DISTRIBUTION OF MAXIMAL LUMINOSITY OF GALAXIES IN THE SLOAN DIGITAL SKY SURVEY

E. REGŐS, A. SZALAY, Z. RÁCZ, M. TAGHIZADEH, K. OZOGANY
CERN (Geneva), Johns Hopkins U. (Baltimore), ELTE (Budapest)

Extreme value statistics (EVS) is applied to the pixelized distribution of galaxy luminosities in the Sloan Digital Sky Survey (SDSS). We analyze the DR6 Main Galaxy Sample (MGS), divided into red and blue subsamples, as well as the Luminous Red Galaxy Sample (LRGS). A non-parametric comparison of the EVS of the luminosities with the Fisher-Tippett-Gumbel distribution (limit distribution for independent variables distributed by the Press-Schechter law) indicates a good agreement provided uncertainties arising both from the finite size of the samples and from the sample size distribution are accounted for.

1 Introduction

Extreme value statistics analyzes the behavior of the tails of distributions. The distribution of extreme values for i.i.d. (independent, identically distributed) variables converge to a few limiting distributions depending on the tail behavior of the parent population (Fisher-Tippet-Gumbell, Weibull, Fisher-Tippet-Frechet). The onset of the scaling behavior is quite slow, therefore requires very large samples. This is the reason why astronomy has seen very few applications of EVS to date. The galaxy samples in the SDSS redshift survey may be just large enough to attempt such an analysis, and we present here a study of the distribution of maximal luminosities (selected from the luminosities in a given direction and solid angle).

Since EVS is well known only for i.i.d. variables, we try to minimize the correlations between luminosities and positions by selecting the maximal luminosities from the batches of galaxies in elongated conical regions (defined by the footprints of HEALPix cells on the sky).

The shape of the galaxy luminosity function is important for the EVS. This function is well described by the Schechter function, functionally similar (and motivated by) the theoretically derived Press-Schechter formula, with a power law distribution and an exponentially falling tail. Such a tail would imply a Fisher-Tippett-Gumbel (FTG) EVS distribution, with corrections for the finite sample sizes depending on the power law at low luminosities. In this analysis we will show that there is an excellent agreement with these expectations, implying that the Schechter function extends to very high luminosities, i.e. there is no indication for a sharp cutoff at a high but finite luminosity.

In order to arrive at the above results, it was important to notice that the galaxies can be divided into three types with significantly distinct luminosities and spatial distributions. It turned out that an EVS analysis of the luminosities for the joint populations of these three types is practically impossible. This is due to the large differences in the luminosity scales of the three types which results from the morphology–density relation combined with the presence of voids and clusters in the galaxy distribution.

Even though the SDSS sample is large, the residual from the FTG distribution can be explained only when we consider the corrections due to both the finite size of the samples and the distribution present in the sample sizes (the number of galaxies in a cone is finite and varies from cone to cone).

In Section 2, we describe our galaxy sample, the data acquisition methods, the details of the separation of the sample according to cuts in a color-magnitude diagram. The distributions of the galaxy luminosities and the galaxy counts in pencil beams are also constructed here. Section 3 discusses extreme value statistics with deviations from the expected limit distributions due to finite number of the galaxies in the pencil beams and the pencil-to-pencil fluctuations in the galaxy counts. Section 4 are the results about the distribution of maximal luminosities with the conclusion that within the uncertainties the Fisher–Tippett–Gumbel distribution is a good fit.

2 Sample Creation

We use data from SDSS-DR6, available in a MS-SQL Server database that can be queried online via CasJobs (`http://casjobs.sdss.org`). The spectroscopic survey renders a complicated geometry defined by `sectors`, whose aggregated area covers 6807.94 $\deg^2$ in the sky.

We explore 4 different galaxy samples: the Luminous Red Galaxies (LRGs) (Eisenstein 2001) and the Main Galaxy Sample (MGS, hereafter MGSall) (Strauss 2002), which is segregated into a blue (MGSblue) and red populations (MGSred). The LRGs are slowly evolving and intrinsically luminous red galaxies composed of old stellar populations, selected for tracing the structure at a higher redshift than MGSs. For selecting the latter we impose a redshift interval of $[0.02, 0.18]$. A total of $N_{\rm T} = 383791$ MGSall and $N_{\rm T} = 66960$ LRG galaxies is obtained in a survey's comoving volume of $V_{\rm S} = 2.75 \times 10^8\,{\rm Mpc}^3$ and $V_{\rm S} = 4.31 \times 10^9\,{\rm Mpc}^3$ respectively, where we used the parameters $(\Omega_{\rm L},\Omega_{\rm M},\Omega_{\rm r},h_0,w_0)$=(0.726,0.274,0.0,0.705,-1).

