

First, second and third massive stars in Open Clusters

Alexey Mints

University of Bielefeld, Germany

Received _____; accepted _____

ABSTRACT

Aims: The goal of this paper is to study possibilities of using first, second and third massive stars in Open Clusters to estimate total cluster mass and membership.

Methods: Estimator functions were built with the use of numerical simulations and analytical approximations.

Results: Precision and error distribution of obtained estimator functions was studied. The distribution of the mass of first, second and third massive stars shows strong power-law tails at the high-mass end, thus it is better to use median or mode values instead of the average ones. It is shown that the third massive star is a much better estimator than the first one — it is more precise and less dependent on parameters such as maximum allowed stellar mass.

1. Introduction

Unfortunately, in most cases it is not possible to get large statistics of star cluster membership, so estimation of cluster mass and/or full number of members is not an easy task. Sometimes when spectroscopical data is available dynamical mass estimates are applied (making use of the virial theorem). But usually only few brightest stars are reliably identified as members (Kharchenko et al. 2003), providing tiny amount of information about the cluster.

Recently Weidner and Kroupa Weidner & Kroupa (2006) discussed the correlation between the mass of most massive star in the cluster and the total cluster mass. Maschberger and Clarke Maschberger & Clarke (2008) discussed correlation between number of stars in the cluster and its most massive member mass.

Here we will try to produce cluster mass (M_{cl}) and membership (N) estimators using masses of three most massive members, and analyse precision of these estimators. We will concentrate on two questions:

- what data provides the most reliable information on the cluster properties;
- what is the best method to extract cluster properties from that data.

Having this two goals in mind, we will neglect at least two very important factors: stellar binarity and evolution. Of course, we can not weight stars directly, we can only estimate their mass, mainly by their brightness, that is usually very uncertain. But, again, as we are interested in the statistical side of the problem, we will postpone astrophysical difficulties for later research. Note, that here we do not have to deal with internal dynamics or spatial distribution of the cluster.

If we assume all stars in the cluster to be single, then (initial) mass of each star depends only on the initial mass function (hereafter IMF). Here we will consider only one IMF — from Kroupa Kroupa (2001). It is built from several power-law parts:

$$\begin{aligned} f(m)dm &= C m^{-\alpha} dm \\ F(m) &= \int_0^m f(m') dm' \end{aligned} \tag{1}$$

IMF parameters was taken from Kroupa Kroupa (2001):

$$\begin{aligned} \alpha_0 &= +0.30 & 0.01 \leq m/M_{\odot} < 0.08 \\ \alpha_1 &= +1.30 & 0.08 \leq m/M_{\odot} < 0.50 \\ \alpha_2 &= +2.35 & 0.50 \leq m/M_{\odot} < m_{\text{max}} \end{aligned} \tag{2}$$

The value of m_{max} is still a matter of debates, Therefore several values will be considered

in this paper ($50, 150$ and $300M_{\odot}$), keeping main attention on $m_{\max} = 150M_{\odot}$. We will try to see, how m_{\max} influences the mass estimator precision.

The following notation will be used throughout the paper: \bar{m}_1 for the average value of m_1 and \tilde{m}_1 for the median value (the same for m_2 and m_3). Another useful value is the position of the peak of $m_{1,2,3}$ distribution for a given M_{cl} or N (mode of the distribution), we will designate them $\hat{m}_{1,2,3}$.

IMF described above has an average stellar mass $\bar{m} = 0.36M_{\odot}$ for $m_{\max} = 150M_{\odot}$.

2. Model

As in (Weidner & Kroupa 2006), three different methods for cluster members generating were used:

Random sampling — N stars are taken randomly from the IMF, with N ranging from 300 to 10000.

Constrained sampling — in this case M_{cl} is fixed first, then stars are taken from the IMF until their total mass surpass M_{cl} . Thus some spread in N is expected in this sample.

