CRYPTIC CROSSWORD PUZZLES: A STATISTICAL ANALYSIS OF ERRORS IN SOLUTION

S. Naranan

A B S T R A C T

****

Cryptic crosswords are very popular linguistic puzzles, intellectually challenging and enjoyable for the solver. I have documented my 'error scores' and presented the statistical distribution of errors for 3404 puzzles over a decade, in 2010. Now the analysis is extended to 5484 puzzles (16 years data). In both, the Negative Binomial Distribution (NBD) is a good fit to data. It was conjectured that NBD will prove to be adequate for all solvers with diverse skills. Data from another solver obtained in 2013 supports the conjecture. A popular model for NBD as a 'mixture distribution' of Poisson and Gamma distributions is appropriate for the distribution of errors. The two free adjustable parameters of NBD (p,k) are measures of average complexity of puzzles for a solver. Simulations of NBD's for different levels of solver skills are presented.

A complementary statistic is the 'complexity of a composer's puzzle', or the study of 'one puzzle, many solvers' which depends on solver feedback to the composer. There exists no systematic study and it is argued that it is much-needed input for the composer. Interestingly, word puzzle lottery, popular 60-70 years ago, is an example of a social experiment for the study of solver behaviour in 'one puzzle, many solvers'. The results conformed to the canonical Binomial distribution. The connection between some commonly occurring statistical distributions is discussed in the appendices.

"I often say that when you can measure what you are speaking about and express it in number, you know something about it; but when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind; it may be the beginning of knowledge but has scarcely in your thoughts advanced to the state of science, whatever the matter be".

Lord Kelvin (1824-1907)

1. INTRODUCTION.

Over the last hundred years crossword puzzles have continued to grow with ever increasing popularity. For the solver, they provide intellectual exercise and entertainment. They are daily features, world-wide in newspapers. Following its first appearance in the *New York World* on 21 December 1913, it quickly spread to U.K; in time it acquired a different hue in the shape of cryptic in Britain. Today the two distinct classes of puzzles are typified by the *New York Times* in the U.S and the *Times* in Britain and they differ in grid structure, word types and cluing pattern. The cryptics have become popular in Europe and India. In particular, *The Hindu* a popular daily in India has its own team of compilers of cryptic. Only on Sundays, *The Hindu* has puzzles syndicated from *The Guardian* of U.K.

I have been a regular solver of cryptics for more than 40 years. I began documenting the number of unsolved clues in each puzzle from 1987. In 2010, I published my statistical analysis of failures in 3404 puzzles over a 10-year period – five years each of the *Times of India* (Bombay) and *The Hindu* (Chennai) – in the *Journal of Quantitative Linguistics* (Naranan 2010). The chief aim of the analysis was to quantify the average complexity of the puzzles for the solver. The data was simply the number of puzzles $N(x)$ with x errors (x = 0,1,2,3.....15). This histogram $N(x)$ *vs* x is referred hereafter as the 'distribution of x'. It is found that the distribution has long tail with x up to 15. As a crude measure of 'complexity' one may adopt the mean (m) and standard deviation (s) of the distribution. But one can do better if a mathematical statistical function can be found to define the entire distribution point by point (x = 0,1,2,3....15). Indeed such a function exists; it is called the Negative Binomial Distribution (NBD) which has two free, adjustable parameters *(p,k)*. *(p,k)* are simply related to *(m,s)* of the distribution.

Why is NBD a good fit to crossword error data? A viable model based on the fact that the individual puzzles span a wide range of complexity resulting in the long tail, was presented in Naranan (2010). It is interesting that in Insurance industry, car accidents in a large group of zones conform to NBD and this fact is used in fixing car insurance tariffs.

But NBD is not alone! It turns out that another well-known distribution called the lognormal distribution (LND2), a variant of the ubiquitous normal (Gaussian) distribution also fits the crossword error results. But it requires three parameters (μ, σ, X_0) rather than two and so is the less preferred of the two. It is conjectured that NBD with (*p,k*) in a different domain, say for another solver, may not be 'equivalent' to LND2.

This paper is a follow-up of the 2010 paper (Naranan, 2010) and has new results of three different categories: (1) an update of my own results to 16 years of data with total number of puzzles 5484 (2) analysis of another solver's (Daniel Peake's) error/success data from 260 puzzles of *The Guardian* in 2013 and (3) a very special kind of word puzzle lottery which was popular in India in 1940's and 1950's. NBD continues to be a good fit to my updated results and for Daniel Peake's data (Sections 2,3). The word puzzle lottery is shown to be a unique social experiment to test the well-known Binomial distribution (Section 7).

It is conjectured that every solver's error scores, can be characterized by some specific (*p,k*) of NBD. To illustrate this, *N(x*) distributions are presented for four sets of *(p,k)* values representing different levels of solver skill (Section 5).

Based on NBD, can one predict the entire error distribution *N(x*) of a solver? The NBD parameters *(p,k)* are derived from the *(m,s)* of the 'raw' data of the solver. In principle it is possible to determine *(p,k)* from *any two* data points. We consider the two values *N(x=0)* and *N(x=1)*. The main caveat here is that the two should be sufficiently large. This criterion is met by my data. In general the predictive power of

NBD is constrained by statistics. The main strength of NBD lies in its capacity to fit the entire distribution *N(x)* with just two parameters (Section 6).

What is lacking is quantifying the complexity of a compiler's puzzle through a study of solvers' inputs of their errors for the same puzzle. This requires a project by an organized group of solvers feeding data to the composer with due attention to uncertainties and biases in solver behaviour (Section 4).

A remarkable feature of the NBD is that superposition of different groups of data with wide-ranging complexities can still conform to NBD. This was demonstrated in Naranan (2010) where four different sets of data individually and collectively conformed to NBD. Using standard methods of statistical theory it can be proved that a superposition of NBD's remains an NBD under some constraints. However NBD remains robust and accommodative of some deviations from the constraints. Several statistical distribution functions figure in this paper. Their interrelationship is discussed in the last Section. The properties of three main discrete distributions, the Binomial, the Poisson and the Negative Binomial are compared. These topics of theoretical interest are presented in the Appendices.

2. CRYPTIC CROSSWORDS: DATA ON FAILURES IN SOLUTION.

My crossword puzzles error (failures) data were first presented in Naranan (2010). They comprised 10 years of data from two dailies, the *Times of India* and *The Hindu*. Now I have six more years of data from *The Hindu*. The cumulated data of 16 years has total sample size $N_T = 5484$.

This data is presented in *Table 1*. *N(x)* is the number of puzzles with x errors (x = 0,1,2,3....15). *N(x)* decreases monotonically with increasing x which extends up to 15 (nearly half the total number of clues). It is instructive to look at the low x and the high x values. *N(0)* and *N(1)* together account for 3746 or 68.3 % of the total. The tail end $x > 8$ contributes 53 or about 1 %. The large sample size has rendered the tail contribution substantial enough for statistical analysis.

The aim of the analysis is to obtain a measure of the complexity of puzzles. Here it pertains to a single solver and the complexity is averaged over a large number of diverse puzzles. For the solver (the author in the present case) the parameters (*m,s*)

CROSSWORD FAILURE
DATA UPDATE 2014

	ALLCW14
x\N(x)	OBS
0	2500
1	1246
2	748
3	394
4	227
5	159
6	67
7	56
8	34
9	19
10	13
11	7
12	6
13	5
14	0
15	3
NTOT	**5484**
m	1.323
s	1.871

ALLCW14:
Times of India (1987-91)
The Hindu (2004-14)

Table 1

provide a measure of complexity. m is the average number of failures and s is the spread around the average. (m,s) = (1.323, 1.871).

