The Burr Distribution

Jackson Curtis

Brigham Young University

Abstract. This project will explore the properties of the Burr distribution, a positive-support distribution useful for modeling household incomes. Maximum likelihood and other estimators are discussed, real data is analyzed, and a simulation study is performed.

1 Introduction

The Burr distribution is a two-parameter, continuous probability distribution with positive support. Its probability density function is given by:

$$f(x|c,k) = ck\frac{x^{c-1}}{(1+x^c)^{k+1}} \tag{1}$$

The Burr distribution is often used for modeling household income. Singh and Maddala (1976) derive the Burr distribution from properties of its hazard function, and argue that it fits income data in both tails better than the log-normal or gamma distributions. In this paper we will explore the properties of the Burr and demonstrate its use with real data on the household incomes of families in the Philippines.

2 Maximum Likelihood Estimators

In order to find estimators for our two parameters, we will compute the log-likelihood function and use the Newton-Raphson method to derive the maximum likelihood estimators. The distribution of n independent Burr random variables is given by:

$$f(x_1, x_2, ..., x_n|c,k) = \prod_{i=1}^{n} ck\frac{x_i^{c-1}}{(1+x_i^c)^{k+1}} \tag{2}$$

From this, the log-likelihood becomes:

$$\ell(x_1, x_2...x_n, c, k) = nlog(c) + nlog(k) + \sum_{i=1}^{n}(c-1)log(x_i) - \sum_{i=1}^{n}(k+1)log(x_i^c+1) \tag{3}$$

We will need the partial derivatives with respect to c and k in order to find the maximum likelihood estimators:

$$\frac{\partial \ell}{\partial c} = \frac{n}{c} + \sum_{i=1}^{n} log(x_i) - \sum_{i=1}^{n}(k+1)\frac{x_i^c log(x)}{x_i^c+1} \tag{4}$$

$$\frac{\partial \ell}{\partial k} = \frac{n}{k} - \sum_{i=1}^{n} log(x_i^c + 1) \tag{5}$$

By setting (5) equal to 0 and solving for k, we can obtain the maximum likelihood estimator for k in terms of c:

$$\hat{k} = \frac{n}{\sum_{i=1}^{n} log(x_i^{\hat{c}} + 1)} \tag{6}$$

$\hat{c}$ cannot be solved for analytically and thus we implement the Newton-Raphson method to calculate it numerically. In order to do that we will need the second partial derivative of c, which is given by:

$$\frac{\partial^2 \ell}{\partial c^2} = \frac{n}{c^2} - (k-1) \sum_{i=1}^{n} \frac{x_i^c log(x_i^2)}{(x_i^c + 1)^2} \tag{7}$$

3 Percentile Matching

Another way to obtain estimates for the parameters c and k is through percentile matching. Percentile matching is similar to the method of moments, but instead of setting sample moments to theoretical moments, we set calculated quantiles to the distribution's theoretical quantiles. We can calculate the theoretical quantiles using the inverse CDF:

$$f^{-1}(Q) = ((\frac{1}{1-Q})^{\frac{1}{k}} - 1)^{\frac{1}{c}} \tag{8}$$

where Q is the quantile of interest. By using the median and the 75th quantile, solving for c and k, and using a root finding algorithm to find an approximation for c, we can obtain estimators for the parameters.

4 Bayes Estimator

A third way to obtain parameter estimates is through Bayesian estimation. In order to obtain estimators using Bayes rule we must assign priors to our parameters of interest, c and k. For this I chose to model both c and k with a Gamma(2.5, .5) using the rate parameterization. This puts the mean of the distribution at 5 and spreads the bulk of the density between 0 and 10. This allowed for a diffuse prior over the range of values of interest.

Using these priors and our Burr likelihood we are able to obtain a posterior density:

$$\pi(c, k | x_1, x_2, ...) \propto (ck)^{n+1} e^{-.5(c+k)} \prod_{i=1}^{n} \frac{x_i^{c-1}}{(x_i^c + 1)^{k+1}} \tag{9}$$

Using this posterior we can implement a Metropolis-Hastings algorithm which allows us to obtain draws from the posterior without knowing the normalizing constant. We use the mean of the posterior draws as our point estimate.