Sorted sampling — M_{cl} is also fixed, then $N' = M_{\text{cl}}/\bar{m}$ stars are taken from the IMF. If $M' = \sum_{N'} m_i$ smaller than M_{cl} , then $\Delta N = (M_{\text{cl}} - M')/\bar{m}$ stars are added to the cluster, repeating the procedure until cluster mass outreaches M_{cl} . Then stellar masses are sorted. If $|M' - M_{\text{cl}}|$ is larger then $|M_{\text{cl}} - (M' - m_1)|$ then the heaviest star is removed from the set.

According to (Weidner & Kroupa 2006), random sampling is the least realistic model, but it is the easiest one to be modeled and described analytically.

For each set of parameters (sampling, m_{\max} and M_{cl} or N), 30000 clusters were simulated, saving for each of them five values: cluster mass M_{cl} , number of stars in the cluster N , and masses of three most massive stars of the cluster — m_1, m_2, m_3 .

The goal is to build a method to find M_{cl} and/or N , knowing m_1, m_2 and m_3 . It seems natural to find functions $M_{\text{cl}}(\bar{m}_{1,2,3})$, $M_{\text{cl}}(\tilde{m}_{1,2,3})$ and $M_{\text{cl}}(\hat{m}_{1,2,3})$ (as well as $N(\bar{m}_{1,2,3})$, $N(\tilde{m}_{1,2,3})$ and $N(\hat{m}_{1,2,3})$), we will call them hereafter *mass estimators (ME)*: average ME, median ME and mode ME.

3. Analitics

Random sampling

The probability for the most massive star to have mass $m_1 \in (m, m + dm)$ can be written (see e.g. (Arnold et al. 1992)) as a probability for a given star to have mass in $(m, m + dm)$ multiplied by the probability that all other stars have masses below m and by number of stars N (because any star can be the most massive one):

$$\begin{aligned} P(m_1 \in (m, m + dm)) &= N f(m) [F(m)]^{N-1} \\ &= N f(m) \left[1 - \int_m^{m_{\max}} f(m') dm' \right]^{N-1} \end{aligned} \quad (3)$$

Of course, m_1 should be smaller then m_{\max} .

Substituting 1 into 3 and integrating we get (for $a \neq 1$):

$$P(m_1 \in (m, m + dm)) = NC m^{-\alpha} \left[1 - \frac{C}{1-\alpha} (m_{\max}^{1-\alpha} - m^{1-\alpha}) \right]^{N-1} \quad (4)$$

where C is normalization constant. We can safely use here $m > 0.5M_{\odot}$ part of the Kroupa IMF, as most massive stars are usually much heavier.

If N is large, than we can use exponent instead of square brackets:

$$P(m_1 \in (m, m + dm)) \simeq NCm^{-\alpha} \exp\left(-\frac{NC}{1-\alpha} (m_{\max}^{1-\alpha} - m^{1-\alpha})\right) \quad (5)$$

Maximum of this distribution (or the mode of distribution) is located at the point

$$\hat{m}_1 = \left(\frac{NC}{\alpha}\right)^{1/(\alpha-1)} \quad (6)$$

For $\hat{m}_1 \geq m_{\max}$ obviously maximum is at the point $\hat{m}_1 = m_{\max}$. This puts an upper limit on the cluster mass that can be estimated with this formula. By inverting this equation we can get an estimate for N and M_{cl} from \hat{m}_1 :

$$\begin{aligned} N &= \frac{\hat{m}_1^{\alpha-1} \alpha}{C} \\ M_{\text{cl}} &= \frac{\bar{m} \hat{m}_1^{\alpha-1} \alpha}{C} \end{aligned} \quad (7)$$

Note that m_{\max} is hidden within the constant C in this equations, although the dependence is weak.

For n 'th massive star, if $n \ll N$ we can use the expression:

$$P(m_n \in (m, m + dm)) \simeq (1 - F(m))^{n-1} P(m_1 \in (m, m + dm)) \quad (8)$$

Finding average and median values for the equation 5 is not that easy.

Building analytical expressions for other samplings is much more complicated task, so we will not focus on it here.

4. Results

4.1. Random sampling

The random sampling model has as natural parameter the number of stars in the cluster — N . Here N ranges from 300 to 10000, with 30000 clusters being simulated for

each value of N .