2.1. NEGATIVE BINOMIAL DISTRIBUTION (NBD).

Parameters (m,s) provide only a gross measure of complexity. One can do better by finding a suitable mathematical function that fits the distribution over the entire range of x. This function will have one or more free adjustable parameters. The search for such a function begins by looking at (m,s). We consider the data as an example of random count data of integer values (x = 0,1,2.....). The first choice (as the default) is the Poisson distribution which depends only on one parameter λ.

$$\text{PD: } Prob\,(x) = P(x) = \exp(-\lambda)\, \lambda^x / x! \qquad (\lambda > 0,\ x = 0,1,2....) \qquad (1)$$

λ completely determines the distribution. For PD, m and s^2 are both equal to λ.

$$m = s^2 = \lambda$$

Here m/s^2= 0.378 much less than 1. This ratio is a measure of over-dispersion; here it indicates a long tail. To account for this tail, an additional parameter besides λ is required. The Negative Binomial Distribution (NBD) is usually the next choice (Feller 1972). The distribution function is

$$NBD\text{: } P(x) = \{\Gamma(k+x)/[\Gamma(k)\, x!]\}\ p^k q^x \qquad (k > 0,\ x = 0,1,2......) \qquad (2)$$

where (p,k) are the two parameters and $q = 1 - p$. Here Γ (k) is the Gamma function defined by

$$\Gamma(k) = (k-1)\,\Gamma(k-1) \qquad (k > 1)$$

When k is an integer

$$\Gamma(k) = (k-1)(k-2)..........1 = (k-1)\,! \,, \quad \Gamma(1) = 1$$

The NBD parameters (p,k) are easily determined from (m, s).

$$p = m/s^2 \qquad\qquad k = m\,p/(1-p)$$

For any given x equation (2) gives $P(x)$. When we wish to tabulate $P(x)$ for all values of x it is easier to use the following recurrence relations implied by eq (2).

$$P(0) = p^k \quad (x=0) \qquad (3a)$$

$$P(x+1) = P(x)\,(1-p)\,[(k+x)/(1+x)] \qquad (x = 0,1,2,3....) \qquad (3b)$$

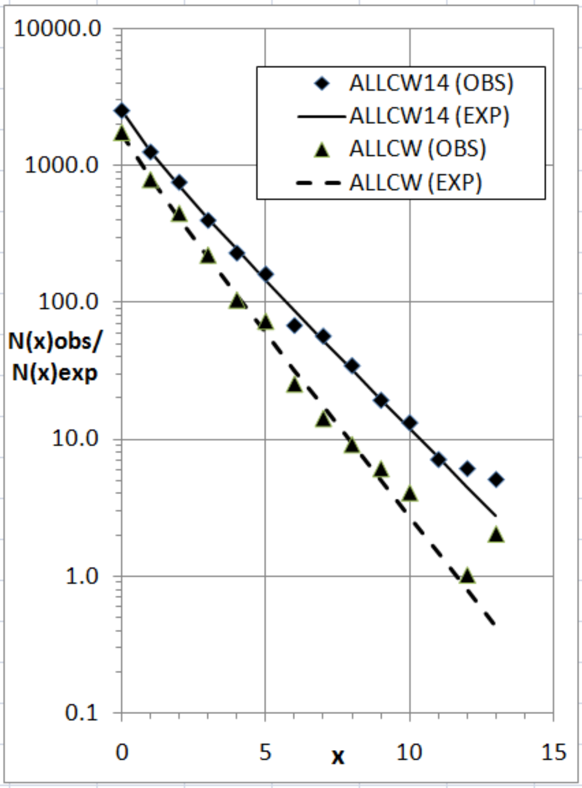
These probabilities are multiplied by N_T to obtain $N(EXP)$.

CROSSWORD FAILURE DATA UPDATE 2014

NEGATIVE BINOMIAL FITS TO DATA

x\N(x)	ALLCW14 OBS	ALLCW14 EXP	ALLCW OBS	ALLCW EXP	
0	2500	*2509.8*	1725	*1716.0*	
1	1246	*1253.9*	781	*813.2*	
2	748	*703.3*	443	*414.4*	
3	394	*408.9*	219	*216.1*	
4	227	*241.9*	103	*114.0*	
5	159	*144.6*	72	*60.5*	
6	67	*87.0*	25	*32.3*	
7	56	*52.6*	14	*17.3*	
8	34	*31.9*	9	*9.3*	
9	19	*19.4*	6	*5.0*	
10	13	*11.9*	4	*2.7*	
11	7	*7.2*		*1.4*	
12	6	*4.4*	1	*0.8*	
13	5	*2.7*	2	*0.4*	
>13	3	*4.3*			
NTOT	**5484**	*5484*	**3404**	*3404*	
m	1.323		1.042	Mean	
s	1.871		1.514	Std Dev	
p	0.378		0.455	NBD	
k	0.803		0.869	parameters	
CHISQ	**11.6**		**9.7**		
NDF	11		8		
Prob 0.1 for chisq					
	>17.3		> 13.4		

NEGATIVE BINOMIAL FITS TO DATA



ALLCW: Times of India (1987-1991)
The Hindu (2004-2008)

ALLCW14: Times of India (1987-1991)
The Hindu (2004- 2014)

Figure 1

The observed values *N(OBS)* and those calculated (or expected) *N(EXP)* for NBD are presented in the Table in *Figure 1*(cols 2,3). Even a cursory comparison of the two sets of numbers suggests very good agreement within the expected sampling errors. For comparison the original data from Naranan (2010) for 10-year data are also shown (cols 4,5). The *(p,k)* values are (0.378, 0.803) for ALLCW14 (16-years) and (0.455, 0.869) for ALLCW (10-years).

The 'chi-squared' (χ^2) statistic is a measure of 'goodness of fit' to NBD. It depends on the deviations *N(OBS) – N(EXP)* for all *x* (Cramer 1955). The χ^2 depends on '*ndf*' the number of degrees of freedom which is $n - 3$ where *n* is the number of data points (the pairs OBS and EXP). For ALLCW14 χ^2 *(ndf)* = 11.6 (11) and for ALLCW 9.7 (8). From the χ^2 tables (Cramer 1955) one finds that for *ndf* = 11, the χ^2 will exceed 17.3 with probability 0.1 and for *ndf* = 8 it will exceed 13.4. The observed χ^2, 11.6 and 9.7 respectively imply the NBD hypothesis is 'acceptable' or more precisely 'cannot be rejected'.

Compared to ALLCW, the ALLCW14 results show an increase in both *m* and *s* implying the puzzles are getting harder to solve. This confirms my general impression that the mix of puzzles of varying difficulty is now weighted more in favour of hard puzzles.

The figure next to the Table, charts the results from all the four columns. Notice that the *y*-axis is logarithmic and for large *x* the plots are nearly straight lines. This is because when $k << x$, the ratio $P(x+1)/P(x)$ tends to $q = 1 - p$, a constant (equation 3b). This is characteristic of a geometric series or exponential decrease of *N(x) vs x.*

Why is NBD a good model for distribution of crossword errors? The basic statistical description is a random counting process which is Poissonian. The different puzzles comprising the total sample have different characteristic Poisson parameters λ's and this fact is modelled by a distribution of the λ' s $g(\lambda)$. If $g(\lambda)$ is a Gamma distribution, the superposed puzzle sample has a Negative Binomial distribution (Feller 1972, Naranan 2010). For details see the Appendix A1. The NBD is therefore a mixture distribution (MD), a convolution of Poisson and Gamma distributions.

