

FITTING BINOMIAL DISTRIBUTION TO ONLINE RATINGS OF GREEK TOURISM ATTRACTIONS

Yiannis Dimotikalis^{1*} and Stelios E. Papadakis²

¹Dept. of Accounting and Finance, T.E.I. of Crete, Estavromenos, Heraklion, Crete, Greece

²Dept. of Business Administration, T.E.I. of Crete, Lakonia, Aghios Nikolaos, Crete, Greece

*jdimit@staff.teicrete.gr

ABSTRACT

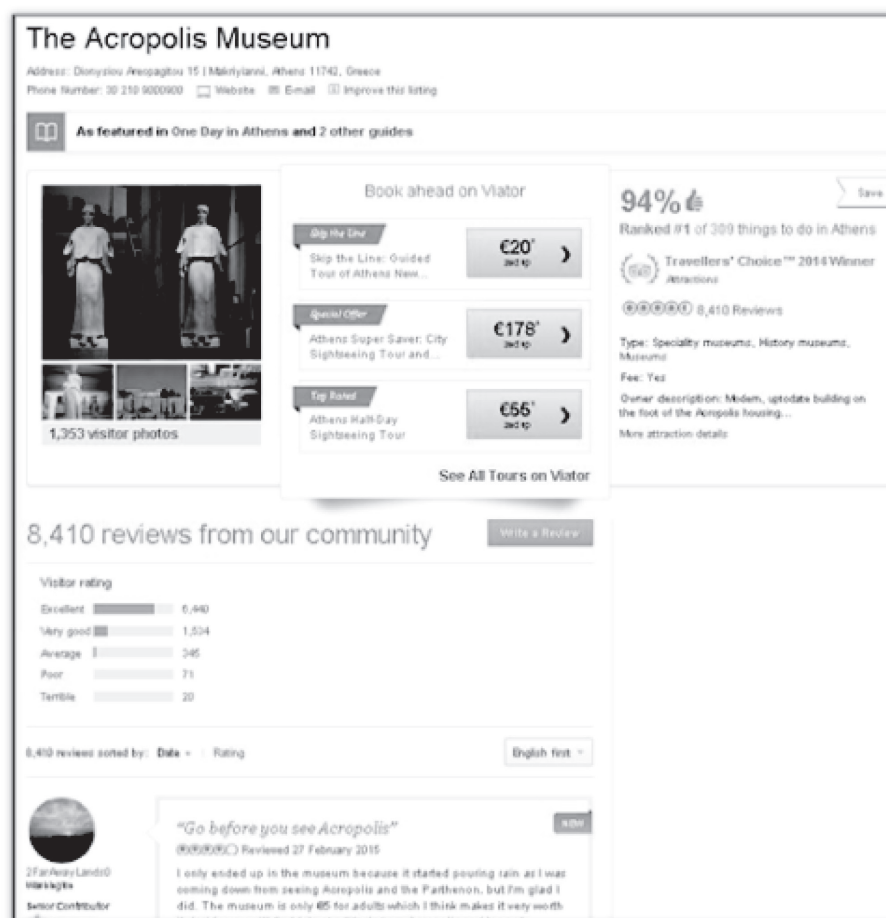
In this paper, we identify the distribution of the visitor's online ratings for 100 tourism attractions in Greece. Provided that the visitors' rating follows the Binomial distribution, we identify the parameter of Binomial distribution by using Particle Swarm Optimization (PSO). The objective of modeling is to investigate the possible bias on user ratings and the performance of fitting method compared to traditional Least Squares and Maximum Likelihood estimation. A novel statistical error measure, namely, "False Rate" is used for the rating data. We used a real world dataset of 100 Greek Tourism Attractions, downloaded from TripAdvisor website. The implications of the tourist attractions, the fitting performance and the five stars rating system is also being discussed. Keywords: Online Rating, Binomial Distribution, Fitting, Greek Attractions, TripAdvisor, Particle Swarm Optimization, False Rate.

1. Introduction

Today, the five stars rating system is a typical scheme, used in online rating of products and services. The fast growing of User Experience Sites or User Content Sites is transformed into a business model capable of attracting visitors, advertising revenues and then profits. Visitors of those internet sites are asked to give their reviews and ratings about books, hotels, articles, electronics, computer applications, etc., both for products and services. The 5 stars rating, before internet growth and diffusion in the last two decades, was known by the car safety rating system, hotels categories, etc. The use of a 5 point scale to answer an opinion question, established by a Likert scale [14] in questionnaires of opinion surveys.

According to the Likert Scale System with 5 points (choices-items), a rating question must have a middle neutral choice, two positive choices (good, very good) and two negative ones (bad, very bad) to express the responder's opinion about the question issue. For example, the implementation of five star rating system by TripAdvisor is presented in Figure 1. Users (members) of the TripAdvisor community write a review and rate an attraction, selecting one of the five available linguistic values: "Excellent", "Very good", "Average", "Poor", and "Terrible". Those 5 values symbolize the opinion of each visitor from very positive to very negative with a neutral choice ("Average") in the middle. The rate of each individual is based on his personal skill, ideas, culture, education and other factors and conditions. In right down side of Figure 1 displayed that 8.410 members' reviews and rates the specific attraction (Acropolis Museum). Follows a graphical representation of the "score" (Nr of rates distribution) each one of the five choices received. On the right upside of the Figure 1 all those 8.410 ratings are converted to a scheme of 5 circles (five stars system) with four and half of the fifth circle filled, meaning that the overall rating is about 4.5 stars out of 5. In this area of Figure 1 the 94% overall rating is promoted (a hand up icon displays that 94% of the reviewers "recommend" the attraction). This 94% overall percentage rate based on the ratings of all 8.410 reviewers. The calculation and reliability of this average percentage rate are the main subject of this paper.