2.1 Color and Magnitude Cuts

In order to segregate the MGSall sample into MGSblue and MGSred, we construct a color-magnitude diagram (CMD) of MGSall and follow the work done in Baldry (2004). The color $C_{ur} \equiv u - r$ is calculated using model magnitudes, whereas the absolute magnitude M_r is derived from the petrosian magnitudes. All magnitudes are galactic extinction and k-corrected. The k-corrections are calculated by using a non-negative least squares fitting method against 30 templates drawn from the Bruzual & Charlot catalogue. Fig.1 shows a smooth separation between MGSblue and MGSred, resembling a bimodal gaussian in color space.

We use the naive Bayes classifier under the classes *Red* and *Blue* in order to construct a color-separating curve $C_S(M_r)$, on top of which the galaxies have the same probability of belonging to either populations. The separator leads to $N_T = 188354$ and $N_T = 195437$ galaxies in the MGSblue and MGSred samples respectively, whose luminosity functions are in Fig.1.

We use 13 JackKnife regions defined as HEALPix cells in the sky, which are part of a $N_{side} = 4$ low resolution HEALPix map of the SDSS-DR6 footprint. Each region has at least 90% of its area inside the footprint.

2.2 Footprint and HEALPix based pencil beams, Distributions of galaxy counts in a HEALPix cell

In order to have close-to-i.i.d realizations (batches) from which to draw the maximal luminosities, we tessellate the sphere into regions defined by individual HEALPix cells (Gorski 2005), all of

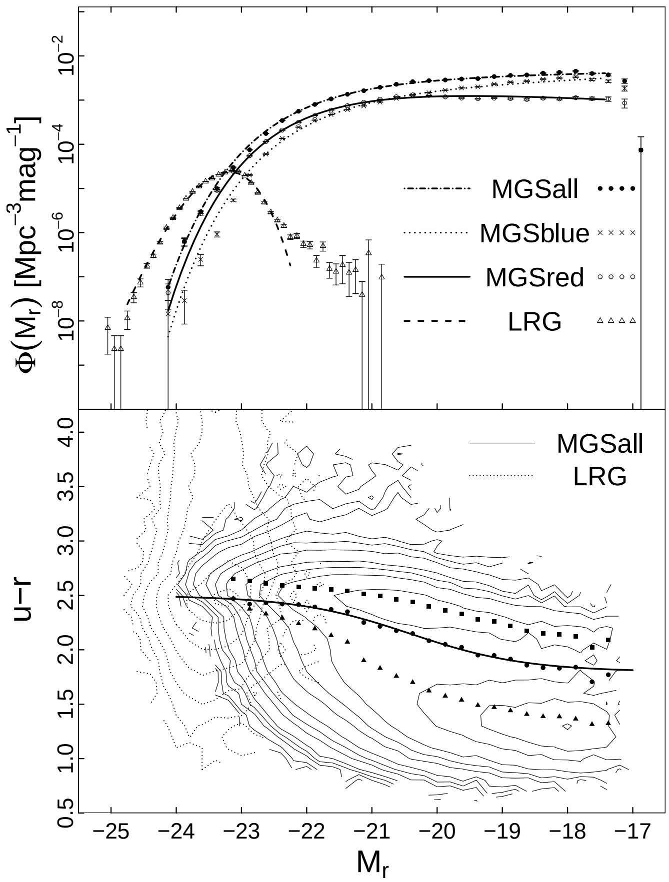

Figure 1: Top: Luminosity functions of the different galaxy samples, together with the best Schechter and generalized gamma fits. The fits are shown only in the range of M_r inside which they were calculated. The magnitude bin widths are $\Delta M_r = 0.25$ and 0.1 for the MGS and LRG samples respectively. Bottom: Color magnitude diagrams of the MGSall and LRG samples. The contour level curves are in $\log_{10}$ scale, set for MGSall at $\{1, 1.8, 3.5\} \times 10^n$ ($n = 0, ...4$); and at $\{1, 3\} \times 10^n$ ($n = 0, 1, 3$) for the LRG sample. The triangles and squares indicate the fitted values of μ for the blue and red populations respectively, whereas the solid thick line is the population separator fitted to the filled circles.

which have the same area. This creates 3-dimensional pencil-like beams that sample the galaxy populations across different redshifts.

