Astronomy & Astrophysics

Crucial aspects of the initial mass function

I. The statistical correlation between the total mass of an ensemble of stars and its most massive star

M. Cerviño[1,2], C. Román-Zúñiga[3], V. Luridiana[2,4], A. Bayo[5,6], N. Sánchez[7], and E. Pérez[1]

[1] Instituto de Astrofísica de Andalucía (IAA-CSIC), Glorieta de la Astronomía s/n, 18008 Granada, Spain
e-mail: mcs@iaa.es
[2] Instituto de Astrofísica de Canarias, c/ vía Láctea s/n, 38205 La Laguna, Tenerife, Spain
[3] Instituto de Astronomía, Universidad Académica en Ensenada, Universidad Nacional Autónoma de México, Ensenada BC, 22860 Mexico, Mexico
[4] Departamento de Astrofísica, Universidad de La Laguna (ULL), 38205 La Laguna, Tenerife, Spain
[5] European Southern Observatory, Casilla 19001, Santiago 19, Chile
[6] Max Planck Institut für Astronomie, Königstuhl 17, 69117 Heidelberg, Germany
[7] S. D. Astronomía y Geodesia, Fac. CC. Matemáticas, Universidad Complutense de Madrid, 28040 Madrid, Spain

ABSTRACT

Context. Our understanding of stellar systems depends on the adopted interpretation of the initial mass function, IMF $\phi(m)$. Unfortunately, there is not a common interpretation of the IMF, which leads to different methodologies and diverging analysis of observational data.
Aims. We study the correlation between the most massive star that a cluster would host, m_{max}, and its total mass into stars, $\mathcal{M}$, as an example where different views of the IMF lead to different results.
Methods. We assume that the IMF is a probability distribution function and analyze the $m_{max} - \mathcal{M}$ correlation within this context. We also examine the meaning of the equation used to derive a theoretical $\mathcal{M} - \hat{m}_{max}$ relationship, $\mathcal{N} \times \int_{\hat{m}_{max}}^{m_{up}} \phi(m)\, dm = 1$ with $\mathcal{N}$ the total number of stars in the system, according to different interpretations of the IMF.
Results. We find that only a probabilistic interpretation of the IMF, where stellar masses are identically independent distributed random variables, provides a self-consistent result. Neither $\mathcal{M}$ nor the total number of stars in the cluster, $\mathcal{N}$, can be used as IMF scaling factors. In addition, $\hat{m}_{max}$ is a characteristic maximum stellar mass in the cluster, but not the actual maximum stellar mass. A $\langle \mathcal{M} \rangle - \hat{m}_{max}$ correlation is a natural result of a probabilistic interpretation of the IMF; however, the distribution of observational data in the $\mathcal{N}$ (or $\mathcal{M}$) $- m_{max}$ plane includes a dependence on the distribution of the total number of stars, $\mathcal{N}$ (and $\mathcal{M}$), in the system, $\Phi_{\mathcal{N}}(\mathcal{N})$, which is not usually taken into consideration.
Conclusions. We conclude that a random sampling IMF is not in contradiction to a possible $m_{max} - \mathcal{M}$ physical law. However, such a law cannot be obtained from IMF algebraic manipulation or included analytically in the IMF functional form. The possible physical information that would be obtained from the $\mathcal{N}$ (or $\mathcal{M}$) $- m_{max}$ correlation is closely linked with the $\Phi_{\mathcal{M}}(\mathcal{M})$ and $\Phi_{\mathcal{N}}(\mathcal{N})$ distributions; hence it depends on the star formation process and the assumed definition of stellar cluster.

Key words. stars: statistics – Galaxy: stellar content – methods: data analysis

1. Introduction

In recent literature, the term initial mass function (IMF) is used to indicate three different types of distributions: (1) the distribution by number of the stellar masses observed in a particular star ensemble; (2) a normalized version of (1), i.e., the frequency distribution of the stellar masses observed in a particular star ensemble; and (3) the theoretical probability density function $\phi(m)$ of the stellar masses that can be formed in a generic star ensemble. In this work, following Scalo (1986), we adopt the third definition and explore some consequences of mixing these definitions.

In the following, we leave distribution (2) out of the discussion and focus, for simplicity, only on distributions (1) and (3)[1]. These two distributions are different but closely related to each

other, as statistics and probability are. Probability deals with predicting the likelihood of possible events in a system with known properties; statistics consists in analysing the distribution of real events with the aim of determining some unknown property of the system. Probability addresses the direct problem, while statistics addresses the inverse problem. In our case, distribution (3) describes the underlying probability distribution from which stellar masses can be drawn, while distribution (1) describes an actual stellar sample from which we wish, ideally, to recover the parameters of the underlying probability distribution.

The relation between the shape of (1) and the shape of (3) depends crucially on the size of the sample, that is, the number of stars $\mathcal{N}$; when $\mathcal{N}$ values are large, the two shapes tend to be similar. This similarity can mislead one into believing that (1) is just a scaled-up version of (3), with $\mathcal{N}$ being the scale factor. This would be very wrong since, as explained above, the physical meanings of both distributions are intrinsically different.

---

[1] However, because distribution (2) is an scaled version of distribution (1), the conclusions derived from (1) also apply to (2).

This paper is dedicated to exploring the implications of such difference.

A major drawback of the distribution-by-number view (number (1) above) is that the very definition of a stellar sample necessarily implies some (hidden or explicit) assumption on the star formation (SF) process that originated the sample. For example, an embedded, open, or globular cluster, an OB associations, and so on, are coeval and cospatial samples; field stars, which are used to study galaxy structure, are neither coeval nor cospatial; the stars in a galaxy that were born at a given time, which are a sample suitable for stellar populations studies, are coeval but not cospatial. These examples make clear that, when a sample is selected, some predefined spatial and time scales are implicitly assumed, and these scales may influence the distribution by the number of the stellar masses. Rephrasing Scalo (1986), when talking about the IMF, we are left in the uncomfortable position of having no means to define an empirical sample that corresponds to a consistent definition of IMF and that can be directly related to the theories of SF without introducing major assumptions.

The probability distribution function (PDF) view (number (3) above) is actually an abstraction used to describe the general universe of initial masses that a star would have. This interpretation implies that we have to use a probability framework in order to make a description of the problem and inferences from observed data sets. One implicit requirement of such an approach is that the stellar mass is an identically independent distributed (iid) variable, and therefore, any realization of the IMF is a random sample[2]. Within this framework, all the empirical samples are included naturally as far as they are particular realizations of the theoretical distribution. Although it is possible to include conditions representing particular SF scenarios, it is generally assumed that the IMF has no memory of the SF event: that is, the SF details have no major impact on the IMF itself, although they can have an impact on the resulting IMF realization once the corresponding conditions are included in the derivation. It is a surprising fact that there is no clear observational evidence that the IMF varies strongly and systematically as a function of different SF scenarios (Bastian et al. 2010).

Throughout this paper, we consider several pieces of work based on a distribution-by-number interpretation of the IMF. The specific way in which the IMF is represented varies depending on the considered paper. Some authors assume that the IMF is a continuous law that returns, for each mass value, the number of stars of that mass; others consider that it returns the number of stars in each mass bin. Some assume that the stars are distributed in a predefined way and the mass of a star depends on the mass of the other stars; others consider that the stars are distributed independently from each other. In the following, we give examples of this and emphasize the differences between the various distribution-by-number interpretations and the PDF view of the IMF.

Naturally, the equations involving the IMF depend on the interpretation of the IMF. More importantly however, the cluster-related quantities inferred from manipulations of the IMF are interpreted differently according to the initial assumptions. One case in which the different views of the IMF lead to dramatically diverging interpretations is the modeling of the correlation between the total stellar mass in a cluster, $\mathcal{M}$, and the mass, $m_{\max}$,

of its most massive star, which we investigate in this series of papers.

There are many facets to the study of the $\mathcal{M} - m_{\max}$ correlation. One is the correlation obtained theoretically from manipulations of the IMF functional form, which is the subject of this paper. Another is the inference of $\mathcal{M}$ from partial information of the system. The lack of information makes this inference deeply dependent on the IMF interpretation (this aspect is discussed in Cerviño et al. 2013, hereafter Paper II). A third issue is the comparison between theory and observational data. This point also depends on the interpretation of the IMF (and is studied in Jimenez-Donaire et al., in prep., from now on Paper III)

The structure of the paper is as follows: in Sect. 2 we present our basic framework for a probabilistic interpretation of the IMF. Section 3 is devoted to analyzing in a probabilistic context the meaning of the basic equation commonly used in the literature relating $\mathcal{M}$ and $m_{\max}$. In Sect. 4 we discuss the different methodologies and assumptions used by other authors to obtain a $\mathcal{M} - m_{\max}$ correlation. We include a discussion on iid stellar masses and on the connection of the IMF with the SF. Finally, we briefly discuss the composition of different IMFs to obtain an integrated galaxy IMF (IGIMF). Our conclusions are described in Sect. 5.

2. Formal probabilistic formulation

Let us start by framing the problem in a formal probabilistic framework:

1. The IMF, $\phi(m) = \mathrm{d}N/\mathrm{d}m$, is a PDF, that provides the probability of finding a star in a given mass range by its integration in such mass range. The mass limits of the PDF, m_{low} and m_{up}, are given by stellar theory and must fulfill $\int_{m_{\mathrm{low}}}^{m_{\mathrm{up}}} \phi(m)\mathrm{d}m = 1$; that is, we are certain that any possible star has a mass between m_{low} and m_{up}. This is the first fundamental difference with respect to the distribution-by-number interpretation: the IMF cannot be arbitrarily normalized to $\mathcal{M}$ or $\mathcal{N}$, since it does not provide numbers of stars with a given mass but the probability for a star to be born with a given mass independently of how many stars are in the cluster or the cluster total mass. In this interpretation of the IMF, there is neither an implicit sample nor predefined space or time scales.

The IMF so defined may have values larger than one, provided its integral over any mass range is lower than one. This is the second fundamental difference with respect to the distribution-by-number interpretation when described in terms of frequencies (case 2 in the Introduction) where no value larger than one is possible by construction.

In this paper we use the Kroupa IMF (Kroupa 2001, 2002) as used in Weidner & Kroupa (2006)[3] and subsequent works, except for the value of m_{up} which we set equal to 120 $M_{\odot}$. Although a larger value would probably be more realistic according to recent studies (Crowther et al. 2010, see also the contributions to the Up2010 conference published by Treyer et al. 2011), this choice is motivated by the fact that the m_{up} value of most public stellar tracks used in most $m_{\max}$ estimations is 120 $M_{\odot}$. In Fig. 1 we show the $\phi(m)$ used in this

---

[2] Random sample means that every possible sample has a calculable chance of selection. This is a requirement of any statistical and probabilistic study (Kendall & Stuart 1977).

[3] We note that Weidner & Kroupa (2004) use $\alpha_2 = 2.30$ in their parametrization of the IMF and that Weidner & Kroupa (2006) use $\alpha_2 = 2.35$.
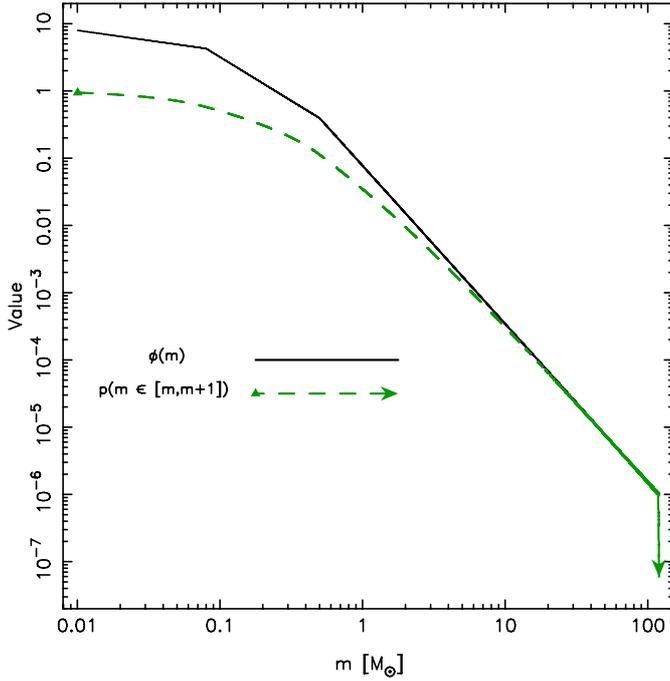
Fig. 1. IMF used in the present work (solid line), as in the parametrization by Kroupa (2001, 2002) and Weidner & Kroupa (2006). Being a PDF, it can have values larger than one; the probabilities are given by the integral over the PDF. We also plot the probability that a star has a mass in the $m, m+1\ M_\odot$ range, which is lower than one (dashed line). This probability declines rapidly when m is larger than $m_{\mathrm{up}} - 1\ M_\odot$.

paper and the probability for a star of having a mass in the range $m, m+1\ M_\odot$.

