An Intriguing Correlation between the Distribution of Star Multiples and American Adults

Alon Retter, 86a/6 Hamacabim St., P.O. Box 4264, Shoham, 60850, Israel, alonretter@gmail.com

It is a known fact that like people, many stars are single while most others tend to couple in binaries, yet the two distributions have not been compared so far. The distribution of 4559 brightest nearby stars was matched with that of American adults at the age of 18-65 years. It was found that the mean values of the distributions of star multiples and US households are almost identical (2.04 vs. 2.03). Moreover, a strong resemblance between the two curves is evident. Monte Carlo simulations suggest that this result is significant at a confidence level higher than 97%. Apparently, there should be no connection between the two populations, thus this striking result may supply some clues about the way Nature works.

Astronomy is the observational study of stars. Sociology is the scientific or systematic study of human societies. Evidently, there should not be any relation between the two fields. Yet, it is known that many stellar systems are coupled in binary stars (*1*) similar to people. Despite this resemblance, to our knowledge there has not been any attempt to compare the distributions of stellar and human multiples. Recent observational data of several thousand brightest nearby stars and the expert research analysis followed supplied a unique opportunity for such a comparison.

The multiplicities of stars were collected for a set of 4559 bright stars with Hipparcos (*2*). The observed sample contained multiplicities up to 7. Taking into account the observational biases, it was concluded that the actual distribution of stars in 1, 2…7 multiples is respectively 1459, 2179, 517, 202, 101, 44, and 48 (*3*), which are 32.1, 47.9, 11.4, 4.4, 2.2, 1 and 1%. Note that there were only 4550 stars in the simulated data. We compared these numbers with the figures of the USA adult population in 2009 (*4*). The numbers of 1, 2…5+ members in the age interval of 18-65 years old in 1000 units were 14900, 43479, 9190, 2878 and 739 for all households. These data correspond to 20.9, 61.1, 12.9, 4.1 and 1% of 1, 2…5+ adults in household, and yield a mean of about 2.03 adults per household, which is remarkably consistent with the average stellar multiplicity – 2.04. In Fig. 1 we plot the two distributions. The resemblance between them is outstanding. Indeed, we estimated from extensive Monte Carlo simulations that the two distributions are consistent with each other with a probability level higher than 97% (Appendix).

The results presented in this paper are quite strange and difficult to believe and to understand for any astrophysicist, scientist or layman. While it is a known anecdote that many stars are found in binaries, this work presents a remarkable numeric similarity between the distributions of star multiples observed in the night skies and humans, which is statistically significant. The data used in this work were taken from partial samples. For persons, only American adults were considered and for stars, the 4559 brightest nearby stars were analyzed. In addition, the distributions of star multiples was built using the observational data of stars and a theoretical analysis of the observational biases, which may suffer from some uncertainties. Yet, to date these samples are the best available, and it seems that the surprising similarity between the two distributions requires some explanation. It can be argued that there is a general Nature rule that states that individuals tend to couple in a certain way with a peak at two, however, it is clear that the distribution of certain animal species (e.g. fish) is clearly different.

