Atlantis Studies in Uncertainty Modelling, volume 3
Joint Proceedings of the 19th World Congress of the International Fuzzy Systems Association (IFSA), the 12th Conference of the European Society for Fuzzy Logic and Technology (EUSFLAT), and the 11th International Summer School on Aggregation Operators (AGOP)

Possibilistic Granular Count: Derivation and Extension to Granular Sum

Corrado Mencar
Department of Computer Science
University of Bari Aldo Moro, Bari, Italy
corrado.mencar@uniba.it

Abstract

Counting data in presence of uncertainty leads to granular counts that can be represented in terms of possibility distributions. The formula of granular count is derived on the basis of two weak assumptions that can be applied in a wide variety of problems involving uncertain data. The formulation is further extended to introduce the granular sum of counts, by taking into account the interactivity of granular counts. Numerical results show the differences in terms of specificity between granular sum and a direct application of the extension principle to sum granular counts.

Keywords: Granular count, Granular sum, Possibility theory, Uncertain data.

1 Introduction

Counting is a fundamental operation that is preliminary to many subsequent statistical operations. Through counting, absolute and relative frequencies of discrete events are computed, which are further used in statistical inferences to provide insights on the phenomena under consideration. Thus, any system of frequency computation requires input from a counting system, but the latter requires some mechanisms that "parse the world into countable entities" [2].

In complex environments, counting is not trivial. As a concrete example, in the domain of Bioinformatics the evaluation of gene expression[1] in cells requires to assess the presence and quantity of mRNA in a biological sample at a given moment [11]. Such an assessment can be performed by technologies like Next Generation Sequencing[2], which enable to identify mRNA fragments called "reads", which are then matched to known genes sono as to count the number or reads associated to each gene. However, this process must face technological and biological challenges, because reads are just small fragments of mRNA that may match different genes; also sequencing may be contaminated by errors that lower the quality of associations [3]. The result is uncertain data on gene expression.

Data uncertainty requires proper management [1]. The simplest approach is to ignore uncertainty by estimating a precise value for each observation (e.g., arbitrarily selecting one gene for each read), but this simplistic approach, though of immediate application, can lead to unwanted bias in the subsequent processing stages, which is difficult to detect. A more comprehensive approach should propagate uncertainty throughout the entire data processing flow. In this way, the results of data processing reveal their uncertainty, which can be evaluated to assess their ultimate usefulness. The need of processing uncertain, imprecise and partially true data, gave rise to Granular Computing as a paradigm of information processing [10].

Several theories can be applied to represent and process uncertain data, which are more or less appropriate on the basis of the nature of uncertainty. In particular, Possibility Theory [5] deals with uncertainty due to incomplete information, e.g. when the value of an observation cannot be precisely determined: we will use the term *uncertain data* to denote data characterized by this specific type of uncertainty, therefore we adopt the possibilistic framework in this paper. In the previous example, the uncertainty in associating reads and genes can be properly modeled throgh Possibility

[1]Gene expression is the process by which the information encoded in a gene is used to direct the assembly of a protein molecule. (source: https://www.genome.gov/genetics-glossary/Gene-Expression)

[2]https://www.sciencedirect.com/topics/medicine-and-dentistry/next-generation-sequencing

Theory because reads are incomplete representations of the expressed genes.

Recently, a definition of granular count through Possibility Theory was proposed [9]. Based on such definitions, algorithms for exact and approximate granular counting were devised [8], as well methods for incremental counting [6] and bounded approximate counting [7]. Granular count was defined on the basis of arguments that are consistent with intuition. In this paper it is shown that granular count can be actually derived on the basis of two weak assumptions, thus confirming the correctness of the definition and the consequent properties.

The result of granular counting is a collection of fuzzy sets (in particular, fuzzy intervals in the domain of natural numbers) representing granular counts, which can be further used in arithmetic operations like addition. However, each fuzzy set should be interpreted as the possibility distribution of a variable representing the uncertain count of an object or referent. In this paper it is proved that such variables are actually interactive (i.e. knowing the value of a variable affects the possibility distributions of the others), therefore a direct application of the extension principle to arithmetic operations results in an overestimation of the results. In this paper, the most specific solution of the sum of a set of granular counts is derived by taking into account the interactivity among variables.

In Sec. 2, some preliminary definitions, notation and terminology are settled. Sec. 3 reports the derivation of granular count on the basis of two weak assumptions. Sec. 4 extends the proof by deriving the possibility distribution of any sum of granular counts. Sec. 5 reports the numerical results on a couple of synthetic datasets. Finally, Sec. 6 draws some concluding remarks with notes on future research.