The probability for a random star of having a mass lower than a given value m_{a} is given by

$$p(m < m_{\mathrm{a}}) = \int_{m_{\mathrm{low}}}^{m_{\mathrm{a}}} \phi(m)\,\mathrm{d}m, \qquad (1)$$

while the probability for a random star of having a mass equal to or larger than m_{a} is given by

$$p(m \geq m_{\mathrm{a}}) = \int_{m_{\mathrm{a}}}^{m_{\mathrm{up}}} \phi(m)\,\mathrm{d}m. \qquad (2)$$

In this work, the integrals over the IMF will always be read as equal to or larger than the lower limit and lower than the upper limit. The use of lower than instead of equal to or lower than in the upper limit and the complementary in the lower limit is just a convention. However, equal cannot be used simultaneously in both equations: no star can simultaneously belong to two independent intervals. The convention we use implies that the nominal value m_{up} cannot be formally reached, although values very close to it are possible.

2. Different observational scenarios can be described by adding constraints to the IMF. For instance, we may explicitly include the limit imposed on $m_{\max}$ by the total mass of the sample we are analyzing, that is, $m_{\max} = \min\{m_{\mathrm{up}}, \mathcal{M}\}$. In this case, we must define an a posteriori PDF, related to the IMF, that includes such a condition:

$$\phi(m|m < m_{\max}) = \frac{\phi(m)\,\mathrm{H}(m_{\max} - m)}{p(m < m_{\max})}, \qquad (3)$$

where $\mathrm{H}(m_{\max} - m)$ is the Heaviside function[4], which ensures that no star equal to or larger than $m_{\max}$ can be present in the cluster. We note that $\phi(m|m < m_{\max})$ is also a PDF. The mean mass of such distribution is

$$\langle m|m < m_{\max}\rangle = \frac{\int_{m_{\mathrm{low}}}^{m_{\mathrm{up}}} m\,\phi(m)\,\mathrm{H}(m_{\max} - m)\mathrm{d}m}{p(m < m_{\max})}. \qquad (4)$$

More elaborated constrained-IMF can be formulated, always keeping in mind that conditions are imposed ad hoc and produce a PDF whose functional form differs from $\phi(m)$.

3. The PDF describing ensembles with a total number of stars $\mathcal{N}$ (formally conditioned to have $\mathcal{N}$ stars) can be calculated as successive convolutions of the corresponding PDF for one star. For instance, the PDF for the total mass, $\Phi_{\mathcal{M}}(\mathcal{M}|\mathcal{N})$, is the result of convolving the IMF $\mathcal{N}$ times with itself (see Cerviño & Luridiana 2006; Selman & Melnick 2008):

$$\Phi_{\mathcal{M}}(\mathcal{M}|\mathcal{N}) = \overbrace{\phi(m) \otimes \phi(m) \otimes\ ....\ \otimes \phi(m)}^{\mathcal{N}}. \qquad (5)$$

A property of self-convolution is that simple relations link the mean value and the high-order moments of $\phi(m)$ and $\Phi_{\mathcal{M}}(\mathcal{M}|\mathcal{N})$ (see, e.g., Cerviño & Luridiana 2006). As an example, the mean integrated mass of $\Phi_{\mathcal{M}}(\mathcal{M}|\mathcal{N})$, $\langle\mathcal{M}|\mathcal{N}\rangle$, is related to the mean stellar mass of the IMF, $\langle m\rangle$, through the relation

$$\langle\mathcal{M}|\mathcal{N}\rangle = \mathcal{N} \times \langle m\rangle = \mathcal{N} \times \int_{m_{\mathrm{low}}}^{m_{\mathrm{up}}} m\,\phi(m)\,\mathrm{d}m. \qquad (6)$$

However, we note that $\Phi_{\mathcal{M}}(\mathcal{M}|\mathcal{N}) \neq \mathcal{N} \times \phi(m)$ and that the actual total mass cannot be obtained, but only an estimate of it. This is the third fundamental difference with the distribution-by-number interpretation, which assumes that for a given $\mathcal{N}$ there is one, and only one, $\mathcal{M}$ value, given by $\mathcal{M}(\mathcal{N}) = \mathcal{N} \times \langle m\rangle$.

3. Relating the number of stars with the most massive star in the sample

According to the law of large numbers, in a sample of $\mathcal{N}$ stars drawn from an underlying PDF, $\phi(m)$, the typical number of stars N_{a} with $m \geq m_{\mathrm{a}}$ is given by $N_{\mathrm{a}} = \mathcal{N} \times p(m \geq m_{\mathrm{a}})$. Particularizing this equation, we can define a characteristic maximum value of $m_{\max}$, $\hat{m}_{\max}$, for which there is typically only one star with mass equal to or larger than $\hat{m}_{\max}$ through

$$1 = \mathcal{N} \times p(m \geq \hat{m}_{\max}) = \mathcal{N} \times \int_{\hat{m}_{\max}}^{m_{\mathrm{up}}} \phi(m)\,\mathrm{d}m. \qquad (7)$$

This is the basic equation used by several authors as the determination of the actual mass of the most massive star in a system (as examples: Elmegreen 1997, 1999, 2000; Kroupa & Weidner 2003; Weidner & Kroupa 2004, 2006). However, we can also obtain a mean value of $m_{\max}$ (Oey & Clarke 2005) or a median value of $m_{\max}$ (Weidner et al. 2010). So the question is: does the definition of the characteristic value $\hat{m}_{\max}$ indeed provide the actual $m_{\max}$ extreme value or only an estimate of it? And if it is an

---

[4] We use here the Heaviside function as a distribution to define the domain of $\phi(m)$, including constraints. In this situation the value of $\mathrm{H}(0)$ is not defined, but it is assigned a posteriori to be consistent with the convention used in the integral limits. In the case of Eq. (3), $\mathrm{H}(0) = 0$.
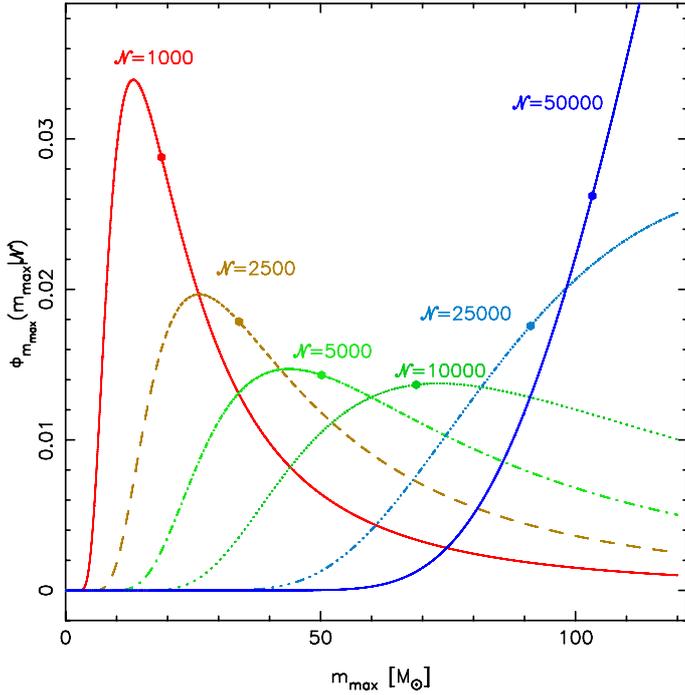
Fig. 2. Distribution of the maximum stellar mass, $\Phi_{m_{max}}(m_{max}|\mathcal{N})$ for different values of $\mathcal{N}$. The circle on each curve is the position of the characteristic value $\hat{m}_{max}$.
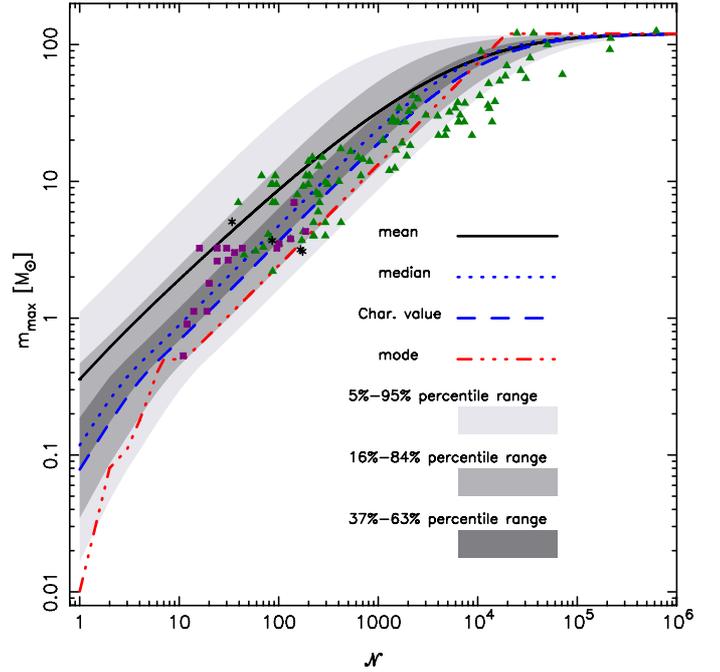

Fig. 3. Percentile analysis around the median of $\Phi_{m_{max}}(m_{max}|\mathcal{N})$ as a function of $\mathcal{N}$ (shaded areas). The figure includes as a reference the position of the characteristic value, median, mean, and mode of the distribution. Small triangles: compilation by Weidner et al. (2010) of observational values of m_{max} and inferred values of $\mathcal{N}$ obtained from observations; squares: observed values of $\mathcal{N}$ and m_{max} from Kirk & Myers (2011); stars: observed values of $\mathcal{N}$ and m_{max} in the field for the four observed regions from Kirk & Myers (2011).

estimate, what is its exact meaning? Let us seek the answer in a probabilistic context[5].

We consider a set of $\mathcal{N}$ stars with unknown stellar masses, m_i, drawn from the IMF. For any given mass m_a, the probability of having at least one star with mass m_i equal to or larger than m_a in the sample, $\mathcal{P}(\exists i \in [1, \mathcal{N}] \,|\, m_i \geq m_a)$, is the complementary probability that all stars have a mass lower than m_a, $\mathcal{P}(m_i < m_a, \forall i \in [1, \mathcal{N}])$. Since the stellar masses are iid drawn from the same distribution $\phi(m)$, the probability $\mathcal{P}(m_i < m_a, \forall i \in [1, \mathcal{N}])$ is the result of multiplying $p(m < m_a)$ by itself $\mathcal{N}$ times[6]:

$$\mathcal{P}(m_i < m_a, \forall i \in [1, \mathcal{N}]) = [p(m < m_a)]^{\mathcal{N}}$$
$$= [1 - p(m \geq m_a)]^{\mathcal{N}}. \qquad (8)$$

Thus,

$$\mathcal{P}(\exists i \in [1, \mathcal{N}] \,|\, m_i \geq m_a) = 1 - \mathcal{P}(m_j < m_a, \forall j \in [1, \mathcal{N}])$$
$$= 1 - [1 - p(m \geq m_a)]^{\mathcal{N}}. \qquad (9)$$

This relation is valid for any value of m_a and any distribution function.

If we now set $m_a = \hat{m}_{max}$, we can replace $p(m \geq \hat{m}_{max})$ in Eq. (9) by $1/\mathcal{N}$ by virtue of the $\hat{m}_{max}$ definition. The probability that there is at least one star with $m \geq \hat{m}_{max}$ in a sample of $\mathcal{N}$ stars is thus given by

$$\mathcal{P}(\exists i \in [1, \mathcal{N}] \,|\, m_i \geq \hat{m}_{max}) = 1 - \left[1 - \frac{1}{\mathcal{N}}\right]^{\mathcal{N}}, \qquad (10)$$

which has an asymptotic value $1 - 1/e \sim 0.63$ for large $\mathcal{N}$ values, with 0.63 being a reasonable approximation for, say, $\mathcal{N} > 100$.