For each value of N the distributions of m_1, m_2 and m_3 were obtained. An example of these distributions is shown on figure 1. From the figure it can be seen that the theoretical estimates given by eqn. 5 and 8 are matching the data well. Note the long power-law tails of distributions, specially for m_1 . This tail leads to significant differences between the average and the median value, making the average much higher. Thus averages are not so well suitable for making cluster mass estimators.

Now the task is to build a method to find M_{cl} and/or N , knowing m_1, m_2 and m_3 . We will try to find functions $M_{\text{cl}}(\bar{m}_{1,2,3}), M_{\text{cl}}(\tilde{m}_{1,2,3})$ and $M_{\text{cl}}(\hat{m}_{1,2,3})$. These functions for M_{cl} are shown in figure 2. They can be approximated with functions of the shape:

$$\begin{aligned} M_{\text{cl}}(m_{1,2,3}) &= am_{1,2,3}^b (m_{\text{max}} - m_{1,2,3})^c \\ N(m_{1,2,3}) &= am_{1,2,3}^b (m_{\text{max}} - m_{1,2,3})^c \end{aligned} \quad (9)$$

so function rises as power law for small m and then “saturates” when m goes to $150M_{\odot}$ (m_{max}).

The parameters of the fits are shown at the table 1. In this table first column refers to one of the functions from equation 9, and second, third and fourth columns are for parameters of this functions — a , b and c , respectively. For $N(\hat{m}_{1,2,3})$ equations 7 can be used. Note that b is always close to $\alpha_2 - 1 = 1.35$ and c is close to -1 , although $f(m) = am^{1.35}(150 - m)^{-1}$ is a bad fit. Note that the value of c decreases from $f(m_1)$ to $f(m_3)$ — this is caused by the fact, that the values of m_3 are much smaller than m_1 and are well separated from m_{max} , therefore “saturation” is not so important for them, and thus $f(m_3)$ is less sensitive to the value of m_{max} .

Now, having this approximations in hand, we can turn back to the initially simulated data, and check, how good these approximations are. Namely, we will substitute $m_{1,2,3}$ for

Table 1: Parameters of fits (see eq. 9)

Fits based on average values			
Function	a	b	c
$M_{\text{cl}}(m_1)$	167.90	1.66	-1.08
$M_{\text{cl}}(m_2)$	1279.49	1.40	-1.07
$M_{\text{cl}}(m_3)$	563.81	1.39	-0.77
$N(m_1)$	915.99	1.59	-1.17
$N(m_2)$	8159.25	1.34	-1.20
$N(m_3)$	4925.55	1.34	-0.97
Fits based on median values			
$M_{\text{cl}}(m_1)$	211.92	1.37	-0.80
$M_{\text{cl}}(m_2)$	228.72	1.37	-0.64
$M_{\text{cl}}(m_3)$	55.20	1.40	-0.27
$N(m_1)$	913.85	1.32	-0.86
$N(m_2)$	1674.88	1.32	-0.80
$N(m_3)$	538.08	1.35	-0.49

each cluster into mass estimators (see eq. 9) to get M_{cl} and N which than can be compared with the real values. This will produce some distributions of estimated M_{cl} and N , together with estimates errors, as far as the “real” values are known from the simulations. The sample of the result for $N(m)$ is shown in the figure 3 for a cluster with pre-defined number of stars ($N = 1000$). Note, that there is a large power-law tail at the high-mass side, that is highest for $N(m_1)$ and smallest for $N(m_3)$ in all cases. Generally $N(m_3)$ shows a smaller spread than other estimators. The distribution of $N(m_3)$ also peaks closer to the real value $N = 1000$.

Tables 2 and 3 summarise the relative errors of mean and relative dispersions for various estimators and samplings. The average estimator is the worst one, giving the highest error of the average value in almost all cases — sometimes up to 40%. The best one seems to be the median estimator, with errors of less than 3%. As expected, the result is due to the power-law tail of distributions, to which median (and mode) values are less sensible. There is a high probability for m_1 to be close to m_{max} , where estimator functions (see eq. 9) are very sensible to m_i , thus producing a higher error. The same reason causes extremely high dispersions for the estimates based on m_1 . Dispersions (and errors of the mean) are smaller for estimators based on m_2 and m_3 , as the slope of the power-law tail is significantly higher. Relative dispersions are smallest for the mode estimator, but it has higher relative error of the mean, comparing to the median estimator. Note, that in most cases estimators based on m_3 show best results in both error of the mean and dispersions.