2 Preliminaries

We assume that data are manifested through *observations*, which refer to some objects or *referents*. The relation between observations and referents—which is called *reference*—may be uncertain in the sense that an unequivocal reference of the observation to one of the referents is not known. We model such uncertainty with Possibility Theory [5] as we assume that uncertainty is due to lack of information in the observation, i.e. the observation is not complete enough to make reference unequivocal.

Given a set O of m observations, a set R of n referents, the reference relation is defined by a possibility distribution for each $o \in O$, i.e.

$$\pi_o : R \mapsto [0,1]$$

such that $\exists r \in R : \pi_o(r) = 1$. The value $\pi_o(r) = 0$ means that it is impossible that the referent r is referred by the observation o, while $\pi_o(r) = 1$ means that the referent r is completely possible (though not certain). Intermediate values of $\pi_o(r)$ stand for gradual values of possibility, which quantify the completeness of information resulting from an observation. (More specifically, the lower the possibility degree, the more information we have to *exclude* a referent.) The possibility distributions of all observations can be arranged in a *possibilistic assignment table*, as exemplified in Table 1. For example, each observation may correspond to a read, while each referent to a gene: the higher possibility the better the matching, however a matching of high quality does not involve certainty of the association.

	r_1	r_2	r_3
o_1	1	0.3	0.54
o_2	0.8	1	0.6
o_3	1	0	0
o_4	0.86	0.91	1
o_5	1	0	0
o_6	0.5	1	0.64
o_7	1	0.8	1
o_8	0.2	0.5	1
o_9	1	0	0
o_{10}	0.6	1	0.78

Table 1: Example of possibilistic assignment table. The j-th row represents a possibility distribution π_{o_j}.

Given a possibilistic assignment table, a possible question is to count the number of observations referring to a given referent. It is worth noting that this count does not coincide with the cardinality of a fuzzy set resulting from each column of the possibilistic assignment table, because each column represents a possibility distribution, hence an epistemic fuzzy set representing incomplete information about the reference relation [4]. As it is partially unknown which obervations refer to each referent, couting the number of observations must be granular to reflect the uncertainty of the reference.

In [9], the granular count was defined through a possibility distribution for each referent $r \in R$ as:[3]

$$C_r(x) = \max_{O_x \subseteq O} \pi_{O_x}(r) \tag{1}$$

for $x \leq m$ and $C_r(x) = 0$ for $x > m$, where $O_x \subseteq O$ is

[3] A slightly different notation w.r.t. [9] is used here to denote the granular count.

any subset of $x \in \mathbb{N}$ observations and

$$\pi_{O_x}(r) = \min\left\{\min_{o\in O_x}\pi_o(r), \min_{o\notin O_x}\max_{r'\neq r}\pi_o(r')\right\} \quad (2)$$

is the possibility that O_x is the subset of *all and only* the observations of the referent r.[4] Intuitively, x is a possible count value for r if there exists at least one set of x observations that refer to r while all the other observations refer to other referents.

3 Derivation of granular count

In this section, it is proved that granular count as defined in (1), can be actually derived starting from weak assumptions on a possibilistic assignment table. The assumptions are:

A1 (functionality) Each observation refers to one referent only. This is a basic assumption of possibilistic models;

A2 (non-interactivity) Knowing the referent of an observation does not affect the uncertainty on the other observations.

The functionality assumption can be formalized by restricting the reference relation to a function, as follows.

Definition 1. Given a set O of observations and a set R of referents, a *reference function* is defined as a mapping

$$\rho : O \mapsto R$$

The reference function formalizes the assumption that an observation cannot refer to more than one referent, and each observation refers to a referent. The uncertainty represented by the possibility distributions π_o propagates to the uncertainty on the reference function, which is defined by a possibility degree $\pi(\rho) \in [0,1]$.

The possibility degree $\pi(\rho)$ is determined by the joint possibility of each observation o referring to reference $\rho(o) \in R$. Based on the non-interactivity assumption, the following definition can be stated.

Definition 2. Given the space of reference functions R^O the possibility measure Π is defined as:

$$\Pi(P) = \max_{\rho\in P}\pi(\rho) \quad (3)$$

for each $P \subseteq R^O$, where

$$\pi(\rho) = \min_{o\in O}\pi_o(\rho(o)) \quad (4)$$

[4]The conventions $\min\emptyset = 1$ and $\max\emptyset = 0$ are followed throughout the paper.

Therefore, ρ is maximally possible if all observations refer to the maximally possible referent; on the other hand, ρ is impossible if at least one referent is assigned to an impossible referent. Eq. (4) defines a possibility measure Π on the space of reference functions R^O.