5 Random Number Generation

To generate draws from a Burr distribution with specific parameter values, we can use the probability integral transformation. By specifying c and k and plugging in uniform(0,1) draws into equation (8), we can obtain draws from any Burr distribution.

6 Household Income in the Philippines

To demonstrate the estimators of a Burr distribution, I obtained a data set containing annual incomes for 41,000 households in the Philippines. The data was collected by the Philippine Statistics Authority and is part of the public domain. The data is in Pesos (approximately $.02 USD), so I divided all incomes by 100,000 for computational feasibility when calculating estimators.

If we are interested in making inferences about household incomes, we need a model that accurately represents the variation between families. We fit a Burr distribution with our three different estimators to see whether this is an accurate model of incomes.

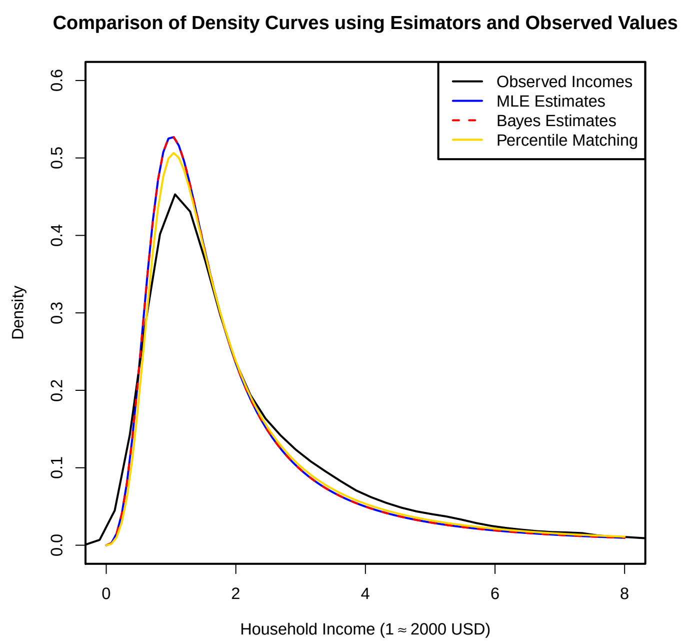

Fig. 1: Observed data and three fitted Burr distributions

The first observation from the plot is that our Burr models fail to capture the weight of the right tail of the distribution, and so assigns too much probability at the mode of our distribution. Secondly, in this case (a large sample size), percentile matching seems to be as good, if not a little better, than our other methods of estimation. Finally, our MLE and Bayes estimates are almost inseparable. This can likely be attributed to our diffuse prior and our large sample size.

Table 1 shows our point estimates for our parameters as well as 95% confidence intervals for the parameters. For the Bayes estimate we obtain confidence interval by sampling from the posterior distribution of c and k. For the other two estimators we obtain confidence intervals by bootstrap sampling the data.

Estimator	Parameter	Point Estimate	95% Confidence Interval
Maximum Likelihood	c	3.473	(3.34, 3.52)
	k	0.401	(0.40, 0.41)
Percentile Matching	c	3.641	(3.45, 3.91)
	k	0.350	(0.33, 0.37)
Bayes Estimator	c	3.346	(3.34, 3.51)
	k	0.401	(0.40, 0.41)

Table 1: Parameter estimation and confidence intervals for income data.

7 Simulation

Our large, real data set showed many similar results in our three methods of obtaining estimates. To get to a better understanding of the methods strengths and weaknesses, we can perform a simulation study. To perform the study we will generate random samples from the Burr distribution under three different parameter combinations and three different sample sizes and then see how well our estimators replicate the underlying parameters.

The following table summarizes the different scenarios and sample sizes:

	c	k
Scenario 1	2	8
Scenario 2	7	3
Scenario 3	5	5

Table 2: Each scenario will be simulated 1000 times at n=20, 100, and 1000

We chose the parameter values to investigate when k is bigger than c, when c is bigger than k, and when they are even. Our sample sizes where chosen to reflect a small sample, medium sample, and large sample of data.