2.2. LOGNORMAL DISTRIBUTION

Is it possible that some other distribution besides the NBD, can also fit the *N(x)* distribution? The χ^2-test merely certifies a hypothetical test distribution as 'not rejected'. It leaves open the possibility of other hypothetical distributions too meeting the test criteria. In the present case an alternative to NBD exists and NBD is not alone! It is the lognormal distribution (LND), a variant of the normal distribution (ND).

$$ND: P(x) = (1/\sigma\sqrt{2\pi})\ \exp[-(x-\mu)^2/2\sigma^2] \qquad -\infty < x < +\infty \qquad (4)$$

where μ is the mean and σ is the standard deviation. In LND, an important variant of ND, *ln x* is normally distributed.

$$LND1: P(x) = (1/\sigma\sqrt{2\pi})(1/x)\ \exp[-(\ln x-\mu)^2/2\sigma^2] \qquad x > 0 \qquad (5)$$

Whereas ND is symmetric (bell shaped) the LND has long tails.

Yet another version of LND is the 3-parameter function in which $ln(x+X_0)$ is normally distributed.

$$LND2: P(x) = (1/\sigma\sqrt{2\pi})[1(/x+X_0)]\ \exp\{-[\ln(x+X_0)-\mu])^2/2\sigma^2]\} \qquad x > 0 \qquad (6)$$

X_0 is a constant and can be positive or negative (Cramer 1955). When $x = 0$, LND2 reduces to LND1. The method of determining the parameters (X_0 , μ , σ) is given in Naranan (2010).

The 'expected' values of *N(x)* for LND2 are given in *Figure 2* col 3 for ALLCW14 and col 5 for ALLCW. The LND2 and NBD parameters are compared in *Table 2*. Compared to ALLCW, in ALLCW14, the *(p,k)* values are lower and the (X_0, μ, σ) values higher. The percentage deviations are given in the last row. χ^2 *(ndf)* for ALLCW14 and ALLCW are 11.0 (11) and 9.3 (8) respectively.

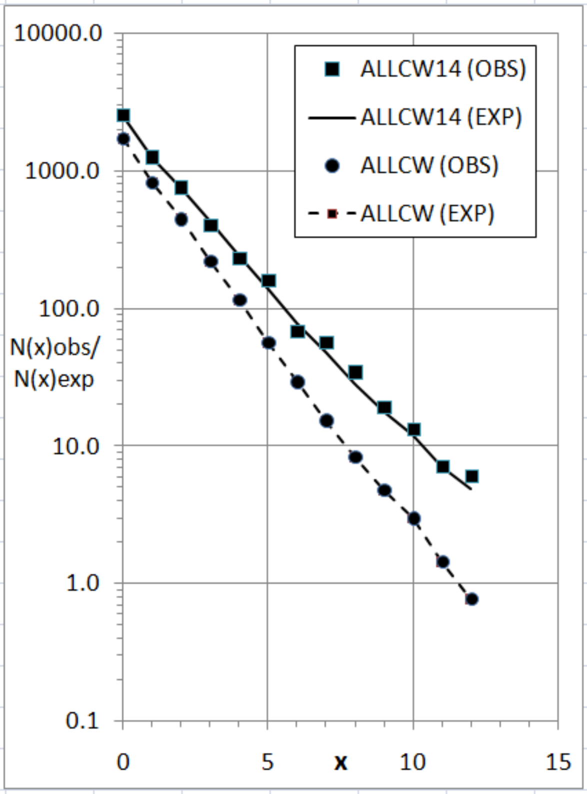
Some insight into the reason for similar behaviour of NBD and LND2 can be obtained by comparing them at large *x*. We compare the expected ratio *P(x+1)/P(x)* for NBD and LND2 for $x \geq 7$ for ALLCW14. In the interval $7 \leq x \leq 13$ the ratio is 0.607 – 0.613 and 0.614 - 0.677 respectively. The nearly constant value for NBD (0.610) is due to the fact that according to equation 3(b) for large *x* and $k << x$, the ratio tends to *1-p*, a constant, in this case 0.622. (Further, when $k = 1$, the ratio is exactly *1-p* for all *x*). It is very plausible that for a different range of *(p,k),* the closeness of NBD and LND2 is unlikely. NBD is therefore likely to be the preferred

CROSSWORD FAILURE DATA UPDATE 2014

LOGNORMAL DISTRIBUTION (LND2) FITS TO DATA

	ALLCW14		ALLCW	
x\N(x)	OBS	EXP	OBS	EXP
0	2500	*2482.6*	1725	*1689.7*
1	1246	*1231.2*	781	*818.0*
2	748	*750.2*	443	*441.8*
3	394	*427.8*	219	*218.5*
4	227	*242.9*	103	*114.7*
5	159	*140.9*	72	*56.5*
6	67	*79.0*	25	*29.3*
7	56	*48.8*	14	*15.3*
8	34	*28.5*	9	*8.3*
9	19	*18.1*	6	*4.8*
10	13	*12.1*	4	*3.0*
11	7	*7.1*		*1.4*
12	6	*4.9*	1	*0.8*
>12	8	*9.9*	2	*0.4*
NTOT	**5484**	*5484*	**3404**	*3403*
m	1.323		1.042	Mean
s	1.871		1.514	Std Dev
SIGMA	0.585		0.536	LND2
MU	0.975		0.895	para-
X0	2.470		2.432	meters
CHISQ	**11.0**		**9.3**	
NDF	11		8	
Prob 0.1 for chisq				
	>17.3		>13.4	

LOGNORMAL DISTRIBUTION (LND2)



ALLCW: Times of India (1987-1991)
The Hindu (2004-2008)
ALLCW14: Times of India (1987-1991)
The Hindu (2004- 2014)

Figure 2

TABLE 2: COMPARISON OF NBD AND LND2 PARAMETERS

Data	NEG BINOMIAL (NBD)		LOGNORMAL (LND2)		
Set	p	k	Xo	MU	SIGMA
ALLCW14	0.378	0.803	2.47	0.975	0.585
ALLCW	O.455	0.87	2.432	0.895	0.536
Deviation %	-16.9	-7.7	1.56	8.9	9.1

Table 2

distribution applicable to all solvers.

As for a theoretical model for LND one can invoke the ‘theory of proportional effect’ (Aitchison and Brown 1957) based on stochastic multiplicative process. Its relevance for crossword puzzle errors is explained in Naranan (2010).

3. CROSSWORD PUZZLES: DANIEL PEAKE’S DATA.

All the data so far is from a single solver. Will NBD apply for a different solver, especially with different skills, which will be reflected in the mean *m* and the standard deviation *s* (or equivalently *p,k*)? Fortunately such data is available.

Daniel Peake undertook a project “2500 clue challenge”. The aim was to solve at least 2500 clues in the year 2013 of cryptics from *The Guardian*. He constantly updated the data on his website

http://danielpeake.com/blog1/2500-clue-challenge/

For every one of the 260 weekday puzzles, the number of failures and successes were given. From this, histograms of the number of successes and failures were made by Shuchismita Upadhyay who operates a blog

http://www.crosswordunclued.com

She drew my attention to Peake’s work and sent me the histograms suggesting that Peake’s data “can be used to refine your study on crossword failures”.

To test NBD for Peake’s data, the *(p,k)* are determined from the mean and variance *m* and s^2 of the distributions. For ‘failures’ *(m,s)* values are (18.6, 6.0). In *Figure 3*, the observed and expected (calculated) distributions are shown. Clearly NBD is a poor fit to the distribution of failures. The reason is fairly obvious. The Poisson distribution and the NBD are valid for ‘rare events’, but here the mean 18.6 is a major fraction of the total number of clues in a grid (typically 30).