Let P_r^x the set of all reference functions where referent r is referred by x observations, $x \leq m$. Formally:

$$P_r^x = \left\{\rho \in R^O : \left|\rho^{-1}(r)\right| = x\right\} \quad (5)$$

Then, the following theorem holds.

Theorem 3. *The possibility measure of P_r^x coincides with the degree of the granular count of r, i.e.:*

$$\Pi(P_r^x) = C_r(x)$$

Proof. By definition (2) and because of associativity of "min":

$$\begin{aligned}\Pi(P_r^x) &= \max_{\rho\in P_r^x}\min_{o\in O}\pi_o(\rho(o)) \quad (6)\\ &= \max_{\rho\in P_r^x}\min\left\{\min_{o\in\rho^{-1}(r)}\pi_o(r), \min_{o\notin\rho^{-1}(r)}\pi_o(\rho(o))\right\}\end{aligned}$$

We define the following relation between reference functions that share the same preimage:

$$\rho\mathscr{R}\acute{\rho} \iff \rho^{-1}(r) = \acute{\rho}^{-1}(r) \quad (7)$$

It is easy to observe that $\mathscr{R}$ is an equivalence relation on P_r^x. Therefore, P_r^x can be partitioned in its equivalence classes. Given an equivalence class K, there exists a set O_x of x observations such that, for each $\rho \in K$, $\rho^{-1}(r) = O_x$, i.e. O_x is shared among all reference functions in the equivalence class K. On the other hand, two different equivalence classes K, K' have different sets of observations O_x, O'_x though with the same cardinality x. Moreover, given a set $O_x \subseteq O$ of x observations there exists a set of reference functions that have O_x as their preimage of reference r, thus forming an equivalence class. For such reasons, an equivalence class will be denoted as $[O_x]$ so as to highlight the common preimage of all the functions in the class.

Based on the equivalence relation $\mathscr{R}$ and the associative property of "max", the possibility measure Π can be rewritten as:

$$\Pi(P_r^x) = \max_{[O_x]\in P_r^x/\mathscr{R}}\ \max_{\rho\in[O_x]}\min\{A, B\} \quad (8)$$

where

$$A \equiv \min_{o\in O_x}\pi_o(r)$$

and

$$B \equiv \min_{o\notin O_x}\pi_o(\rho(o))$$

Thanks to the distributive property of min and max functions, eq. (8) can be rewitten as:

$$\Pi\left(P_r^x\right) = \max_{[O_x]\in P_r^x/\mathscr{R}} \min\left\{\max_{\rho\in[O_x]} A,\ \max_{\rho\in[O_x]} B\right\}$$

It is worth noting that, for each $\rho \in [O_x]$, the quantity $\min_{o\in O_x}\pi_o(r)$ does not change, therefore:

$$\max_{\rho\in[O_x]} A = \max_{\rho\in[O_x]}\min_{o\in O_x}\pi_o(r) = \min_{o\in O_x}\pi_o(r)$$

On the other hand, given $\rho \in [O_x]$ and $o \notin O_x$, it is necessary that $\rho(o) \neq r$ (otherwise $o \in O_x$) and, for each $r' \neq r$ there exists $\rho' \in [O_x]$ such that $\rho'(o) = r'$. Therefore:

$$\max_{\rho\in[O_x]} B = \max_{\rho\in[O_x]}\min_{o\notin O_x}\pi_o(\rho(o)) = \min_{o\notin O_x}\max_{r'\neq r}\pi_o\left(r'\right)$$

Putting the results together, eq. (8) can be rewritten as:

$$\Pi\left(P_r^x\right) = \max_{[O_x]\in P_r^x/\mathscr{R}} \min\left\{\min_{o\in O_x}\pi_o(r),\ \min_{o\notin O_x}\max_{r'\neq r}\pi_o\left(r'\right)\right\} \tag{9}$$

With the further notice that eq. (9) does not depend on assignment functions but only on sets $O_x \subseteq O$ of x observations, we get

$$\Pi\left(P_r^x\right) = \max_{O_x\subseteq O}\min\left\{\min_{o\in O_x}\pi_o(r),\ \min_{o\notin O_x}\max_{r'\neq r}\pi_o\left(r'\right)\right\} \tag{10}$$

which is exactly the value of $C_r(x)$. □

In consequence of the theorem, the granular count as defined in (1) is a direct consequence of the assumptions of functionality and non-interactivity, which are formalized by definitions 1 and 2.

