) = \frac{1}{2}\left(1 - \frac{Z_x\left(h_x^{max}\right)}{F_x}\right); \qquad \nu_x(h) = \frac{1}{2}\left(1 + \frac{Z_x\left(h_x^{min}\right)}{F_x}\right); \tag{11.1}$$

$$\mu_y(h) = \frac{1}{2}\left(1 - \frac{Z_y\left(h_y^{max}\right)}{F_y}\right); \qquad \nu_y(h) = \frac{1}{2}\left(1 + \frac{Z_y\left(h_y^{min}\right)}{F_y}\right); \tag{11.2}$$

Thus, the resulting columns "near" and "cheap" on Figure 5 represent the distribution of house entries with respect to the vague terms "near" and "cheap". For example, according to the first row of the table on Figure 5, there are 536 houses that belong to an intuitionistic fuzzy set that represents "cheap houses" with a degree of membership 0.16 and degree of non-membership 0.81. The estimated selectivities of the predicates "house is near downtown" and "house is cheap" would then be valued with the intuitionistic fuzzy pairs below, which in fact correspond to the operator $W$ (see [1, 2]) applied over the intuitionistic fuzzy sets "downtown houses" and "cheap houses". Now each selectivity is applied to the histogram $H$, therefore it takes into account the implied uncertainty from accuracy loss, hence it is represented as an intuitionistic fuzzy pair, where the degree of indefiniteness corresponds to uncertainty:

$$sel_{near}\left(H\right)=\left\langle \frac{1}{N}\sum_{h\in H}h_n*\mu_x\left(h\right),\frac{1}{N}\sum_{h\in H}h_n*\nu_x\left(h\right)\right\rangle; \tag{12.1}$$

$$sel_{cheap}\left(H\right)=\left\langle \frac{1}{N}\sum_{h\in H}h_n*\mu_y\left(h\right),\frac{1}{N}\sum_{h\in H}h_n*\nu_y\left(h\right)\right\rangle. \tag{12.2}$$

| count | km [min, max] | price [min, max] | | "near" <μ, ν> | "cheap" <μ, ν> | "near&cheap" <μ, ν> |
|---|---|---|---|---|---|---|
| 536 | [3.9, 4.4] | [3419, 3516] | | <0.78, 0.19> | <0.16, 0.81> | <0.12, 0.85> |
| 405 | [14.8, 15.3] | [1194, 1290] | IFE | <0.23, 0.74> | <0.90, 0.06> | <0.21, 0.76> |
| 110 | [4.5, 5.1] | [1774, 1871] | → | <0.74, 0.22> | <0.71, 0.26> | <0.53, 0.42> |
| 4 | [16.1, 16.7] | [3806, 3903] | | <0.17, 0.81> | <0.03, 0.94> | <0.01, 0.98> |
| ... | ... | ... | | ... | ... | ... |

Figure 5. Relational view of a sample of a bidimensional histogram. Additional columns represent the intuitionistic fuzzy estimations (IFE) of predicates "near" and "cheap" computed on the range bounds of values of the attributes *km* and *price* respectively. The estimation of the conjunctive predicate "near & cheap" is obtained by applying the "product" T-norm.

Let us consider an intuitionistic fuzzy set defined by a composite predicate like "cheap houses near the city center". The membership of histogram entries is then defined by a conjunction between the "near" and "cheap" memberships. Applying the "product" T-norm to compute the degrees of membership (with the corresponding co-norm for the degrees of non-membership) results in the distribution of house entries with respect to the condition "near and cheap" (the last column on Figure 5). The joint selectivity is then evaluated with an intuitionistic fuzzy pair as follows:

$$sel_{near\&cheap}\left(H\right)=\langle\frac{1}{N}\sum_{h\in H}h_n*\mu_x\left(h\right)\mu_y\left(h\right),\frac{1}{N}\sum_{h\in H}h_n*\left(\nu_x\left(h\right)+\nu_y\left(h\right)-\nu_x\left(h\right)\nu_y\left(h\right)\right)\rangle. \tag{13}$$

Thus, the correlation now is not an exact value, but evaluated within an interval [$\mathrm{corr}_{\min}$, $\mathrm{corr}_{\max}$] with bounds computed linearly from the degrees of truth and falsity of (13), as follows:

$$\mathrm{corr}_{\min}\left(x,y\right)=F_xF_y\left(4\mu_{sel_{near\&cheap}(H)}-1\right); \tag{14.1}$$

$$\mathrm{corr}_{\max}\left(x,y\right)=F_xF_y\left(4\left(1-\nu_{sel_{near\&cheap}(H)}\right)-1\right). \tag{14.2}$$

The width of this interval corresponds to the precision of the selectivity estimation. When lower number of bins are used to build the histogram, the ranges represented by histogram entries are wider, which introduces more uncertainty, hence the correlation interval is also wider, i.e., the estimation is less precise. Respectively, higher number of bins results in more precise estimation, which would take however more time to be computed.