This suggests that one can try to fit to NBD the distribution of ‘successes’. From *(m,s)* = (9.9, 6.1), the *(p,k)* are determined as (0.263, 3.548). The observed and the calculated distributions are shown in *Figure 4*. The two seem to be in good agreement. For a χ^2-test of the goodness of fit to NBD, one has to group the data in

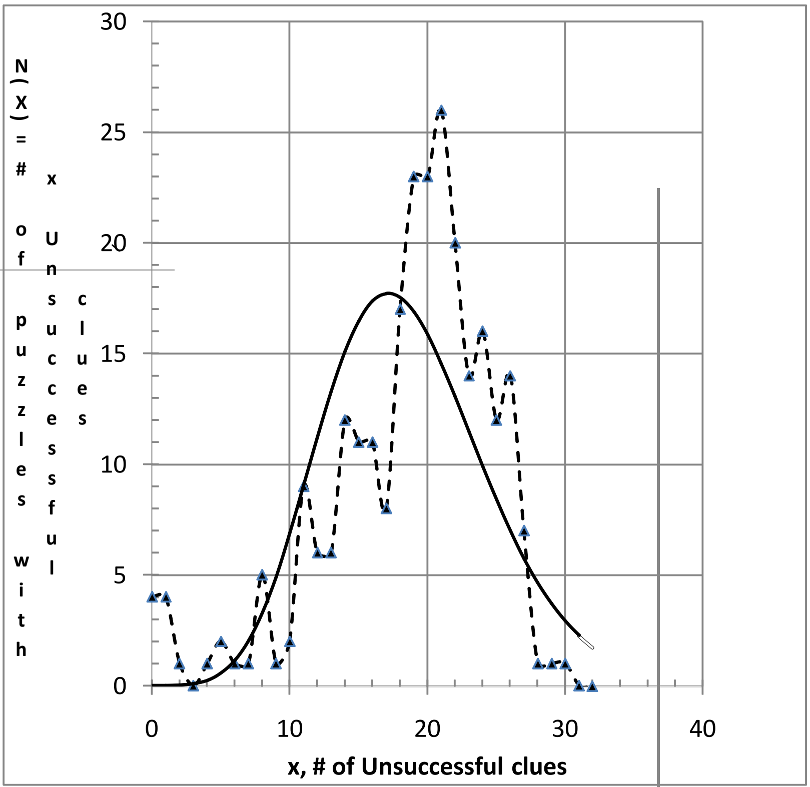
DANIEL PEAKE'S CROSSWORD DATA - UNSUCCESSFUL CLUES

X = # OF **UNSUCCESSFUL** CLUES

N(X) = NUMBER OF PUZZLES WITH X UNSUCCESSFUL CLUES

TABLE 1

x	N(x) obs	N(x) cal
0	4	0.001
1	4	0.005
2	1	0.025
3	0	0.088
4	1	0.244
5	2	0.564
6	1	1.130
7	1	2.022
8	5	3.285
9	1	4.921
10	2	6.871
11	9	9.023
12	6	11.224
13	6	13.304
14	12	15.101
15	11	16.483
16	11	17.364
17	8	17.708
18	17	17.530
19	23	16.885
20	23	15.858
21	26	14.547
22	20	13.057
23	14	11.484
24	16	9.910
25	12	8.401
26	14	7.003
27	7	5.747
28	1	4.646
29	1	3.704
30	1	2.914
31	0	2.264
32	0	1.738
NTOT	260	255.1
m	18.581	
s	5.989	
p=m/s*s	**0.518**	
k=mp/(1-p)	**19.973**	



Diamonds are the observed values. Solid Line is the best fit for values according to the Negative Binomial Distribution

TABLE 2

X	N(X) OB	N(X) CA
0-2	9	0.030
3-5	3	0.896
6-8	7	6.437
9-11	12	20.814
12-14	24	39.629
15-17	30	51.556
18-20	63	50.273
21-25	88	57.399
26-32	24	28.017
NTOT	260	255.1

N(X)OB = NUMBER OBSERVED

N(X)CA = NUMBER CALCULATED AS PER

NEGATIVE BINOMIAL DISTRIBUTION (NBD)

Figure 3

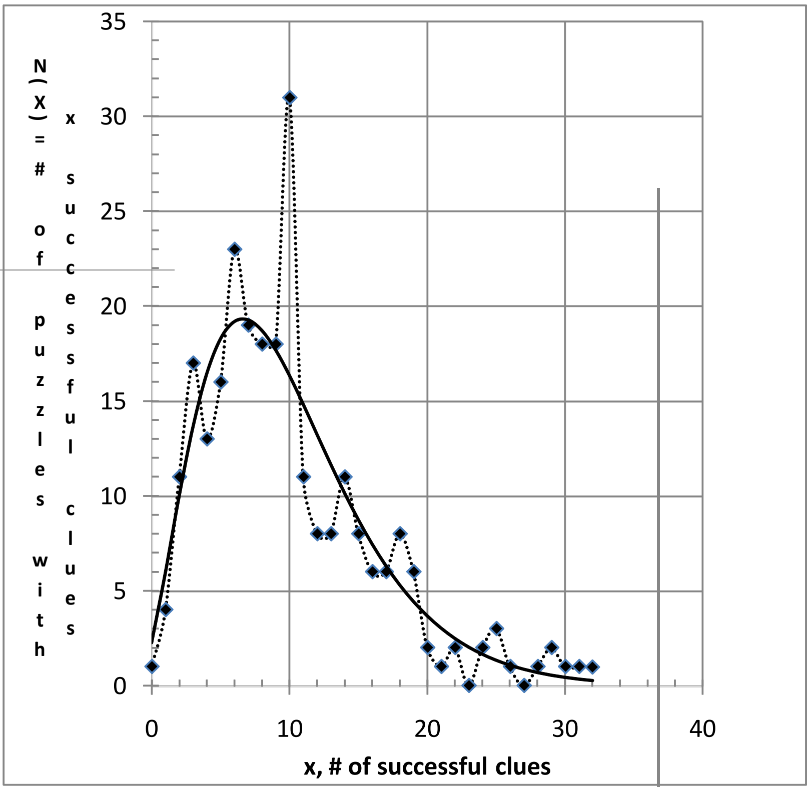
DANIEL PEAKE'S CROSSWORD DATA - SUCCESSFUL CLUES

X = # OF **SUCCESSFUL** CLUES

N(X) = NUMBER OF PUZZLES WITH X SUCCESSFUL CLUES

TABLE 1

x	N(x) obs	N(x) cal
0	1	2.287
1	4	5.977
2	11	10.011
3	17	13.637
4	13	16.444
5	16	18.286
6	23	19.190
7	19	19.281
8	18	18.726
9	18	17.699
10	31	16.360
11	11	14.842
12	8	13.255
13	8	11.677
14	11	10.167
15	8	8.762
16	6	7.482
17	6	6.337
18	8	5.329
19	6	4.452
20	2	3.697
21	1	3.054
22	2	2.510
23	0	2.054
24	2	1.674
25	3	1.358
26	1	1.099
27	0	0.886
28	1	0.712
29	2	0.570
30	1	0.456
31	1	0.363
32	1	0.289
NTOT	260	258.9
m	9.923	
s	6.138	
p=m/s*s	**0.263**	
k=mp/(1-p)	**3.548**	



Diamonds are the observed values. Solid Line is the best fit for values according to the Negative Binomial Distribution

TABLE 2

X	N(X) OB	N(X) CA
0-2	16	18.3
3-5	46	48.4
6-8	60	57.2
9-11	60	48.9
12-14	27	35.1
15-17	20	22.6
18-20	16	13.5
21-25	8	10.6
26-32	7	4.4
NTOT	260	258.9

N(X)OB = NUMBER OBSERVED

N(X)CA = NUMBER CALCULATED AS PER

NEGATIVE BINOMIAL DISTRIBUTION (NBD)

Figure 4

intervals. In *Table 2* in the figure, the grouped data are shown. With nine groups the *ndf* is 6 and the χ^2 = 7.9, making the NBD hypothesis acceptable.

The contrasting behaviour of NBD for 'successes' and 'failures' is a vindication of the rationale for the choice of NBD. The model for NBD as 'mixture distribution' – a superposition of many Poisson distributions with their parameters λ's (the mean values) distributed as a Gamma distribution – is now valid for two sets of data (of two solvers) in widely differing domains of the *(p,k)* space.