Here we emphasize once again, that errors are distributed in a significantly non-Gaussian way in this problem. Using median values minimizes error for highest fraction of the data, while for the small fraction errors remain large.

Table 2: Relative error (in procents) of mean value of estimated masses

	Random sampling			Constrained sampling			Sorted sampling		
Value	$f(m_1)$	$f(m_2)$	$f(m_3)$	$f(m_1)$	$f(m_2)$	$f(m_3)$	$f(m_1)$	$f(m_2)$	$f(m_3)$
				Median estimator					
N	0.88	0.28	0.41	0.48	0.32	0.27	1.77	1.89	1.90
M	1.37	0.99	1.05	0.36	0.20	0.12	0.36	0.23	0.13
				Average estimator					
N	23.23	20.13	15.29	12.44	13.66	11.43	18.09	10.41	6.43
M	23.24	20.14	15.29	12.48	13.69	11.45	19.08	11.99	8.44
				Mode estimator					
N	74.97	43.85	25.87	19.44	15.66	12.99	7.41	6.35	4.17
M	74.82	43.73	25.76	21.27	9.29	7.79	7.92	7.63	4.67

Table 3: Relative dispersion (in procents) for mass estimations

	Random sampling			Constrained sampling			Sorted sampling		
Value	$f(m_1)$	$f(m_2)$	$f(m_3)$	$f(m_1)$	$f(m_2)$	$f(m_3)$	$f(m_1)$	$f(m_2)$	$f(m_3)$
Median estimator									
N	46.02	1.80	0.74	13.56	1.84	0.66	122.50	1.17	0.43
M	39.00	1.62	0.70	12.97	1.77	0.65	111.41	1.13	0.43
Average estimator									
N	264.68	3.38	0.82	35.47	3.11	0.73	1296.57	1.99	0.46
M	259.82	3.34	0.81	34.34	3.04	0.73	547.78	1.00	0.36
Mode estimator									
N	1.08	0.84	0.58	0.41	0.60	0.42	0.77	0.44	0.32
M	1.08	0.84	0.58	0.38	0.56	0.39	0.79	0.46	0.33

4.2. Constrained sampling

For constrained sampling almost the same algorithm of ME construction was applied as for the random sampling. The only change is that we don't have an analytical formula for $\hat{m}_{1,2,3}$, so we have to use fits to the simulated data of the shape $f(\hat{m}) = a\hat{m}^b$.

From the figure 4 one can see that difference between random and constrained samplings are not very large in most cases. The distribution for constrained sampling rises and falls not slower then the one for random sampling. The faster decrease at the distribution high end for small cluster (4, top panel) is due to the fact, that when during the simulation the total mass comes close to the desired M_{cl} , massive stars are preferentially rejected from the sample, as far as adding them will make the cluster too massive. Obviously, this effect vanishes for higher M_{cl} , as one can see from the bottom panel of 4.

4.3. Sorted sampling

Sorted sampling should suppress the probability of high-mass star formation even more than the constrained one. This can be seen on figure 5: distribution of m_1 for sorted sampling is almost like the one for m_2 for random sampling at the high-end.

4.4. Estimator reliability

First let us try to compare various estimators by comparing the data, on which they are constructed. Let us turn back to the figure 2. It is obvious, that the curves are very close to each other. This can also be seen from the similarities in parameters a, b, c of fits 9. So one might expect that predictions made with different estimators for the same value of m will not differ from each other significantly. This difference can be even smaller then

the difference between estimated and real value, as predictions can deviate from the real value in the same direction. We calculate average difference as:

$$\Delta f(m_i) = 100\% * \left\langle \left| \frac{f(m_i) - f_{\text{random}}(m_i)}{f_{\text{random}}(m_i)} \right| \right\rangle \quad (10)$$

where $i = 1, 2, 3$ and m_i goes from 3 to $140M_{\odot}$. Thus it shows how far away the estimators are from each other on average.

