Where should I submit my work for publication? An asymmetrical classification model to optimize choice

A. Ferrer-Sapena, J.M. Calabuig,
L.M. García-Raffi, E. A. Sánchez-Pérez

Instituto Universitario de Matemática Pura y Aplicada,
Universitat Politècnica de València,
Camino de Vera s/n, 46022 Valencia, Spain.

{anfersa,jmcalabu,lmgarcia,easancpe}@upv.es

Abstract

Choosing a journal to publish a work is a task that involves many variables. Usually, the authors' experience allows them to classify journals into categories, according to their suitability and the characteristics of the article. However, there are certain aspects in the choice that are probabilistic in nature, whose modelling may provide some help. Suppose an author has to choose a journal from a preference list to publish an article. The researcher is interested in publishing the paper in a journal with a rank number less than or equal to k. For this purpose, a simple classification model is presented in order to choose the best journal from the list, from which some fundamental consequences can be deduced and simple rules derived. For example, if the list contains 100 journals and is ordered using 2-year Impact Factor, the rule "send to the journal at the $k - 10$ position" is adequate.

1 Introduction

Scientific authors are often constrained to publish their research work in a closed list of journals having the following properties: they are specifically devoted to the scientific field in which they work, and they are well-considered – or at least accepted – by the institutions that will evaluate the authors' work. This fact leads to the problem of deciding where a just finished research paper must be sent for publication. Or, as a more planned way of working, choose a suitable journal before starting the research work. The problem is not in general easy to solve, apart from the case that the journal is completely determined by the topic or by some concrete circumstance (for example, the publication of a special issue of a journal). In case the authors have different options for publication, a general decision rule does not exist. This "best option" must be understood to be the one that provides best benefits in terms of the evaluation of the research. In fact, the definition of the best option should refer, for example, to publication in journals that are read by leading researchers, which could be transferred to the elaboration of an evaluation list that favours publication in those journals. The aim of this paper is to show such a decision rule for helping the authors to make the best election based on some probabilistic arguments.

The question about the reasons of an editor for accepting a paper for publication and the capacity of the authors to know in advance which is the best option is a complicated business. There are a lot of factors that condition the acceptance of a paper, and publication of scientific results has become a difficult task for the research groups. This circumstance is not new, and was already well-known at the beginning of this century (see Lawrence (2003)). As a first approximation to the problem, the authors may consider more specific criteria that can be found in general works or handbooks that are written for helping researchers to publish their work (see for example Gibbs (2016)). Among other things, experts in scientific publication advise that the journal where the paper will be sent for publication must be chosen in advance of doing the research work (see for example Ch.6 in Gastel and Day (2016)). Even if this is the case, the model presented in this paper may be helpful when there are several journals – in fact, an ordered list – in which the author would be tempted to try to publish their works.

A different way to address this problem is to use a recommendation system specifically designed for article publishing. In particular, an author could use knowledge-based recommendation techniques, based on his/her expertise and experience of his/her research environment, to develop such a recommendation tool (see Lu et al. (2015), Section 2.3). Although this option makes sense, it would be used if a complete institution is interested in such a system: for a single author it does not seem to be the best option. A classic model like the one we propose here has the advantage of providing an overview of the process trends based on its mathematical structure, which leads to simple rules and arguments that the author can use.

Thus, in short, the aim of this paper is to provide a practical tool to help authors choose the best journal for their research article. In other words, we in-

tend to show a procedure of classifying the list of journals in terms of suitability. We are interested in addressing the problem using a mathematical model that should be applied after all considerations on the relevance of the work to the scientific community have been made, including the authors' self-assessment of the paper, and the scope of the journals have been taken into account.

The context for a suitable application of our model should be the following: authors have an ordered list of journals – all of them interesting to publish their research – and they have to choose one of them, bearing in mind that the order is given by the ease with which an article can be accepted. Obtaining the ordered list can be done through different procedures: the authors themselves can create it using their own knowledge of the topic. Another method could be given using external advice, or some bibliometric index. The last option will be used in some of the examples shown in the second part of the work. The main criticism that might be done for using this procedure is that there is of course a lot of different ways of ordering a list of journal by their "relevance", "impact" or any other criterion that should provide an index – see for example Bradshaw and Brook (2016). We must say that this ordering criterion must be chosen by the author who wants to use the model: the mathematical model does not depend on the particular ordering.

Consequently, the assumptions for our model to work are independent of other typical arguments that an author takes into account when presenting a paper. In fact, given an ordered list of journals that the author trusts, self-evaluation of the work is done when the author chooses the "level" when wanting to publish. After all the considerations are taken into account by the author – including considering only a subset of the global list, the process of submitting an article for publication is similar to investing in stock options when no market-specific information is known, but only general trends. From this point of view, concepts as interest rate risk and volatility have counterparts in the scientific publication world. Therefore, the mathematics presented here are more related to some financial tools than to classical bibliometric studies, and this is the main novelty of our work. Although we have not followed any special financial model, our arguments can be understood in economic terms: the interested reader can find more information about such models for example in Hol (2013) and the references therein.

We will assume the following simplifications for our work:

a) We will consider ordered journal lists that are dynamic, in the sense that the order may change each fixed period of time – typically, every year, and the journals appearing in it are – more or less fixed. The canonical example of such a list is the one given by the Journal Citation Reports of Clarivate Analytics – which is one of the best known quality indicator in research activity[1] – for a fixed subject category; it is based on the 2-year Impact Factor, among other indicators, which measures the number of citations in the first two years after publication. However, it should be noted that these lists have not retained their main properties over time, and therefore

---

[1] https://clarivate.com/products/journal-citation-reports/

cannot be expected to be an irrefutable reference. On the contrary, for example, they have been found to have experienced a process of inflation in the number of cites, motivated precisely by their use as an evaluation tool – see Althouse et al. (2008); see also Neff and Olden (2010).

b) The order induced in the list of journals could also be defined on the basis of the author's experience in publishing in the field, by assigning an order to the list of journals in which he/she usually publishes – or wants to publish – his/her work.

c) For the sake of simplicity, the change in the order of the journals on the list is considered to be independent of changes in the order of the other publications on the list. Specifically, it is assumed that the "jumps in order" of each journal follow approximately a normal distribution.

d) Order numbers of top journals are more stable than the rest. That is, the variation of the order in the successive lists of a given series is more pronounced for journals in low positions than for publications in top positions. Thus, the variance of the corresponding normal distributions is assumed to be increasing with the order number. In other words, it is supposed to have a power-law behavior.

