INFORMETRICS 89/90
L. Egghe and R. Rousseau (Editors)

A MODEL FOR THE FULL CIRCULATION DATA

Isola AJIFERUKE and Jean TAGUE

School of Library and Information Science,
University of Western Ontario, Elborn College, London,
Ontario N6G 1H1, Canada

Abstract

Three models of library circulation, the beta-binomial, negative binomial and generalized negative binomial, are applied to fourteen large circulation data sets. Both the generalized negative binomial and the beta-binomial provided good fits to the empirical data. However, a comparison of their performances in predicting future circulations is needed before either one can be chosen as the model for the full circulation data.

INTRODUCTION

Circulation data are often analyzed when decisions are to be made about relegation or weeding of materials. This has led many researchers [1-7] to develop probability models for the yearly frequency of circulation. The most common of these models is the negative binomial distribution. However, this model has not been substantiated by empirical data.
Ravichandra Rao [2] tried several models on 203 sub-data sets (the annual data for 5 Canadian university libraries were subdivided by discipline and by user type) and found the negative binomial to be the best. However, it only fitted about 45 % of the sub-data sets and did not fit any of the 17 annual circulation data sets.
Burrell, though admitting that the negative binomial did not fit the 4 annual circulation data sets for University of Sussex library [4], assumed the model for the base year and used it to develop a model for predicting future circulations of an item given the number of times it circulated in the base year [5]. However, Tague and Ajiferuke [8] have shown significant differences between his model's predictions and actual circulation figures.
The failure of the negative binomial is mainly due to its poor fitting of the upper tail of empirical distributions [2,3,7]. For this reason, Gelman and Sichel [7] suggested the beta-binomial but tested the zero-truncated version of the model on only three data sets and found it to fit two of them. Their reason for excluding the zero class was that they distrusted such figures. However, since the advent of automated circulation systems, it has become much easier to obtain reliable count of the items in the zero class. Any circulation analysis that excludes the zero class, which accounts for a large percentage of the collection (at least 50 % in most libraries), is not very useful because it is mostly the items in this category that would end up being relegated or weeded.
Hence, the objective of this study is to test the beta-binomial model on several full circulation data sets and to compare its performance with those of the negative binomial and generalized negative binomial, which is obtained by compounding the negative binomial with another parameter and is expected to provide a better fit than the latter.

THE MODELS

a) Beta-Binomial (BBD)

$$p(x) = \binom{s}{x} \frac{\Gamma(\alpha+x)\,\Gamma(s+\beta-x)\,\Gamma(\alpha+\beta)}{\Gamma(s+\alpha+\beta)\,\Gamma(\alpha)\,\Gamma(\beta)}$$

for $x = 0,1,\ldots,s$; $\alpha,\beta > 0$

The maximum likelihood method of estimating the parameters of this model is very complex. Hence, in this paper, we use the moment method as proposed by Skellam [9].
If $\mu'_{(j)}$ is the j^{th} factorial moment about the origin and $R_j = \mu'_{(j)}/\mu'_{(j-1)}$, then the estimate of

(i) s is given by the nearest integer to the positive root of the equation

$$As^2 + Bs + C = 0$$

where $A = R_3 - 2R_2 + R_1$;
$B = R_3R_2 - 2R_3R_1 + R_1R_2 - R_3 + 4R_2 - 3R_1$;
$C = 2R_1(R_3 - R_2 + 1)$

(ii) α is given by

$$\hat{\alpha} = \frac{R_1R_2 - (\hat{s}-1)R_1}{(\hat{s}-1)R_1 - \hat{s}R_2}$$

(iii) β is given by

$$\hat{\beta} = \hat{\alpha}[(\hat{s}/R_1) - 1]$$

b) Negative Binomial (NBD)

$$p(x) = \frac{\Gamma(v+x)}{x!\,\Gamma(v)}\,\alpha^v\,(1-\alpha)^x$$

where $0 < \alpha < 1$, $v > 0$ for $x = 0,1,\ldots$

The maximum likelihood estimator for v is given by the root of the equation

$$\log(1 + \bar{x}/\hat{v}) - \sum_{j=1}^{x_{max}} \frac{F_j}{(\hat{v}+j-1)} = 0 \ ,$$