4 Sum of granular counts

Given a referent r, the corresponding granular count C_r is a fuzzy interval over the set of natural numbers $\mathbb{N}$ [9]. As a consequence, the granular counts of two referents r', r'' can be added by using standard fuzzy arithmetic: it is easy to show that the sum of two or more granular counts is always a fuzzy interval. However, the functionality assumption makes the granular counts *interactive*, i.e. knowing the exact count of a referent impacts on the uncertainty of the counts of the remaining referents. In the extreme case, summing all counts of all the referents in R should result in a single integer value m, corresponding to the cardinality of O, without uncertainty. This results cannot be achieved through standard fuzzy arithmetic, because the latter assumes non-interactivity of the involved variables. As a consequence, applying fuzzy arithmetic to sum granular counts results in a fuzzy set that is less specific than the sum that can be inferred if interactivity is taken into account.

In order to take into account interactivity, a summation operator $\oplus$ must be defined accordingly. It is possible to take profit of the proof schema of Theorem 3 to define such a function. Let $P_{r',r''}^{x',x''}$ the set of all reference functions where referent r' is referred by x' observations and r'' is referred by x'' observations, with $r' \neq r''$ and $x' + x'' \leq m$. Formally:

$$P_{r',r''}^{x',x''} = \left\{\rho \in R^O : \left|\rho^{-1}\left(r'\right)\right| = x' \wedge \left|\rho^{-1}\left(r''\right)\right| = x''\right\} \tag{11}$$

The following lemma holds.

Lemma 4. *The possibility measure of $P_{r',r''}^{x',x''}$ is:*

$$\Pi\left(P_{r',r''}^{x',x''}\right) = \max_{O_{x'},O_{x''}}\min\left\{\begin{array}{l}\min_{o\in O_{x'}}\pi_o\left(r'\right),\\ \min_{o\in O_{x''}}\pi_o\left(r''\right),\\ \min_{o\notin O_{x'}\cup O_{x''}}\max_{r\notin\{r',r''\}}\pi_o(r)\end{array}\right\} \tag{12}$$

where $O_{x'} \subseteq O$, $O_{x''} \subseteq O$ and $O_{x'} \cap O_{x''} = \emptyset$.

Proof. By the functionality assumption,

$$\rho^{-1}\left(r'\right) \cap \rho^{-1}\left(r''\right) = \emptyset$$

Similarly to the proof of Theorem 3, the set $P_{r',r''}^{x',x''}$ can be partitioned into a number of equivalence classes according to the relation

$$\rho\mathscr{R}\acute{\rho} \Longleftrightarrow \rho^{-1}\left(r'\right) = \acute{\rho}^{-1}\left(r'\right) \wedge \rho^{-1}\left(r''\right) = \acute{\rho}^{-1}\left(r''\right)$$

Each equivalence class is characterized by a pair of disjoint sets $(O_{x'}, O_{x''})$ with cardinality x', x'' respectively, which correspond to the common preimages $\rho^{-1}(r'), \rho^{-1}(r'')$ for all reference functions ρ of the same equivalence class. Following the same line of reasoning as in Theorem 3, it is possible to derive (12). □

The extension to more referents is straightforward.

Theorem 5. *Let $\mathbf{r} = r_1, r_2, \ldots, r_k$ a sequence of referents, $r_i \in R$ and $r_i \neq r_j$ for all $i, j = 1, 2, \ldots, k$. Let $\mathbf{x} = x_1, x_2, \ldots, x_k$ a sequence of k integer numbers such that $x_i \geq 0$ and $\sum_{i=1}^k x_i \leq |O|$. Let*

$$P_{\mathbf{x}}^{\mathbf{r}} = \left\{\rho \in R^O : \left|\rho^{-1}(r_i)\right| = x_i, i = 1, 2, \ldots, k\right\} \tag{13}$$

Then:

$$\Pi\left(P_{\mathbf{x}}^{\mathbf{r}}\right) = \max_{\mathbf{O}}\min\left\{\begin{array}{l}\min\left\{\pi_o(r_i) : o \in O_i, i = 1, \ldots k\right\},\\ \min_{o\notin\bigcup O_i}\max_{r\notin\mathbf{r}}\pi_o(r)\end{array}\right\} \tag{14}$$

where $\mathbf{O} = O_1, O_2, \ldots, O_k$ *with* $O_i \subseteq O$, $|O_i| = x_i$ *and* $O_i \cap O_j = \emptyset$ *for* $i \neq j$.