For each scenario and sample size combination, I ran 1000 simulations and calculated the mean squared error and bias from the estimates produced. The following graphs summarize my findings:

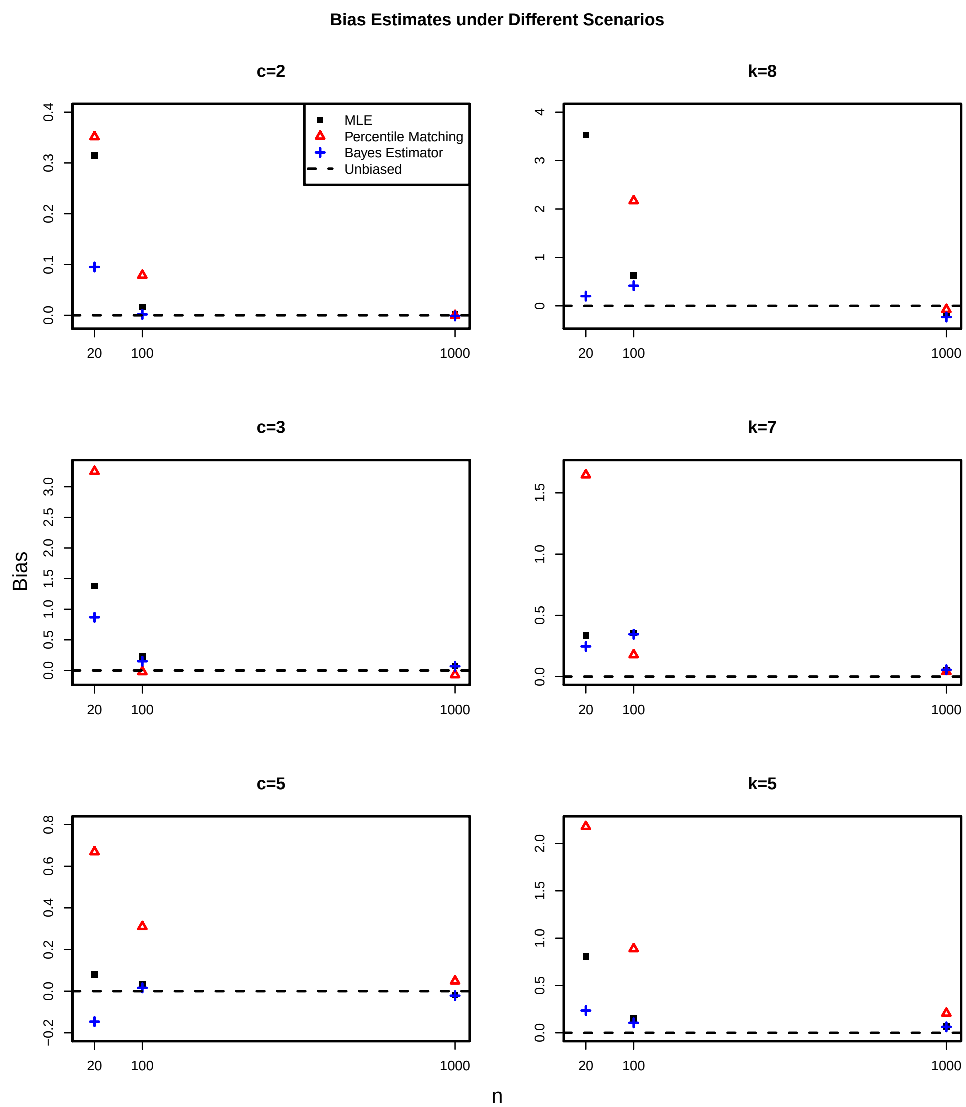

Fig. 2: Simulated bias estimates for parameter estimation (each row is one scenario)