where F_j = proportion of x's which are greater than or equal to j;

and

$$\hat{\alpha} = \hat{v}/(\bar{x}+\hat{v})$$

c) Generalized negative binomial (GNBD)

$$p(x) = \frac{v\,\Gamma[v+\beta x\ ]}{x!\ \Gamma[v+\beta x-x+1]}\ \alpha^{v+\beta x-x}\ (1-\alpha)^{x}$$

where $v > 0$, $0 < \alpha < 1$, $|\alpha\beta| < 1$ for $x = 0,1,\ldots$

Note that $p(x) = 0$ for $x \geq v/(1-\beta)$.
The negative binomial with parameters α and v is a special case of this model and is obtained when $\beta = 1$ [10]. The binomial distribution is also a particular case and is obtained when $\beta = 0$.

The maximum likelihood method of estimating the parameters of this model is also very complex. Hence, we use the method of moments to estimate the values of α, β and v. These are given by

$$\hat{\alpha} = A/2 - (A^2/4 - 1)^{1/2}$$

$$\hat{\beta} = \frac{1}{1-\hat{\alpha}}\ [1 - (\bar{x}\hat{\alpha}/\mu_2)^{1/2}]\ ;\ \text{and}$$

$$\hat{v} = \frac{\bar{x}\ [1 - (1-\hat{\alpha})\hat{\beta}]}{1-\hat{\alpha}}$$

$$\text{where } A = -2 + \frac{[\bar{x}\mu_3 - 3\ \mu_2^2]^2}{\bar{x}\mu_2^3}$$

and μ_2 and μ_3 are the second and third central moments respectively.

DATA ANALYSIS

To compare the three models, fourteen data sets were used. The first ten are the same as the collection used in Tague and Ajiferuke's study [8]. This group *consists of the yearly frequency circulation from* 1968-69 to 1977-78 of items in the University of Saskatchewan library which had circulated at least once in the year 1967-68. The eleventh data set is the frequency of circulation of items in the Hillman library of Pittsburgh University during 1974, while the last three are for Sussex University library during 1976-77, 1978-79 and 1979-80 sessions. We note here that some of the last four data sets were also considered by Burrell [1,4], Bagust [3], and Gelman and Sichel [7].

The fits of the three models to the data sets are shown in Tables 1 to 14. *The generalized negative bionomial passed the Chi-square test 8 times at 1 %* level of significance, beta-binomial 6 times and negative-binomial once. The generalized negative binomial provided the best fit to 9 data sets while the beta-binomial provided the best fit to the other five data sets. Both the generalized negative binomial and the beta-binomial tend to fit the upper tail of the observed distribution very well while the negative binomial tends to overestimate. Thus, it seems that the generalized negative binomial, with its additional parameter, is capable of solving the upper tail problem normally encountered in fitting the negative binomial to the annual circulation statistics. *A possible explanation for the better performance of the generalized* negative binomial over the negative binomial is that the latter has only a single mode of variation and any extra variations present in the data are assumed to be random whereas the extra parameter of the generalized negative binomial takes into account the variations in the variance/mean ratio [10].
Though the generalized negative binomial performed slightly better than the beta-binomial, it is necessary to compare how well the two of them predict future circulations before a judgement can be made on which is the better model for the full circulation data. The authors hope to work on this in the nearest *future*.

Table 1 : Annual circulation for 1968-69, University of Saskatchewan library (Tague & Ajiferuke [8])

Circulation per year	Observed Number of Items Circ.	Expected Number of Items Circulated BBD	NBD	GNBD
0	51992	51943.50	51906.42	51924.52
1	7674	7786.81	8091.81	7738.59
2	3576	3607.37	3575.21	3645.54
3	2087	2023.69	1920.40	2048.32
4	1250	1222.48	1123.05	1228.14
5	748	763.61	688.87	758.08
6	485	483.55	435.68	473.99
7	320	306.70	281.48	297.55
8	181	193.16	184.73	186.46
9	115	119.91	122.71	116.15
10	73	72.86	82.28	71.66
11	39	43.01	55.61	43.62
12	23	24.45	37.82	26.08
13	12	13.23	25.86	15.24
14	5	6.72	17.76	8.63
15-19	10	5.05	33.57	8.49
Total	68590			
Mean	0.573			
Variance	1.922			
s		19		
α		0.190	0.272	0.185
β		6.125		0.939
v			0.214	0.165
χ^2		12.286	111.959	8.686
d.f.		12	13	12
$p(\chi^2)$		0.423	0.77E-17	0.730