Figure 1. The Acropolis Museum presentation in TripAdvisor Website



TripAdvisor [19] announced more than 200 million reviews covering more than 4.5 million accommodations, restaurants and attractions. All those millions of ratings on TripAdvisor and also millions of ratings on Amazon, Google Play, Yelp, etc., attract the public interest, especially in cases where those rating used as part of a business model. TripAdvisor uses the rating and reviews to attract visitors (travelers) to his web pages and advertising revenues from businesses related to rated objects, usually from the tourism industry domain.

The importance of ratings for business success caused phenomena like ratings and reviews fraud, questions about their trustiness and reliability. The public interest voiced by newspaper articles in W.S.J. [6] and scientific magazines like MIT Sloan Management Review [2] about rating fraud. In scientific domain the area of accommodation systems (or collaborative filtering) is affected and start to study large dataset of reviews and rating, their impact on sales, profitability, product or service or company reputation, etc.[17], [20]. Especially in the data mining field a lot of papers [3] trying to describe the relation of the text review with ratings or to discover possible fraud reviews.

Probably because of Likert Scale literature popularity, the distribution of ratings in the 5 item scale was not studied in depth. Typically in the analysis of Likert Scale questionnaires (IRT: Item Response Theory), Normal Distribution of the response distribution suggested and some statistical parameters calculated like Cronbach A [21], [16].

The distribution of ratings to rating scale is critical from another point of view. In the presence of false or biased ratings (rating manipulation) by reviewers, the examination of ratings distribution can be used as a tool to investigate and examine the biasness and manipulation. To this direction, in literature, there are published works based on review text inspection and the associated rating possible biasness [8], [9]. Also discussed from statisticians the deviation of ratings distribution from Normal Distribution, which is visible to the naked eye, typically the majority of ratings located in the two positive choices resulting to an average in the area of 4 stars. An explanation of this situation is that only people who like an attraction (or generally a rated object) review it and rate it, causing to observed reality of high overall rating.

Another interesting point [7] is the known “j-shaped” distribution of ratings observed in Amazon and other sites. The “j-shaped” term used to explain that in the ratings scale of 5 stars, very often observed an increasing number of ratings moving from the center of 3 stars in both directions of 5 stars (positive) and 1 star (negative), that seems to remind the j letter and specifically two modes (maxima) of the distribution curve.

The rest of the paper organized as follows: In section 2 present the Binomial Distribution, suitable fitting methods and some error criteria for the fitting. In section 3 Binomial Distribution fitted to a dataset of 100 Greek tourism attractions ratings¹, rated by TripAdvisor members. Finally, in section 4 some concluding remarks and suggestions for future work discussed.

2. Binomial Distribution and Fitting to Online Ratings

The rating of a TripAdvisor user is one of five linguistic values (“Excellent”, “Very good”, “Average”, “Poor”, “Terrible”), mathematically represented as a Discrete Random Variable X taking the values $x_i=1,2,3,4,5$ with probability $P(X=x_i)=f_i$, where f_i is the relative % frequencies of attraction rating by reviewers. Setting N as the number of users who rate the attraction and N_i , $i=1,2,3,4,5$ the number of users rate with value x_i , then $P(X=x_i)=f_i=N_i/N$.

Because X is defined as a Discrete Random Variable, the Expected Value of X , $E(X)$ calculated by the known equation:

$$E(X) = \sum_{i=1}^5 x_i P(X = x_i) = \sum_{i=1}^5 x_i f_i = \sum_{i=1}^5 x_i \frac{N_i}{N} \quad (1)$$

Obviously $E(X)$ is a value between 0 and 5 and is the overall rating of attraction (a weighted “average”) by all reviewers. Setting $p=E(X)/5$ (because 5 is the maximum rate by user) the obtained value p is a percentage in the range 0%-100% can be seen as overall rate %, in probability theory, explained as the **probability of attraction likeness** or **average likeness of attraction**.

For the random variable X described above, the next modeling step is to find a theoretical distribution that X follows, e.g. a statistical probability distribution function. Choosing a specific distribution is possible to test if the real data sets of ratings follow this distribution (a mathematical function).