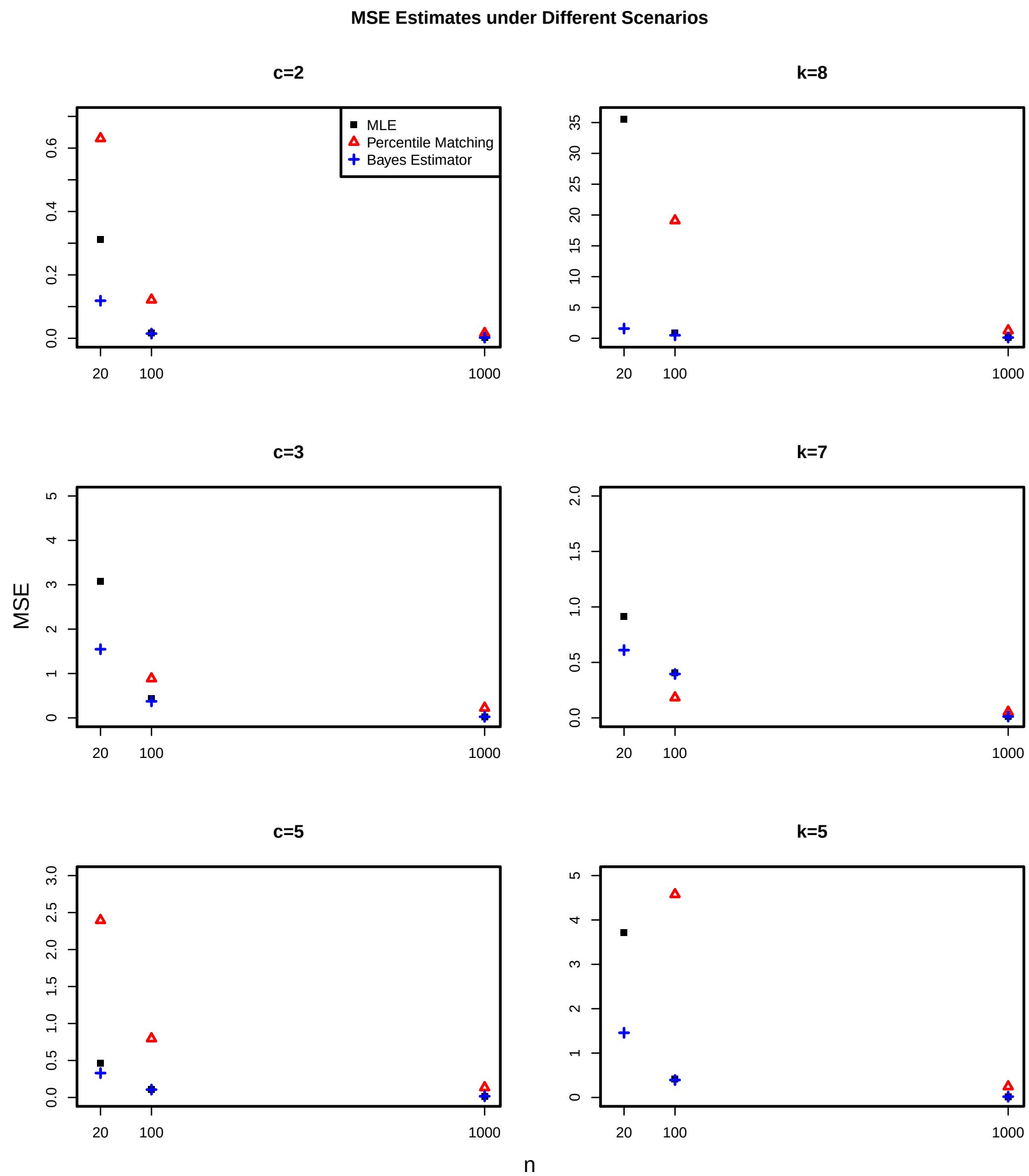

Fig. 3: Simulated MSE estimates for parameter estimation (Note: when percentile matching does not appear on the graph it is because it was significantly higher than any other point on graph)

An obvious conclusion from our simulation is that percentile matching performs far worse than the other two methods in terms of bias and mean squared error, particularly for a sample size of twenty. One explanation for this is that our quantiles are extremely uncertain with that few data points, and so they don't well represent the distribution. Another observation is that while both MLE and Bayes perform quite well with 100 samples, Bayes tends to do better for the smaller sample size. This might be because our priors were quite close to the actual values of our parameters. At 1,000 data points our estimators do extremely well at estimating the parameters. Interestingly, almost all measures of bias overestimated the parameters, in only a few scenarios was our bias negative.