Table 2 : Annual circulation for 1969-70, University of Saskatchewan library (Tague & Ajiferuke [8])

Circulation per year	Observed Number of Items Circ.	Expected Number of Items Circulated BBD	NBD	GNBD
0	55884	55675.10	55788.38	55582.21
1	5354	5803.43	6032.60	5826.19
2	2754	2714.75	2650.38	2765.95
3	1758	1579.59	1457.04	1602.08
4	988	999.06	881.43	997.80
5	657	656.00	562.41	641.94
6	456	437.08	371.28	418.98
7	293	291.22	250.95	274.68
8	183	191.79	172.59	179.70
9	106	123.44	120.29	116.72
10	71	76.64	84.72	74.93
11	43	45.12	60.18	47.31
12	17	24.54	43.05	29.20
13	13	11.80	30.98	17.47
14-17	13	5.73	59.23	18.74
Total	68590			
Mean	0.471			
Variance	1.750			
s		15		
α		0.124	0.229	0.145
β		3.817		0.938
v			0.140	0.109
χ^2		72.316	269.616	69.72
d.f.		11	12	11
$p(\chi^2)$		0.44E-10	< E-20	0.14E-9

Table 3 : Annual circulation for 1970-71, University of Saskatchewan library (Tague & Ajiferuke [8])

Circulation per year	Observed Number of Items Circ.	Expected Number of Items Circulated BBD	NBD	GNBD
0	60060	59951.10	60031.42	59958.17
1	4007	4243.63	4312.96	4197.70
2	1882	1823.46	1777.30	1840.85
3	1037	991.13	933.96	1004.80
4	637	590.28	543.75	595.43
5	355	367.35	335.07	366.93
6	210	233.61	214.08	230.70
7	158	149.86	140.25	146.47
8	96	96.14	93.58	93.31
9	54	61.25	63.33	59.37
10	41	38.53	43.33	37.60
11	17	23.79	29.92	23.62
12	17	14.33	20.81	14.67
13	8	8.37	14.57	8.97
14	4	4.69	10.25	5.38
15-21	7	4.65	16.06	5.83
Total	68590			
Mean	0.290			
Variance	1.056			
s		21		
α		0.089	0.248	0.177
β		6.339		0.947
v			0.096	0.078
χ^2		29.089	78.478	20.635
d.f.		12	13	12
$p(\chi^2)$		0.004	0.21E-10	0.056

Table 4 : Annual circulation for 1971-72, University of Saskatchewan library (Tague & Ajiferuke [8])

Circulation per year	Observed Number of Items Circ.	Expected Number of Items Circulated BBD	NBD	GNBD
0	58073	57982.20	58026.52	57961.56
1	5305	5564.63	5713.12	5547.86
2	2467	2327.36	2253.96	2354.36
3	1242	1200.88	1111.68	1209.82
4	644	668.05	603.15	664.38
5	394	382.35	345.10	374.67
6	214	219.81	204.27	213.05
7	111	124.90	123.79	120.86
8	62	69.20	76.32	67.86
9	47	36.87	47.69	37.45
10	16	18.59	30.11	20.14
11	10	8.67	19.17	10.45
12-16	7	5.33	30.90	8.51
Total	68590			
Mean	0.318			
Variance	0.935			
s		15		
α		0.127	0.309	0.213
β		5.864		0.929
v			0.143	0.109
χ^2		29.078	111.488	24.783
d.f.		9	10	9
$p(\chi^2)$		0.0006	0.26E-18	0.0032

Table 5 : Annual circulation for 1972-73, University of Saskatchewan library (Tague & Ajiferuke [8])