4. COMPLEXITY OF A COMPOSER'S PUZZLE.

One solver of many puzzles by different composers, provides an average measure of complexity of puzzles for the solver. One can consider the complementary problem of a composer estimating the complexity of his puzzle(s) by soliciting data from his solvers. In general, a composer has no quantitative data about the complexity of his puzzle for the solvers. We propose an experiment to get a distribution *N(x) vs x* where *N(x)* is the number of solvers with *x* failures/successes.

A solver sends an SMS

Puzzle #, *# of failures* *# of successes*

an unsigned, therefore anonymous message to the composer. A suitable cap on the time allowed – say a week – is suggested.

There are practical problems, mainly related to uncertainties in solver behaviour and possible biases (e.g. poor solvers may not want to communicate). We return to the topic later in Section 8.

From the distribution of errors *(m,s)* can be calculated. These two parameters give a broad estimate of the complexity of the puzzles: the mean *m* defines the average and the standard deviation *s* the spread or the tail of the distribution. These are model-independent measures of complexity. But *(m,s)* does not define the details of the distribution over the entire range of errors. The NBD offers the solution. The *(p,k)* are determined from *m* and *s* as

$$p = m / s^2 \qquad \text{and} \qquad k = m\,p / (1-p)$$

The entire error distribution can be calculated and its conformity to data can be decided by a χ^2-test as described in Section 2.1.

Ideally the NBD is a good fit when the mean *m* of errors is small. It will also work when the mean of the successes is small. But in practice, NBD seems to a robust distribution, tolerant of deviations from the theoretical constraints. But the hypothesis of NBD can be tested and rejected based on observations.

5. SIMULATIONS OF NBD: DIFFERENT LEVELS OF SOLVERS' SKILL.

How does NBD appear for wide-ranging complexity as parameterized by *(p,k)*? In *Figure 5* NBD is plotted for four sets of *(p,k)* values representing increasing levels of complexity. As a convenient measure of solver skill we adopt X^* the value of x that corresponds to the maximum probability or $P(X^*)$ is a maximum. Since X^* takes only discrete values $P(X^*+1) = P(X^*)$. From the NBD recurrence relation (eq. 3b)

$$P(X^*+1)/P(X^*) = (1 - p)\,(k+X^*)/(1+X^*)$$

Setting the above equal to 1 and solving for *k*

$$k = (1+pX^*)/(1 - p)$$

For any desired X^* and a specified value of p , k can be calculated. For example for Level 1, X^* is taken as 9. If $p = 0.2$ then $k = 3.5$. This gives the Level 1 parameters.

Level	X*	(p,k)
1	9	(0.2, 3.5)
2	5	(0.5, 6.5)
3	3	(0.8, 17.0)
4	1	(0.95, 39.0)

Note $P(X^*) = P(X^*+1)$. In Figure 5, one can check the positions of the maxima X^* (9, 5, 3, 1) corresponding to Levels 1 to 4.

6. PREDICTIVE POWER OF NBD.

The two parameters that fully characterize NBD, *(p,k)* are derived from *(m,s)* of the entire range of observed distribution of errors. However, it is in principle possible to obtain *(p,k)* from *any two* given probabilities for errors, say *P(0)* and *P(1),* i.e. the error probabilities for no errors and one error. From the recurrence relations (equation 3)

$$P(0) = p^k \qquad \text{and} \qquad P(0)/P(1) = [1/k\,(1 - p)] \qquad (7\text{ a,b})$$

Explore NBD with wide range of 'complexity' of puzzles

SIMULATION OF n(x) EXP FOR A WIDE RANGE OF NBD PARAMETERS (p,k)

SAMPLE SIZE N = 10000

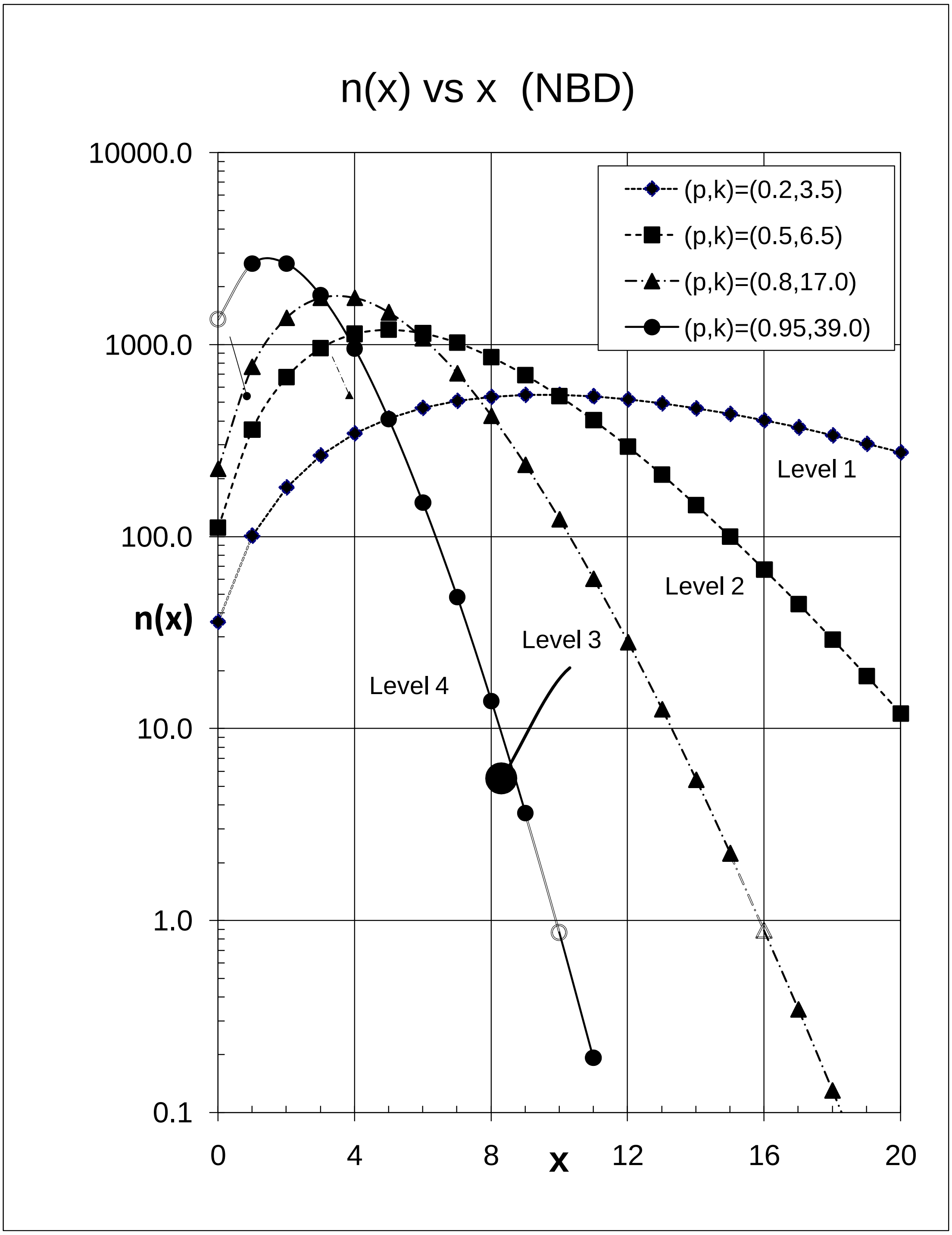


Figure 5

Putting

$$P(0)/P(1) = r_{01} \qquad and \qquad G = r_{01} \log P(0)$$

and eliminating *k*

$$p = 10 \wedge [G\,(1 - p)] \qquad (8)$$

Here *log* is logarithm to base 10. The above equation can be solved for p since G is known. Since p appears in the exponent on the right and also on the left side, the solution is obtained by the method of successive approximations.