¹ The data set is available for download at: <https://sites.google.com/site/ydimotikalis/home/onlinering/greekattractions>

2.1 Binomial and other Distributions

Let X a random variable which describes the rate of an attraction by a visitor. There are five alternatives to reviewers, X take the values $x_i=1,2,3,4,5$. An appropriate theoretical statistical distribution, is the Multinomial Distribution $M(k=5, p_1, p_2, p_3, p_4, p_5)$, where $k=5$ are the five alternative values and p_i the probability of each alternative. Although this is an optimal distribution in the case of ratings, a serious problem appears. In the real rating datasets the application of $M(k=5, p_1, p_2, p_3, p_4, p_5)$ simply implies that p_i estimated by observed relative frequencies f_i . Only if are known the theoretical values of p_i a priori, from other sources, it is possible to test for error in the real data sets, but this is not the case in this work. Multinomial Distribution is an optimal choice theoretically, but without practical implications to study the real data sets.

A special case of Multinomial Distribution is for the value of $k=2$, the Binomial Distribution $B(m, p)$, where m is the number of trials and p the probability of success in each trial. By definition in Binomial $B(m, p)$ the one alternative called success with probability p and the other failure with probability $q=1-p$. For the five stars rating the random variable $X \sim B(m=4, p)$, the acceptance of Binomial $B(4, p)$ means that each user rating is the result of 4 Bernoulli trials, each one with probability of success p , which remain constant on all 4 trials and each trial is independent with others. To explain the case, each reviewer in his/her rate selection (mathematically one of the numbers 0,1,2,3,4) thinks of 4 equally important issues and if none of 4 is satisfied then select 0, if one of 4 satisfied select 1, etc.

The classical example in statistics for $B(m, p)$ is the coin tossing procedure. If a "fair" coin tossed 4 times, then the probabilities of results in Heads (success) and Tails (failure) is given by $B(m=4, p=0.50)$. The number of successes (Heads) ranges from 0 to 4, with most probable value 2, because the probability of success is $p=0.50=50\%$ and $p \cdot m = 0.50 \cdot 4 = 2$. Back to ratings, the probability of success p could be explained as probability of attraction likeness p and each visitor rate is expected to depend on this value. An important point is that, as in the "fair" coin tossing, the result of 4 trials is not always as expected: 2 Heads and 2 Tails, but this is the most likely result. All the other alternatives (events in statistics): 3 Heads-1 Tail, 1 Head-3 Tails, 0 Heads-4 Tails, 4 Heads-0 Tails, is possible to observed, but with different probabilities (frequencies in a series of coin tossing experiments). This means that, as in the coin tossing example, for ratings is expected to observe not only one particular rate selected, but all the five alternatives with different frequencies (probabilities), which are dependent on probability of attraction likeness parameter p defined before.

If a random variable $X \sim B(m, p)$, follows the Binomial Distribution with parameters m, p , then, the formula of the Binomial Distribution probabilities is:

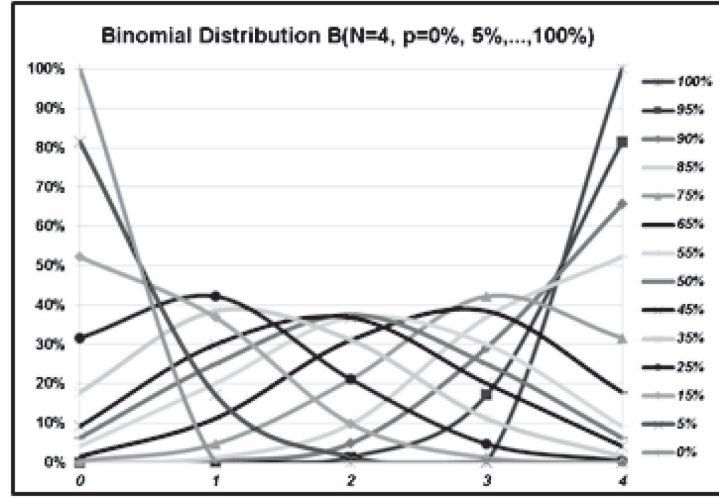
$$P(X = x) = \binom{m}{x} p^x (1 - p)^{m-x} = \frac{m!}{x!(m-x)!} p^x (1 - p)^{m-x} \quad (2)$$

Where m is the number of trials, x the number of successes and p the probability of success in each trial. In our case $m=4$ and $x=0,1,2,3,4$ (corresponding to 1,2,3,4,5 stars respectively) and p the probability of attraction likeness. Then for each value of x eq. 1 gives the following set of equations:

$$\begin{aligned} P(1 *) &= P(X = 0) = (1 - p)^4 \\ P(2 *) &= P(X = 1) = 4p(1 - p)^3 \\ P(3 *) &= P(X = 2) = 6p^2(1 - p)^2 \\ P(4 *) &= P(X = 3) = 4p^3(1 - p) \\ P(5 *) &= P(X = 4) = p^4 \end{aligned} \quad (3)$$

Where the parameter p is the probability of attraction likeness. In Figure 2 we present the values of the above equations 3, Binomial Distribution $B(m=4, p)$, for values of $p=0\%, 5\%, 10\%, \dots, 95\%, 100\%$. The Binomial is a discrete distribution; in each colored curve of the graph there are only five values (points); the connecting smooth lines are used for improved optical illustration and understanding. The curves of parameters' values $p\%$ and $1-p\%$ are symmetrical around the average value of 2, only the curve for $p=50\%$ is symmetrical. For $p=50\%$ is well known the Normal approximation of Binomial Distribution.