The information required in order to adapt the model to each particular situation might be obtained from the known behavior of the usual impact factor lists.

This information is usually given by the statistical analysis of the impact factor lists, mainly concerning the variation of the distribution of the order values of the journals in the list. Nowadays there is a lot of information of this type, since several researchers have centered their attention in this kind of analysis. Probably the most comprehensive recent study concerning stability of the impact lists has been published by Pajić (2015). In the same line, the paper by Ferrer-Sapena et al. (2017) provides an index easy to understand for measuring the stability of the impact factor lists. More information about the topic can be found in the papers by Xu et al. (2014), Haghdoost et al. (2014), Althouse et al. (2009), Black (2012) and Mansilla et al. (2007). The interested reader can obtain from these papers and the references therein the data that are needed for computing the parameters of the model, in the sense we will explain in the present work. Some studies on the stability of the journals in the impact factor ordered lists can be found in the papers by Ferrer-Sapena et al. (2015), (2016) and (2017) (see also the references therein).

2   The problem

The ideas that we use for the model are based on some classical arguments related to the distribution of probability of some processes of transmission of information. The interested reader can find the roots of these mathematical

tools in the classical papers of Kelly (1956) and Shannon (1948). In particular, we will consider lists assuming that all the journals in them are interesting for the author, taking into account his/her own motivations and restrictions. For example, young researchers might assume that they could have more difficulty publishing in high-level journals, and would prefer to publish their article in "medium" journals rather than risk a long and ultimately failed process. *The author is interested in finding a journal in a list to submit his/her article that is at least in the k position after a time t_0 for a fixed order number k that he/she considers "good enough" for the work.*

2.1 The general model.

The basic assumption of the model is that the order of the list is dynamic, that is, it changes over time. To simplify the calculations in the model we declare the position of a journal n as a continuous variable with values in $[0, N_{max}]$. This approach facilitates mathematical management, and is not so risky, as we are considering lists with a relatively large number of journals (for example, greater than or equal to 20). Therefore, for the sake of simplicity, we assume that the random variable has a continuous probability density function.

1) Fix a list of scientific journals ordered by an impact index having N_{max} elements. An author – or a group of authors – "a" wants to publish in a journal placed in one of the first k positions on the list in a time period $[0, t_0]$, where $t = 0$ is the moment when the paper is submitted. The probability of the event that the journal s – that is in position n of the list – accepts the paper after the time t_0 together with the fact that the journal has a ranking position less or equal than k is given – after normalization – by a function $P_{a,k}^{t_0}(s)$; note that we consider the best position in the ranking the order value 1. The assumptions are that t_0 and the order k are fixed. The variable n – that is the ordering number of the journal in a fixed list – is the one with respect we want to find the optimal value of the function. That is, we want to find the value of n that maximizes the function $P_{a,k}^{t_0}(n)$.

2) In our model, we suppose that the function $P_{a,k}^{t_0}(n)$ is the product of two probability functions $\Phi_a^{t_0}(n)$ and $\Psi_k^{t_0}(n)$, that is,

$$P_{a,k}^{t_0}(n) = \Phi_a^{t_0}(n) \cdot \Psi_k^{t_0}(n), \quad n = 1, ..., N_{max}.$$

   a) The first function $\Phi_a^{t_0}(n)$ represents the probability of the journal of giving a positive answer to the author a in a time belonging to the interval $[0, t_0]$. We can accept that it has two multiplicative factors, $d(n, t)$ and $a(n)$. $a(n)$, is the probability associated to the event that the journal in the position n accepts the paper presented by a; this should depend on the reputation of the researcher a, or simply on the properties of the paper a – novelty, quality,.... The factor $d(n, t)$ – the

5

time decision function – represents the probability that the journal gives an answer – "accept" or "reject" – in a time before the time t

$$\Phi_a^{t_0}(n) = d(n, t_0) \cdot a(n), \quad n = 1, ..., N_{max}.$$

b) The second function $\Psi_k^{t_0}(n)$ gives an estimate of the probability of the event that the journal n has an order value bigger or equal than k in the time t_0 of giving a positive answer to the author "a" in a time belonging to the interval $[0, t_0]$. We assume that this value is independent of the probability distribution for other values of the variable n. Therefore, if $g_n^{t_0}(s)$ is the probability density function for the journal indexed by n, we have that

$$\Psi_k^{t_0}(n) := \int_0^k g_n^{t_0}(s) \, ds.$$

3) Therefore, the function that – after normalization – gives the probability of the author a to publish the paper in a journal in the position n and this journal is at a position $n \leq k$ at the time t_0, is

$$P_{a,k}^{t_0}(n) = d(n, t_0) \cdot a(n) \cdot \int_0^k g_n^{t_0}(s) \, ds, \quad n = 1, ..., N_{max}.$$

In our construction we use the product to consider together in the final distribution the probability functions associated with the variables that concern the process. We accept this standard assumption for the description of the general model. However, in general we cannot expect these probability functions to be independent; as we have said, this is a simplification of a complex problem. To improve our approach in particular cases, certain current developments in statistics and other related mathematical fields would help. In this direction, there are several techniques that could be used. The first is the use of appropriate copulas to mix distributions. Copulas provide a method for linking separate statistical distributions that affect the same event; see Nelsen (1999) and Klement et al. (2001). Selecting an accurate copula that fits the nature of the model would provide an improvement of our method. Other "mixing" methods could also be considered for this purpose. For example, the use of aggregation functions, which allow us to find a suitable function adapted to the model for mixing independent distributions, provides another way of obtaining such an improvement, Grabisch et al. (2009). Another interesting mathematical context that would help to improve the model in this direction is that of fuzzy mathematics. The so-called t-norms, see Klement et al. (2000), would also provide a methodological framework for mixing probability functions according to the requirements of the model when it is adapted to a specific situation.