We illustrate the method with observed $P(0)$ and $P(1)$ for ALLCW14 (Table 1). Here

$$P(0) = 2500/5484 = 0.4559 \quad \text{and} \quad P(1) = 1246/5484 = 0.2272.$$

$$r_{01} = 2.0066, \quad G = -0.6845$$

Substituting in equation (8)

$$p = 10 \wedge [-0.6845\,(1 - p)]$$

To solve for p start with an initial value p_0 (say 0.5) on the right to calculate p_1 on the left. Then use p_1 on the right to calculate p_2........ The sequence p_0 p_1 p_2.......rapidly converges to the desired solution. The first few iterates are 0.5, 0.4547, 0.4234, 0.4030. To obtain accuracy to four decimal places, 15 iterations give $p = 0.3712$. k can be obtained from equation 7(b) as $k = 0.7925$.

The parameters (p,k) = (0.371, 0.793) can be used to predict the expected number of x errors for $x > 1$. Comparing with the (p,k) obtained from (m,s) – (0.378, 0.803) - these values are less by < 2 %. The latter values are the more accurate since they use the entire data, not just $P(0)$ and $P(1)$. Both the sets of (p,k) when fitted to data yield comparable χ^2 values when subjected to test of NBD hypothesis, viz. 12.8 and 11.6 for $ndf = 11$.

This 'short-cut' to (p,k) works well only when $P(0)$ and $P(1)$ together account for a large proportion of the sample N_T. In the present case they contribute 68.3 % (3746 out of 5484). Although it is tempting to invoke the predictive power of NBD, its true merit is the fact that it fits the entire data extending over a wide range (x = 0 – 15) with just two free parameters. This fit can be objectively validated by the χ^2-test.

7. ONE PUZZLE, MANY SOLVERS: WORD PUZZLE LOTTERY.

In 1940's and 1950's word puzzle lottery was very popular in India. Most puzzles had the following features. (1) 18 clues each with two equally probable answers, say H or T. (2) An entry consists of all the 18 answers filled in. It is submitted with a fee. The promoter receives millions of entries. (3) Each entry is compared with a 'correct' list of answers with the promoter. Cash prizes are offered for entries with no errors and errors 1, 2 and 3.

Usually the entries are filled by 'solvers' in a random fashion, like flipping a coin for each clue to fill in H or T. But this is a hypothesis that needs to be tested.

Here we review the results presented in my paper '*Word Puzzles and the Laws of Chance'* (http://vindhiya.com/snaranan/wordpuzzlelottery/wordpuzzlelottery-2ss.pdf) in 2011. Treating the flips as Bernoulli trials with two possible outcomes with probabilities u and $(1-u)$, the Binomial distribution gives the relative probabilities of entries with number of errors $x = 0,1,2,3........18$. In the present case $u = 0.5$. (See *Figure 6*). The mean is 9.0 and the standard deviation is 4.5. The distribution is symmetric about $x = 9$ with most weight in the range $x = 7 – 11$. Less than 0.4 % contribution is in the low x tail ($x < 4$).

Published results are available only for $x < 4$, for which prizes are offered. The total number of different possible entries is 2^{18} or 262,144. Of these the number of entries with $x = 0,1,2,3$ are in the proportion 1:18:153:816. Results from 12 puzzles are presented in *Table 3*. They are in broad agreement with the predictions.

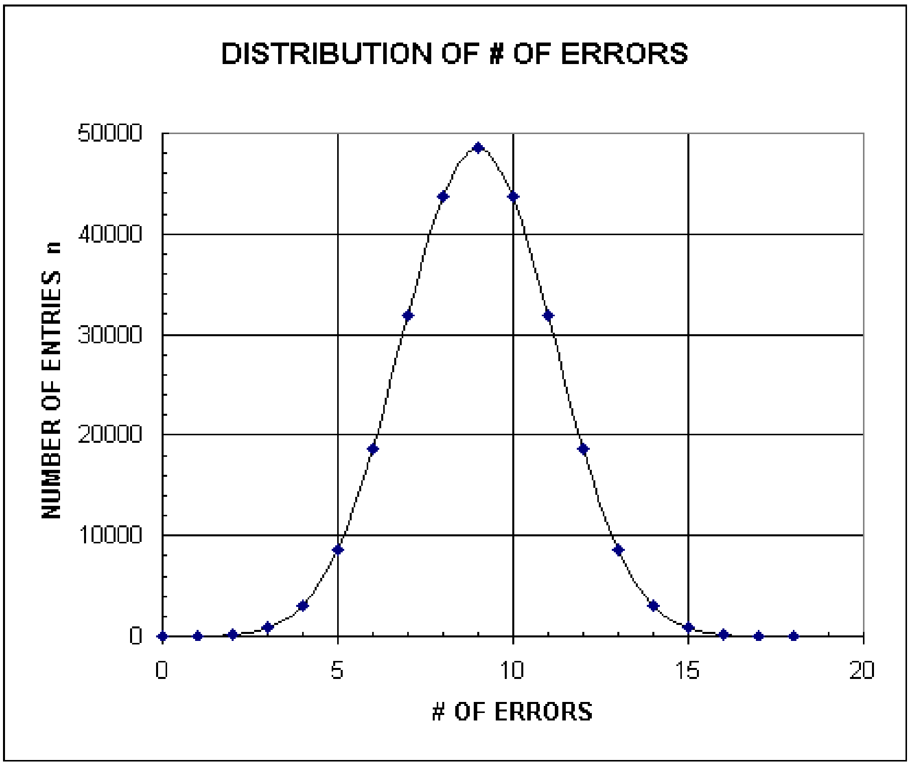
In Section 4, we had posed the problem of 'one puzzle, many solvers' for the cryptic puzzle, with the intent of quantifying the complexity of a composer's puzzles. Various difficulties, known and unknown make the project a complex one to implement.

In contrast, in word puzzle lottery, we have a 'text-book example' of a social experiment tailor-made to test the Binomial distribution, since it has well defined probabilities of success and failure. In a string of n Bernoulli trials, the probability of x failures and $n-x$ successes is given by the Binomial distribution:

$$BD: \; P(x,n) = \binom{n}{x} u^x \; v^{(n-x)} \quad (x = 0,1,2,3.....n), \tag{9}$$

Table

X	n(X)
0	1
1	18
2	153
3	816
4	3060
5	8568
6	18564
7	31824
8	43758
9	48620
10	43758
11	31824
12	18564
13	8568
14	3060
15	816
16	153
17	18
18	1
TOT	262144



BINOMIAL DISTRIBUTION

Figure 6

WORD PUZZLE LOTTERY

Each puzzle has 18 clues

Each clue has two equally probable answers

A puzzle filled with all the 18 answers is an entry

x = # of 'incorrect' answers in an entry

N(x) = # of entries with x incorrect answers

Table gives the N(x) for x = 0,1,2,3 aggregated for 12 puzzles

N(x<4) = Total number of entries (x=0,1,2,3).

x	N(x) OBS	N(x) EXP	OBS-EXP	% DEV
0	43	*83*	-40	-48.193
1	1353	*1502*	-149	-9.9201
2	12826	*12749*	77	0.60
3	68124	*68012*	112	0.1647
N(x<4)	**82346**	***82346***		

THE ESTIMATED TOTAL NUMBER OF ENTRIES IS ABOUT 22 MILLION

Table 3

Here $\binom{n}{x}$ is the number of combinations of x different objects from a collection of n objects, the particular order of the selected objects being irrelevant.

$$\binom{n}{x} = n(n-1)(n-2)(n-3).......(n-x+1)/x!$$

The number of clues n (=18) is not large and the probabilities are not small ($u = v = 0.5$), but the Binomial distribution works for all n and u and no approximation is involved as in the case of Poisson and Negative Binomial distributions.