Proof. The proof follows the same line of reasoning of lemma 4, by noticing that now by O_i it is denoted a set of observations with cardinality x_i. Each sequence $O_1, O_2, \ldots, O_k$ characterize an equivalence class in $P_{\mathbf{x}}^{\mathbf{r}}$. The associative property of "min" is applied to simplify the final formula. □

The previous theorem establishes the possibility measure of a set of reference functions. It is worth observing that a set $P_{\mathbf{x}}^{\mathbf{r}}$ represent the joint event that referent r_1 is referred by x_1 observations, r_2 is referred by x_2 observations, and so on up to referent r_k. Its possibility distribution can be used to define a summation function that takes into account the interactivity of granular counts.

The extension principle applied to the addition operator between two fuzzy sets X, Y results in the following fuzzy set $X + Y$:

$$(X+Y)(z) = \max_{x,y:x+y=z} \min\{X(x), Y(y)\} \tag{15}$$

This equation assumes that, when X and Y are the possibility distributions of two variables, they are non-interactive. In the case of interactive variables, eq. (15) must be extended so as to take into account the relation between variables, as follows:

$$(X+Y)(z) = \max_{x,y:x+y=z} \mathfrak{R}_{XY}(x,y) \tag{16}$$

It is possible to exploit (16) to define the sum of granular counts. To this end, it may be convenient to introduce an arbitrary order on the referents in R. With $\overrightarrow{\mathbf{r}}$ it is denoted the ordered sequence of all elements of R; given any subset $S \subseteq R$, the ordered sub-sequence of elements of S is denoted by $\overrightarrow{\mathbf{s}}$.

Definition 6. Given a subset of referents $S \subseteq R$ with $|S| = k$, the sum of granular counts is defined as:

$$\oplus S(z) = \max_{\mathbf{x}_z} \Pi\left(P_{\mathbf{x}_z}^{\overrightarrow{\mathbf{s}}}\right) \tag{17}$$

where $\mathbf{x}_z = (x_1, x_2, \ldots, x_k) \in \mathbb{N}^k$ such that $\sum_{i=1}^{k} x_i = z$.

An interesting property of the sum operator defined in (17) is that the sum of the granular counts of all referents is exactly equal to the number of observations.

Proposition 7. *Let* $m = |O|$*, then:*

$$\oplus R \equiv m$$

that is, $\oplus R(m) = 1$ *and* $\oplus R(z) = 0$ *for* $z \neq m$.

Proof. By definition, for each observation $o \in O$ there exists a referent r such that $\pi_o(r) = 1$. Let $\overrightarrow{\mathbf{r}} = (r_1, r_2, \ldots, r_n)$ the ordered tuple of referents in R and let $(O_1, O_2, \ldots, O_n)$ a tuple of subsets of O such that the following properties are satisfied:

1. $\forall i \forall o \in O_i : \pi_o(r_i) = 1$;
2. $O_i \cap O_j = \emptyset$ for $i \neq j$;
3. $\bigcup_{i=1}^{n} O_i = O$

(Sets O_i can be defined by construction: for each observation o, it is assigned to the subset corresponding to the referent r_i where $\pi_o(r_i) = 1$; in the case that more than one referent satisfies this condition, assignment is random.)

By property 1, it can be observed that

$$\min\{\pi_o(r_i) : o \in O_i, i = 1, \ldots k\} = 1$$

Also, property 3 assures there are no observations outside $\bigcup O_i$. Therefore:

$$\min \left\{ \begin{array}{l} \min\{\pi_o(r_i) : o \in O_i, i = 1, \ldots k\}, \\ \min_{o \notin \bigcup O_i} \max_{r \notin \overrightarrow{\mathbf{r}}} \pi_o(r) \end{array} \right\} = 1$$

Let $x_i = |O_i|$ and $\mathbf{x}_m = (x_1, x_2, \ldots, x_n)$. By construction, $\sum_i x_i = m$. It can be observed that the subsets O_i and numbers x_i are legitimate to apply the right-hand side of (14), therefore, by theorem 5:

$$\Pi\left(P_{\mathbf{x}_m}^{\overrightarrow{\mathbf{r}}}\right) = 1$$

As a consequence, by definition 6:

$$\oplus R(m) = 1$$

Let $z < m$ and $\mathbf{x}_z = (x_1, x_2, \ldots, x_n)$ such that $\sum_i x_i = z$. Suppose, *ad absurdum*, that $P_{\mathbf{x}_z}^{\overrightarrow{\mathbf{r}}} \neq \emptyset$. Let $\rho \in P_{\mathbf{x}_z}^{\overrightarrow{\mathbf{r}}}$, then: (i) ρ assigns x_i observations to each referent $r_i \in R$, and (ii) since $\sum_i x_i < m$ there is a non-empty subset of observations not assigned to any referent in R. This is absurd because, by definition, ρ assigns each observation to a referent. Therefore, for all $\mathbf{x}_z$:

$$\Pi\left(P_{\mathbf{x}_z}^{\overrightarrow{\mathbf{r}}}\right) = \Pi(\emptyset) = 0$$

hence $\oplus R(z) = 0$ for $z < m$. For $z > m$, again $P_{\mathbf{x}_z}^{\overrightarrow{\mathbf{r}}} = \emptyset$ because, by definition, the preimages of any assignment function partitions the set of observations, i.e. $\sum_{r \in R} \left|\rho^{-1}(r)\right| = m$. □