# 3 Validation

To validate the concept, we use Python to generate a synthetic dataset of about 1 million records of house data, with prices per $m^2$ in the *pr* column between 1000 and 4000 EUR and distances from city center in the *km* column between 0 and 20. The generator introduces dependencies between the two attributes, so that a negative correlation be observed; in fact, the graphs on Figure 3 and Figure 4 are drawn from a histogram built on the generated dataset. The dataset is stored in a PostgreSQL table with the following structure:

```
create table data (
      id integer primary key,
      km double precision,
      pr double precision
);
```

Then, the generator builds the bidimensional histogram of the dataset. To study the performance-accuracy trade-off, we use two different histogram sizes:

(a) $H_1$ splits data into 53 bins per dimension, resulting in histogram size of 2809 bidimensional ranges;

(b) $H_2$ splits data into 101 bins per dimension, i.e., a total of 10201 ranges.

In both cases, the histogram is stored into another PostgreSQL table with the following structure:

```
create table hist2d (
      kmmin double precision, kmmax double precision,
      prmin double precision, prmax double precision,
      cnt bigint
);
```

The summary metrics $N$, $M_x$, $S_x$, and $F_x$, as per (1), (2), (3), and (4), are collected from the `data` table and stored in a table named `stats` with the following SQL statements:

```
create table stats (n bigint,
      km_avg double precision, km_std double precision,
      pr_avg double precision, pr_std double precision,
      km_coef double precision, pr_coef double precision
);

insert into stats (n, km_avg, km_std, pr_avg, pr_std)
select count(*), avg(km), stddev(km), avg(pr), stddev(pr)
from data;

update stats set
km_coef = (select max(abs((km-km_avg)/km_std))
           from data, stats),
pr_coef = (select max(abs((pr-pr_avg)/pr_std))
           from data, stats);
```

The summary metrics are then used for defining a view on the histogram that returns the z-scores of range bounds, adjusted to the interval [–1, 1]:

```
create view zhist2d as
select cnt,
    (prmin-pr_avg)/pr_std/pr_coef as z_prmin,
    (prmax-pr_avg)/pr_std/pr_coef as z_prmax,
    (kmmin-km_avg)/km_std/km_coef as z_kmmin,
    (kmmax-km_avg)/km_std/km_coef as z_kmmax
from hist2d, stats;
```

The view near_and_cheap below applies the intuitionistic fuzzy estimations, as proposed in Section 2.3, to obtain from the histogram the distribution of houses with respect to the condition "near and cheap".

```
create view near_and_cheap as
select cnt,
      (1-z_prmax)*(1-z_kmmax)/4 as mu,
    1-(1-z_prmin)*(1-z_kmmin)/4 as nu
from zhist2d;
```

To use the histogram to compute the intuitionistic fuzzy joint selectivity of the condition "near and cheap" as per (13), together with the corresponding correlation interval as per (14), we use the following query $Q_H$:

```
select sel_mu, sel_nu,
    (4*sel_mu-1)*km_coef*pr_coef as corr_min,
    (4*(1-sel_nu)-1)*km_coef*pr_coef as corr_max
from stats, (
    select sum(cnt*mu/n) as sel_mu,
           sum(cnt*nu/n) as sel_nu
    from near_and_cheap, stats) sel;
```

To use the entire dataset to compute the fuzzy joint selectivity and exact correlation between the attributes *km* and *pr*, we use the query $Q_D$ below:

```
select
    sum((1+z_pr/pr_coef)*(1+z_km/km_coef)/4/n) as sel,
    sum(z_pr*z_km/n) as corr
from stats, (
    select (pr-pr_avg)/pr_std as z_pr,
           (km-km_avg)/km_std as z_km
    from data, stats) zdata;
```

Table 1 shows the result of executing query $Q_H$ against the histogram, in the two cases for the histogram size, and a comparison with executing query $Q_D$ against the dataset regarding response time. Computing joint selectivity and correlation from data is accurate, but expensive. Using histograms reduces significantly the response time, but introduces uncertainty that corresponds to the degree of indefiniteness in the intuitionistic fuzzy estimation of the joint selectivity, which also affects the width of the correlation interval. The performance–accuracy trade-off can be fine-tuned through the histogram size, i.e., the number of bins per dimension: a higher number of bins leads to narrower correlation interval (i.e., more precise estimation) but slower computing, while a lower number of bins results in faster computing but less precise estimation.