2.2   The specific model: the inverse proportionality criterion.

To estimate the probability function $P_{a,k}^{t_0}$ some simplifications are needed. Let us fix a unitary time period $[0, t_0]$ – for example, one year. The following

considerations can be adopted.

(i) We may assume that the time decision function $d(n, t)$ is not known, but it does not vary too much for different journals in the list. Therefore, we suppose that $d(n, t_0) = d = constant$.

(ii) The function $a(n)$ depends on a main term and a factor $F(a)$ that depends, broadly speaking, on the reputation of a and the quality of the paper. The main term is given by a function that describes the expected behavior of the journals in a list. Recall that by assumption, this list have been ordered roughly speaking by some criterion based on the decreasing difficulty of publication. In this sense, and as a first approximation, it can be associated to an impact type factor, so that the function can be given by an increasing bounded function that fits such a list. For example, an inverse proportionality function $K/(N_{max} + 1 - n)$ would play the role. It would represent that in general publication in a top journal is more difficult, and the bigger the order number n, the bigger the probability of acceptance. We would obtain then the following simplified model,

$$\Phi_a^{t_0}(n) = d \cdot F(a) \cdot \frac{K}{N_{max} + 1 - n}, \quad n = 1, ..., N_{max}.$$

However, the function $\Phi_a^{t_0}(n)$ might be also given by other increasing bounded function. A degree-2 polynomial fitting a cloud of points will play the role in Section 3.1.

(iii) We propose also as a reasonable and user-friendly model that the possible changes in the position of a given journal n in the list follows a Gaussian type distribution localized in the domain $[0, N_{max}]$ and with center in n. The "variance" of this distribution is variable and depends on n.

This leads to a model in which the distribution of the probability that in the time t_0, the journal that is in the position n changes to the position s is given by the equation

$$\frac{1}{N(n)} e^{-\alpha(n)(s-n)^2}, \quad s \in [0, N_{max}],$$

where $N(n)$ is the normalization constant $N(n) = \int_0^{N_{max}} e^{-\alpha(n)(s-n)^2} ds$.

Some recent bibliometric analysis on the stability of the positions in a journal list using the most common order criterion – the two-year impact factor IF2 – suggest that the changes in the positions of the journals in an impact factor list are bigger for bigger values of the order number n (see Pajić (2015), Ferrer-Sapena et al. (2017) and the references therein). That is, the bigger the order value, the bigger the variance of the distribution. This dependence might be represented for example by the simple model

$$\alpha(n) = \frac{b}{n+1}, \quad n \in [0, N_{max}],$$

7

where b is a positive parameter and $n+1$ is written instead of n for assuring that the maximum of the function in the given domain is b. Removing the normalization constants of the equation, we obtain that

$$e^{-\alpha(n)(s-n)^2} = e^{-\frac{b}{n+1}(s-n)^2},$$

and so, comparing with the general equation of a Gaussian distribution we get

$$\frac{1}{2\sigma^2} = \frac{b}{n+1}, \quad \text{and then} \quad b = \frac{n+1}{2\sigma^2}, \tag{1}$$

where σ^2 is the variance of the distribution.

After normalization, we obtain the equation

$$\Psi_k^{t_0}(n) = M \frac{\int_0^k e^{-\alpha(n)(s-n)^2} \, ds}{\int_0^{N_{max}} e^{-\frac{b}{n+1}(s-n)^2} \, ds}, \quad n \in [0, N_{max}],$$

where M is a normalization constant. For the aim of simplifying the explanation of the model, we will assume in the rest of the paper that according to the comments above, $\alpha(n) = b/(n+1)$. However, note that this can be adapted to any particular situation, in case this function does not describe the expected behavior of the journals of the list.

(iv) The final equation for the model, summing up all the contributions, is

$$P_{a,k}^{t_0}(n) = (d \cdot F(a) \cdot M) \cdot \frac{K}{N_{max} + 1 - n} \cdot \frac{\int_0^k e^{-\frac{b}{n+1}(s-n)^2} \, ds}{\int_0^{N_{max}} e^{-\frac{b}{n+1}(s-n)^2} \, ds}, \quad n \in [0, N_{max}].$$

Since the aim of this paper is to obtain the value on n for which this function attains the maximum, we can assume that $d \cdot F(a) \cdot M \cdot K = 1$, and so

$$P_{a,k}^{t_0}(n) = \frac{1}{N_{max} + 1 - n} \cdot \frac{\int_0^k e^{-\frac{b}{n+1}(s-n)^2} \, ds}{\int_0^{N_{max}} e^{-\frac{b}{n+1}(s-n)^2} \, ds}, \quad n \in [0, N_{max}]. \tag{2}$$

Observe that this function only depends on a scalar parameter b, which can be computed using formula (1). Therefore, we need a good estimate of the variances of the normal distributions of the changes in the position of the journal with different order values n. We will show in the next section a study of the possible range of the values of b, and some estimates of its real value using some recent bibliometric analysis of the Journal of Citation Report (JCR) list of Mathematics. Note that, since we are using a model assuming that the variable (the order number) take real values, we can normalize every list to $N_{max} = 100$ in order to compare different lists. This will be done in the following sections.

0.05	0.10	0.10	0.15	0.18	0.21	0.27	0.38	0.45	0.56
0.05	0.10	0.12	0.15	0.20	0.24	0.30	0.42	0.50	0.58
0.08	0.10	0.13	0.18	0.20	0.27	0.35	0.42	0.56	0.64

Table 1: Rates of acceptance of the journals of the list.