Hence, the characteristic mass, $\hat{m}_{max}$, obtained by solving Eq. (7) is the value of m that is not reached or exceeded[7] with a probability 0.37 in a sample of $\mathcal{N}$ stars. This means that in a large enough set of clusters, all of them with $\mathcal{N}$ stars, typically in 63% of the clusters the mass of the most massive star will be equal to or larger than $\hat{m}_{max}$, while in 37% of the clusters it will be lower than $\hat{m}_{max}$. So the $\hat{m}_{max}$ value obtained in Eq. (7) does not provide the mass m_{max} of the most massive star in a cluster of $\mathcal{N}$ stars, contrary to what is stated in several astrophysical papers[8].

Actually, for any possible value $\hat{m}_{max}$ lower than m_{up} that we would use as a proxy of the actual value of m_{max}, there is a probability larger than 90% that the most massive star in the system is more massive than such $\hat{m}_{max}$ value (see Appendix A for details).

3.1. The PDF of m_{max} for a known $\mathcal{N}$, $\Phi_{m_{max}}(m_{max}|\mathcal{N})$

Actually, there is no unique value of m_{max} for a total number of stars $\mathcal{N}$, but the possible values of m_{max} are distributed following

---

[5] The discussion in this section is mainly based on Sornette (2004), Kendall & Stuart (1977), and Gumbel (1958), although the same formulae can be found in other works.

[6] Here we use p to represent probabilities on the IMF (cf., Eqs. (1) and (2)) and $\mathcal{P}$ to represent probabilities on the sample with $\mathcal{N}$ stars.

[7] We note that, depending on the reference and the convention used in Sect. 2, this value can be defined either as reached or exceeded or just as exceeded.

[8] The characteristic largest value defined by Eq. (7) is related to the estimation of the number of events we must record to have an event larger than a given value m_a (which is called return period in extreme value theory). If the events are taken in a regular time interval, for instance, it could be the estimation of the number of years between earthquakes larger than a given magnitude, the number of years between economy crashes, and so on.
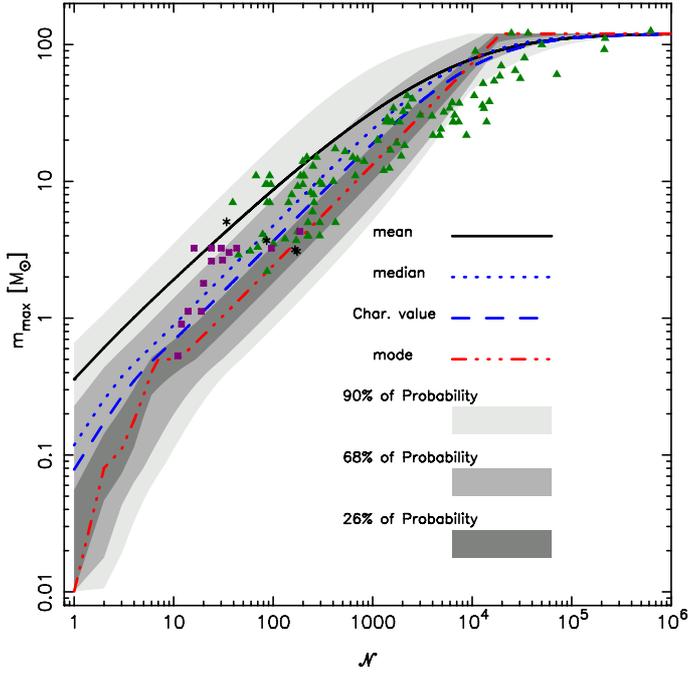
Fig. 4. Confidence interval analysis of $\Phi_{m_{max}}(m_{max}|\mathcal{N})$ as a function of $\mathcal{N}$ (shaded area). Lines and symbols have the same meaning as in Fig. 3.
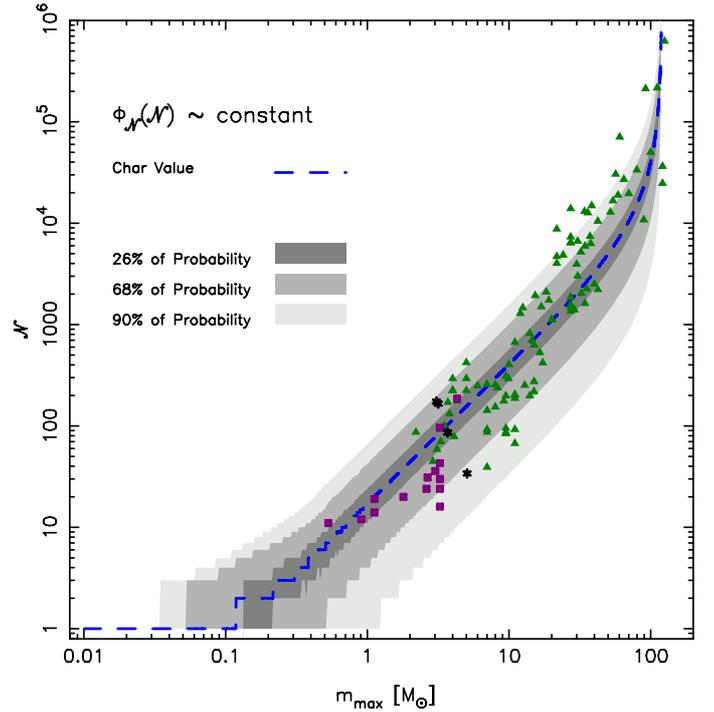

Fig. 5. Confidence interval analysis of $\Phi_{\mathcal{N}}(\mathcal{N}|m_{max})$ as a function of m_{max} for a $\Phi_{\mathcal{N}}(\mathcal{N})$ = constant. Symbols have the same meaning as in Fig. 3.

the probability function

$$\Phi_{m_{max}}(m_{max}|\mathcal{N}) = \mathcal{N}\,\phi(m_{max})\,p(m < m_{max})^{\mathcal{N}-1} \qquad (11)$$

$$= \mathcal{N}\,\phi(m_{max})\left(\int_{m_{low}}^{m_{max}} \phi(m)\,\mathrm{d}m\right)^{\mathcal{N}-1}, \qquad (12)$$

as deduced by Gumbel (1958), Sornette (2004), van Albada (1968), Oey & Clarke (2005), Maschberger & Clarke (2008), Pflamm-Altenburg & Kroupa (2008), among others.

In Fig. 2 we show the distribution $\Phi_{m_{max}}(m_{max}|\mathcal{N})$ for different values of $\mathcal{N}$. The circle on each PDF corresponds to the position of the characteristic value $\hat{m}_{max}$, which divides the PDF in two areas: the left one containing the 37% of the probability and the right one containing the 63% of the probability. We note that $\Phi_{m_{max}}(m_{max}|\mathcal{N})$ is highly asymmetrical. Given the shape of the distribution, it cannot be described only by their parameters (mean, variance, and so on); we must consider the whole distribution for any comparison with the observational data. This can be done in two ways, by a percentile analysis (analysis around the median) and by a confidence interval analysis around the mode[9] (the maximum value of the distribution, which is related to the most common value obtained in a set of observations).

Figure 3 shows a percentile analysis of the distribution. The figure also includes the position of the mean, mode, and characteristic values of the distribution for reference. The position of the mean, $\langle m_{max}|\mathcal{N}\rangle$, mostly falls between the 63% and 84% percentile, i.e., far from the median of the distribution. On the other hand, $\hat{m}_{max}$ corresponds, as predicted, to the 37% percentile. Finally, the mode of the distribution lies in the lowest percentile range. The figure also shows the $(m_{max}, \mathcal{N})$ values compiled by Weidner et al. (2010), in which m_{max} is determined from observations and $\mathcal{N}$ is inferred from star counting in a given mass

range[10]. It also shows the data from Kirk & Myers (2011), who quote the observed masses of individual stars of 14 young stellar groups in four different regions (m_{max}, $\mathcal{N}$, and $\mathcal{M}$ were obtained from their tabulated data). We also show the corresponding m_{max} and $\mathcal{N}$ values of field stars in each region analyzed by Kirk & Myers (2011), which are in agreement with the general trend of the correlation.

The confidence interval around the mode analysis takes into account the distribution shape and the range of probability of any region in the diagram. This is done by sorting the contributions to the probability in decreasing order and finding the m_{max} range that contains some specified amount of probability. Different confidence intervals are obtained by adding the sorted probabilities, taking into account their associated m_{max} values. This methodology is extensively used in the analysis of redshifts in photometric surveys (see Fernández-Soto et al. 2002, for more details). The situation is illustrated in Fig. 4, which includes the 90, 68, and 26% confidence intervals.

3.2. The PDF of $\mathcal{N}$ for a known m_{max}, $\Phi_{\mathcal{N}}(\mathcal{N}|m_{max})$

In Sect. 3.1 we discussed the estimation of m_{max}, given the number of stars $\mathcal{N}$. Alternatively, we can also investigate the opposite case, the estimation of $\mathcal{N}$ from a known m_{max} (that is, the determination of the $\Phi(\mathcal{N}|m_{max})$ distribution). To address this problem, we can use the Bayes' theorem:

$$\Phi_{\mathcal{N}}(\mathcal{N}|m_{max}) = \frac{\Phi_{m_{max}}(m_{max}|\mathcal{N})\,\Phi_{\mathcal{N}}(\mathcal{N})}{\int \Phi_{m_{max}}(m_{max}|\mathcal{N})\,\Phi_{\mathcal{N}}(\mathcal{N})\,\mathrm{d}\mathcal{N}}. \qquad (13)$$

---

[9] The analyses based on the parameters of the distribution, on the percentile, and on confidence intervals around the mode are equivalent only in the Gaussian case, where 1σ is almost equivalent to the percentile range 16−84% and the 68% confidence interval.

[10] Except in a few cases, Weidner & Kroupa (2004) and Weidner et al. (2010) obtain $\mathcal{N}$ by extrapolating to the full IMF range the number of stars N_a observed above a specified mass or within a specified mass range. Then, $\mathcal{M}$ is obtained by means of $\mathcal{M} = \mathcal{N} \times \langle m\rangle$. We obtained the plotted $\mathcal{N}$ values by division of the $\mathcal{M}$ values quoted in their tables by $\langle m\rangle$.
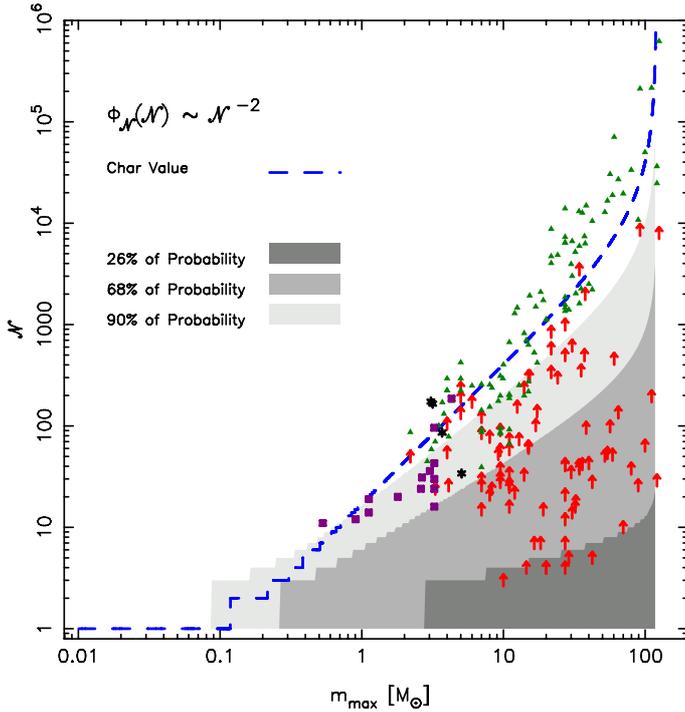
Fig. 6. Confidence interval analysis of $\Phi_{\mathcal{N}}(\mathcal{N}|m_{\max})$ as a function of $m_{\max}$ for a $\Phi_{\mathcal{N}}(\mathcal{N}) \propto \mathcal{N}^{-2}$. Arrows: data points by Weidner et al. (2010) using $\mathcal{N}_{\mathrm{obs}}$ without correction of incompleteness due to unobserved stars. Other symbols have the same meaning as in Fig. 3.