The result is shown in table 4. Note that in most cases $\Delta f(m_3) < \Delta f(m_1)$. Nevertheless, differences remain large, at the order of 20%, which is much larger than the relative errors of the mean value from table 2.

Another check for the reliability of the obtained estimators is to try to apply them to the “wrong” dataset, for example — using the median estimator from sorted sampling (see section 4.3) to estimate masses for the random one (see section 4.1) or use estimator from dataset with $m_{\text{max}} = 150M_{\odot}$ to estimate masses for $m_{\text{max}} = 300M_{\odot}$ dataset and so on. As far as $m_3 < m_1 < m_{\text{max}}$, one can expect, that m_3 will be a better source for estimates in this case as well. At least it should be less dependent on m_{max} .

In this case precision is much worse, of course. It increases to 30 – 100% for ME based on m_1 and to 5 – 20% for ME based on m_3 . Thus the latter are, as expected, much less sensible to any m_{max} uncertainty.

5. Conclusions

Several mass estimators for cluster mass from the first, second and third most massive stars are defined in this paper. Their precision is estimated. Estimators based on the mass of the third massive member m_3 give the best results (approximately 3-5 times better than those based on m_1), and are less dependent on the maximum allowed stellar mass m_{max} and assumed way of star formation (form of sampling from the IMF). It is also better to

Table 4: Difference (in procents, see eq. 10) between estimators for random and other sampling

Sampling	$\Delta f(m_1)$	$\Delta f(m_2)$	$\Delta f(m_3)$
Average estimator			
Constrained	24	17	18
Ordered	92	38	21
Median estimator			
Constrained	20	14	14
Ordered	89	55	45
Mode estimator			
Constrained	71	66	67
Ordered	34	45	47

build estimators on the median or mode values of m_i instead of the average values. The reason the strong power-law tails in the m_i distributions, that make average value a less representative parameter.

The most important parameter remaining is the assumed algorithm describing how the cluster mass is distributed among stars.

As far as stellar evolution was not taken into account, this results cannot yet be applied to most of real clusters. Inclusion of the evolution into this model is a subject for further work, but here it was shown that m_3 is a good candidate for building mass estimators. Error analysis was also done, showing power-law tail in error distribution, and proving that the median (or mode) values are much better sources for mass estimators than the average values.

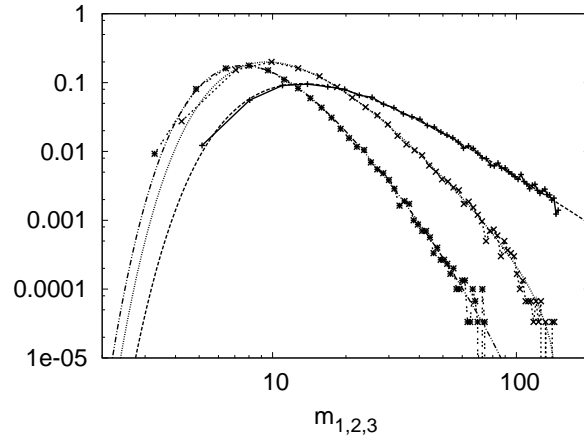


Fig. 1.— Distribution of m_1 (plus signs), m_2 (crosses) and m_3 (stars) with theoretical estimates from eq. 5 and 8 (long-dashed, short-dashed and dot-dashed, respectively) for $N = 1000$