There is only one correct solution in the lottery as in the cryptic, but the distinction is that in the lottery every permutation of 18 answers (H or T) has an equal chance of being the correct solution whereas in cryptic there is no ambiguity in the clue answers. Being a lottery, the number of entries is very large (typically millions) ideal for testing the tails of BD. But error data are available only for $x = 0,1,2,3$ the low end tail of BD which accounts for only 0.4 % of the total.

8. CONCLUDING REMARKS.

The motivation for the analysis on the distribution of the number of unsolved clues in cryptic puzzles was to put a 'numerical label' to the average puzzle which is a measure of the ease/difficulty of its solution by the solver. In doing so, one goes beyond the qualitative labels 'easy', 'moderate', 'hard'. Lord Kelvin, as quoted in the opening of this paper, noted that by expressing in numbers, our 'knowledge' advances to the state of 'science'. The Absolute Temperature scale (^{o}K) used in physics is named after Lord Kelvin. The zero of the scale (0^{o}K) is -273.16 ^{o}C (Celsius). It is fundamental to the science of Heat and Thermodynamics.

From the raw data on the distribution of errors one obtains the mean and standard deviation *(m,s)* which serve as numerical measures of complexity. But they under-utilize the available data. In finding the Negative Binomial Distribution (NBD) as an appropriate function to describe the whole range of errors (x = 0-15), one has advanced one more step towards the 'science'. Science allows an exploration of the genesis of the statistical distribution that leads to an understanding of the solver behaviour.

The data base of 5484 puzzles for error distribution presented here is perhaps unique in the study of solver behaviour. It is only for one solver. I was seeking data from other solvers. Daniel Peake's data based on 260 puzzles, not only provided one more solver but also a different domain of average complexity. NBD is a good fit also for Peake's results.

But NBD is not alone ! A 3-parameter lognormal distribution (LND2) is also an adequate fit to the error data. This dichotomy may be valid only 'rare events' regime and not for all levels of solver skill.

There remains the complementary problem of 'compiler behaviour'. Does a composer or compiler of a puzzle have any knowledge of the complexity of his puzzle for the average solver? To my knowledge, there is not adequate solver feedback to the compiler that can help put a numerical label on the puzzle. In fact, the compiler needs to have a measure of the complexity of his puzzles as an indispensable aid in the construction of his puzzle. It is not difficult to make a puzzle extremely hard that can confound the solver and turn him away from such puzzles. A good compiler is one who balances clues of different levels of difficulty while avoiding crosswordese, arcane words that seem to be especially suited to construct a grid of interlocking words. Just as in any social enterprise, in crossword puzzles too, there is a 'code of conduct' for the compiler. In the words of Don Putnam, a popular compiler of 'plain and novelty' cryptics: "the compiler, whenever he prepares for battle, must prepare for a battle which he eventually intends to lose. Puzzles which are too hard for his average solver to complete are comparative failures". A proposal for assessing compiler complexity was given in Section 4. It requires a well coordinated effort by a group of devoted solvers who voluntarily provide the feedback to the compiler. With the large number of solvers as in the daily *The Hindu*, one can hope for large sample size far exceeding what a single solver can achieve.

A significant result from my analysis is that the total sample need not be homogeneous, i.e. it can comprise of diverse sources spread over space and time – newspaper dailies across continents and across ages – puzzles of varying levels of difficulty. The model proposed (NBD) for the error distribution indeed incorporates such diversity. Models come with some theoretical constraints, but the results of analysis show robustness tolerant of flexibility in them.

The compiler's data base is the feedback from his solvers, the distribution of errors; *in this case N(x) is the number of solvers with x errors*. Tolerance to diversity noticed in the case of different puzzles can also be assumed for different solvers. For example, it is not necessary the solver should communicate regularly; there can be random skipping of puzzles. There need be no time limit for solving a puzzle; the solver can decide when he wants to declare closure on his efforts and settle for the final error score. (It is my experience that one reaches a grid-lock that cannot be unravelled in any reasonable time). But certain amount of discipline among solvers is desirable especially in avoiding biases, e.g. not communicating a poor score. It is easy to communicate by cell-phone or e-mail in anonymity if desired. It is the task of the compiler to collect the data and process it to the point of tabulating *N(x)* vs. *x*. Further analysis – getting *(m,s)*, *(p,k)* and testing goodness of fit to NBD – can be left to those interested.

Crossword puzzle failure data are not just trivia. The number tags on errors provide the tools for progressing from mere 'knowledge' to 'science' in the spirit of Lord Kelvin. The formulation of NBD for my error data was extended to another observer, providing significant support for the conjecture of universality of NBD for all solvers. The robustness of NBD encourages further organized efforts by groups of solvers, despite the inherent inhomogeneities in solving behaviour. Thus the scientific approach widens the horizons of area of research, while at the same time bringing about unity in diversity and a simplification of the underlying mechanisms. This is illustrated by the finding that failures in clue solution are like flips of a random assortment of biased coins, with varying degrees of probability, say of tails (failures). To the rhetorical question "What is common to car accidents, purchase of branded products and crossword error scores?", the answer is NBD! The common underlying mechanism is the mixture of Poisson and Gamma distributions.

Cryptics and word puzzle lotteries are two extremes on the scale of puzzles graded by skill. Cryptics require different levels of linguistic skill whereas the word puzzle lottery is devoid of any solver skill. However the results are predictable in both. NDB for cryptics and BD for lotteries both have a common origin as reflected in their titles; this is unification.

It is revealing to ponder this common origin. The BD's progenitor is the Binomial theorem which was originally formulated for a positive integral value of the power n in $(u+v)^n$. In the BD the number of Bernoulli trials n is a positive integer. But as the Binomial theorem is extended to negative and not necessarily integral values, it is clear that Bernoulli trial model would not work. But the NBD with a negative real index ($-k$) does have a real-life application as described in the paper. This illustrates the sense of wonder articulated by Einstein. "How can it be that mathematics being after all a product of human thought, is so admirably adapted to the objects of reality?"

To go deeper into the science, one needs the tools of mathematics and statistics. Some of these topics are relegated to the Appendices which duplicate parts of the main text, to ensure that they can be read independently. In particular the last appendix '*Interrelationship of some statistical distributions*' has no direct relevance for the article and is added to illustrate the unity underlying disparate entities.

A P P E N D I C E S

A1. SUPERPOSITION OF NBD'S.

In Section 2.1, it was stated that NBD is a good model for distribution of errors, because it incorporates the diversity of the puzzles as reflected in differing degrees of complexity of the puzzles. This was demonstrated in Naranan (2010), and here we review the proof. The next step is to consider superposition of different groups of data, each of which is an NBD. It is shown that the result is again an NBD.

In Naranan (2010) it was shown that NBD arises as 'mixture' of Poisson and Gamma distributions. The essential steps are as follows. Poisson distribution (PD) has one free parameter λ

$$\text{PD: } Prob\,(x) = P(x) = \exp(-\lambda)\, \lambda^x / x! \qquad (\lambda > 0, x = 0,1,2....) \qquad \text{(A1)}$$

Consider many PD's superposed with a spectrum of λ values which can be expressed as a Gamma distribution $\Gamma(\lambda)$:

$$GD: \; P(\lambda) = \exp(-\lambda/\beta)\, \lambda^{\alpha-1} / [\Gamma(\alpha)\, \beta^{\alpha}] \qquad (\alpha, \beta > 0) \qquad \text{(A2)}$$

It has two parameters α the shape parameter and β the scale parameter. The GD is structured similar to PD and is an appropriate choice for the distribution of λ. The resulting mixture distribution is

$$MD: \; P(x) = \int_0^{\infty} Poisson\,(x|\lambda)\, \Gamma(\lambda)\, d\lambda$$

MD has the same form as NBD

$$NBD: \; P(x) = \{\Gamma(k+x)/[\Gamma(k)\, x!]\}\; p^k q^x \qquad (k > 0, \; x = 0,1,2......) \qquad \text{(A3)}$$

with $q = 1\text{-}p$

$$k = \alpha \qquad \text{and} \qquad p = 1/(\beta + 1)$$

The mixture distribution model mirrors closely the real world of crossword puzzles. The basic character of errors is Poissonian, but the total sample of puzzles is a mixture of puzzles with varying complexity (λ).