3   How to use the model.

The main elements for the construction of the model have been given in the previous section. However, sometimes mathematical expressions do not suggest attractively how the model can be used, although it provides a wide range of potential applications. This section shows some of these specific uses.

3.1   Using the personal experience of the members of the research group

Let us present a simple – and in a sense canonical – example of use of our model. Suppose that an author is interested in using it for the optimization of the publication time of his/her papers. He belongs to a research institute that is specialized in a particular scientific field, and they have elaborated a list of scientific journals in which articles from this area are regularly published. They also have statistical information about the acceptance or rejections of papers with the concrete scope of the investigation that is developed in the institute. The list contains 30 journals, and the rate of acceptance of papers of the institute by each of them is known. This rate is defined as

$$r_i = \frac{Number\ of\ times\ that\ a\ paper\ sent\ to\ the\ journal\ ``i"\ has\ been\ accepted}{Number\ of\ times\ that\ a\ paper\ has\ been\ sent\ to\ be\ published\ in\ ``i"}.$$

We can assume without loss of generality that the list is ordered using the increasing value of the rate r_i, and so the index i represents the journal appearing in the position i of the list. Table 1 shows the rates of the list of journals.

After fitting these values to a 2-degree polynomial, we obtain the function

$$R(n) = 0.07 + 6.24 \times 10^{-4} \times n^2, \quad n = 1, ..., 30,$$

that represents the value of the rate of acceptance for the journal n. Figure 1 shows the representation of the data and the fitting function R.

On the other hand, the variation of the position of the journals in the list when it is updated after a year follows the following general rule: the first journals tend to preserve their positions, while the last journals change a lot. It is estimated that the publication process will take at least one year, so the final list is not the one that we know for the present year. That is, we have to assume a change of position of the journals, that is affected by a probabilistic law. Therefore, following our model a fixed journal J follows a Gaussian distribution
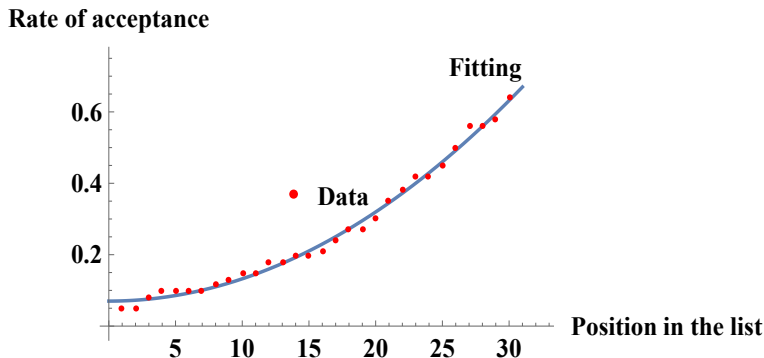
9

Figure 1: Representation of the rates of acceptance of the list of journals.

which variance increases with the inverse of the position of J. The constant b appearing in the variance function is assumed to be equal to 1. Summing up all these arguments, we obtain the following equation for the probability of publishing in a journal that is at least in the position $k \in \{1, ..., 30\}$ after the year that is needed for finishing the process.

$$P_{a,k}(n) = \frac{1}{N}\left(0.07 + 6.24 \times 10^{-4} \times n^2\right) \cdot \frac{\int_0^k e^{-\frac{1}{n+1}(s-n)^2}\,ds}{\int_0^{30} e^{-\frac{1}{n+1}(s-n)^2}\,ds}, \quad n \in [0, 30],$$

where N is a normalization constant.

The following arguments are accepted by the author to choose the best option. His main interest is to diversify the journals in which his publications appear, as a strategy to improve the visibility of his work. In addition, a research assessment committee may find journals with low acceptance rates more valuable: often "difficulty of publication in a journal" and "quality of the work to be published" are considered directly related facts. Consequently, *he wants to publish in journals that are in the top part of the list,* although all journals might be acceptable.

In particular, the *author's goal is to publish the article in a journal that is in the top half of the list.* The following Figure 2 provides a representation of the probability function when $k = 15$ – top half part of the list – is fixed. It can be seen that the best chance of successfully publishing the paper is when it is sent to the journal in 12th position.

3.2 How to use the JCR list for applying the model in Mathematics

We fix a list of 98 journals of mathematics appearing in the JCR list Mathematics and the associated order numbers, considering a time series of 15 years
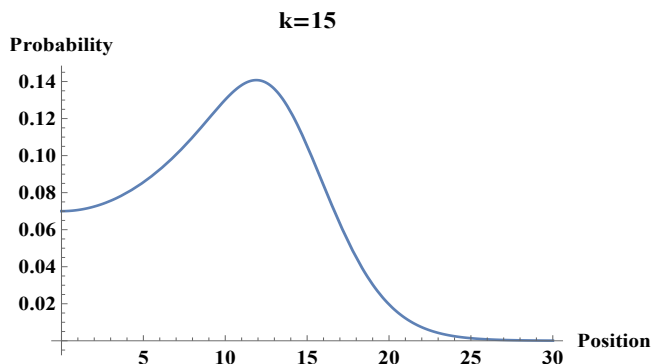
Figure 2: Probability of success when $k = 15$ is fixed.

starting in 1997. Let us obtain an estimate of the parameter b appearing in the model given by equation (2). We choose ten journals uniformly distributed in the list and apply formula (1) for finding a sequence of variances. After that, we will compute the mean of all the values, using it as the desired estimate.

In order to do that, we first consider the order number of each journal in all the years. The order number is computed as the ratio among the position of the journal and the total number of journals each year in the Journal Citation Reports of Thomson-Reuters – Web of Science (WOS), now Clarivate Analytics – multiplied by 100 for getting a percent representing the position of the journal in the list. This allows to simplify the calculations by using this number directly and assuming that the total number of journals is 100; this is an advantage of considering a continuous model instead of a discrete one. The list of values is given in Table 2. The journals that have been considered are – in the same order as in the table – Journal of Functional Analysis, Journal of Differential Equations, Studia Mathematica, European Journal of Combinatorics, Mathematische Nachrichten, Journal of Mathematical Analysis and Applications, Annals of Fourier Analysis, Nagoya Journal of Mathematics, Osaka Journal of Mathematics and Acta Mathematica Hungarica.