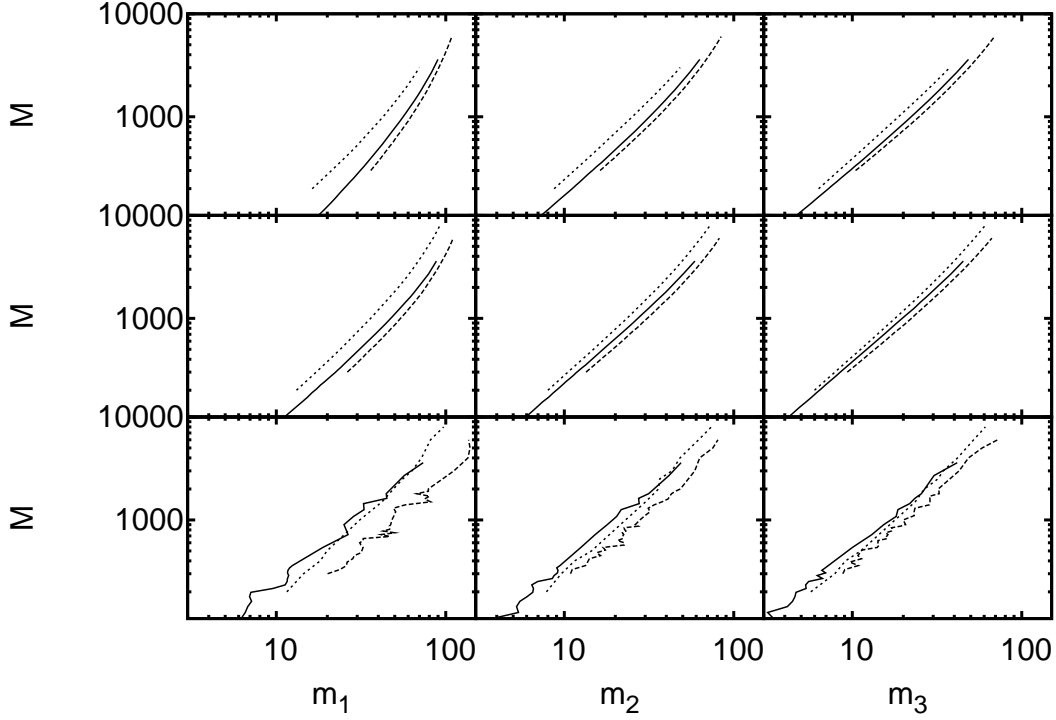


Fig. 2.— Estimators data: dependencies of M_{cl} on m_1 (left), m_2 (middle) and m_3 (right) for random (solid lines), constrained (long-dashed) and sorted (short-dashed) samplings. Top row is for average values, middle is for medians and bottom for the mode.

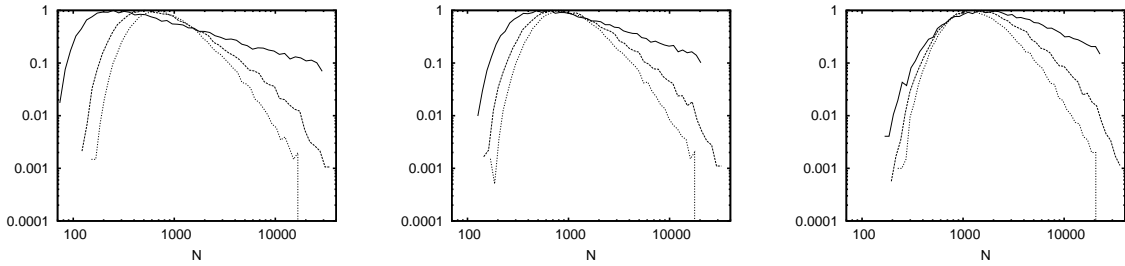


Fig. 3.— Distribution of estimates for a number of stars in the cluster $N(m_{1,2,3})$ (solid, long-dashed and short-dashed lines, respectively) for a cluster with pre-defined 1000 stars. Estimators are based on average (left), median (middle) and mode (right) values.