We create the entire SDSS-DR6 spectroscopic footprint with resolution $N_{side} = 512$ ($\sqrt{\Omega_{pix}} \simeq 6.87'$). The total area of the HEALPix footprint is 6806.61 deg^2, close to the joint area of all sectors (6807.94 deg^2).

We degrade the footprint to 3 lower resolution maps defined by $N_{side} = 16, 32$ and 64, creating thus the cells that define the pencil beams. We use only the group of cells which satisfy that their fractional area occupancy inside the footprint $f \geq 0.98$. HEALPix maximal luminosity maps are shown for MGSall in Fig. 2.

3 Theory of Extreme Value Statistics

Extreme value statistics (EVS) is concerned with the probability of the largest value in a batch of N measurements. For us, they are galaxy luminosities in a given solid angle of the sky and N is the number of galaxies in the given angle.

The results of the EVS are simple for i.i.d. variables. The limit distribution belongs to one of three types and the determining factor is the large-argument tail of the parent distribution. Frechet type distribution emerges if the parent distribution f decays as a power law, Fisher-

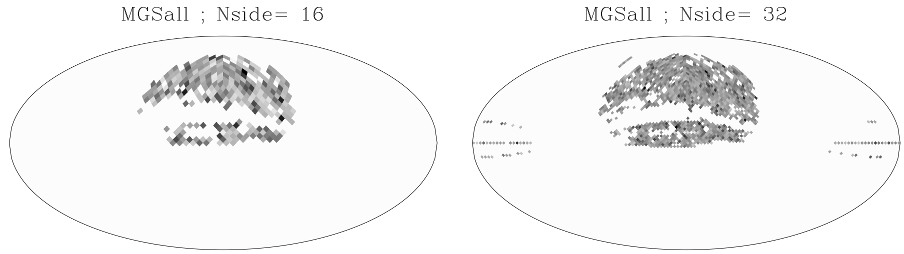

Figure 2: HEALPix Maps of maximal luminosities (in linear scale) for the MGSall galaxy sample at different values of N_{side}. Darker color means higher luminosity. The SDSS-DR6 footprint becomes easily recognizable at resolution $N_{side} = 64$.

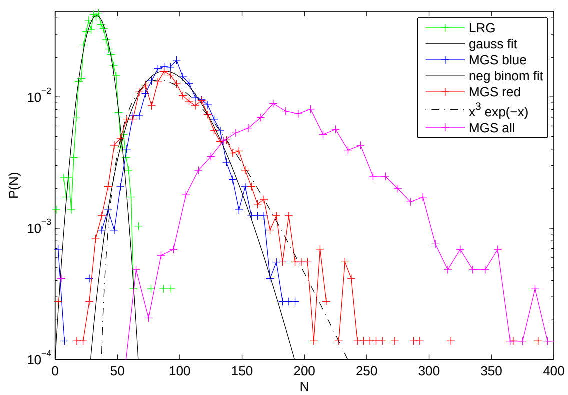

Figure 3: Distribution $P(N)$ of the number of galaxies N in the pencil beams for the case of $N_{side} = 32$. The results are for the MGS-red, MGS-blue, and the LRG samples (red, blue, and green lines, respectively) and for the sample containing all the galaxies (magenta) are plotted on the top panel. Fits to the empirical data on MGS-red, MGS-blue and LRG data are also shown.

Tippett-Gumbel (FTG) distribution is generated by fs which decays faster than any power law and parent distributions with finite cutoff and power law behavior around the cutoff yield the Weibull distribution. All the above cases can be unified as a generalized EVS whose integrated distribution $F_N(v)$ is given in the $N \to \infty$ limit by $F(v) = \exp\left[-(1+\xi v)^{-1/\xi}\right]$ where $\xi > 0, = 0, < 0$ correspond to the Frechet, FTG, Weibull classes, respectively, with the parameter ξ being the exponent of the power law behavior.