In a second step, we consider the year series of values of each journal separately, in order to fit them into a Gaussian model centered in the mean value. We have to take into account now that the variable n in the model has to be computed as $100 - p$, where p is the order value given in the table: the first journal corresponds to the value $n = 0$ in the model, and to 100 in the table. After this transformation, we obtain the values of the mean and the variance for each journal, that are presented in Table 3.

We can use now equation 1 – that provides the relation among the variance and the coefficient b – in the model is given by $b_i = (mean_i + 1)/(2var_i)$, with the values of the means and the variations ($mean_i$ and var_i, respectively) in Table 3; each index $i = 1, ..., 10$ corresponds to the journal in the position i in this table., we obtain a list of estimates of the constant b.

11

	JFA	JDEQ	STUD	EJCO	MANA	JMAA	AFOU	NAJM	OSJM	AMHU
1997−	92.647,	83.823,	2.794,	63.235,	54.411,	53.676,	56.617,	30.882,	24.264,	3.676
1998−	88.405,	83.333,	50.7246,	55.797,	59.420,	63.768,	76.811,	37.681,	20.289,	57.97
1999−	88.275,	86.896,	38.620,	57.93,	39.310,	59.310,	72.413,	44.137,	32.413,	17.93
2000−	92.307,	87.179,	59.615,	26.282,	48.717,	52.564,	70.512,	40.384,	16.025,	17.94
2001−	90.062,	91.925,	49.689,	36.024,	40.372,	59.006,	69.565,	55.279,	21.739,	13.043
2002−	92.352,	91.176,	45.882,	58.823,	44.117,	55.882,	82.941,	34.117,	42.352,	22.352
2003−	91.379,	86.206,	47.126,	44.827,	48.850,	55.172,	70.689,	76.436,	28.735,	35.057
2004−	91.712,	87.292,	65.745,	27.624,	53.591,	62.983,	60.773,	18.232,	11.049,	22.651
2005−	81.215,	88.397,	66.298,	25.414,	50.276,	67.955,	60.221,	29.281,	33.149,	13.259
2006−	82.887,	93.048,	51.871,	73.262,	32.620,	75.935,	72.727,	44.919,	49.732,	33.155
2007−	85.024,	92.270,	56.521,	68.115,	33.333,	82.608,	74.879,	76.328,	25.1208,	24.637
2008−	86.511,	94.883,	22.325,	61.860,	45.116,	85.581,	72.558,	20.465,	40.9302,	9.767
2009−	90.157,	92.519,	51.574,	72.047,	53.543,	88.582,	76.771,	31.102,	21.653,	34.645
2010−	88.889,	90.681,	46.953,	66.667,	59.498,	88.531,	69.892,	10.394,	12.544,	38.709
2011−	88.541	92.708	54.166	63.541	64.931	86.111	47.917,	27.431	26.736	35.764

Table 2: Representation of the coefficients "b" in the model for the journals in the example.

	JFA	JDEQ	STUD	EJCO	MANA	JMAA	AFOU	NAJM	OSJM	AMHU
Mean	88.691,	89.489,	51.994,	53.430,	48.540,	69.178,	69.019,	38.471,	27.115,	25.371,
Variance	12.18,	12.72,	147.31,	287.14,	90.22,	192.29,	81.24,	363.89,	122.33,	194.96,

Table 3: Representation of the coefficients "b" in the model for the journals in the example.

Figure 3 shows a representation of the values obtained for b_i, $i = 1, ..., 15$. It can be seen that the obtained values almost belong to the interval $[0.1, 0.5]$. A good estimate for fitting all of them to a constant value is the mean of all the means of the journals, that is

$$mean = 0.236, \quad \text{and the distribution has a variance} \quad var = 0.023.$$

Therefore, we choose as an estimate of the coefficient in the distribution function
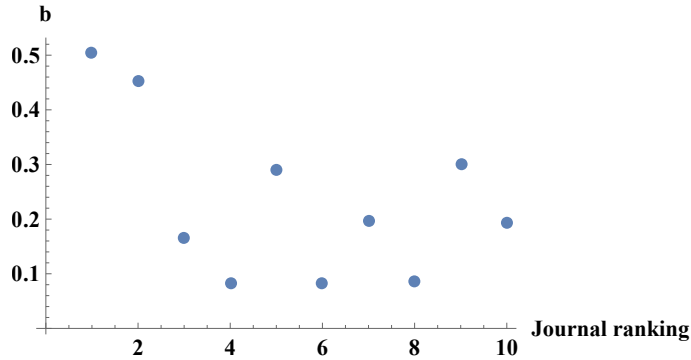
$$\mathbf{b = 0.236}.$$

Figure 3: Representation of the coefficients "b" in the model for the journals in the example.

The function in the model for the subset of 100 journals – 98 re-scaled to 100 – in the JCR list MATHEMATICS, for $t_0 = 1\, year$ and a generic author a is then given by the equation

$$P_{a,k}^{t_0}(n) = \frac{1}{101 - n} \cdot \frac{\int_0^k e^{-\frac{0.236}{n+1}(s-n)^2} \, ds}{\int_0^{100} e^{-\frac{0.236}{n+1}(s-n)^2} \, ds}, \quad n \in [0, 100]. \tag{3}$$

Let us analyze the behavior of the model by showing the graphics of the function for several values of k.