Figure 2. Binomial Distribution Curves for different P parameter values



Another Discrete Distribution is the Uniform Distribution $U(m)$, where m is the number of alternatives. This distribution is practically appropriate when all the alternatives have equal probability (like a fair dice). Iannario and Piccolo [10] use Uniform as a part of a mixture model, to take into account the case where part of raters makes random choices. In Likert scale literature the Normal Distribution of Item Response Rate is the standard hypothesis, but this explained mainly because of the several questions analyzed in a questionnaire. Also, it is known that Normal Distribution is the limit of Binomial Distribution when the number of trials (alternatives in rating) increases and is greater than 30, practically. In our opinion the 5 star ratings do not expect to follow Normality, as a result of only 5 alternatives (stars) used.

2.2 Distribution Fitting Methods to Real Rating Data

To fit a mathematical model to real data, Least Squares estimation is the standard mathematical method. In the case of theoretical statistical distributions fitting to frequencies data, from a statistical point of view, Maximum Likelihood Estimation (M.L.E.), proposed by Fisher [5], is the preferable method. In the Binomial distribution case the M.L.E. of p is the expected value $E(X)=p$ and calculated as the weighted average of the relative observed frequencies (see eq. 1). We call this value p_{MLE} , estimation of p parameter by M.L.E. method.

Also the goodness of the fit X^2 statistical test, proposed by Karl Pearson [17] in 1900, is a method to examine the goodness of fit of a statistical distribution function to real data. The value of the test statistic X^2 given by:

$$X^2 = \sum_{i=0}^4 \frac{(O_i - E_i)^2}{E_i} = \sum_{i=0}^4 \frac{(f_i - P(X=i))^2}{P(X=i)} \quad (4)$$

where O_i are the observed and E_i the expected frequencies. Pearson paper include some examples of “big datasets” like the throw of 12 dices 26306 times by prof. Weldon and computation the theoretical frequencies of the observed 5 or 6 using Binomial Distribution $B(m=12, p=1/3)$. In some way our analysis of ratings is the opposite: given the observed frequencies of each 5 stars scale, calculate the probability p .

Least Squares method is the minimization of the Sum of Squared Errors (S.S.E.):

$$S.S.E. = \sum_{i=0}^4 (P(X=i) - f_i)^2 \quad (5)$$

where f_i are the observed (real data) relative frequencies of ratings and $P(X=i)$ is the probabilities (theoretical frequencies), calculated by Binomial distribution function (eq. 3). The modern Spreadsheets solver tool can be used to find the value of parameter p that minimizes the S.S.E. We call the estimated value of parameter p by Least Squares p_{LSE} .

The source of error in every real data set, generally called noise, in rating can be viewed as false (unexpected) rates from some of the visitors of the attraction. A detailed analysis of this error or noise is needed.

2.3 Fitting Criteria and Error of Rating Data

By fitting Binomial Distribution to real data a value of p parameter calculated and then by eq. 3 the theoretical frequencies estimated (associated with the observed frequencies f_i). Using theoretical (Binomial) $P(X=i-1)$ and observed frequencies (Data) f_i several error criteria are available.

After the fitting (estimation of Binomial p parameter) calculated 5 error values $e_i = P(X=i-1) - f_i$, $i=1,2,3,4,5$ and is known that the sum of all errors in fitting is zero $\sum e_i = 0$. Either all the errors are zeros (perfect fit) or some of them are positive the others negative and the sum is zero. When $e_i \geq 0$ then $P(X=i) \leq f_i$ the Binomial overestimate and when $e_i \leq 0$ then $P(X=i) \geq f_i$ the Binomial underestimate real data. But because of the $\sum e_i = 0$ in every data set the overestimation errors (positive) are equal (in absolute value) to underestimation errors (negative). If only one rater moves his vote from one point of scale to another then he produces an underestimation error to the first point and an overestimation error to the second point. Every one false rate by a visitor produces a double error, a positive one and a negative one at the same time. This remarkable property of error in rating can used to define the False Rate (F.R.) as the number of users who provided false (unexpected) rates. Mathematically, it is related to the Sum of Absolute Error (S.A.E.) defined by:

$$S.A.E. = \sum_{i=0}^4 |e_i| = \sum_{i=0}^4 |P(X=i) - f_i| \quad (6)$$

where f_i are the observed relative frequencies of ratings and $P(X=i)$ is the theoretical frequencies, calculated by the Binomial distribution function (eq. 3). Because of the equality of positive and negative errors e_i , S.A.E. will be twice the False Rate %, thus $F.R.\% = \frac{1}{2} S.A.E. = \sum |e_i| / 2$. This F.R.% is easily explained to everyone as the percentage of the whole N users whose rates are false (different than expected by the Binomial Distribution). The source or cause of false rate can be random noise (not a specific reason) or biasness (an intention to change the expected average rate of the attraction).

By F.R. value, checking the particular point (value) where the expected rates are more than expected it is possible to discover the outcome of false rate. If rate moved to a higher point of scale it is expected to increase the average value of the rating parameter p , or the opposite if moved to lower points. Both cases can happen for certain reasons. Our approach to set the F.R. error as an indication of fraud or manipulation is very clear. Because of the probabilistic nature of the rating procedure, small error values on ratings is expected, bigger values of error are suspicious for manipulation. The examination of the scale values where the error observed helps to discover possible biasness.