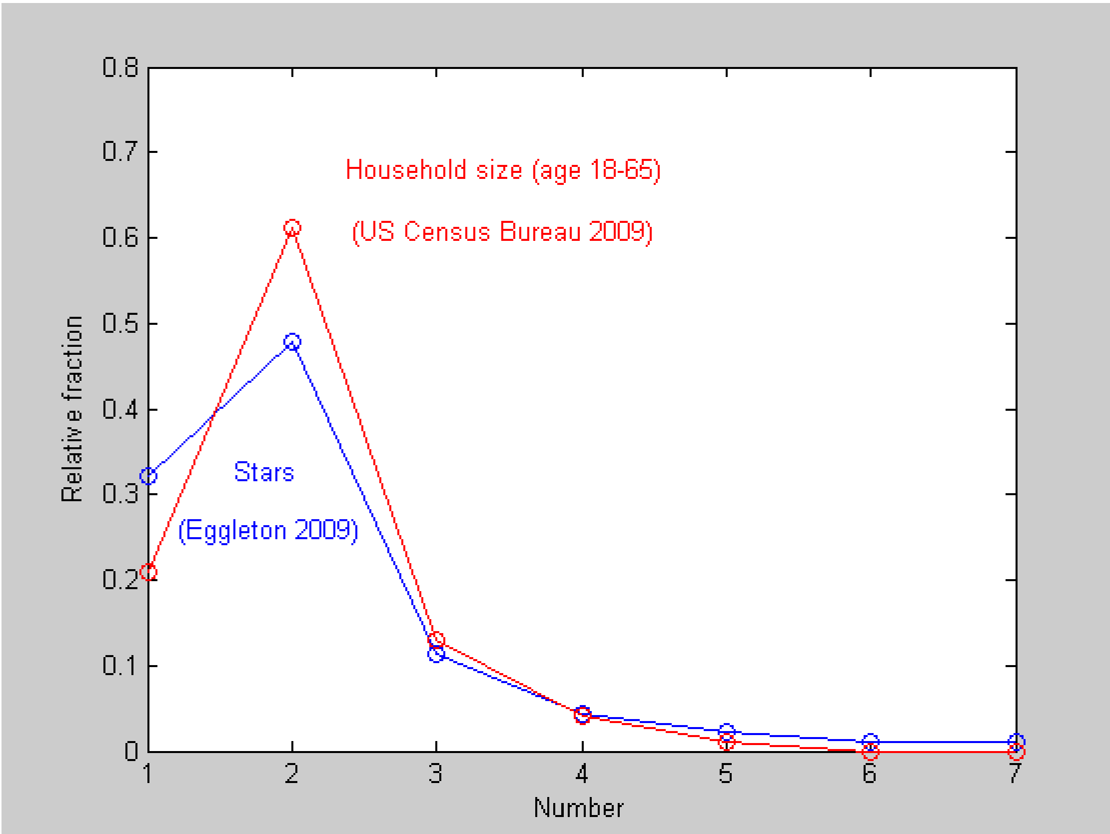

Fig.1 – A comparison between the distributions of star multiples and American households in 2009. There is a remarkable similarity between the two curves. The mean value in humans is 2.03, while that of stars was estimated as 2.04 based on observations and a thorough analysis of selection effects. From numerical simulations we deduced that there is less than 3% chance probability to randomly achieve this result (Appendix).

The similarity between humans and stars is significant at a confidence level higher than 97%, so there is still ~3% chance that it is only a coincidence. Taken into account the uncertainties involved in the theoretical analysis of the observational selection effects in the observations of stars, this number could even be larger. It can also be asked why only adults in the age of 18-65 years and only regular stars were considered in the comparison between astronomical objects and humans. The reason is quite simple. We believe that people older that ~65 years old should be compared with old stars – white dwarfs, neutron stars and black holes and that similarly children should be matched with planets. Finally, this paper actually only presents a glimpse of our ideas, which we admit sounds absurd. Some similarity between the distributions of children and planets was found as well, although for a small sample of planets (*5-6*). Combining these results, the significance level is higher, reaches about 99.9% and cannot be regarded as an anecdote. These results should be re-examined in the future when larger data samples are available, but the current strange picture which arises from them (*7*), is quite disturbing.

It is a pleasure to thank Shlomo Heller (Kartar) for many stimulating discussions.

REFERENCES

1. Warner, B. Cataclysmic Variable Stars, Cambridge: Cambridge Univ. Press (1995**)**

2. Eggleton, P. P., & Tokovinin, A. A. *MNRAS*, **389**, 869-879, (2008).

3. Eggleton, P. P. *MNRAS*, **399**, 1471-1481, (2009).

4. US Census Bureau, [http://www.census.gov/population/www/socdemo/hh-fam/cps2009.html], table F1, (2009).

5. Retter, A. http://vixra.org/abs/1004.0129, (2010a).

6. Retter, A. http://vixra.org/abs/1004.0130, (2010b).

7. Retter, A., & Heller, S. http://vixra.org/abs/1004.0128, (2010).

8. Press, W.H., Teukolsky, S.A., Vetterling, W. T., & Flannery, B.P. Numerical Recipes. Cambridge University Press, Cambridge, (1992).