The parent distribution for galaxy luminosities is known, it is the Gamma-Schechter distribution. For this the limit distribution of extremal luminosities belongs to the FTG class $(\xi \to 0)$

$$P(v) = \frac{dF(v)}{dv} == \frac{1}{b}\exp\left[-\frac{v-a}{b} - \exp(-\frac{v-a}{b})\right] . \qquad (1)$$

where the parameters can be fixed by setting $\langle v \rangle = 0$ and $\sigma = \sqrt{\langle v^2 \rangle - \langle v \rangle^2} = 1$. This choice ($a = \pi/\sqrt{6}$ and $b = \gamma_E \approx 0.577$) leads to a parameter-free comparison with the experiments provided the histogram of the maximal luminosities $P(v)$ is plotted in terms of the variable $x = (v - \langle v \rangle_N)/\sigma_N$ where $\langle v \rangle_N$ is the average the maximal luminosity while $\sigma_N = \sqrt{\langle v^2 \rangle_N - \langle v \rangle_N^2}$ is its standard deviation. The resulting scaling function is the universal function (1) in the limit $L \to \infty$: $\Phi_N(x) = \sigma_N P_N(\sigma_N x + \langle v \rangle_N) \to \Phi(x)$.

3.1 Deviations from the ideal case

In addition to the assumption of i.i.d. variables, there are two additional problems with comparing data with theory. A notorious aspect of EVS is the slow the convergence to the limit distribution. Second, the batch size N (the number of galaxies in a given solid angle) varies with the direction of the angle. Thus the histogram of the maximal luminosities $P_N(v)$ is built from a distribution of Ns. Both effects introduce corrections to the limit distribution.

3.2 Finite size corrections

Finite size corrections in EVS have been studied with the conclusion that to first order the scaling function can be written as $\Phi_N(x) \approx \Phi(x) + q(N)\Phi_1(x)$, where $q(N \to \infty) \to 0$ and the shape correction $\Phi_1(x)$ is universal function. Both the amplitude q and the shape correction Φ_1 are known for Press-Schechter type parent distributions. The convergence to the limit distribution is slow since $q(N) = -\theta/\ln^2 N$. The value of θ is roughly 1 thus for characteristic range of $N \approx 10 - 200$, the amplitude is of the order of 0.2-0.04. Thus one can expect a 20-4% deviations coming from finite-size effects.

The finite-size shape correction is $\Phi_1(x) = [M_1(x)]'$ where

$$M_1(x) = \Phi(x) \left[\frac{ax^2}{2} - \frac{\zeta(3)x}{a^2} - \frac{a}{2} \right] . \tag{2}$$

3.3 Variable batch size

Variable sample size raises basic questions about EVS.

We use various approximations for the $F(N)$, including using the experimentally observed one. We carried out simulations for the following $N_{side} = 32$ cases: (a) $F(N)$ is approximated as a Gaussian for the LRG sample, (b) $F(N)$ of the MGS red sample is fitted to $(aN+b)^3 \exp(-(aN+b))$, (c) $F(N)$ of MGS blue sample is described by a binomial distribution whose parameters were determined from the empirical values of the average and mean-square fluctuations of N and (d) $F(N)$ is just the exact empirical distribution for the MGS red sample.

Combining the finite-size effects with the effects coming from the variable batch-size can produce the features observed in the deviations from the FTG limit.

4 Distribution of Maximal Luminosities, Discussion

When considering only cells with $N > 10$, leads to a good agreement for the MGSall sample.

If the distributions obtained are FTG distributions then we can claim that at large distances, r, the luminosity-luminosity correlations should decay with an exponent $C(r) \sim r^{-\sigma}$ with $\sigma > 2$.

The FTG-s found suggest weak correlations, and the noise, or more precisely, the correlated features of the noise are, in principle, explainable by the combination of finite-size and variable batch-size effects.

References

[Bruzual & Charlot (2003)] Bruzual A.,G. & Charlot S. 2003, MNRAS 344, 1000

[Baldry et al. (2004)] Baldry, I. K. et al. 2004, ApJ, 600:681-694

[Fisher Tippet (1928)] Fisher, R.A and Tippet, L.H.C. 1928. Proc. Camb. Phil. Soc. 24,180-190