Proposition 8. *The granular sum is an extension of granular count, i.e.*

$$\forall r \in R : \oplus\{r\} = C_r$$

Proof. When $S \subseteq R$ is a singleton $\{r\}$, the right-hand side of eq. (17) simplifies to $\Pi(P_r^z)$ which is equal to $C_r(z)$ in accordance to theorem 3. □

5 Numerical examples

5.1 Binary possibilities

As a first example, the possibilistic assignment table reported in Table 2 is considered. The example consists of three observations and three referents and binary possibility distributions. According to the table, observation o_1 could refer to any referent r_1, r_2, r_3, while o_2 can only refer to r_1 while o_3 could refer to r_1 or r_2 but not to r_3.

	R		
O	r_1	r_2	r_3
o_1	1	1	1
o_2	1	0	0
o_3	1	1	0

Table 2: Binary possibilistic assignment table.

The application of (1) leads to the the following granular counts:[5]

$$C_{r_1} \equiv \{1,2,3\}$$
$$C_{r_2} \equiv \{0,1,2\}$$
$$C_{r_3} \equiv \{0,1\}$$

In fact, referent r_1 must be referred at least once (by o_2) and could be referred by o_1 and o_3, hence the count of r_1 spans from one to three. On the other hand, r_2 could be referred by no observations, or one observation or two, but not three (because o_2 cannot refer to r_2). Similarly, r_3 may not be referred by any observation (because o_1 is the only observation that could refer to r_3, but it could also refer to r_1 or r_2) but it cannot be referred by more than one observation.

The application of (17) yields the following granular sums:

$$\oplus\{r_1,r_2\} \equiv \{2,3\}$$
$$\oplus\{r_1,r_3\} \equiv \{1,2,3\}$$
$$\oplus\{r_2,r_3\} \equiv \{0,1,2\}$$
$$\oplus\{r_1,r_2,r_3\} \equiv \{3\}$$

In fact:

1. The observations o_2 and o_3 must refer to either r_1 or r_2 (in particular, o_2 can only refer to r_1), therefore the granular sum of r_1 and r_2 is at least 2; since o_1 could also refer to r_1 or r_2 (but not necessarily), the granular sum is at most 3.

[5] Since the possibility distributions in Table 2 are binary, the resulting granular counts are characteristic functions of subsets of natural numbers. Thus, they are equivalent ($\equiv$) to sets.

2. The granular sum of r_1 and r_3 cannot be 0 because o_2 must refer to r_1.

3. For the same reason, the granular sum of r_2 and r_3 cannot be 3.

4. The sum of all the three referents is exactly equal to 3, as anticipated by proposition 7.

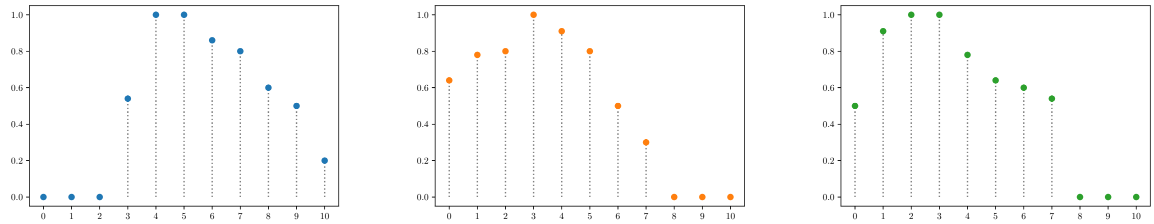
5.2 Fuzzy possibilities

Figure 1 depicts the granular counts of the three referents occurring in table 1. It is possible to notice that all granular counts are convex and normal, i.e. they define fuzzy intervals in the domain of natural numbers.

Granular counts are consistent with the assumptions of functionality and non-interactivity. For example, from fig. 1a it is observed that $C_{r_1}(3)$ is less than 1 (more precisely, it is 0.54); in fact, while it is true that three observations must necessarily refer to r_1 (namely, o_3, o_5, o_9), the observation o_1 has small possibility to refer to other referents than r_1(precisely: 0.3 for r_2 and 0.54 for r_3); therefore, it is more possible that r_1 is referred by four observations than by three: this is captured by the granular count C_{r_1}.