We know all terms on the right-hand side of this equation, except $\Phi_{\mathcal{N}}(\mathcal{N})$, which is the probability of having a system with a given total number of stars, i.e., an initial number-of-stars-per-cluster function (an initial cluster number function, ICNF). If $\Phi_{\mathcal{N}}(\mathcal{N})$ is a power-law distribution in a similar fashion to the initial cluster mass function (ICMF), $\Phi_{\mathcal{N}}(\mathcal{N}) = A\mathcal{N}^{-\beta}$ with A a normalization value, we find

$$\Phi_{\mathcal{N}}(\mathcal{N}|m_{\max}) = A'\, p(m < m_{\max})^{\mathcal{N}-1}\, \mathcal{N}^{1-\beta}, \tag{14}$$

where A' is a normalization value that includes A.

The mode of $\Phi_{\mathcal{N}}(\mathcal{N}|m_{\max})$, $\mathcal{N}^{\mathrm{mode}}$, is obtained by equaling to zero its first derivative with respect to $\mathcal{N}$, which yields[11]

$$\mathcal{N}^{\mathrm{mode}} \approx \frac{\beta - 1}{\ln p(m < m_{\max})}. \tag{15}$$

This equation has an acceptable solution only for $\beta < 1$; in particular, for a flat distribution of $\mathcal{N}$ (i.e., $\beta = 0$) the result is approximately $1/p(m \geq m_{\max})$. This justifies the name of $\hat{m}_{\max}$ as the characteristic value, since it provides $\mathcal{N}^{\mathrm{mode}}$ as a function of the most extreme value of the distribution under the hypothesis of a flat $\Phi_{\mathcal{N}}(\mathcal{N})$[12]. In Fig. 5 we plot the confidence intervals of the $\Phi_{\mathcal{N}}(\mathcal{N}|m_{\max})$ distribution as a function of $m_{\max}$. We note that

---

[11] $\mathcal{N}$ is not a continuous variable; hence it cannot have been derivated and $\mathcal{N}^{\mathrm{mode}}$ must be an integer number. Thus, the formulae provide only an approximation.

[12] In Paper II we show that this assumption is implicit when $\mathcal{N}$ is inferred from the number N_{a} of massive stars in the $(m_{\max}, m_{\mathrm{a}})$ range by using the relation $\mathcal{N} = N_{\mathrm{a}} \times p(m \geq m_{\mathrm{a}})$. Similarly, the assumption is implicit when $\mathcal{M}$ is inferred by multiplying the mean stellar mass by $\mathcal{N}$; it is a general assumption found in the literature and, in particular, is the method used to infer $\mathcal{M}$ in the Weidner et al. (2010) compilation.

the axes of the plot have changed with respect to the figures in the previous section, since $m_{\max}$ is now the variate. We also plot the data points from Weidner et al. (2010) and Kirk & Myers (2011).

However, Eq. (15) results in a negative value without astrophysical meaning if the ICNF is similar to the ICMF; $\Phi_{\mathcal{N}}(\mathcal{N}|m_{\max})$ is a decreasing function for all $\mathcal{N}$, and the most probable $\mathcal{N}$ corresponds to the maximum of $\Phi_{\mathcal{N}}(\mathcal{N})$, i.e., the lower limit of the $\Phi_{\mathcal{N}}(\mathcal{N})$ distribution. Hence, $\Phi_{\mathcal{N}}(\mathcal{N})$ modifies the confidence interval analysis of $\Phi_{\mathcal{N}}(\mathcal{N}|m_{\max})$, as shown in Fig. 6.

It seems surprising that, depending the independent variable used ($m_{\max}$ or $\mathcal{N}$), one has to take into account $\Phi_{\mathcal{N}}(\mathcal{N})$. Where is the $\Phi_{\mathcal{N}}(\mathcal{N})$ dependence in Figs. 3 and 4? Actually, we must be aware that Figs. 3−6 are not representations of $\Phi_{\mathcal{N},m_{\max}}(\mathcal{N}, m_{\max})$, which would be the one to be compared with observational data. Instead, they are a representation of the probability for fixed values in the x-axis, i.e., the figures can be only interpreted making vertical (discrete or infinitesimal) slices. Hence, for comparison with data, the x-axis on Figs. 3 and 4 must be weighted by $\Phi_{\mathcal{N}}(\mathcal{N})$, and the x-axis on Figs. 5 and 6 must be weighted by $\phi(m)$. Obviously, such a weight process changes the probability density in the $\mathcal{N} - m_{\max}$ plane.

3.3. Which information does the $\mathcal{N}$ (or $\mathcal{M}$) − $m_{\max}$ plane contain?

All the quantities considered here, $m_{\max}$, $\mathcal{N}$, and $\mathcal{M}$, have their own distributions, $\phi(m)$, $\Phi_{\mathcal{N}}(\mathcal{N})$, and $\Phi_{\mathcal{M}}(\mathcal{M})$. So, any uncertainty of data points in the $\mathcal{N}$ (or $\mathcal{M}$) − $m_{\max}$ plane would be minimized or amplified by such distributions, and neither $\Phi_{m_{\max}}(m_{\max}|\mathcal{N})$ nor $\Phi_{\mathcal{N}}(\mathcal{N}|m_{\max})$ (or their $\mathcal{M}$ counterparts) are suitable descriptions. The only suitable distribution of data points is given by $\Phi_{m_{\max},\mathcal{N}}(m_{\max}, \mathcal{N})$[13] (or their $\mathcal{M}$ counterpart, see below). This PDF is shown in Fig. 7 for the case of a $\Phi_{\mathcal{N}}(\mathcal{N}) \propto \mathcal{N}^{-2}$. However, the use of $\Phi_{m_{\max},\mathcal{N}}(m_{\max}, \mathcal{N})$ imposes some important caveats.

The first of these caveats affects any test on the $\mathcal{N}$ (or $\mathcal{M}$) − $m_{\max}$ correlation. Such a test can only be done at a distribution level and not in a data-point-by-data-point analysis. This means that we need a quantitative characterization of the uncertainty associated to each data point and must combine the corresponding uncertainties to obtain a density map in the $\mathcal{N}$ (or $\mathcal{M}$) − $m_{\max}$ plane.

The second caveat refers to the plane to be used: $\mathcal{N} - m_{\max}$ or $\mathcal{M} - m_{\max}$? It includes two different aspects. The first is that any $\mathcal{M}$ inference implicitly includes an $\mathcal{N}$ inference, and in most of the cases (all where $\langle m \rangle$ is used), it is actually an $\mathcal{N}$ inference itself but expressed as $\langle \mathcal{M} \rangle$ (i.e., the plane to be used is actually $\mathcal{N} - m_{\max}$). The second aspect is that the distribution of data points in the $\mathcal{N} - m_{\max}$ plane includes $\phi(m)$ and $\Phi_{\mathcal{N}}(\mathcal{N})$ and the distribution of data points in the $\mathcal{M} - m_{\max}$ plane also includes $\Phi_{\mathcal{M}}(\mathcal{M})$. This means that some hypothesis about the relation between $\mathcal{N}$ and $\mathcal{M}$ is always required when the $\mathcal{M} - m_{\max}$ plane is used.

We conclude this section with a brief discussion about the falsification of the random sampling of the IMF claimed by Weidner et al. (2010) in view of the results presented here, that

---

[13] That is:

$$\begin{aligned} \Phi_{m_{\max},\mathcal{N}}(m_{\max}, \mathcal{N}) &= \Phi_{m_{\max}}(m_{\max}|\mathcal{N})\,\Phi_{\mathcal{N}}(\mathcal{N}) \\ &= \Phi_{\mathcal{N}}(\mathcal{N}|m_{\max})\,\phi(m_{\max}). \end{aligned}$$
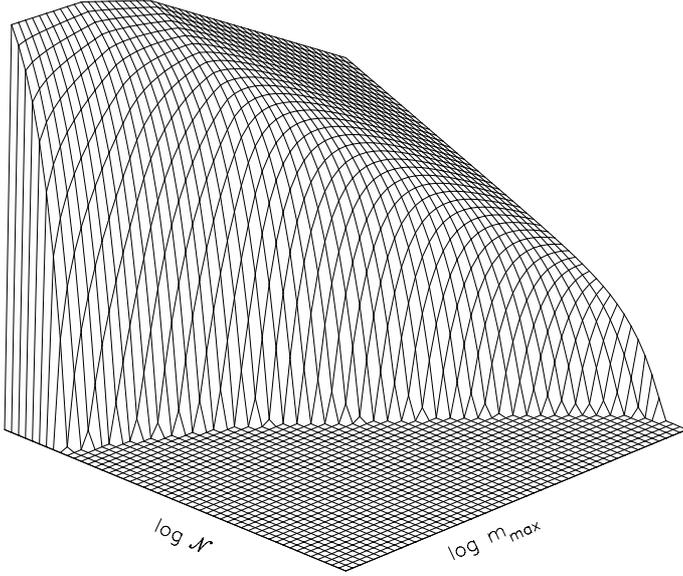
Fig. 7. 3D representation of $\log \Phi_{m_{max}, \mathcal{N}}(m_{max}, \mathcal{N})$ distribution for a $\Phi_{\mathcal{N}}(\mathcal{N}) \propto \mathcal{N}^{-2}$.

is, the dependence on $\Phi_{\mathcal{N}}(\mathcal{N})$ and $\Phi_{\mathcal{M}}(\mathcal{M})$ in the distribution of data points in the $\mathcal{N}$ (or $\mathcal{M}$) − m_{max} plane.

First, random sampling is an axiom in statistics and probability. It is not a hypothesis. Statistical tests evaluate the compatibility of a hypothetical distribution with a given sample. There can be two main reasons for the incompatibility of both entities: (a) the assumed distributions are not a correct representation of the sample; (b) the sample is biased or not randomly chosen. In the present case, the hypothesized distributions are the IMF, the ICNF, and the ICMF, where the ICMF and the ICNF are linked not trivially by Eq. (5). We would assume a universal IMF, but still need an ICMF (or ICNF) characterization. The very definition of the ICMF (or ICNF) leads to an uncomfortable situation similar to the case of the IMF: we have no means of defining an empirical sample that can be directly related to SF theories without introducing a major assumption, that is, the cluster definition. Can a single star be considered as a valid cluster? How do we define a single cluster formation event in a giant molecular cloud? Is there a difference between the ICMF defined over a random set of clusters and the one defined over a group of clusters that would have a common origin in a large-scale star-forming event?

Hence, the results obtained by Weidner et al. (2010) can be interpreted in different ways:

- The clusters in the sample do not follow the assumed IMF.
- The clusters in the sample do not follow the assumptions about the ICMF or ICNF.
- The sample is biased due to selection effects (including the definition of what a cluster is).
- The sample is incomplete, so no conclusions about the preceding items can be obtained.

We will discuss these issues in more detail in Papers II and III.

4. Discussion

In the previous sections we have established the formal probabilistic interpretation of the IMF and the propagation of this interpretation in the correlation between m_{max} and $\mathcal{N}$. We can now explore the implications of such an interpretation and (a) compare it with the implications of concurrent interpretations

(Sect. 4.1); and (b) discuss the random-sampling assumption of this work and its implications for the relation between the IMF and the SF (Sect. 4.2).

4.1. Literature on the $\mathcal{M} - m_{max}$ and the $\mathcal{N} - m_{max}$ correlations

There are copious studies related to the existence and modeling of a $\mathcal{M} - m_{max}$ correlation (for instance, Reddish 1978; Larson 1982; Vanbeveren 1982; García-Vargas & Díaz 1994; García-Vargas et al. 1995; Elmegreen 1997, 1999, 2000; Larson 2003; Kroupa & Weidner 2003; Weidner & Kroupa 2004; Oey & Clarke 2005; Weidner & Kroupa 2006; Parker & Goodwin 2007; Selman & Melnick 2008; Maschberger & Clarke 2008; Weidner et al. 2010; Kroupa et al. 2011). Some of these articles give an explicit formulation of this relation, while others propose that it is a physical relation that links both quantities. Others even argue that the relation is not physical but only an effect of the size of samples. As we will see, the difference among the various $\mathcal{M} - m_{max}$ relationships and their meaning does not depend on the relation itself, but rather on how each author interprets the IMF.

One common assumption is that the $\mathcal{N} - m_{max}$ and the $\mathcal{M} - m_{max}$ correlations are theoretically equivalent. With this idea in mind, the first correlation is preferred by Selman & Melnick (2008) and Maschberger & Clarke (2008), who argue that $\mathcal{N}$ is the natural independent variable for testing the random-sampling hypothesis. The second one is preferred by Weidner et al. (2010) because, with the two quantities inferred, the possible error in $\mathcal{N}$ is larger than the error in $\mathcal{M}$. Only a few authors (Selman & Melnick 2008) explore the question of whether they are indeed formally equivalent or not. As we have seen previously, in a probabilistic framework they are not equivalent (cf. Eq. (5)).