Each one of the equations 4, 5, 6 based on error e_i , can be used as a criterion for estimating the p parameter. Eq. 4 can be seen as a weighted S.S.E. (eq. 5), where $1/P(X=i)$ are the weights. The question of the convergence to the same parameter p value is discussed in our results. Because of direct relation of F.R. to S.A.E., the S.A.E. is used as a preferred fitting criterion.

2.4 Calculation of Optimal

The shape of a binomial distribution given in the form of Eq. 2 is parametrically defined by the value of parameter. The optimal value of parameter is the one which minimizes the error norm:

$$E(p) = |f_i - P_i(X=i)|_L \quad (7)$$

where $L = 0.1, 2, \dots, \infty$. In this work we used $L=1$, which is equivalent to the Sum of Absolute Errors (S.A.E.) and $L=2$ which is equivalent to the Sum of Squared Errors (S.S.E.). Since the equality constraint $\sum_{i=0}^4 P_i = 1$

must be hold, the problem of calculating the optimal value of p can be formulated as a typical constraint optimization problem in the form of Eq. 8:

$$\begin{aligned} & \text{minimize : } E(p) \\ & 0 \leq p \leq 1 \\ & \text{s.t.} \\ & \sum_{i=0}^4 P(X=i) = 1 \\ & 0 \leq P(X=i) \leq 1, i = 0, 1, 2, 3, 4 \end{aligned} \quad (8)$$

Actually, $P(x=i)$ values depend on the value of parameter p as it is explained by Eq. 2. Generally, the fulfillment of equality constraints is a hard job for any optimization algorithm. Typically, a constraint optimization problem can be transformed into an unconstrained one by imposing a penalty function to deteriorate the infeasible solutions (the solutions which violate the constraint). We adopt the following penalty function:

$$PF(p) = a \cdot \left| 1 - \sum_{i=1}^4 P(x=i) \right| \quad (9)$$

Where a is an arbitrarily large, non-negative, real number. Please note that the inequality constraints in Eq. 8 are guaranteed, by definition. The equality constraint was dealt with by using the penalty function of Eq. 9 to penalize the infeasible solutions. The constraint optimization problem of Eq. 8 is transformed into the unconstrained optimization problem in the form of Eq. 10:

$$\begin{aligned} \text{minimize : } E(p) + a \cdot \left| 1 - \sum_{i=1}^4 P(x=i) \right| \\ 0 \leq p \leq 1 \end{aligned} \quad (10)$$

The problem of Eq. 10 can be solved by using any non-linear optimization method. In this work we use a particle swarm optimization method. Particle swarm optimization, or PSO for short, was initially introduced by Kennedy [12]. It is a stochastic, evolutionary, derivative free, optimization technique with strong ability of escaping from local optimum as well as plateau regions of the search space. Moreover, the simplicity of PSO's implementation was an additional reason for employing it.

2.5 PSO Description

A particle swarm optimizer includes a population of particles. Each particle is an entity consisting of a *position* and a *value of importance*. The position of each particle encodes a solution of the problem being solved, while the value of importance is an evaluation of the respective solution. The position of particles is updated through simple equations (motion equations) by using the particle's previous information (local knowledge) as well as information from other particles (social knowledge). The objective of update is the improvement of particles' position, that is, the discovery of better solutions. The motion equations for a particular particle are given by Eq. 11:

$$\begin{aligned} V(t+1) &= \alpha(t) \cdot V(t) + c_1 \cdot u(0,1) \cdot [X(t) - x(t)] + c_2 \cdot u(0,1) \cdot [G(t) - x(t)] \\ X(t+1) &= X(t) + \beta(t) \cdot V(t+1) \end{aligned} \quad (11)$$

, where: t denotes the current time step of evolution; $\alpha(t), \beta(t)$ namely, inertia and restriction factor respectively, could be either time decreasing or constant real numbers; $c_1, c_2 \in (0,1] \subseteq \mathbb{R}$ are predefined, constant, real numbers, namely, *learning factors*; $x(t)$ is the current position of particle at time step t ; $X(t)$ is particle's best known position, while $G(t)$ is the population's best known position; $u(0,1)$ is a random number in the range $[0,1]$, generated by a uniform random $[0,1]$, number generator. $V(t)$ is the velocity of particle at time t . At the beginning ($t=0$) of evolution all particles have random positions: $x(0)=u(x_{min}, x_{max})$, where x_{min}, x_{max} are the minimum and the maximum allowed value of particle's position. Each particle is evaluated according to the value of the objective function for the specific position (solution) and the best position over all particles ($G(t), t=0$) of the population is memorized. The local best ($X(t), t=0$) of each particle is also memorized. At the next time step ($t=1$) a new position for each particle is calculated by Eq. 11. The new positions are reevaluated and both local ($X(t), t=1$) and global ($G(t), t=1$) are updated (if a better position was achieved). The evolution is terminated at time t_{stop} when a predetermined *stopping criterion* is reached. Usually, the stopping criterion is either a predefined number of time steps (T_{max}), or a predefined number of successive no improvements of $G(t)$. The final solution is the best attained position $G(t_{stop})$.