[Gnedenko (1943)] Gnedenko, B.V 1943. Ann. Math. 44, 423-453

[Górski et al. (2005)] Górski, K. M. et al. 2005, ApJ, 622:759 , 109, 301

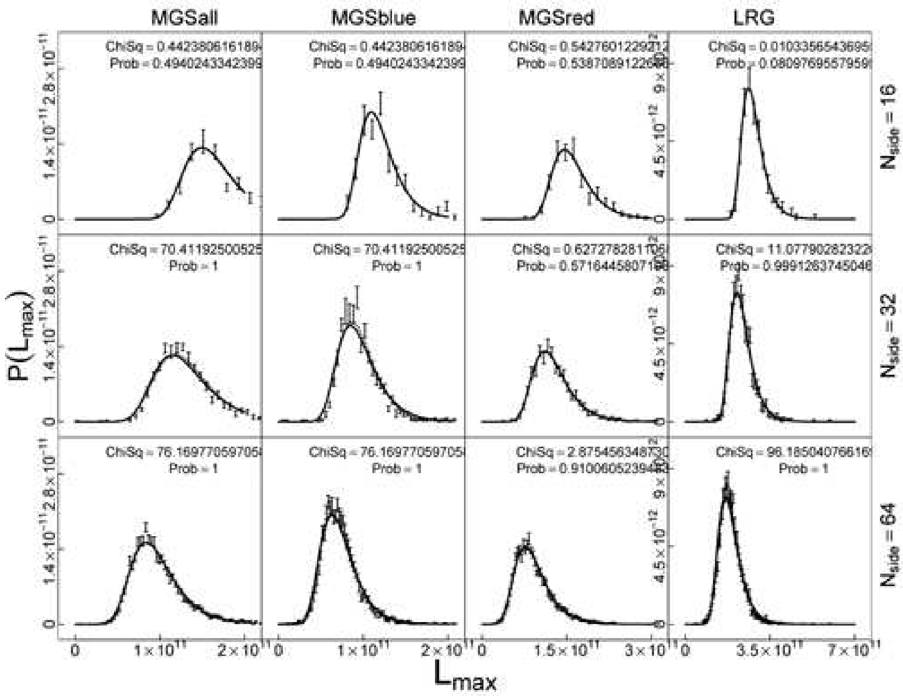

Figure 4: Histogram of the maximal luminosities L_{max}. The continuous line corresponds to the GEV fit, whereas the dotted shows the Gumbel model. The χ^2 values are for the Likelihood ratios between the 2 models, with the corresponding probability of accepting the Gumbel model.

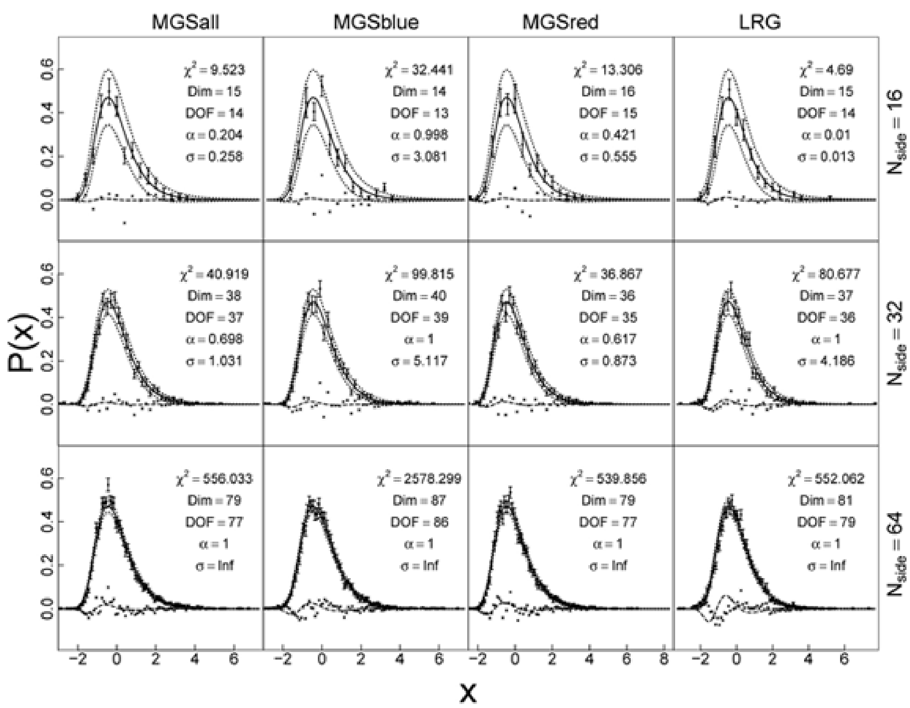

Figure 5: The filled dots represent the normalized maximum luminosity histograms for N_{side} = 16, 32, 64 and for the 4 galaxy samples. The continuous line is the theoretical FTG universal curve with $\sigma = 1$ and the dotted curve is its first order correction. The crosses are the residuals.

9. Posters