4.1.1. The IMF as an exact analytical law

Let us consider the case of García-Vargas & Díaz (1994) and García-Vargas et al. (1995) as an example of this interpretation. They assume that the IMF is not a probability distribution but an exact analytical law, $\phi_{GV}(m) = k(\mathcal{M}) \times \phi(m)$, where $k(\mathcal{M})$ is a renormalization constant that, because $\mathcal{M}$ is the exact value of the amount of gas transformed into stars, verifies

$$\mathcal{M} = \int_{m_{low}}^{m_{up}} m \, \phi_{GV}(m) \, dm = k(\mathcal{M}) \int_{m_{low}}^{m_{up}} m \, \phi(m) \, dm, \qquad (16)$$

where $\phi(m)$ is the standard functional form of the IMF. The exact number of stars with mass m_a in the cluster is given by $N_a = \phi_{GV}(m_a)$, which implies that $k(\mathcal{M}) = \mathcal{N}$. Taking into account that stars are discrete entities, they propose a scenario in which only the stellar masses that verify $\phi_{GV}(m) \geq 1$ represent acceptable physical solutions (the so-called richness effect). Given that $\phi_{GV}(m)$ decreases with m, the most massive star in the cluster is the one that verifies

$$\phi_{GV}(m_{max}) = \mathcal{N} \times \phi(m_{max}) = 1. \qquad (17)$$

For a power-law IMF, $\phi(m) = A \, m^{-\alpha}$, this leads to a $\mathcal{M} - m_{max}$ relationship with the form:

$$m_{max} \propto \mathcal{M}^{\frac{1}{\alpha}}. \qquad (18)$$

According to the scenario proposed, the cluster forms stars in a sorted way, in which the stars with an associated larger value
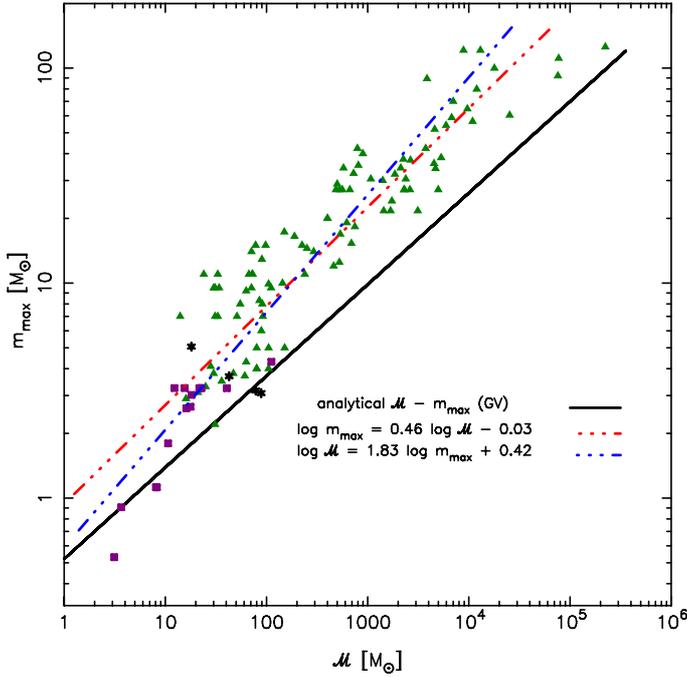
Fig. 8. $\mathcal{M} - m_{\max}$ relationship resulting from the analytical formulation of the IMF of García-Vargas & Díaz (1994); García-Vargas et al. (1995). The figure includes data points from Weidner et al. (2010) and Kirk & Myers (2011), where symbols have the same meaning as in Fig. 3 and the result of two linear fits to the data from Weidner et al. (2010) and Kirk & Myers (2011) using either $\log \mathcal{M}$ or $\log m_{\max}$ as the independent variable.

of $\phi_{\mathrm{GV}}(m)$ take precedence over stars with associated lower values of $\phi_{\mathrm{GV}}(m)$. So, the most massive star (the one with the lowest $\phi_{\mathrm{GV}}(m_{\max})$ value) is conditioned to the formation of a large enough number of lower mass star (the richness effect). Stated otherwise, the mass of this most massive star is determined by the amount of gas that remains after all possible lower mass stars have been formed with relative numbers established by the IMF. We note that the relevant point here is that there must be a certain amount of mass transformed into stars with mass $m < m_{\mathrm{a}}$ in order to have a star with mass m_{a}.

A similar $\mathcal{M}_{\mathrm{cloud}} - m_{\max}$ relationship is found by Larson (1982, 2003). However, Larson's results come from fitting the observational data of cloud masses, $\mathcal{M}_{\mathrm{cloud}}$, with respect to $m_{\max}$, and they are quoted as a statistical correlation, not a physical law. We note that a correlation between $\mathcal{M}_{\mathrm{cloud}}$ and $m_{\max}$ does not imply the same correlation between $\mathcal{M}$ and $m_{\max}$, since an efficiency factor is required (see Shadmehri & Elmegreen 2011, for a more detailed discussion).

In Fig. 8 we show the resulting $\mathcal{M} - m_{\max}$ relationship under these assumptions on the IMF and assuming the functional form of the IMF used in this work. The figure includes data points from Weidner et al. (2010) and Kirk & Myers (2011). We have included the result of two linear fits to the data from Weidner et al. (2010) and Kirk & Myers (2011) using either $\log \mathcal{M}$ or $\log m_{\max}$ as the independent variable. The theoretical relation is off toward larger $\log \mathcal{M}$ values.

This interpretation of the IMF stems from stellar counting procedures. Since $\phi_{\mathrm{GV}}(m)$ is a continuous function, it cannot return a natural number N_{a} for any mass value m_{a}; because stars are discrete entities, this approach can only be an approximate description. This alone is sufficient to invalidate Eq. (17) as a way to obtain the actual most massive star, since $\mathcal{N}$ may

(unphysically) turn out to be a non-natural number. A consequence, this equation can only provide an approximation.

This situation implies that continuous functional forms of the IMF can only be directly related to the number of stars with a given mass interval, and not to the number of stars with a given mass. This possibility is explored in the next interpretation case.

4.1.2. The IMF as a distribution of the number of stars

One alternative view of the IMF is that it can be arbitrarily normalized and provide the exact number of stars in a given mass range. This is the case assumed by Reddish (1978), Vanbeveren (1982), Elmegreen (1997, 1999, 2000), Kroupa & Weidner (2003), Weidner & Kroupa (2004), Elmegreen (2006), Weidner & Kroupa (2006), Weidner et al. (2010) and Kroupa et al. (2011). We refer to these articles as those that use the IMF de facto as a distribution of the number of stars. Their interpretation is that the number of stars between m_{a} and m_{b}, with $m_{\mathrm{a}} < m_{\mathrm{b}}$, is given by

$$N(m \in [m_{\mathrm{a}}, m_{\mathrm{b}}]) = \int_{m_{\mathrm{a}}}^{m_{\mathrm{b}}} \phi_{\mathrm{Elm}}(m)\,\mathrm{d}m, \qquad (19)$$

where $\phi_{\mathrm{Elm}}(m) = k \times \phi(m)$ with k a normalization constant. This equation is the general case of Eq. (7), that is, the definition of $\hat{m}_{\max}$, described above. The difference with the previous case is that the total number of stars in the cluster is now given by

$$\mathcal{N} = \int_{m_{\mathrm{low}}}^{m_{\mathrm{up}}} \phi_{\mathrm{Elm}}(m)\,\mathrm{d}m, \qquad (20)$$

so, $k = \mathcal{N}$. The actual total mass is given by integration of $m \times \phi_{\mathrm{Elm}}(m)$ within the same mass limits. However, how the limits are written and what interpretation is given to them varies according to the author. Here we use the formalization by Elmegreen (1997, 1999, 2000, 2006):

$$\mathcal{M} = \int_{m_{\mathrm{low}}}^{m_{\mathrm{up}}} m\,\phi_{\mathrm{Elm}}(m)\,\mathrm{d}m = \mathcal{N} \times \int_{m_{\mathrm{low}}}^{m_{\mathrm{up}}} m\,\phi(m)\,\mathrm{d}m, \qquad (21)$$

and postpone to the next subsection the discussion of the special case of Weidner & Kroupa (2004, 2006), Weidner et al. (2010), and Kroupa et al. (2011). Whatever the normalization is, we need an additional assumption to obtain the actual maximum stellar mass in the cluster from Eq. (19). We have to assume ad hoc that the most massive star $m_{\max}$ is the result of solving Eq. (7) (i.e., that $\hat{m}_{\max}$ is the actual $m_{\max}$). To do so, external arguments, similar to the richness effect, are required.

For a power-law IMF and $m_{\mathrm{up}} = \infty$, the $m_{\max} - \mathcal{M}$ correlation is

$$m_{\max} \propto \mathcal{M}^{\frac{1}{\alpha-1}} \propto \mathcal{N}^{\frac{1}{\alpha-1}}. \qquad (22)$$

Elmegreen (1997, 1999, 2000) argue that, since the cluster is filled through random sampling, the inferred $m_{\max}$ can only be an estimate of the actual value. Only Vanbeveren (1982) states that it is possible to obtain the actual $m_{\max}$ value.

In Fig. 9 we show the resulting $\mathcal{M} - m_{\max}$ correlation under these assumptions using the functional form of the IMF employed here. The curve is completely equivalent to the $\langle \mathcal{M} \rangle - \hat{m}_{\max}$ correlation obtained in the PDF case. The figure includes data points from Weidner et al. (2010) and Kirk & Myers (2011) just for comparison. We also included the result of a linear fit of $\log \mathcal{M}$ as a function of $\log m_{\max}$ obtained from the data.
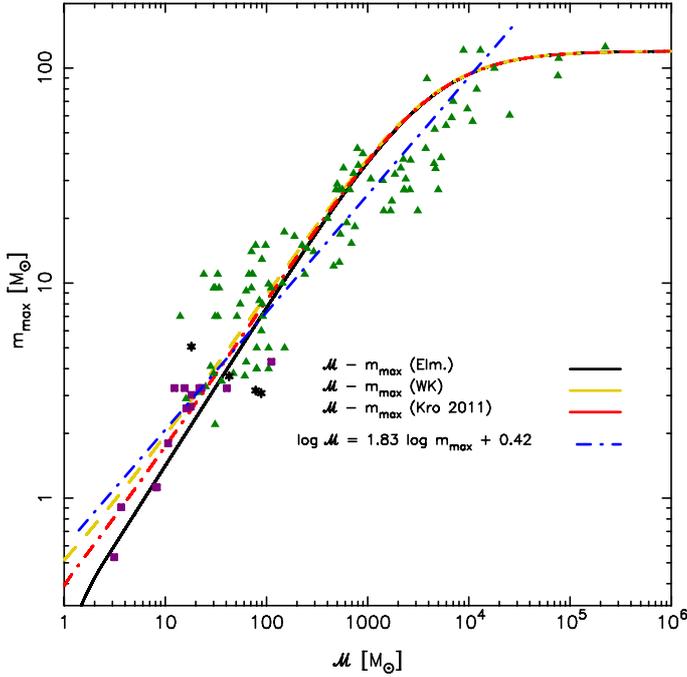
Fig. 9. $\mathcal{M} - m_{max}$ relationship resulting from the distribution function formulation of the IMF of Elmegreen (1997, 1999, 2000), the formulation of Weidner & Kroupa (2004, 2006), and the optimal sampling formulation of Kroupa et al. (2011). The figure includes data points from Weidner et al. (2010) and Kirk & Myers (2011) and the result of the linear fit of the data to $\log \mathcal{M}$ as a function of $\log m_{max}$.