In this work we employed a PSO with a population of ten particles. The position of each particle was a real number in the range $[0,1]$, representing the value of parameter p . The evaluation function was the objective function given by Eq. (10). According to the norm in Eq. 7 we used either $L=1$ or $L=2$ with equivalent results. The stopping criterion was a predefined number of successive no improvements for 100 time steps. The total time for 100 attractions was 75sec, or less than a second per attraction.

3. Fitting Binomial Distribution to Greek Attractions Rating Dataset

3.1 The Greek Attractions Dataset

The website of TripAdvisor is organized to present tourism destinations worldwide. Each destination includes a number of attractions, mainly suggested by site users. From the TripAdvisor website, we collected the data of 100 Greek Attractions ratings at 1-10/2/2015, a sample of more than 1000 rated attractions of the most popular 30 Greek tourism destinations. Our sample includes 33 beaches, 21 archeological sites, 11 museums, 31 natural or other interest sites, 2 water parks and 2 aquariums, totally rated by 95897 members of TripAdvisor. Each attraction was rated by a number of visitors. In our sample the average was 959 rates/attraction, fluctuating from a minimum of 173 to a maximum of 8153 with a median of 651. The average rating given by TripAdvisor to our dataset of attractions was 87.70% having: min of 56%, max 99% and median 89.50%.

3.2 Statistical Analysis of Fitting Results

The fitting statistics of the parameter p value, for the whole dataset of 100 attractions, is presented in Table 1. The second column presents the values of p provided by TripAdvisor (see Figure 1, upper right side). The values of $p_{M.L.E.}$ (3rd column of table 1) calculated by eq. 1 (Maximum Likelihood Estimation for Binomial Distribution). The next two columns of table 1 present $p_{S.S.E.}$ and $p_{F.R.}$ calculated by the P.S.O. by minimizing the S.S.E. and F.R.%, respectively. The average p value of F.R. 87.32% is slightly smaller than that of S.S.E. and about 1% smaller than that of M.L.E.

Table 1. Fitting Statistics by Method in 100 Greek Attractions

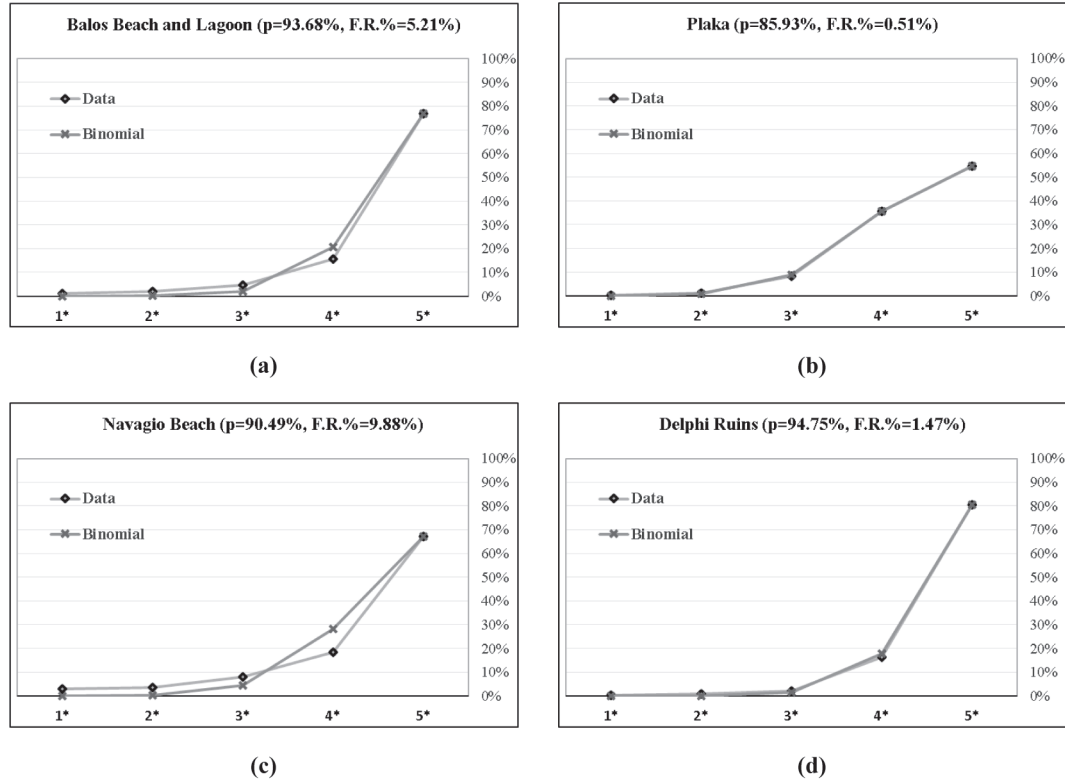
Method (Criterion)	TripAdvisor	M.L.E.	S.S.E.	F.R.
Average P	87.50%	88.63%	87.56%	87.32%
Min P	56.00%	72.10%	67.21%	68.36%
Max P	99.00%	98.32%	98.13%	98.06%
Median P	89.00%	89.29%	88.60%	88.51%
False Rate Average	-	6.27%	3.95%	3.64%
False Rate Min	-	0.97%	0.08%	0.08%
False Rate Max	-	13.98%	12.99%	11.51%
False Rate Median	-	5.65%	3.20%	2.92%