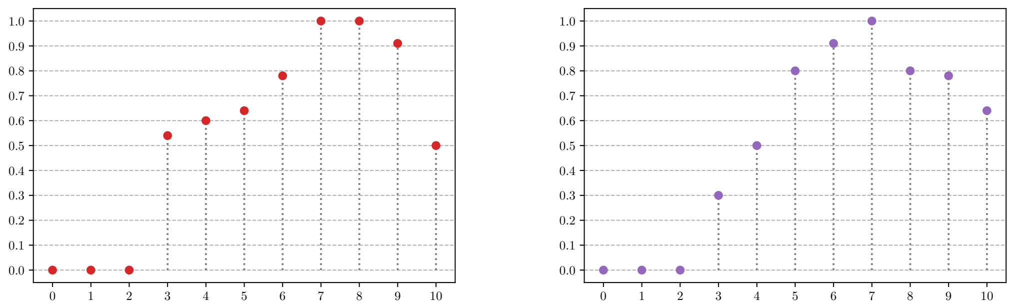
Figure 2 reports the four possible granular sums of the three referents under consideration. It is worth noting that the sum of r_1 and r_3 has one value only with maximum possibility (see fig. 2b) notwithstanding that the granular counts of both referents have two values with maximal possibility each (see figs. 1a and 1c). In particular, observations o_1, o_3, o_5, o_7, o_9 can be assigned to r_1 with full possibility while o_4, o_7, o_8 can be assigned to r_3 with full possibility, hence the possibility of 8 as sum of observations would have been maximal if a naive application of the extension principle had been performed. However o_7 cannot be shared by both referents (according to the functionality assumption) and the best possible assignment of 8 observations is o_1, o_2, o_3, o_5, o_9 for r_1 and o_4, o_7, o_8 for r_3; the possibility of such a combination is 0.8.

Similarly, the possibility that the sum is 6 amounts to 0.91 because the most possible assignment of observations is o_1, o_2, o_3, o_9 to r_1 and o_7, o_8 to r_3. It should be noticed that all these assignments have maximal possibility to their respective referents but it must be also noticed that observation o_4 can be assigned to r_2 with possiblity 0.91, which is slightly less than the maximum. This means that o_4 has more possibility to be assigned to another referent (r_3 in particular), therefore the possibility that the sum of observations for r_1 and r_2 is 6 (i.e. excluding o_4) is lower than the possibility that the sum is 7 (i.e., including o_4). Again, this property is fully captured by the granular sum.

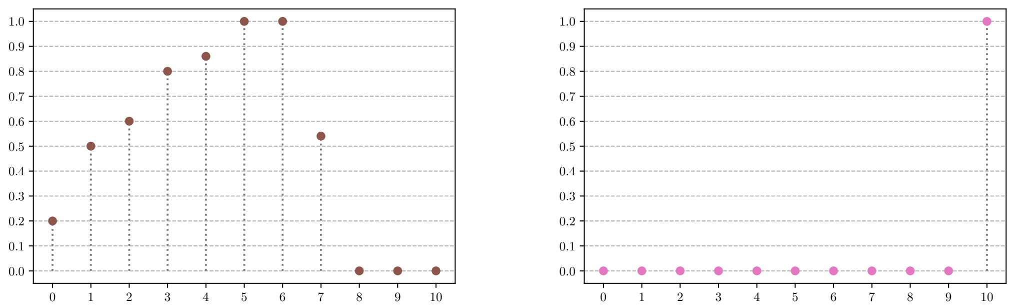


(a) C_{r_1} (b) C_{r_2} (c) C_{r_3}

Figure 1: Granular counts of the referents of the possibilistic assignment table in Table 1.



(a) $\oplus\{r_1, r_2\}$ (b) $\oplus\{r_1, r_3\}$



(c) $\oplus\{r_2, r_3\}$ (d) $\oplus\{r_1, r_2, r_3\}$

Figure 2: Granular sums of the referents in Table 1.

As a final remark, fig. 3 reports the granular sum of r_1 and r_3 in comparison with the result of applying the extension principle for summing the granular counts of the referents (“fuzzy sum”). It is immediate to observe that granular sum is significantly more specific than fuzzy sum. In particular, the latter has a support in the interval $[3, 17]$ while granular sum has a support limited in $[3, 10]$ (as expected because there are no more than 10 observations). Also, the possibility degree of granular sum is always less than or equal to fuzzy sum. In particular, when the possibility degree of granular sum is 1, then the possibility degree of fuzzy sum is also 1; on the other hand, when the possibility degree of fuzzy sum is 0, then the possibility degree of granular sum is zero too. The same considerations apply for any other set of referents.