Appendix – significance estimates

The distributions of adults and stars were discussed above and it was concluded that they are alike (Fig. 1). The purpose of this appendix is to check the significance of this result. One may try to use the Kolmogorov-Smirnov (KS) probability test (*8*) to check whether two different distributions are consistent with each other. However, this test is adequate for a large number of points that can get continuous values, while our relevant distributions only have a few discrete points. Therefore, it was decided to check the significance of the result by extensive Monte Carlo simulations.

Given a distribution, Pa = [Pa_1, Pa_2 … Pa_n] for bins [1, 2…n] that complies Pa_1+ Pa_2+…+ Pa_n=1, we posed the question: "what is the chance probability to obtain by random a second vector distribution, Pb = [Pb_1, Pb_2… Pb_n] with Pb_1+ Pb_2+ Pb_n =1?" We defined a difference parameter δ (Pa_cum, Pb_cum) = sqrt (Σ (Pb_cum_i- Pa_cum_i)^2), for i=1…n between the corresponding cumulative distributions: Pa_cum = [Pa_cum_1, Pa_cum_2…Pa_cum_n] = [Pa_1, Pa_1+Pa_2 … Pa_1+Pa_2+…+Pa_n] and Pb_cum = [Pb_cum_1, Pb_cum_2…Pb_cum_n] = [Pb_1, Pb_1+Pb_2 … Pb_1+Pb_2+… +Pb_n]. For the cumulative distributions of the pair - adults ([0.209, 0.820, 0.949, 0.990, 1]) and stars ([0.321, 0.800, 0.914, 0.958, 1]), n=5 and δ=0.123. Note that we re-arranged the stars data into 5 bins in order to fit the second data set. For the test we built one million random distribution samples with noise taken from the data using a few different methods. First, the mean and standard error of the Pa data were found, and then for every simulation we raffled Gaussian distributed noise and obtained n random numbers around the data mean. Negative numbers were shifted upwards and given a random number around 0.01, and the total simulation vector was normalized to 1, so Ps_1+ Ps_2+… +Ps_n=1. From this initial vector the cumulative distribution was calculated to obtain the final probability vector Ps_cum = [Ps_cum_1, Ps_cum_2… Ps_cum_n] = [Ps_1, Ps_1+Ps_2…Ps_1+Ps_2+…+Ps_n]. The difference parameter between this simulated distribution and the first given distribution, δ (Pa_cum, Ps_cum), was calculated. For one million simulations, one million values of this parameter were obtained. The significance level was defined as the ratio between the number of values higher than the observed difference parameter, δ (Pa_cum, Pb_cum), calculated above, to the total simulations number. This test suggested that there is 98.9% chance probability that the pair is significant.

The highest peak in both distributions is at the second bin (2). We repeated the simulations, giving a preference for the highest probability value in each simulation to be either at bin 1 or 2, while all other n-1 figures were randomly shuffled in the remaining bins. The results of these simulations were that there is 97.4% chance probability that the two distributions are consistent with each other. This value was adopted in the paper and it means that the pair is highly significant. This is a conservative approach, because a priori given the distribution of adults, the distribution of stars could be completely different, say with all multiples above n=5, and there is no reason why the highest peak in the astronomical distribution would be either at 1 or 2.

We performed another test that clearly underestimated the significance level of the results. We imposed the highest probability value exactly as observed - in the second bin. The resulting significance level was 90.8%. These simulations confirmed that once the distribution of adults is given, there is a very low chance probability to randomly obtain the observed distribution of stars.

Another test we applied was to differently model the data. We either fitted a 2D or 3D polynomial to them. The standard errors were found from the difference between the fit and the data. Then we raffled random numbers according to the standard error and added them to the fit to obtain n random numbers. Negative values were given random numbers around 0.01, and the total simulation vector was normalized to 1. The data bins were either randomly shuffled or given some preferences as discussed above. The cumulative distribution was calculated to obtain the final simulation vector, and the difference parameters, δ (Pa_cum, Ps_cum), was calculated. The outcome of these simulations was very similar to the results obtained above with typical differences of only tenths percent.