Observing the values of False Rate Average (6th line of table 1) by F.R. criterion, the improvement over M.L.E. is significant (42.86% on average from 6.27% to 3.64%) and about 7.85% smaller than S.S.E. (3.64% from 3.95%). The minimum value of F.R.% is 0.08% in the 100 attractions and corresponds to Sacred Monastery of Arkadi in Crete. The value of F.R. is zero in Nr of raters (all 486 users rates were exactly as it was expected!). Only 4 attractions have F.R.% values greater than 10% including 2 beaches and 2 sites. The 2 beaches and 2 sites, which were very suspicious for manipulation. TripAdvisor announces yearly the top 20 beaches of the world voting, as a promotional issue of his site to mass media, possibly this is the cause of the observed beaches rating manipulation. The case is illustrated in Figure 4(c) where in the 4th point the difference the actual and the expected rates was high. This is an indication of manipulation for increasing attraction's overall rate (likeness).

The value of F.R. in the attractions, illustrated in Figure 4, is presented in parenthesis of each chart title with the best value of the fitted parameter p (by minimization of F.R.%). In Figure 5 the histogram (empirical distribution) of False Rate % error to our 100 datasets is illustrated.

By M.L.E. criterion the curve of F.R.% value distribution is centered around the average of 6%, it is not precisely symmetrical, but follows a bell shaped curve. The other two curves, F.R.% by S.S.E. and S.A.E. criterion are clearly better concentrated on the lowest error values 1-2-3%. The S.S.E. and S.A.E. (F.R.%) curve are very close each other, but have a serious difference with the third of M.L.E., the first two are of "binomial" type, with higher frequencies to lower values of False Rate%. An explanation of this characteristic of M.L.E. curve explained by observing eq. 1, the computation of $p=E(X)$ treats all the observed data equally, even those that are false (manipulated). Then the calculated $p_{M.L.E.}$ value, a priori affected by false rates. This is an undesirable property for ratings data, but it is known in principle that M.L. Estimation is the best estimator of the observed data $E(X)$.

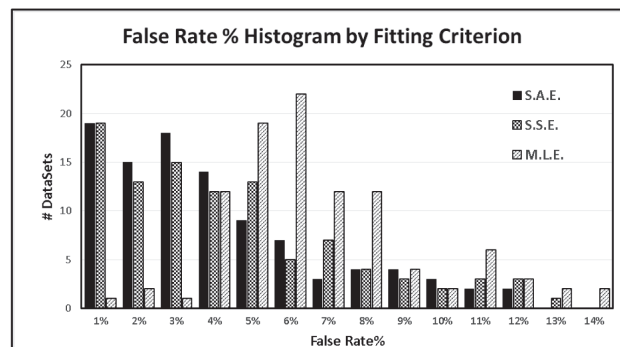
Figure 4. Fitting to Greek Tourism Attractions Examples



Finally, to test the linear relation of several parameters of ratings data, some linear regressions performed by using the sample of 100 data sets values of the following “parameters”:

- I. **Y: False Rate % vs X: Nr of Attraction Rates**, the equation of linear regression is: $y=2E-07x+0.0363$, $R^2=6E-05$, it is clear that there is not a linear correlation between the 2 variables. In our sample the Nr raters do not affect the False Rate, i.e. even with small nr of raters the False Rate performs well.
- II. **Y: False Rate % vs X: TripAdvisor Rating**, the equation of linear regression is: $y=-0.2606x+0.2645$, $R^2 = 0.5292$, the value of R^2 is greater than 50% and the x coefficient negative, means that in average greater values of TripAdvisor Rating corresponds to lower False Rate. This is the stronger linear relation found. Explanation: TripAdvisor treats the rates and uses some unknown algorithm to adjust the calculated rate. Our False Rate is able to partially uncover the unknown TripAdvisor algorithm.
- III. **Y: False Rate % vs X: $p_{S.A.E.}$** , the values of linear regression are: $y = -0.2496x + 0.2544$, $R^2 = 0.1965$, the value of R^2 is low and the x coefficient negative but not significant. It means that False Rate% does not linearly dependents on $p_{S.A.E.}$ value.

Figure 5. False Rate % by fitting criterion



4. Conclusions and Suggestions

The approximation of Greek Attractions ratings is generally good or very good by a simple mathematical model. The model is the Binomial Distribution with one parameter p , the parameter is explained as the attraction likeness probability. The probabilistic empirical distribution of ratings appears to follow simple statistical rules. The known “j-shaped” distribution of ratings phenomenon [7] is not present or more precisely is very limited to affect the fitting in our dataset, seriously. The best fitting in terms of F.R.% is significantly better than that of M.L.E. on attractions ratings data. This was also previously observed to other data sets from TripAdvisor (Hotels) in [4].