In general, it can be seen that for top part of the list (k in $[0, 75]$), the model shows that the best valued (the optimum for the probability of acceptance in a journal in a position $n \geq k$), is to choose a journal that is in a position before k in the list. How to choose it depends on the value of k. However, in the last part of the list ($k \geq 75$) it seems better to consider any journal below k in the impact list, namely the last one. Before this value 75, the option of choosing a journal with a position $n \leq k$ is still in competition with this "last position rule" of the list". In the interval in which the model shows a "standard" behavior – $k \in [25, 75]$ – , the best option is to send the paper to a journal that is in the
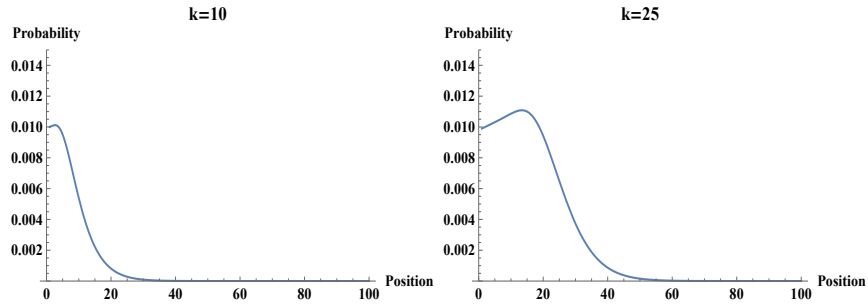
Figure 4: Model for $k = 10$ (left) and $k = 25$ (right) in the example.

position $k - r$, where $7 \leq r \leq 15$. The greater the value of k, the greater must be r.

Figure 4 shows that the model for the first values of k, – that is, the author wants to publish in a top journal – , is to send to a journal in the position $k - 4$. Since the distributions is rather flat in the first part, it is reasonable to simply send to the second journal in the list, for example, as a first option.
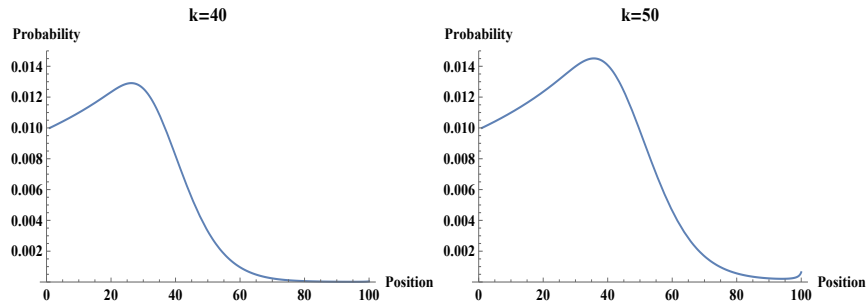

Figure 5: Model for $k = 40$ (left) and $k = 50$ (right) in the example.

However, Figure 5 shows that the model works in a uniform way when the values of k are situated in the central part of the list. In these cases, –starting with the case $k = 25$ shown in Figure 4 and considering the values $k = 40$ and $k = 50$ – , it can be seen that the journal to which the paper must be sent follows the formula $k - n$, where n starts having the value 12 and increases as k increases. However, these increments have a more or less fixed value. Thus, for $k = 40$ and adequate value should be $n = 14$ – and so the author must send the paper to the journal $40 - 14 = 26$ – , and for $k = 50$, n must be equal to 15, and the author must consider the journal in the position $50 - 15 = 35$ for submitting the paper. The rule

"for publishing in a journal at a position at least equal to k,

14

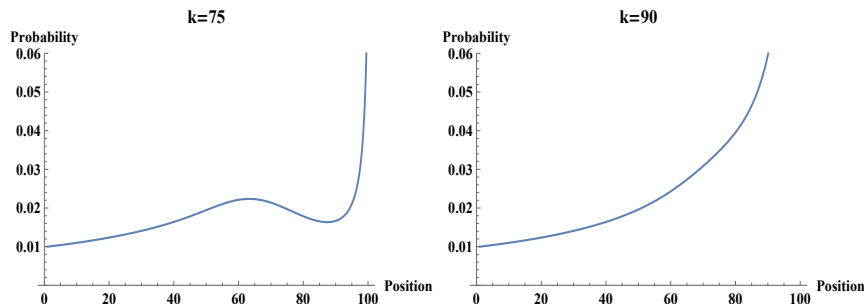seems to work as a simplification of the general behavior of the model.

Figure 6: Model for $k = 75$ (left) and $k = 90$ (right) in the example.

Figure 6 shows a different behavior of the model. After the value of $k = 75$, it seems better to improve the probability of publication to send the paper to a journal at the end of the list. The high variance of these journals, together with the increase of probability of being published in a journal with low factor, makes a better decision to send to the last journals instead of following the rule written before.

3.3   How to use the JCR list for applying the model in Applied Physics

Using the same procedure as in the case of Mathematics, we have also considered the JCR list of Applied Physics. In this case, the journals that we have chosen are Nanotechnology, Journal of Applied Physics, Applied Physics B, Solid State Electronics, Physica C, European Journal of Applied Physics, European Physics Journal, International Journal of Modern Physics and Journal of Nonlinear Optics Physics and Materials. Although the total number of journals of the original list in which we took this sample is 54, we normalize by taking the order number to compare the graphics with the case of Mathematics.

After the same computations that have been explained in Section 3.2, in Figure 7 we show the representation for the coefficient b of all the journals in the case of the list for Applied Physics. As the reader can see, the values are concentrated around the mean value $b = 0.334$, that we fix as a parameter for the model.

The mean value – and so our estimate – for b is in this case bigger than in the case of Mathematics. Indeed, it was equal to 0.236 for Mathematics, while its value for Applied Physics is 0.334. This fits with the already known fact that the list of Applied Physics is more stable than the one of Mathematics. Indeed, as can be seen in equation 1 the value of b is inversely proportional to
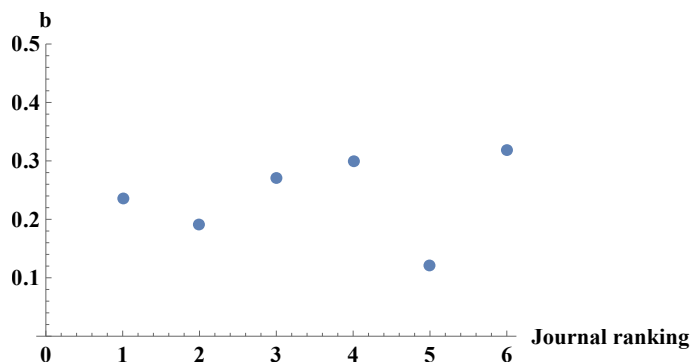
Figure 7: Representation of the values of the parameter "b" for the journals in the list of Applied Physics.

the variance of the distribution: the bigger the parameter b, the narrower the distribution.