6 Conclusions

Theorem 3 shows that granular count is a consequence of two assumptions: functionality and non-interactivity. Functionality requires that each observation must refer to one referent only, although it may

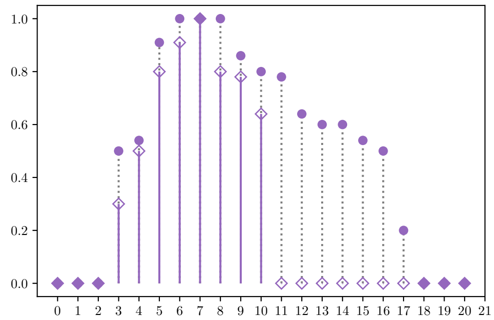


Figure 3: Comparison of granular sum of r_1 and r_3 (empty diamonds ♢) and fuzzy sum of the respective granular counts (filled circles ●). Filled diamonds (♦) represent coinciding values.

be not known in advance; non-interactivity requires independence of two observations: knowing the referent of an observation does not affect the knowledge of the referent of the other. These assumptions are very weak and may be applied in a wide variety of problems involving counting in presence of uncertainty.

Definition 6 defines the sum of granular counts as the application of the extension principle applied to addition, but taking into account the interactivity of granular counts. In force of theorem 5, the sum of granular counts is the most specific fuzzy set that can be obtained. The numerical results show that the difference—in terms of specificity—between the defined sum and the naive application of the extension principle (i.e. without taking into account interactivity), can be significant.

The direct application of granular count (1) and granular sum (17) leads to exponential-time procedures, which are intractable for problems involving a large number of observations. As concerning granular count, a quadratic-time exact counting procedure, and a linear-time approximate counting procedure are available [9]. The search for a polynomial-time procedure for granular sum is subject of ongoing research. This would shed light on future extensions involving granular statistical techniques for data analysis in presence of uncertainty.

Acknowledgement

The author wishes to express his thankfulness to Dr. Ciro Castiello of the University of Bari Aldo Moro, for his helpful suggestions. The research is partially supported by the GNCS project "Computational Intelligence methods for Digital Health". The author is member of the INdAM Research group GNCS.

References

[1] N. Boukhelifa, M.-E. Perrin, S. Huron, J. Eagan, How Data Workers Cope with Uncertainty: A Task Characterisation Study, in: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, ACM, 2017, pp. 3645–3656.

[2] G. L. Brase, L. Cosmides, J. Tooby, Individuation, counting, and statistical inference: The role of frequency and whole-object representations in judgment under uncertainty., Journal of Experimental Psychology: General 127 (1) (1998) 3.

[3] A. Consiglio, C. Mencar, G. Grillo, F. Marzano, M. F. Caratozzolo, S. Liuni, A fuzzy method for RNA-Seq differential expression analysis in presence of multireads, BMC Bioinformatics 17 (S12:345) (2016) 167–182.

[4] D. Dubois, H. Prade, Gradualness, uncertainty and bipolarity: Making sense of fuzzy sets, Fuzzy Sets and Systems 192 (2012) 3–24.

[5] D. Dubois, H. Prade, Possibility Theory, in: Computational Complexity, Springer New York, New York, NY, 2012, pp. 2240–2252.

[6] C. Mencar, An incremental algorithm for granular counting with possibility theory, in: IEEE International Conference on Fuzzy Systems, IEEE, Glasgow, United Kingdom, 2020, pp. 1–7.

[7] C. Mencar, Possibilistic Bounds for Granular Counting, in: Information Processing and Management of Uncertainty in Knowledge-Based Systems. IPMU 2020., Vol. 1239 of Communications in Computer and Information Science, Springer, Cham., 2020, pp. 27–40.

[8] C. Mencar, W. Pedrycz, GrCount: Counting method for uncertain data, MethodsX 6 (2019) 2455–2459.

[9] C. Mencar, W. Pedrycz, Granular counting of uncertain data, Fuzzy Sets and Systems 387 (2020) 108–126.

[10] W. Pedrycz, A. Skowron, V. Kreinovich, Handbook of Granular Computing, Wiley & sons, 2008.

[11] D. Ramsköld, E. Kavak, R. Sandberg, How to analyze gene expression using RNA-sequencing data, in: Next Generation Microarray Bioinformatics, Springer, 2012, pp. 259–274.