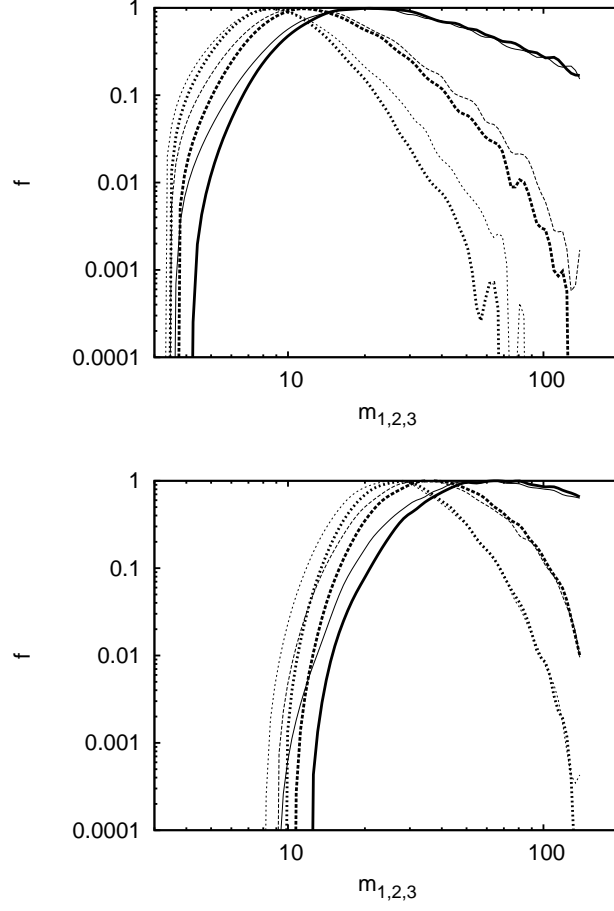


Fig. 4.— Normalized distribution of m_1 (solid), m_2 (long-dashed) and m_3 (short-dashed) for constrained sampling (thick lines) and for random sampling (thin lines, same as at figure 3). Top panel: $N = 1000$; $M_{\text{cl}} \approx 300$. Bottom panel: $N = 4500$; $M_{\text{cl}} \approx 1500$.

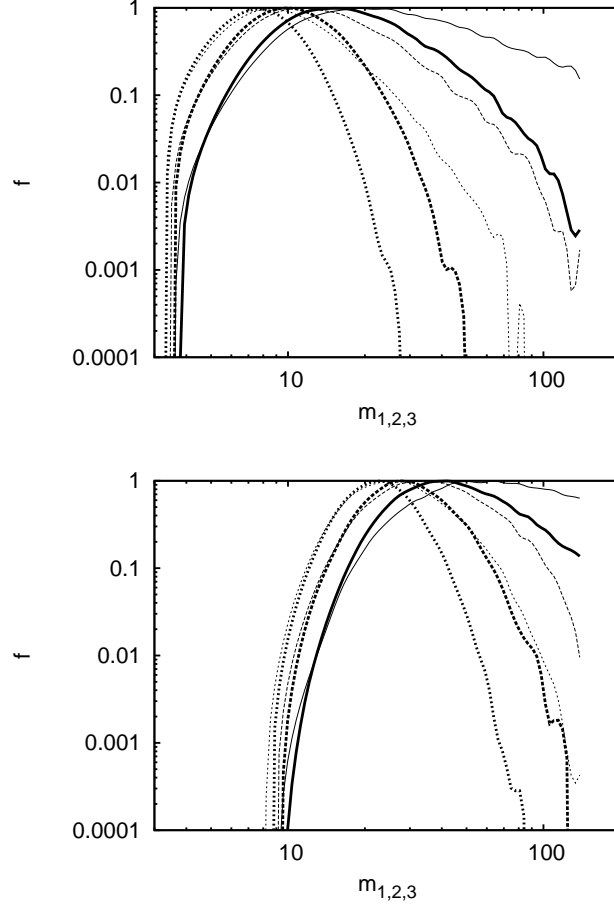


Fig. 5.— Normalized distribution of m_1 (solid), m_2 (long-dashed) and m_3 (short-dashed) for sorted sampling (thick lines) and for random sampling (thin lines, same as at figure 3). Top panel: $N = 1000$; $M_{cl} \approx 300$. Bottom panel: $N = 4500$; $M_{cl} \approx 1500$.

REFERENCES

- Arnold B. C., Balakrishnan N., & Nagaraja H. N., 1992, John Wiley & Sons, New York
- N. V. Kharchenko, L. K. Pakulyak, & A. E. Piskunov, 2003, VizieR Online Data Catalog, 808:291
- P. Kroupa. MNRAS, 2001, 322:231–246
- T. Maschberger and C. J. Clarke, 2008, MNRAS, 391:711–717
- C. Weidner and P. Kroupa. 2006, MNRAS, 365:1333–1347