Let us show the result of the model for the cases $k = 40$ and $k = 75$ in Figure 8.
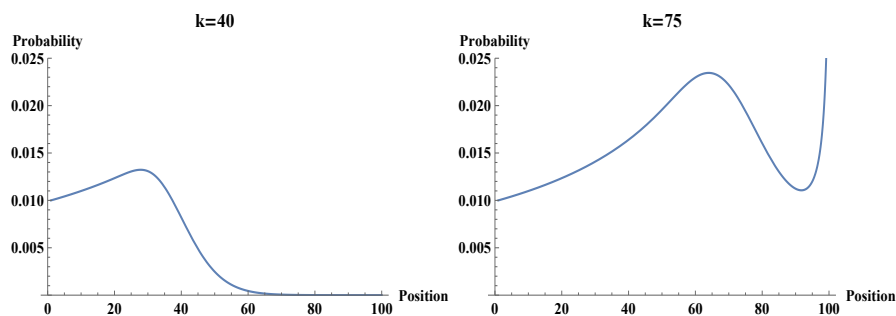

Figure 8: Model for $k = 40$ (left) and $k = 75$ (right) for Applied Physics.

To finish with the examples, let us show together the results of the model for Mathematics and Applied Physics for $k = 50$. As can be seen in Figure 9 the graphics are similar. However, for the case of Applied Physics the maximum is attained for an approximate value of 38 – that corresponds to $n = 12$ – , while for the case of Mathematics, the maximum value is attained at 35 – that corresponds to $n = 15$ – , as we have shown in Section 3.2.

As the reader may notice the behavior for both cases is similar. So we can propose as a reasonable common value of b for both models $b = 0.28$. Taking into account that these ordered journal lists – mainly the one of Mathematics – are particularly unstable, we consider $b = 0.35$ as a suitable value in case that
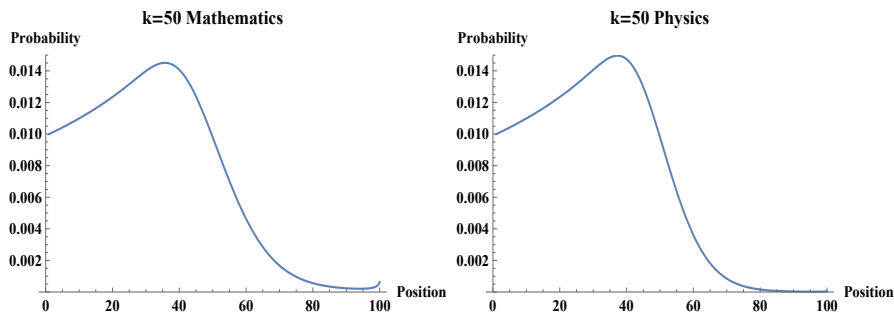
16

Figure 9: Model for $k = 50$ for Mathematics (left) and Applied Physics (right).

we have no other specific information on the list to which we can apply the model.

3.4 How can the method be applied by researchers

As we have explained in previous sections, the model can be adapted to the particular situations faced by an author, in terms of optimizing the dissemination of his work. We have mainly shown two different ways of solving the problem, the first using the personal experience of the author or his research institution in the publication of the results, and the second using a ranking obtained through a JCR list. However, other criteria can also be applied to find a ranking of journals, which would also include adequacy to the results of the author's research using, for example, a taxonomy approach such as that presented in Murtagh et al. (2018). Complementary information, e.g. the position of the researcher's ranking, can also be added to the model. This would be done by changing the distribution function, so that the probability of publication in a well-placed journal in an ordered list is higher when the author is well positioned, according to the taxonomy of his/her research area. It must be understood that our method provides an additional tool in a decision-making process that is complex and in which a large number of variables have to be taken into account.

Once the ranking is decided, the author must choose his own distribution function, which can be calculated using external rankings (such as the one provided by a JCR list), or his own ranking, calculated from his personal experience in publishing in the field of research. The reader can find a detailed algorithm of how to do it in the Appendix. We have used Mathematica for the calculations, but any other standard mathematical software, such as Matlab or R, can be used. The ranking information should be entered as a data matrix, and then a simple fitting procedure based on the model functions presented in the previous sections should be used.

17

4   Conclusions

We have faced the classical problem of the authors about deciding where to send a just finished research paper for publication. Of course, the first argument is related with the contents of the paper and the list of journals in which the article may be published according to the topic. The second argument is given by a self-evaluation of the authors in relation with the paper, what should give some feeling about the adequacy of a particular journal. Other aspects – for example, number of citation in the article to other papers in a given journals – , must also be considered.

Once these arguments have been applied, the authors might produce an ordered list – with the help for example of their personal experience in the research field, the Clarivate Analytics's 2-year impact factor or the Scopus SNIP – and apply our model.

As we have shown in Section 3.1, the first procedure for doing that is to construct the list just by using the expertise of the author himself/herself. In this case, he/she must analyze using its own publication statistics the increasing function of probability of acceptance and the variance of the "jumps distribution" for each journal. A degree-2 polynomial and the function $1/N_{max} - n + 1$ have been used in this example.