Circulation per year	Observed Number of Items Circ.	Expected Number of Items Circulated BBD	NBD	GNBD
0	59287	59160.30	59259.80	59171.66
1	5195	5466.32	5451.43	5428.30
2	2134	2046.51	1984.96	2063.17
3	1004	944.10	902.81	952.45
4	471	471.09	451.57	472.18
5	260	243.32	238.15	241.71
6	110	127.37	129.92	125.52
7	75	66.75	72.56	65.49
8	28	34.72	41.23	34.11
9	12	17.81	23.74	17.66
10	4	8.95	13.81	9.04
11	3	4.39	8.10	4.56
12-23	7	3.73	10.97	4.14
Total	68590			
Mean	0.253			
Variance	0.655			
s		23		
α		0.123	0.364	0.294
β		12.448		0.939
v			0.145	0.121
χ^2		35.048	62.254	28.385
d.f.		9	10	9
$p(\chi^2)$		0.58E-4	0.14E-8	0.0008

Table 6 : Annual circulation for 1973-74, University of Saskatchewan library (Tague & Ajiferuke [8])

Circulation per year	Observed Number of Items Circ.	Expected Number of Items Circulated BBD	NBD	GNBD
0	60236	60181.70	60215.12	60185.23
1	5167	5330.58	5378.35	5306.70
2	1914	1795.81	1730.93	1811.06
3	744	731.87	691.23	734.83
4	293	316.56	302.82	314.62
5	135	138.68	139.70	136.72
6	63	60.00	66.62	59.16
7	20	25.19	32.50	25.16
8	11	10.10	16.13	10.39
9-15	7	5.67	14.33	6.09
Total	68590			
Mean	0.200			
Variance	0.161			
s		15		
α		0.147	0.446	0.345
β		10.824		0.914
v			0.161	0.123
χ^2		16.510	42.569	12.673
d.f.		6	7	6
$p(\chi^2)$		0.0113	0.40E-6	0.0485

Table 7 : Annual circulation for 1974-75, University of Saskatchewan library (Tague & Ajiferuke [8])

Circulation per year	Observed Number of Items Circ.	Expected Number of Items Circulated BBD	NBD	GNBD
0	60565	60546.50	60544.84	60527.15
1	5177	5251.04	5361.88	5262.79
2	1753	1715.98	1625.91	1723.22
3	688	664.54	609.38	657.64
4	263	265.58	250.19	259.74
5	90	103.43	108.09	101.82
6	36	37.67	48.25	38.60
7	13	12.28	22.03	13.77
8-10	5	4.14	17.32	5.67
Total	68590			
Mean	0.184			
Variance	0.362			
s		10		
α		0.146	0.882	0.352
β		7.774		0.891
v			0.171	0.120
χ^2		4.739	45.716	5.049
d.f.		5	6	5
$p(\chi^2)$		0.4483	0.34E-7	0.4097

Table 8 : Annual circulation for 1975-76, University of Saskatchewan library (Tague & Ajiferuke [8])

Circulation per year	Observed Number of Items Circ.	Expected Number of Items Circulated BBD	NBD	GNBD
0	60610	60584.80	60593.01	60573.69
1	5158	5243.08	5322.84	5249.47
2	1758	1689.93	1617.14	1696.04
3	645	650.74	607.72	645.52
4	246	261.95	250.26	257.27
5	112	104.75	108.46	103.23
6	46	40.34	48.57	40.85
7	11	14.34	22.25	15.69
8-12	4	14.56	18.69	8.34
Total	68590			
Mean	0.183			
Variance	0.365			
s		12		
α		0.149	0.480	0.371
β		9.599		0.905
v			0.169	0.125
χ^2		8.313	37.231	9.424
d.f.		5	6	5
$p(\chi^2)$		0.1399	0.16E-5	0.0933

Table 9 : Annual circulation for 1976-77, University of Saskatchewan library (Tague & Ajiferuke [8])