This interpretation of the IMF relies on stellar counting followed by a binning process. It is by far the most common interpretation and is assumed in a wide range of situations, from IMF determinations to stellar population synthesis. Its main feature is that Eq. (19) provides the actual number of stars and that $\mathcal{M} = \mathcal{N} \times \langle m \rangle$ provides the actual total stellar mass in the cluster (this last feature is also shared by the analytical law interpretation). In this case it may seem that the problem with integer numbers of stars mentioned in the previous case is solved as far as we can always choose a suitable set of bins such that Eq. (19) produce a natural number for any m_a and m_b values. However, the solution is not so trivial: depending on the bin definition, distributions with different shapes are obtained (D'Agostino & Stephens 1986; Maíz Apellániz & Úbeda 2005), but the shape of the IMF is still defined by $\mathcal{N} \times \phi(m)$. Consequently, the bins cannot be defined at will. The only plausible solution is to assume that Eq. (19) (and hence Eq. (21)) is only valid in the limiting case $\mathcal{N} = \infty$ (Cerviño et al. 2002; Fouesneau & Lançon 2010; Piskunov et al. 2011), and that, for finite $\mathcal{N}$ values, they do not provide actual $N(m \in [m_a, m_b])$ or $\mathcal{M}$ values but only estimates of such values. Again, we must understand what exactly this estimate represents.

To summarize this section, no continuous functional form of the IMF can provide the actual number of stars, neither for a given mass nor for a given mass interval, but only an estimate of it. The only way to give meaning to this estimate is by adopting a probabilistic framework. This implies using a probabilistic algebra, which explicitly prevents arbitrary normalizations of $\phi(m)$.

4.1.3. The Weidner & Kroupa case

The studies by Weidner & Kroupa (2004, 2006), Weidner et al. (2010), and Kroupa et al. (2011) are another example of an

interpretation of the IMF in terms of a distribution of the number of stars. However, they deserve special attention since they represent a major effort to include conditions in the IMF.

The equations to find a $\mathcal{M} - m_{max}$ relationship proposed by Weidner & Kroupa (2004, 2006), once corrected by an improper account of m_{max} in $\mathcal{M}$ (Kroupa et al. 2011), are

$$1 = \int_{m_{max}}^{m_{up}} \phi_{WK}(m) \, dm, \tag{23}$$

$$\mathcal{M} - m_{max} = \int_{m_{low}}^{m_{max}} m \, \phi_{WK}(m) \, dm. \tag{24}$$

As in the previous case, Eq. (23) is equivalent to the definition of $\hat{m}_{max}$ given in Eq. (7) and $\phi_{WK}(m)$ has the same functional form (scaled by a constant k_{WK}). A simple inspection shows that $k_{WK} = \mathcal{N}$. The difference with the previous case is in Eq. (24): the upper limit of the integral is m_{max} and not m_{up}. By doing so, Kroupa et al. (2011) aim to constrain the IMF in such a way that Eq. (23) provides the actual m_{max} value rather than an estimate of it.

They justify that Eq. (23) provides such actual value by focusing on how the IMF is sampled. Their first approach was the sorted sampling scenario (Weidner & Kroupa 2006), according to which the IMF is sort-sampled, where the stars with the lowest mass are those that form first. This scenario is physically motivated, based on the hydrodynamical simulations of cluster formation in competitive accretion without the inclusion of possible (positive or negative) feedback of massive stars (Bonnell et al. 2003, 2004). Weidner & Kroupa (2006) presented Monte Carlo simulations to support this model, where clusters with a given total mass $\mathcal{M}$ are drawn from a randomly sampled IMF. The number of stars used in the simulation was estimated from $\mathcal{M}$ divided by the mean stellar mass. After that, the sample is sorted and the desired $\mathcal{M}$ value approximated by accepting or rejecting the most massive star in the cluster. The most recent work (Kroupa et al. 2011) is based on the concept of the optimal sample: sampling is optimal if Eq. (23) is verified and produces the actual value of m_{max}. In both cases, it is argued that the IMF is not random sampled. Figure 9 shows the original and the corrected $\mathcal{M} - m_{max}$ relationship they obtain.

This interpretation is based on a strict vision of the IMF as a stellar counting process involving an individual star, the one with $m = m_{max}$, and a stellar counting plus binning procedure for the remaining $\mathcal{N} - 1$ stars. This can be seen from the treatment of the integral limits or equivalently, the histograms bins, throughout the different versions. In the original set of equations proposed by Weidner & Kroupa (2006), m_{max} was counted twice in two non-overlapping bins. The new version (Kroupa et al. 2011) clearly states the bin where m_{max} is, but now it opens a problem with the $\phi(m)$ definition. We recall that it is mainly a problem of inclusion of conditions, which is not a trivial issue. Let us consider the possible self-consistent cases:

1. We use the criteria of equal to or larger than for lower integral limits and lower than for upper ones to give a physical meaning to Eq. (23). However, if we want m_{max} to appear directly in the computation of $\mathcal{M}$, we must impose it ad hoc, which is done by using $\mathcal{M} - m_{max}$ instead of $\mathcal{M}$. A self-consistent formulation, taking into account the integral limits in Eq. (23), is to write explicitly the mass contribution of the stars in the (m_{max}, m_{up}) range

$$m_{max} = \int_{m_{max}}^{m_{up}} m \, \phi_{WK}(m) \quad \Rightarrow$$

$$\phi_{WK}(m) = \delta(m - m_{max}) + (\mathcal{N} - 1) \times \phi(m | m < m_{max}), \tag{25}$$

where $\delta(m - m_{\max})$ is the Dirac delta function. However, this implies an ad hoc variation of the $\phi(m)$ functional form, which is necessary to impose that $m_{\max}$ is the maximum stellar mass.

2. We use the criteria of larger than for lower integral limits and equal or lower than for upper ones. Then, we can compute $\mathcal{M}$ properly using $m_{\max}$ as the upper integral limit. However, in this case we must change Eq. (23) by

$$0 = \int_{m_{\max}}^{m_{\mathrm{up}}} \phi_{\mathrm{WK}}(m)\mathrm{d}m$$
$$\Rightarrow \phi_{\mathrm{WK}}(m) = k_{\mathrm{WK}} \times \phi(m|m \le m_{\max}), \qquad (26)$$

which means that there is no star more massive than $m_{\max}$. This means, however, that we lose the equation giving $m_{\max}$ value, which must be imposed ad hoc.

Cases (1) and (2) above are the only possible ones, and both constrain ad hoc $m_{\max}$ to be the maximum stellar mass in the cluster. Now, we have shown previously that any description of the IMF as a continuous function implicitly eliminates the dependence with $\mathcal{N}$ (and hence $\mathcal{M}$) and its interpretation as a distribution by number. The Kroupa et al. (2011) case clearly shows that there is no way to include constraints into a distribution-by-number description of the IMF and, at the same time, enjoy the advantages of a continuous distribution representation. Once a continuous functional form for $\phi(m)$ is assumed, only a PDF interpretation is valid, and we implicitly renounce obtaining actual values of stellar masses, actual total masses, or actual values of $m_{\max}$. In particular, it would not be possible to obtain a hidden physical law implicit in the $\phi(m)$ functional form. At most we could obtain statistical correlations like the $\langle \mathcal{M} \rangle - \hat{m}_{\max}$. If there were such physical laws, their origin would be external to the IMF and could only be inferred from detailed simulations, and not from algebraic manipulation of the IMF. That is the price we must pay for the advantages of a continuous formulation of the IMF.

4.1.4. The probabilistic case

The IMF is treated as a probability distribution in Oey & Clarke (2005), Elmegreen (2006), Parker & Goodwin (2007), Maschberger & Clarke (2008), Selman & Melnick (2008), Hass & Anders (2010), among others. Their basic assumption is similar to the one of this paper, and some partial results of the description shown here have been obtained by other authors (including Weidner et al. 2010). Here, we summarize the results from works on the topic in the global context of the formulation given in the previous section. The common point of these works is that, without additional ad hoc conditions, an $\mathcal{M} - m_{\max}$ relationship cannot be defined trivially as a physical law, but only as a statistical correlation. The total mass in the cluster, the total number of stars in the cluster, and the particular number of stars with given stellar masses are not fixed quantities, but distributed ones, and none of them can be obtained univocally from the others. Hence, the use of $\mathcal{M} - m_{\max}$ or the use of $\mathcal{N} - m_{\max}$ is not just a question of choice in terms of observational considerations; it is actually the result of statistical correlations of different distributions.

The probabilistic description of the IMF is included, by construction, in works that make use of Monte Carlo simulations (see Weidner & Kroupa 2006; Elmegreen 2006; Parker & Goodwin 2007; Selman & Melnick 2008; Hass & Anders 2010, as examples), where the IMF is sampled star by star up to a given value of $\mathcal{M}$ or $\mathcal{N}$. Such Monte Carlo simulations have

been devoted to explain and compare different results using different sampling algorithms. Hass & Anders (2010) made an explicit, exhaustive, and detailed study of the issue. As far as we know, only Elmegreen (2006) and Selman & Melnick (2008) have made theoretical studies aimed of describing the relationship of $\mathcal{M} - m_{\max}$ using conditional probabilities.

Most of the theoretical studies have been carried out in terms of an $\mathcal{N} - m_{\max}$ relationship, using $\mathcal{N}$ as variate and $m_{\max}$ as variable and making use of $\Phi_{m_{\max}}(m_{\max}|\mathcal{N})$. They often include an expression for the mean value of the distribution (Oey & Clarke 2005), the mode of the distribution (Gumbel 1958; Kendall & Stuart 1977), or the percentile analysis (Weidner et al. 2010). However, there is almost no study in terms of the $m_{\max} - \mathcal{N}$ relationship nor in the $\Phi_{\mathcal{N}}(\mathcal{N})$ dependence of the $\mathcal{N} - m_{\max}$ correlation (Elmegreen 2006; Selman & Melnick 2008).

So, in the probabilistic case, the $\mathcal{N} - m_{\max}$, $\mathcal{M} - m_{\max}$, $m_{\max} - \mathcal{N}$, and $m_{\max} - \mathcal{M}$ correlations are not equivalent to each other. The $\mathcal{M} - m_{\max}$ correlation requires a $\Phi_{\mathcal{N}}(\mathcal{N}|\mathcal{M})$ distribution which is not required by the $\mathcal{N} - m_{\max}$ correlation. In addition, establishing the $m_{\max} - \mathcal{N}$ and $m_{\max} - \mathcal{M}$ correlations requires some priors about the distribution of $\Phi_{\mathcal{N}}(\mathcal{N})$ and $\Phi_{\mathcal{M}}(\mathcal{M})$ that are not considered in the previous correlations.

The probabilistic formulation offers the advantages of using continuous distributions and including conditions formally. However, this does not mean that any condition can be represented analytically. We have mentioned above that the Weidner & Kroupa (2004, 2006) formulation is a major effort to include conditions in the IMF. Let us rewrite Eq. (25) in statistical terms and give a meaning to such distribution:

$$\phi(m|m_{\max}; \mathcal{N}) = \frac{\delta(m - m_{\max})}{\mathcal{N}} + \frac{\mathcal{N} - 1}{\mathcal{N}} \phi(m|m < m_{\max}). \qquad (27)$$

The above equation describes the constrained IMF for a fixed $m_{\max}$ value in a set of $\mathcal{N}$ stars. This constraint does not imply that a star with $m_{\max}$ is present in the cluster, but just that there are no stars more massive than $m_{\max}$ and that the event $m = m_{\max}$ has a probability of $1/\mathcal{N}$. Since all the arguments of the characteristic value hold here, the associated characteristic value is the fixed $m_{\max}$ value, which is also a cut-off value of the distribution. So, 63% of realizations for clusters with $\mathcal{N}$ stars following such PDF have at least one star with mass $m_{\max}$ (and no stars more massive than $m_{\max}$).

Hence, there is no way to include in an analytical form the condition that the most massive star is actually $m_{\max}$ and that such a star is present in any realization. There is also a similar problem with $\mathcal{M}$, although the problem in this case is more severe since it also requires a $\Phi(\mathcal{N})$ (discrete) distribution. However, there is an infinite number of combinations of stellar masses that are consistent with any reasonable $\mathcal{M} - m_{\max}$ physical law.

The only possible solution at the moment to include a $\mathcal{M} - m_{\max}$ physical law and work with it is to perform a large set of Monte Carlo simulations, which should assume a particular $\Phi(\mathcal{N})$ distribution, and just consider the subset where the chosen $\mathcal{M} - m_{\max}$ physical law is verified. Then, any physical result must be obtained numerically (as opposed to analytically). The advantages of describing $\phi(m)$ as a continuous distribution are thus lost[14].

---

[14] We note that any sampling proposal that aims to reproduce a $\mathcal{M} - m_{\max}$ physical law with a finite number of stars $\mathcal{N}$ is also doomed to this situation: it provides a $\phi(m_i)$ array, but not a continuous $\phi(m)$ distribution.