We have shown in Sections 3.2 and 3.3 that the ordered list can also be defined if we use the JCR list of Clarivate Analytics. In this case, an estimate of the value of the parameter b is needed; since the model does not pretend to give an exact answer and has a probabilistic meaning, a rough approximation should be enough. As an example, we have estimated a value $b = 0.236$ for a subset of 100 journals of the JCR list of the field Mathematics. The conclusion in this case is that, for publishing the paper in a journal that is in the first k positions after a year, the author must submit it to the journal $k - 15$, whenever the value of k is in the middle of the range (say $20 \leq k \leq 75$). In case the author pretends to publish the article in a top journal, he/she must submit it to any of the journals that are in the positions $k \leq 20$; in case he/she wants to publish in a low journal (say $k \geq 75$), any journal in the low part of the list would be a good candidate. A relatively (time-)stable impact list would imply that the value of b is bigger; an estimated range for its values would be $0.1 \leq b \leq 0.5$. In general, for the central values of k, the paper must be sent to a journal with an order number slightly smaller than k. Regarding the parameter of the model, as bigger the b, smaller the difference among k and the order number of the journal to which the article should be submitted.

Our model is just an attempt of giving a practical tool for the authors. As we have shown, the rule $k - 15$ for a 100-papers list of Mathematics improves the chance of publication. If a given list is more stable then the quantity 15 must be reduced: for example, $k - 10$ would give a reasonable value. *The main general conclusion is that the author should always submit his/her work to a journal that is in the first part of the list – with the limit defined by k – but "close" to k.*

Finally, let us remark that our general approach could be improved by using

advanced techniques to mix the probability functions that represent the different aspects concerning the model. In particular, copulas, aggregation functions or t-norms, in the context of fuzzy mathematics, would provide such tools (see Remark 2.1).

References

[1] ALTHOUSE, B. M., WEST, J. D., BERGSTROM, C. T., and BERGSTROM, T. (2009), "Differences in impact factor across fields and over time", *Journal of the Association for Information Science and Technology,* 60(1), 27-34.

[2] BLACK, S. (2012), "How much do core journals change over a decade?" *Library Resources and Technical Services,* 56, 80-93,

[3] BRADSHAW, C. J., and BROOK, B. W. (2016), "How to rank journals", *PloS one,* 11(3), e0149852.

[4] FERRER-SAPENA, A., DIAZ-NOVILLO, S. and SÁNCHEZ-PÉREZ, E.A. (2017), "Measuring Time-Dynamics and Time-Stability of Journal Rankings in Mathematics and Physics by Means of Fractional p-Variations", *Publications,* 5(3), 21.

[5] FERRER-SAPENA, A., SÁNCHEZ-PÉREZ, E. A., GONZÁLEZ, L. M., PESET, F. and ALEIXANDRE-BENAVENT, R. (2015), "Mathematical properties of weighted impact factors based on measures of prestige of the citing journals", *Scientometrics,* 105(3), 2089-2108.

[6] FERRER-SAPENA, A., SÁNCHEZ-PÉREZ, E. A., GONZÁLEZ, L. M., PESET, F. and ALEIXANDRE-BENAVENT, R. (2016), "The impact factor as a measuring tool of the prestige of the journals in research assessment in mathematics", *Research Evaluation,* 25(3), 306-314.

[7] GASTEL, B., and DAY, R. A. *How to write and publish a scientific paper,* ABC-CLIO, Santa Barbara, 2016.

[8] GIBBS, A. (2016), "Improving publication: advice for busy higher education academics", *International Journal for Academic Development,* 21(3), 255-258.

[9] M. GRABISCH, M., MARICHAL, J.L., MESIAR, R. and PAP, E. *Aggregation Functions,* Cambridge University Press, Cambridge, 2009.

[10] HAGHDOOST, A.; ZARE, M.; BAZRAFSHAN, A. (2014), "How variable are the journal impact measures?" *Online Information Review,* 38, 723-737.

[11] HOL, E.M. *Empirical studies on volatility in international stock markets,* Springer, Berlin, 2013.

[12] KELLY Jr, J. L. (1956), "A new interpretation of information rate", Manuscript. Reprinted in: *The Kelly Capital Growth Investment Criterion: Theory and Practice,* (2011), 25-34.

[13] KLEMENT, E.P., MESIAR, R. and PAP, E. *Triangular Norms,* Kluwer Academic Publishers, Dordrecht/Boston/London, 2000.

[14] KLEMENT, E.P., MESIAR, R. and PAP, E. (2001), " Uniform approximation of associative copulas by strict and non-strict copulas", Illinois J. Math. 45(4), 1393–1400.

[15] LAWRENCE, P.A. (2003), "The politics of publication", *Nature* 422.6929, 259-261.

[16] LU, J., WU, D., MAO, M., WANG, W., and ZHANG, G. (2015), "Recommender system application developments: a survey", *Decision Support Systems,* 74, 12-32.

[17] MANSILLA, R., KÖPPEN, E., COCHO, G., MIRAMONTES, P. (2007), "On the behavior of journal impact factor rank-order distribution", *Journal of Informetrics,* 1, 155-160.

[18] MURTAGH, F., ORLOV, M. and MIRKIN, B. (2018), "Qualitative judgement of research impact: Domain taxonomy as a fundamental framework for judgement of the quality of research." *Journal of Classification,* 1–24.

[19] NEFF, B. D., and OLDEN, J. D. (2010), "Not so fast: Inflation in impact factors contributes to apparent improvements in journal quality," *BioScience,* 60(6), 455-459.

[20] NELSEN, R.B. *An Introduction to Copulas,* Springer, New York, 1999.

[21] PAJIĆ, D. (2015), "On the stability of citation-based journal rankings," *Journal of Informetric,* 9, 990-1006.

[22] SHANNON, C. E. (1948), "A mathematical theory of communication", *The Bell System Technical Journal,* 27, 379-423. Reprinted in: *ACM SIGMOBILE Mobile Computing and Communications Review,* 5(1)(2001), 3-55.

[23] XU, H., MARTIN, E., and MAHIDADIA, A. (2014), "Contents and time sensitive document ranking of scientific literature", *Journal of Informetrics,* 8, 546-561.