Circulation per year	Observed Number of Items Circ.	Expected Number of Items Circulated BBD	NBD	GNBD
0	61159	61131.10	61142.05	61124.76
1	5055	5159.39	5224.68	5156.10
2	1613	1515.18	1438.05	1522.18
3	502	520.69	486.79	517.49
4	169	182.75	180.18	180.28
5	71	61.83	70.11	61.64
6	13	19.33	28.17	20.06
7-10	8	6.78	18.45	7.70
Total	68590			
Mean	0.160			
Variance	0.287			
s		10		
α		0.158	0.535	0.405
β		9.748		0.883
v			0.184	0.128
χ^2		13.797	42.064	12.508
d.f.		4	5	4
$p(\chi^2)$		0.008	0.57E-7	0.014

Table 10 : Annual circulation for 1977-78, University of Saskatchewan library (Tague & Ajiferuke [8])

Circulation per year	Observed Number of Items Circ.	Expected Number of Items Circulated BBD	NBD	GNBD
0	63251	63251.50	63245.76	63249.05
1	3976	3991.28	4037.73	3988.57
2	997	968.69	922.98	972.36
3	260	274.56	261.67	272.98
4	67	78.86	81.37	78.13
5	34	21.61	26.64	21.75
6-10	5	6.81	9.02	5.67
Total	68590			
Mean	0.105			
Variance	0.170			
s		10		
α		0.140	0.607	0.485
β		13.131		0.877
v			0.162	0.112
χ^2		11.034	12.250	9.850
d.f.		3	4	3
$p(\chi^2)$		0.0116	0.0101	0.0199

Table 11 : Annual circulation for 1974, Pittsburgh University library (Kent [11])

Circulation per year	Observed Number of Items Circ.	Expected Number of Items Circulated BBD	NBD	GNBD
0	385961	385924.00	385791.50	386120.70
1	63526	63657.40	64427.02	63219.23
2	25653	25442.60	25204.45	25573.38
3	11855	12044.80	11743.91	12140.40
4	6055	6103.16	5910.88	6133.69
5	3264	3195.94	3107.56	3193.60
6	1727	1701.15	1680.20	1689.01
7	931	911.86	926.40	900.36
8	497	489.28	518.19	481.50
9	275	261.71	293.09	257.52
10	124	139.10	167.23	137.42
11	68	73.27	96.10	73.04
12	28	38.16	55.55	38.59
13	13	19.61	32.27	20.25
14	6	9.93	18.83	10.52
15	9	4.49	11.02	5.41
16-36	8	4.53	26.16	5.32
Total	500000			
Mean	0.4342			
Variance	1.1029			
s		36		
α		0.257	0.385	0.323
β		21.025		0.950
v			0.271	0.228
χ^2		22.926	93.942	27.493
d.f.		13	14	13
$p(\chi^2)$		0.0426	0.68E-13	0.0106

Table 12 : Annual circulation for 1976-77, Sussex University Main library (Burrell[4])

Circulation per year	Observed Number of Items Circ.	Expected Number of Items Circulated BBD	NBD	GNBD
0	81146	81047.20	81031.12	80928.82
1	9674	9977.87	10421.46	10048.91
2	4351	4274.90	4111.59	4328.59
3	2275	2206.27	1986.60	2198.81
4	1250	1205.94	1047.92	1174.52
5	663	664.72	580.63	635.36
6	355	358.48	332.01	341.64
7	154	184.06	194.05	180.25
8	72	87.04	115.26	92.14
9	37	36.03	69.31	44.89
10	14	11.85	42.09	20.27
11-14	9	2.39	57.59	10.24
Total	100000			
Mean	0.3788			
Variance	1.0032			
s		11		
α		0.163	0.340	0.217
β		4.574		0.912
v			0.195	0.139
χ^2		40.689	261.146	36.097
d.f.		8	9	8
$p(\chi^2)$		0.24E-5	< E-20	0.17E-4

Table 13 : Annual circulation for 1978-79, Sussex University Main library (Burrell [4])