4.2. Sampling, iid variables, and the relation of the IMF with SF

We have seen that the existence of a physical law linking $\mathcal{M}$ and m_{max} cannot be established through a simple manipulation of the IMF functional form. The current debate on whether the IMF is randomly or non-randomly sampled stems mainly from works by Weidner & Kroupa (2006) and Weidner et al. (2010), where $\hat{m}_{max}$ is interpreted as the exact value of the most massive star in a cluster with a given mass. This debate has been focusing on different sampling proposals. Even if the authors themselves now consider the sorted sampling proposal just as a first approximation (Kroupa et al. 2011), we want to emphasize that the key point of different sampling algorithms is not the sorting process, but the assumed relation between $\mathcal{N}$ and $\mathcal{M}$ (e.g., the sorted sampling proposal uses an $\mathcal{N}$ value estimated by means of $\mathcal{M}$ divided by $\langle m|m < m_{max}\rangle$, which imposes a constraint in $\mathcal{N}$). The situation is actually more clearly described in the richness effect proposed by García-Vargas & Díaz (1994); García-Vargas et al. (1995): a star with mass m_a is formed according to the amount of gas that remains in the system once a certain number of stars with $m < m_a$ have been formed. The sampling problem appears when we try to fix $\mathcal{M}(m < m_a)$ and $\mathcal{N}(m < m_a)$ simultaneously and include it analytically in the $\phi(m)$ functional form.

As we have shown, there is no self-consistent way to do it with the current description of $\phi(m)$. The inclusion of any $\mathcal{M} - m_{max}$ physical law, no matter what its interpretation is, precludes using an analytical functional form for the IMF. The sampling methods proposed by different authors are actually operational methods, not an implementation of the physical process[15].

However, we want to stress that the question on whether the IMF is randomly sampled or not (i.e., whether stars are iids or not) is completely valid, independent of the particular problem motivating the question. So we will not attempt to discuss this question in terms of any specific results from literature, but from a more general perspective.

4.2.1. Identical and independent distributed variables and the relation of the IMF with the star formation

The question we aim to answer is: are stellar masses iid variables, or, at least, can they be treated as if they were? A sample is an iid sample if each random variable has the same identical probability distribution and all of them are mutually independent.

Throughout the paper, we have explicitly excluded a mention to the SF physics. It is now time to take a look at different ways in which the SF and the IMF can be linked and how randomness enters in this game. There are several possible ways. (a) Some physicists prefer to assume a deterministic universe in which one and only one result is obtained for a given set of initial conditions. But there is such a large variety of initial conditions that they can be only described in a probabilistic way. Hence the results of SF events, like the IMF itself, can be only described in a

probabilistic way. (b) We can also assume an universe where determinism, although it exists, is somehow hidden by complexity. Thus we assume accordingly that the SF is a complex process in the mathematical sense: nonlinear and with interconnected components, producing such a large variety of results that they can only be treated in a probabilistic way. (c) We admit that there are intrinsically random variables in nature and that the SF is an intrinsically random process (like turbulence), so its results can only be treated in a probabilistic way. We refer to Shadmehri & Elmegreen (2011), Sánchez et al. (2006), Elmegreen (1999, 2011) as examples where some of these different scenarios are considered.

The feature common to these three cases is that the IMF should be used probabilistically (i.e., stellar masses are randomly sampled), which does not imply that the SF is random. There would be no physical $\mathcal{M}$ and m_{max} relationship at all, or there would be a deterministic physical law linking $\mathcal{M}$ and m_{max}. However, the internal distribution of stellar masses that are physically compatible (in the SF sense) with this physical law would depend on a set of unknown (and variable) initial conditions or intrinsically random characteristics. Then the IMF could only be described by means of a probabilistic formulation. A probabilistic interpretation of the IMF does not contradict a deterministic vision of the physics of SF.

On a large scale, the IMF is the result of all possible SF events and SF modes, although it does not necessarily describe any particular one. Following this argument, we are able to describe probabilistically the incidence of having a star with a given mass that was born at a a given time, the stellar birth rate $\mathcal{B}(m, t)$, as the composition of two independent functions: the star formation history, SFH $\psi(t, \mathcal{M})$ (although $\psi(t, \mathcal{N})$ would be more adequate) and the IMF, $\phi(m)$ (Schmidt 1959, 1963; Tinsley 1980; Scalo 1986). The first function includes all the possible SF modes and provides the time-scale and the amount of gas transformed into stars. The second one describes how a given amount of gas would be distributed among different stellar masses. We recall that the first IMF determinations were done with field stars (Salpeter 1955), so they implicitly averaged a large variety of SF modes.

The separation of $\mathcal{B}(m, t)$ into two independent functions seems to be a valid approach for the study of galaxies and a variety of systems where different modes of star formation coexist; it has been extensively used in extragalactic astronomy and cosmology. One particular characteristic of this approach is the use of single stellar populations (SSP, Renzini & Buzzoni 1986) which corresponds to $\psi(t, \mathcal{N}) = \mathcal{N} \times \delta(t)$. Since any function can be described by a sum of $\delta(t - \tau)$ functions, it allows the SFH to be recovered from observational data or the evolution of galaxies to be described as a composition of SSPs with different intensity. The star formation rate, SFR, can then be defined as a time average of the SFH (da Silva et al. 2012) or as the result of a flat SFH ($\psi(t, \mathcal{N}) = $ const.). Current SF rate indicators are based on SSP modeling with constant SFH (Kennicutt 1998).

The case would be different if we changed the scale to smaller systems. When we restrict the situation to specific SF modes, particular details emerge and have some imprint on the IMF. The more restrictive the mode, the more details are present. In this case we are moving ourselves to particular IMF realizations with given conditions, which may depart from the probabilistic description given by $\phi(m)$. At small scales, the validity of the decomposition of $\mathcal{B}(m, t)$ in two independent functions is not clear. However, the universality of the IMF even at such scales leads one to think that it would be the case (however, see Elmegreen 2011 for an example of possible variations of the

---

[15] The optimal sampling algorithm provided by Kroupa et al. (2011) is based on obtaining bins through the larger than for lower integral limits and equal to or lower than for upper integral limits. These criteria are complementary to those underlying their equations to obtain the $\mathcal{M} - m_{max}$ relationship. In addition, the IMF is filled from m_{max} down to lower masses, contrary to the physical arguments given to justify the sorting sampling algorithm. We stress that it is not a problem of the formulation in as much as the physical formulation of the problem is not linked with the operational mathematical method used to solve the physical equations.

IMF, especially in the low-mass tail, depending on the environmental conditions).

The approach we have presented here when talking about $\mathcal{B}(m, t)$ is a top-down one: $\phi(m)$ is the most generic representation, so that the larger the system, the more valid it is. We note that this vision is mentioned by Vanbeveren (1982), who also claimed existence of a $\mathcal{M} - m_{max}$ physical law. Because there is an universal IMF at a large scale, he says, the IMF varies at small scale.

In this case it is expected the IMF has a quasi universal shape at high scales with possible variations at small scales. Here, we understand that deviations from a universal shape are allowed as far as they are small compared to the global budget. In addition, the incidence of deviations also depends on the size of the system, that is, the integral of the $\psi(t, \mathcal{N})$ over time (see da Silva et al. 2012, for a discussion).

There is also a bottom-up approach when talking about $\mathcal{B}(m, t)$, which is the one proposed by the IGIMF theory. In this case, universality in the IMF functional form is assumed. However, there is a $\mathcal{M} - m_{max}$ physical law that relates $\mathcal{M}$ with m_{max}; hence there is IMF variability in the sense of a variable m_{max} for given $\mathcal{M}$. It is assumed that this physical law operates for all SF modes, or equivalently, that there is one SF mode: star formation in clusters. In this case, the mass distribution of stars depends on where (and when) they were formed, so only stars formed in the same cluster (or clusters with the same $\mathcal{M}$) share the same IMF.

For the study of galaxies or, in general, systems that may contain clusters with different masses, it is necessary to take into account the distribution of the total masses of these clusters: the ICMF. As a result, at a galactic scale there is not one IMF, but a IGIMF that results through the combination of the ICMF and different IMFs. It depends on $\mathcal{M}$ and implies a redefinition of the IMF itself (Kroupa & Weidner 2003). In this case it is not clear if $\mathcal{B}(m, t)$ can be separated into independent functions and how (Cerviño et al. 2011). This implies major revisions of global galactic and extragalactic studies, including the SSP concept, and there is currently a large debate on the issue (Corbelli et al. 2009; Fumagalli et al. 2011; Eldridge 2012). Although a full discussion goes beyond the scope of this paper, we want to point out that there would be a $\langle \mathcal{M} \rangle - \hat{m}_{max}$ physical law, although it must be imposed ad hoc, and that, whatever the case, random sampling and a probabilistic description of the IMF are compatible with it.

5. Conclusions

Having carried out a thorough analysis of different IMF interpretations, with a focus on the question of how information on m_{max} can be extracted from the IMF itself, we are in position to formulate the problem in a different way: what information does the IMF contain? Can we extract information on the SF process from an algebraic manipulation of the IMF? The answers to these questions are driven by the interpretation of the IMF adopted by each author and, in particular, their conclusion as to whether, without direct observations, m_{max} can be exactly determined or just estimated.

Our analysis of the problem has led us to the following main conclusion: Only a probabilistic interpretation of the IMF, where $\phi(m)$ is a PDF (ruling out arbitrary normalizations) and stellar masses are random sampledly iid variables, provides a physical and mathematical self-consistent formulation that explains the $\langle \mathcal{M} \rangle - \hat{m}_{max}$ statistical correlation obtained from IMF algebraic manipulation. We also give plausible arguments that introduce

the IMF as a probabilistic distribution when related with the physics of the star formation process.

Additional conclusions of this work are:

1. The actual total stellar mass of a cluster, $\mathcal{M}$, cannot be inferred from an IMF, $\phi(m)$, with a continuous functional form. A direct IMF integration only provides its mean value, $\langle \mathcal{M} \rangle$, for a given number of stars $\mathcal{N}$:

$$\langle \mathcal{M} \rangle = \mathcal{N} \times \langle m \rangle = \mathcal{N} \times \int_{m_{low}}^{m_{up}} m \, \phi(m) \, dm. \qquad (28)$$

Although some authors do not consider $\mathcal{N}$ as a relevant physical variable (Kroupa et al. 2011), the fact that stars are discrete entities and $\mathcal{N}$ is a natural number are relevant physical constraints that must be included in the treatment of the IMF and in the algebra used to obtain physical results from it.

2. Given the equation defining the most massive star in a system,

$$\frac{1}{\mathcal{N}} = \int_{\hat{m}_{max}}^{m_{up}} \phi(m) \, dm, \qquad (29)$$

the resulting $\langle \mathcal{M} \rangle - \hat{m}_{max}$ correlation is practically independent of the specific IMF interpretation adopted. However, how this equation is understood strongly depends on the framework of the interpretation.

3. In a probabilistic interpretation, Eq. (29) provides a characteristic mass, $\hat{m}_{max}$, that is, the value of m that is not reached or exceeded with a probability 0.37 in a sample of $\mathcal{N}$ stars, but not the actual mass of the most massive star in the sample.

4. For any $\hat{m}_{max} \gtrsim 10 \, M_\odot$ and not close to m_{up}, there is a probability larger than 90% that the most massive star in the system is larger than such $\hat{m}_{max}$ value. Therefore, assuming that Eq. (29) provides the actual mass of the most massive star in the cluster, as argued in the framework of different interpretations of the IMF, is an ad hoc assumption and not a physical fact.

5. $\hat{m}_{max}$ defines the mode of the distribution $\Phi_{\mathcal{N}}(\mathcal{N}|m_{max})$ of the possible $\mathcal{N}$ values inferred from the most massive star in the cluster assuming a flat $\Phi_{\mathcal{N}}(\mathcal{N})$ distribution. A similar dependence in $\Phi_{\mathcal{N}}(\mathcal{N})$ is present when $\mathcal{N}$ is inferred from the number of the N_a most massive stars in the cluster (cf. Paper II). However, the observational evidence is that $\Phi_{\mathcal{N}}(\mathcal{N})$ is a power law (if it is related with the ICMF).

6. When the total cluster mass is inferred through the equation $\langle \mathcal{M} \rangle = \mathcal{N} \times \langle m \rangle$ and $\mathcal{N}$ is obtained assuming a flat $\Phi_{\mathcal{N}}(\mathcal{N})$, the observational data become consistent with a $\hat{m}_{max} - \langle \mathcal{M} \rangle$ statistical correlation. This is indeed the case when $\Phi_{\mathcal{N}}(\mathcal{N})$ is not taken into account explicitly in the $\mathcal{N}$ (and $\mathcal{M}$) estimation (as found in most of the cluster in the Weidner et al. 2010 sample).