Table 1. Comparison of joint selectivity and correlation compute –
from data and histograms

| | **$H_1$ = 53 bins** | **$H_1$ = 101 bins** |
|---|---|---|
| $Q_H$: compute time | 11 ms | 58 ms |
| • Selectivity ⟨$\mu$, $\nu$⟩ from histogram | ⟨0.2203, 0.7613⟩ | ⟨0.2246, 0.7658⟩ |
| • Correlation interval | [–0.735, –0.280] | [–0.628, –0.389] |
| $Q_D$: compute time | 367 ms | |
| • Selectivity from data | 0.2294 | |
| • Correlation from data | –0.509 | |

# 4 Conclusion

We presented a methodology for modelling uncertainty from histograms for fast selectivity estimations with intuitionistic fuzzy evaluations of vague predicates, where accuracy loss from per-range summarization is mapped to degree of indefiniteness. In the case of bidimensional histograms, we discuss intuitionistic fuzzy evaluation of conjunctive predicates on single attributes in a way that joint selectivity can be considered as a measure of correlation. We did experimental validation with a synthetic PostgreSQL dataset on house prices use case.

As subject to further work we will: (a) adapt the method to work on sliding windows of data streams making intuitionistic fuzzy estimators be updated incrementally; (b) extend the method to capture uncertainty from data, e.g. if an imprecise address cannot determine the exact distance from city center; (c) extend the method for multidimensional histograms, e.g. for intuitionistic fuzzy estimations about 3-way attribute interactions; (d) apply the method on real-world datasets.

# References


[1] Atanassov, K. (1999). *Intuitionistic Fuzzy Sets*. Springer-Verlag, Heidelberg.

[2] Atanassov, K. (2012). *On Intuitionistic Fuzzy Sets Theory*. Springer, Berlin.

[3] Atanassova, L., Gluhchev, G., & Atanassov, K. (2010). On intuitionistic fuzzy histograms. *Notes on Intuitionistic Fuzzy Sets*, 16(4), 31–36.

[4] Bureva, V. (2014). Intuitionistic fuzzy histograms in grid-based clustering. *Notes on Intuitionistic Fuzzy Sets*, 20(1), 55–62.

[5] Chor, B., Horn, D., Goldman, N., Levy, Y., & Massingham, T. (2010). Genomic DNA *k*-mer spectra: Models and modalities. *Proceedings of the 14th International Conference on Research in Computational Molecular Biology*, 24–27 April 2010, Lisbon, Portugal, 1, pp. 571–580.

[6] Ioannidis, Y. E. (1996). Query optimization. *ACM Computing Surveys*, 28(1), 121–123.

[7] Ioannidis, Y. (2003). The history of histograms (abridged). *Proceedings of the 29th International Conference on Very Large Data Bases*, 19–30 August 2003, Berlin, Germany, 29, pp. 19–30.

[8] Kind, A., Stoecklin, M. P., & Dimitropoulos, X. (2009). Histogram-based traffic anomaly detection. *IEEE Transactions on Network and Service Management*, 6(2), 110–121.

[9] Kolev, B. (2004). Intuitionistic fuzzy relational databases and translation of the intuitionistic fuzzy SQL. *Proceedings of the 6th International FLINS Conference on Applied Computational Intelligence*, 1–3 September 2004, Blankenberge, Belgium, pp. 189–194.

[10] Kolev, B. (2005). Intuitionistic fuzzy PostgreSQL. *Advanced Studies in Contemporary Mathematics*, 11(2), 163–177.

[11] Kooi, R. P. (1980). *The Optimization of Queries in Relational Databases*. Case Western Reserve University, Cleveland.

[12] Lee, J.-H., Kim, D.-H., & Chung, C.-W. (1999). Multidimensional selectivity estimation using compressed histogram information. *Proceedings of the ACM SIGMOD International Conference on Management of Data*, 1–3 June 1999, Philadelphia, USA, pp. 205–214.

[13] Muralikrishna, M. & DeWitt, D. (1988). Equi-depth histograms for estimating selectivity factors for multidimensional queries. *Proceedings of the ACM SIGMOD International Conference on Management of Data*, 1–3 June 1988, Chicago, USA, pp. 28–36.

[14] Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1), 62–66.

[15] Pass, G. & Zabih, R. (1999). Comparing images using joint histograms. *Multimedia Systems*, 7, 234–240.

[16] Poosala, V., Ioannidis, Y., Haas, P., & Shekita, E. (1996). Improved histograms for selectivity estimation of range predicates. *Proceedings of the ACM SIGMOD International Conference on Management of Data*, 4–6 June 1996, Montreal, Canada, pp. 294–305.

[17] Selinger, P. G., Astrahan, M. M., Chamberlin, D. D., Lorie, R. A., & Price, T. G. (1979). Access path selection in a relational database management system. *Proceedings of the ACM SIGMOD International Conference on Management of Data*, 30 May–1 June 1979, New York, USA, pp. 23–34.