Circulation per year	Observed Number of Items Circ.	Expected Number of Items Circulated BBD	NBD	GNBD
0	170120	169955.00	169813.70	169927.30
1	39326	39327.50	40328.20	39200.53
2	16226	16808.10	16447.12	16931.19
3	8113	8096.97	7641.57	8130.22
4	4304	4044.71	3767.33	4022.63
5	2171	2029.00	1921.49	1998.74
6	1128	1003.99	1001.86	984.90
7	425	483.71	530.49	477.36
8	148	224.27	284.13	225.93
9	60	98.86	153.52	103.47
10	29	40.85	83.53	45.46
11	10	15.53	45.71	18.75
12	5	5.31	25.13	7.01
13-25	10	2.02	31.20	2.51
Total	242075			
Mean	0.5630			
Variance	1.2696			
s		16		
α		0.360	0.422	0.306
β		9.857		0.910
v			0.411	0.299
χ^2		147.513	419.439	169.336
d.f.		10	11	10
$p(\chi^2)$		< E-20	< E-20	< E-20

Table 14 : Annual circulation for 1979-80, Sussex University Main library (Burrell [4])

Circulation per year	Observed Number of Items Circ.	Expected Number of Items Circulated BBD	NBD	GNBD
0	175892	176006.00	175674.40	175922.60
1	36705	36687.40	37972.44	36627.08
2	15283	15364.30	14959.72	15483.32
3	7113	7299.48	6780.23	7307.23
4	3694	3584.36	3273.96	3543.88
5	1881	1752.80	1639.10	1716.76
6	866	834.41	840.04	818.44
7	357	379.38	437.64	379.66
8	114	161.46	230.78	169.30
9	47	62.72	122.83	71.32
10	16	21.45	65.86	27.51
11	12	6.10	35.53	9.03
12-19	5	1.45	42.11	1.92
Total	242075			
Mean	0.5048			
Variance	1.1168			
s		13		
α		0.319	0.428	0.296
β		7.898		0.901
v			0.378	0.262
χ^2		54.148	363.355	71.226
d.f.		9	10	9
$p(\chi^2)$		0.18E-7	< E-20	0.87E-11

CONCLUSION

The generalized negative binomial and the beta-binomial fit the full annual circulation data better than the negative binomial. In particular, the models tend to fit the upper tail of the distribution very well while the negative binomial tends to overestimate. Though the generalized negative binomial performed slightly better than the beta-binomial, the authors hope to compare the performances of the two models in predicting future circulations before deciding on which is the better model for the full circulation data.

ACKNOWLEDGEMENT

The authors wish to thank Professor Quentin Burrell for his comments on the earlier version of this paper.

REFERENCES

[1] Burrell, Q. A simple stochastic model for library loans. Journal of Documentation, 36(2) : 115-132; 1980.

[2] Ravichandra Rao, I.K. Document and user distributions in transaction records of Canadian University libraries. Doctoral thesis. London, Ontario : University of Western Ontario; 1981.

[3] Bagust, A. A circulation model for busy public libraries. Journal of Documentation, 39(1) : 24-37; 1983.

[4] Burrell, Q. A note on ageing in a library circulation model. Journal of Documentation, 41(2) : 100-115; 1985.

[5] Burrell, Q. A second note on ageing in a library circulation model : the correlation structure. Journal of Documentation, 42(2) : 114-128; 1986.

[6] Burrell, Q. A third note on ageing in a library circulation model : application to future use and relegation. Journal of Documentation, 43(1) : 24-45; 1987.

[7] Gelman, E., Sichel, E.S. Library book circulation and the beta-binomial distribution. Journal of the American Society for Information Science, 38(1) : 4-12; 1987.

[8] Tague, J., Ajiferuke, I. The Markov and the mixed-Poisson models of library circulation compared. Journal of Documentation, 43(3) : 212-231; 1987.

[9] Skellam, J.G. A probability distribution derived from the binomial distribution by regarding the probability of success as variable between the sets of trials. Journal of the Royal Statistical Society B, 10 : 257-261; 1948.

[10] Jain, G.C., Consul, P.C. A generalized negative binomial distribution. SIAM Journal of Applied Mathematics, 21(4) : 501-513; 1971.

[11] Kent, A. Use of Library Materials : The University of Pittsburgh Study. New York : Marcel Dekker; 1979.