7. The meaningful distribution to be tested against observational data is $\Phi_{m_{max}, \mathcal{N}}(m_{max}, \mathcal{N})$ and not $\Phi_{\mathcal{N}}(\mathcal{N}|m_{max})$ or $\Phi_{m_{max}}(m_{max}|\mathcal{N})$.

8. Weidner et al. (2010) claim that the results of their analysis falsify the hypothesis of a random sampling of the IMF. Based on the two preceding points, we consider that such claim should be revised, both because of the $\mathcal{M}$ values it relies on and because of the methodological choice of using $\Phi_{m_{max}}(m_{max}|\mathcal{N})$.

9. Different sampling algorithms proposed in the literature are not physical requirements, but convenient mathematical algorithms that try to simplify the implications of such physical law on studies where the IMF is used (as is the case of

stellar population in galaxies). Unfortunately, such simplification is not possible.

10. We cannot exclude that a hard physical law linking $\mathcal{M}$ to m_{max} (the actual values) does indeed exist; but, if this is the case, it must arise from considerations of the problem including a full-fledged SF analysis, which cannot be shortcut through algebraic IMF manipulations. Whatever the case is, the existence of such an $\mathcal{M} - m_{max}$ physical law is compatible with random sampling of stellar masses and a probabilistic interpretation of the IMF.

11. If such a physical law exists, it cannot be incorporated to an analytical IMF functional form, but must rather be approached by computing Monte Carlo simulations and taking into account only the subset of simulations that verify the assumed $\mathcal{M} - m_{max}$ physical law. We note that this approach is fully compatible with the optimal sampling definition provided by Kroupa et al. (2011).

We conclude that a random sampling IMF is not in contradiction to a possible $m_{max} - \mathcal{M}$ physical law. However, such a law cannot be obtained from IMF algebraic manipulation or included analytically in the IMF functional form. The possible physical information that would be obtained from the $\mathcal{N}$ (or $\mathcal{M}$) $- m_{max}$ correlation is closely linked with the $\Phi_{\mathcal{M}}(\mathcal{M})$ and $\Phi_{\mathcal{N}}(\mathcal{N})$ distributions; hence it depends on the SF process and the assumed definition of stellar cluster. In a second paper of this series we will explore the application of the probabilistic description of the IMF formulated in this study. Particularly, we will describe how to use it to make inferences about quantities that characterize some stellar systems, and how observational constraints work as a priori conditions, affecting the sampling distributions of $\mathcal{M}$ and $\mathcal{N}$ that we can infer.

Appendix A: The intensity function

As stated in Sect. 3, $\phi(m)$ cannot provide a value of m_{max} that can be used as the actual maximum stellar mass in a hypothetical cluster. Still, we can calculate the probability for the actual value of m_{max} to be close to the mean, the median, the characteristic value, or the mode of $\Phi_{m_{max}}(m_{max}|\mathcal{N})$. In general, we can evaluate the probability that a value known to be larger that m_b is smaller than $m_b + dm_b$. To do that, we need to introduce the intensity function[16], $\mu(m_b)$:

$$\mu(m_b)\,dm_b = \frac{\phi(m_b)dm_b}{1 - p(m < m_b)} \geq \phi(m_b)\,dm_b. \qquad (A.1)$$

---

[16] We use $\mu(m)$ to follow the notation used by Gumbel (1958). It must not be confused with the definition of the mean value that is used in other papers.
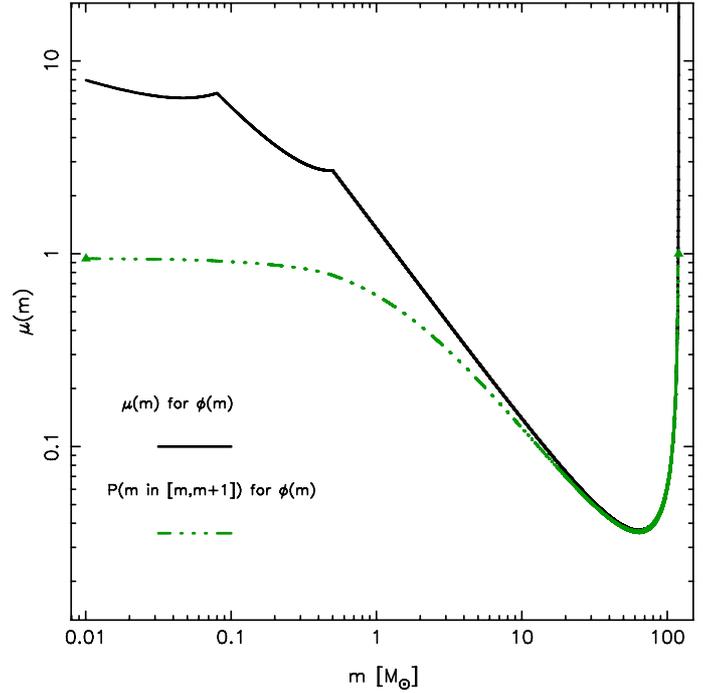

Fig. A.1. Intensity function $\mu(m)$ as a function of m for the IMF. The figure also shows the probability that m will be in the range $(m_b, m_b + 1\ M_\odot)$.

The intensity function is not a PDF; it is independent of $\mathcal{N}$, as implicit in the idd variable hypothesis: the probability of obtaining a value equal to or larger than 5 throwing one dice is 2/6, independently of previous throws. This must not be confused with the case we studied in the previous paragraphs, which would be equivalent to the probability of obtaining at least one throw with a result equal to or larger than 5 in $\mathcal{N}$ draws.

In Fig. A.1 we plot the intensity function for different values of m_b for the case of the IMF used in this work. The figure also shows the probability that a star known to have $m \geq m_b$ will be in the range $[m_b, m_b + 1\ M_\odot)$. The figure shows that $\mu(m_b)$ has a minimum at a value close to m_{up}, and it goes to infinity at m_{up}. The probability of m in the range $[m_b, m_b + 1\ M_\odot]$ decreases with m_b, except for values close to m_{up}. For example, there is only a chance lower than 10% that, given a star in the $m_b - m_{up}$ range, this star has a mass m_b for $m_b \geq 10\ M_\odot$. The situation changes in the extreme case in which m_b is close to m_{up}: if we know that there is one star with mass m_{up} or larger, the mass must certainly be m_{up} (i.e., probability equal to 1), since stars with mass larger than m_{up} do not exist.

This has an interesting implication for the statement that $\hat{m}_{max}$ actually provides the mass of the most massive star in the cluster: assuming that there is one star equal to or more massive than $\hat{m}_{max}$ and that $\hat{m}_{max} \geq 10\ M_\odot$ and is not close to m_{up}, there is a probability larger than 90% that the most massive star is more massive than $\hat{m}_{max}$!

References

Bastian, N., Covey, K. R., & Meyer, M. R. 2010, ARA&A, 48, 339
Bonnell, I. A., Bate, M. R., & Vine, S. G. 2003, MNRAS, 343, 413
Bonnell, I. A., Vine, S. G., & Bate, M. R. 2004, MNRAS, 349, 735
Cerviño, M., & Luridiana, V. 2006, A&A, 451, 475
Cerviño, M., Valls-Gabaud, D., Luridiana, V., & Mas-Hesse, J. M. 2002, A&A, 381, 51

Cerviño, M., Pérez, E., Sánchez, N., Román-Zúñiga, C., & Valls-Gabaud, D. 2011, UP2010: Have Observations Revealed a Variable Upper End of the Initial Mass Function? eds. M. Treyer et al. (San Francisco, CA: ASP), ASP Conf. Proc., 440, 133

Cerviño, M., Román-Zúñiga, C., Bayo, A., et al. 2013, A&A, 553, A32

Corbelli, E., Verley, S., Elmegreen, B. G., & Giovanardi, C. 2009, A&A, 495, 479

Crowther, P. A., Schnurr, O., Hirschi, R., et al. 2010, MNRAS, 408, 731

D'Agostino, R. B., & Stephens, M. A. 1986, Goodness-of-Fit Techniques (New York: Marcel Dekker)

Eldridge, J. J. 2012, MNRAS, 422, 794

Elmegreen, B. G. 1997, ApJ, 486, 944

Elmegreen, B. G. 1999, ApJ, 515, 323

Elmegreen, B. G. 2000, ApJ, 539, 342

Elmegreen, B. G. 2006, ApJ, 486 , 944

Elmegreen, B. G. 2011, ApJ, 731, 61

Fernández-Soto, A., Lanzetta, K. M., Chen, H.-W., Levine, B., & Yahata, N. 2002, MNRAS, 330, 889

Fouesneau, M., & Lançon, A. 2010, A&A, 521, L22

Fumagalli, M., da Silva, R. L., & Krumholz, M. R. 2011, ApJ, 741, L26

García-Vargas, M. L., & Díaz, A. I. 1994, ApJS, 91, 553

García-Vargas, M. L., Bressan, A., & Díaz, A. I. 1995, A&AS, 112, 13

Gumbel, E. J. 1958, Statistics of Extremes (Columbia University Press)

Haas, M. R., & Anders, P. 2010, A&A, 512, 79

Kendall, M., & Stuart, A. 1977, The advanced theory of statistics (London: Griffin), 4th edn.

Kennicutt, R. C., Jr. 1998, ARA&A, 36, 189

Kirk, H., & Myers, P. C. 2011, ApJ, 727, 64

Kroupa, P. 2001, MNRAS, 322, 231

Kroupa, P. 2002, Science, 295, 82

Kroupa, P., & Weidner, C. 2003, ApJ, 598, 1076

Kroupa, P., Weidner, C., Pflamm-Altenburg, J., et al. 2011 [arXiv:1112.3340]

Larson, R. B. 1982, MNRAS, 200, 159

Larson, R. B. 2003, Galactic Star Formation Across the Stellar Mass Spectrum, eds. J. M. De Buizer, & N. S. van der Bliek (San Francisco: ASP), ASP Conf. Ser., 287, 65

Maíz Apellániz, J., & Úbeda, L. 2005, ApJ, 629, 873

Maschberger, T., & Clarke, C. J. 2008, MNRAS, 391, 711

Oey, M. S., & Clarke, C. J. 2005, ApJ, 620, L43

Parker, R. J., & Goodwin, S. P. 2007, MNRAS, 380, 1271

Pflamm-Altenburg, J., & Kroupa, P. 2008, Nature, 455, 641

Piskunov, A. E., Kharchenko, N. V., Schilbach, E. et al. 2011, A&A, 525, 122

Reddish, V. C. 1978, International Series in Natural Philosophy (Oxford: Pergamon)

Renzini, A., & Buzzoni, A. 1986, Spectral Evolution of Galaxies, Astrophysics and Space Science Library, 122, 195

Salpeter, E. E. 1955, ApJ, 121, 161

Sánchez, N., Alfaro, E. J., & Pérez, E. 2006, ApJ, 641, 347

Scalo, J. M. 1986, Fund. Cosm. Phys. 11, 1

Schmidt, M. 1959, ApJ, 129, 243

Schmidt, M. 1963, ApJ, 137, 758

Selman, F. J., & Melnick, J. 2008, ApJ, 689, 816

Shadmehri, M., & Elmegreen, B. G. 2011, MNRAS, 410, 788

da Silva, R. L., Fumagalli, M., & Krumholz, M. 2012, ApJ, 745, 145

Sornette, D. 2004, Critical phenomena in natural sciences: chaos, fractals, self-organization and disorder: concepts and tools, Springer series in synergetics (Heidelberg: Springer)

Treyer, M., Wyder, T., Neill, J., Seibert, M., & Lee, J. 2011, UP2010: Have Observations Revealed a Variable Upper End of the Initial Mass Function? ASP Conf. Proc., 440

Tinsley, B. 1980, Fun. Cosm. Phys., 5, 287

van Albada, T. S. 1968, Bull. Astron. Inst. Netherlands, 20, 57

Vanbeveren, D. 1982, A&A, 115, 65

Weidner, C., & Kroupa, P. 2004, MNRAS, 348, 187

Weidner, C., & Kroupa, P. 2006, MNRAS, 365, 1333

Weidner, C., Kroupa, P., & Bonnell, I. A. D. 2010, MNRAS, 401, 275