The False Rate% introduced and used as error and fitting criterion is very clear and understandable by all the audience. Our fitting dataset includes several Greek attractions where the False Rate achieved is extremely low (in the values of 1-2%). The higher False Rate% values is an indicator of possible rating manipulation in both sides of the 5* rating scale.

Our fitting results are useful to authorities and businesses related to tourism attractions. The F.R.% value is a sign of possible problems and “irregularities”. The examination of biggest error point in 5* scale (visible to fitting chart), in parallel with inspection of text reviews, is capable to help the discovery of the source and the cause of the issue. By our fitting procedure the likeness of the attraction does not change with the addition of few new rates as happen to $E(X)$ value regularly.

The P.S.O. method is perfect as the fitting method, because of the p parameter definition as probability. In standard unconstrained nonlinear regression methods tested, the problem of non-converging starting values occur, a disadvantage not present on P.S.O. method.

Motivating future directions of this work:

- I. The used Binomial Distribution fitting can tested to other rating datasets like that of apps rated on Google Play, products on Amazon, movies on IMDB, restaurants on YELP, etc. Also, large datasets are available and suitable to test the evolution time of parameter p [13]. The time variation of p is an interesting question [15], especially on tourism attractions where the factors and issues evaluated by visitors may vary over time.
- II. Modification of Binomial model and test of similar models (distributions) to improve fitting performance, variants of such models are known in the literature [1], [11].

References

- Allik J., (2014). A mixed-binomial model for Likert-type personality measures, *Frontiers in Psychology*, Vol 5, 371.
- Aral S., (2013). The Problem with Online Ratings, *MIT Sloan Management Review*, 55(2).
- Chung C.-Y., Hsu P.-Y., Huang S.-H., (2013). βP : A novel approach to filter out malicious rating profiles from recommender systems, *Decision Support Systems* 55, 314-325.
- Dimotikalis Y., (2014). Fitting Binomial Distribution to Online Rating Data: TripAdvisor Ratings in Crete Island, *SMTDA2014: 3rd Stochastic Modeling Techniques and Data Analysis International Conference*, 11-14 June 2014, Lisbon Portugal.
- Fisher R. A., (1922). The goodness of fit of regression formulae and the distribution of regression coefficients, *J. Roy. Statist. Soc.* 85, 597-612.
- Grant K. B., (2013). 10 things online reviewers won't say: Why consumers can't trust five-star ratings - or nasty complaints, *MarketWatch, The Wall Street Journal*, Available at: <http://www.marketwatch.com/story/10-things-online-reviewers-wont-say-2013-03-01> accessed: 20/1/2015.
- Hu N., Zhang J., and Pavlou P. A., (2009). Overcoming the J-shaped distribution of product reviews, *Commun. ACM* 52, 10, pp. 144-147.
- Hu N., Liu L., Sambamurthy V., (2011a). Fraud detection in online consumer reviews, *Decision Support Systems* 50, pp. 614-626.
- Hu N., Bose I., Gao Y., Liu L., (2011b). Manipulation in digital word-of-mouth: A reality check for book reviews, *Decision Support Systems* 50, pp. 627-635.
- Iannario M. and Piccolo D., (2010). A New Statistical Model for the Analysis of Customer Satisfaction, *Quality Technology & Quantitative Management*, Vol. 7, No. 2, pp. 149-168.
- Kadane J. B., Krishnan R., Shmueli G., (2006). A Data Disclosure Policy for Count Data Based on the COM-Poisson Distribution, *Management Science*, Vol. 52, No. 10, pp. 1610-1617.
- Kennedy, J. and Eberhart, R. C. (1995). Particle swarm optimization. *Proc. IEEE intl conf. on neural networks* Vol. IV, pp. 1942-1948.
- Leskovec J. and Krevl A., (2014). SNAP Datasets: Stanford Large Network Dataset Collection, <http://snap.stanford.edu/data>.
- Likert R., (1932). A technique for the measurement of attitudes, *Archives of Psychology*, Vol 22, No 140, 55.
- McAuley J. and Leskovec J., (2013). From Amateurs to Connoisseurs: Modeling the Evolution of User Expertise through Online Reviews, *WWW 2013*, Rio de Janeiro, Brazil.
- Norman G., (2010). Likert scales, levels of measurement and the “laws” of statistics, *Adv in Health Sci Educ* 15, pp. 625-632.
- Park D. H., Kim H. K., Choi I. Y., Kim J. K., (2012). A literature review and classification of recommender systems research, *Expert Systems with Applications* 39, pp. 10059-10072.
- Pearson Karl, (1900). “On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling”, *Philosophical Magazine Series* 5, 50 pp. 157-175.
- TripAdvisor (2014). About TripAdvisor, Available at: http://www.tripadvisor.com/PressCenter/c6About_Us.html accessed: 20/1/2015.
- Zhao Y., Yang S., Narayan V., and Zhao Y., (2013). Modeling Consumer Learning from Online Product Reviews, *Marketing Science* 32, 1, pp. 153-169.
- Zumbo B. D., Gadermann A. M., and Zeisser C., (2007). Ordinal Versions of Coefficients Alpha and Theta for Likert Rating Scales, *Journal of Modern Applied Statistical Methods*, Vol. 6, Iss. 1, Article 4.