8 Model Misspecfication

So far our simulations have studied data that was actually generated under a Burr-like process. A natural question is to wonder how the model handles data that may have similar qualities to the Burr (positive support, right-skewed), but come from a different distribution. For this section we compare what would happen if our data generating process was a gamma distribution instead of a Burr.

For the first case we generated 100 samples from a Gamma(1.5,3) distribution. Figure 4 shows the plotted densities of our estimators along with our theoretical and sampled values. Interestingly, the MLE and Bayes estimator seem to capture the true theoretical distribution better while percentile matching seems to capture the observed data better.

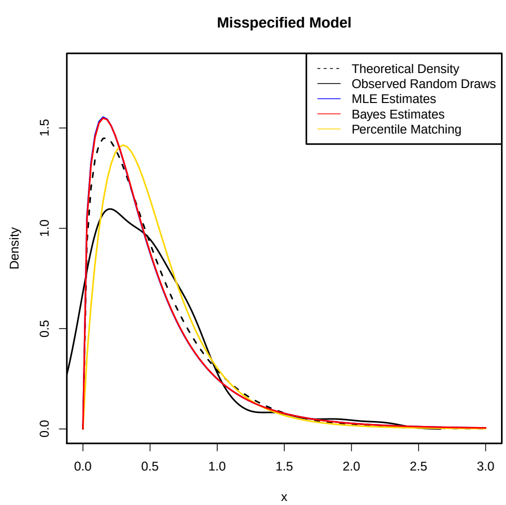

Fig. 4: Burr's imitation compared to gamma density

While the approximations look reasonable for small alpha values like the example above, as alpha gets bigger and becomes larger than beta, the Burr distribution does a poor job of modeling the data. In the following example our data follows a gamma(5,2) and we see that the MLE and Bayes are both very far away from the true distribution of the data. In addition, the percentile matching algorithm is unable to find a solution to its root-finding for the c parameter, and is thus not able to produce an estimate.

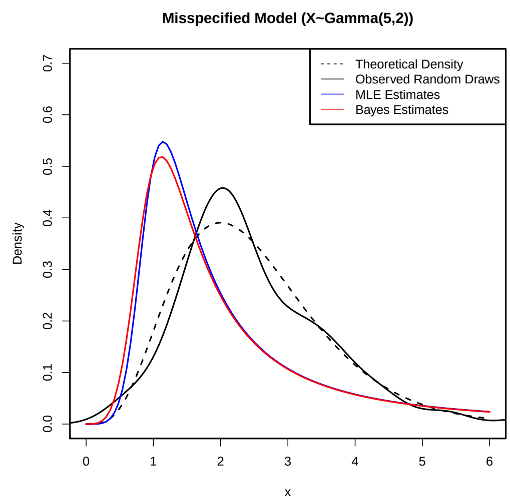

Fig. 5: Burr inaccurately models gamma when $\alpha > \beta$

From this we conclude that extreme caution should be taken when fitting the Burr distribution when the mean increases and the data generating process is uncertain.

9 Conclusion

The Burr distribution may be a useful model to consider when fitting models of household income and similarly skewed data. The relative simplicity of the pdf and derivatives makes it easy to fit the MLE and Bayes estimators. A concern of using the Burr distribution may be its lack of versatility, especially when fitting models that may come from other distributions. However, if the data is properly modeled by the distribution, both a Bayesian analysis and maximum likelihood methods produce good results for making inference.

Bibliography

(2017) Filipino family income and expenditure. URL https://www.kaggle.com/grosvenpaul/family-income-and-expenditure

Singh SK, Maddala GS (1976) A function for size distribution of incomes. Econometrica 44(5):963–970, URL http://www.jstor.org/stable/1911538