A remarkable outcome of the MD is that the NBD parameters (p,k) are completely determined by the GD parameters. A superposition of NBD's, therefore is equivalent to a superposition of GD's with (α_i, β_i) (i = 1,2,3....). It is well known that a superposition of GD's is also a GD, provided all the constituent GD's have the same β, or $\beta_i = \beta$ for all i. There is no restriction on α_i's (Keeping 1962).

Since $\beta = (1 - p)/p$ all the p's of the constituent NBD's must be the same for the superposition to remain an NBD. As $p = m/s^2$, a measure of over-dispersion, both m and s^2 may vary in tandem to keep their ratio constant. The above result can be derived by a standard formal technique of cumulant generating functions to find the distribution of a superposition of multiple distributions.

To summarize, multiple sets of data on error statistics of crossword puzzles, each conforming to NBD, can be lumped together to yield an NBD, provided the constituent sets have the same over-dispersion.

In the real world, observed statistical distributions are known to be robust and stable with respect to the constraints of theory. There is theoretical justification for this robustness (Feller 1972). However part of the robustness can be attributed to the generally limited sample sizes masking the deviations from theoretical expectations.

It was mentioned in the Introduction that the original data of 3404 puzzles consisted of four diverse sets of data. The individual sets as well as the cumulated set conformed to NBD. It can be seen from the (p,k) values that three of the four constituent sets had the same p ($\approx$ 0.460) and the fourth had a higher value (0.629). The cumulated set (ALLCW) had p = 0.455 (Naranan 2000).

A2. INTERRELATIONSHIP OF SOME STATISTICAL DISTRIBUTIONS.

Here we review the connections between several distributions that have figured in our paper. (See Box 1). At the top is the 'mother of all distributions', the Binomial distribution (BD) based on Bernoulli trials. Each trial has only two outcomes: 'failure' with probability u and 'success' with probability $1-u$. In a string of n independent trials the probability of x failures and $n-x$ successes is given by BD.

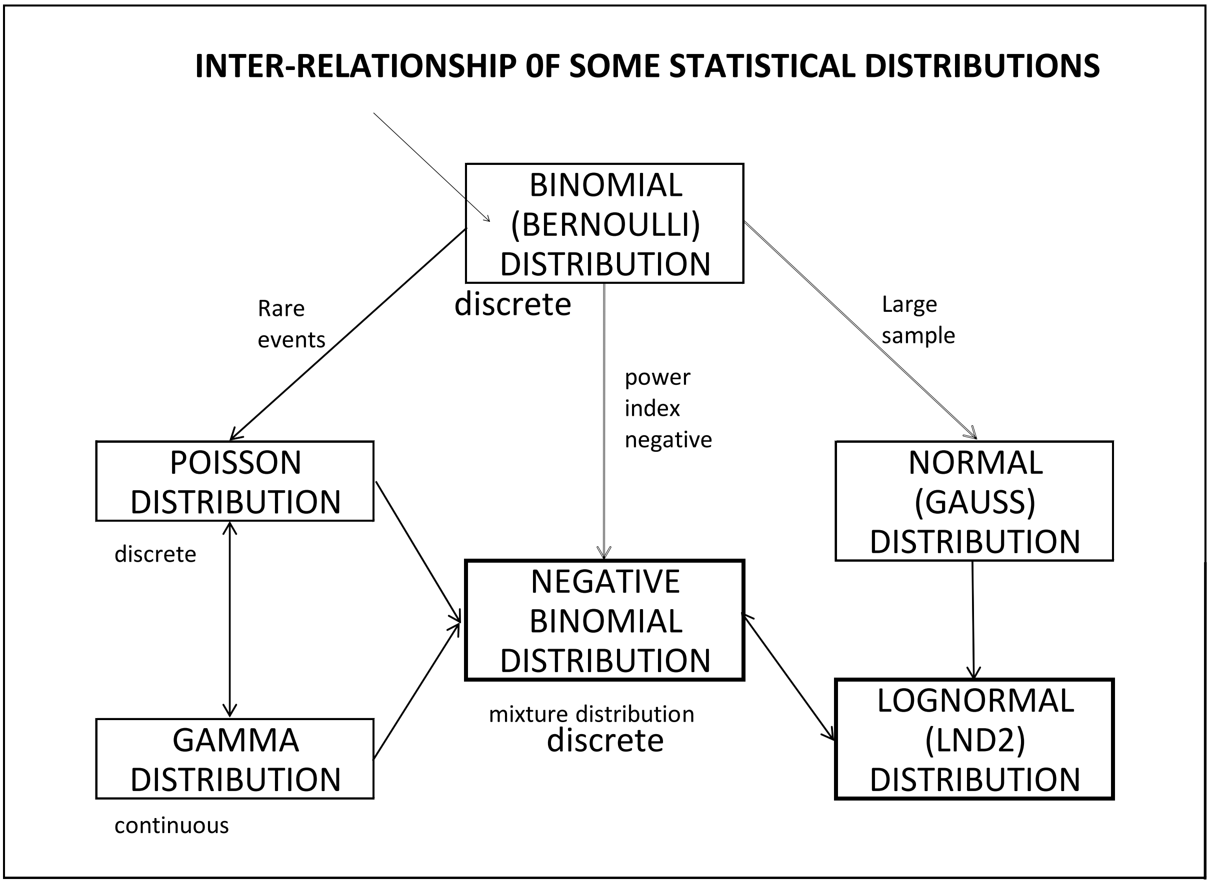
$$BD: \; P(x,n) = \binom{n}{x} u^x \; v^{(n-x)} \quad (x = 0.1.2,3.....n), \qquad \text{(A4)}$$

Here

$$\binom{n}{x} = n(n-1)(n-2)(n-3).......(n-x+1)/x!$$

The name 'Binomial' arises from the fact equation (A4) is the x^{th} term in the binomial expansion $(u + v)^n$. The mean m and variance s^2 of BD are

$$m = n\,u \qquad and \qquad s^2 = n\,u\,(1 - u)$$


INTER-RELATIONSHIP 0F SOME STATISTICAL DISTRIBUTIONS
BINOMIAL (BERNOULLI) DISTRIBUTION
discrete
Rare events
power index negative
Large sample
POISSON DISTRIBUTION
discrete
NORMAL (GAUSS) DISTRIBUTION
NEGATIVE BINOMIAL DISTRIBUTION
mixture distribution
discrete
LOGNORMAL (LND2) DISTRIBUTION
GAMMA DISTRIBUTION
continuous

Box 1

For BD $m > s^2$ since $u < 1$.

When the number of trials n is large and u is low (rare events) such that $nu = \lambda$ is small, then BD becomes PD.

$$\text{PD: } Prob\ (x) = P(x) = \exp(-\lambda)\ \lambda^x / x! \qquad (\lambda > 0, x = 0,1,2....) \qquad \text{(A1)}$$

with $m = s^2 = \lambda$

PD is a discrete distribution. The Gamma Distribution (GD) is Poisson-like but takes continuous values for x

$$GD:\ P(x) = exp\ (-x/\beta)\ x^{\alpha-1} / [\Gamma(\alpha)\ \beta^{\alpha}] \qquad (\alpha, \beta > 0) \qquad \text{(A5)}$$

Here $m = \alpha\,\beta$ *and* $s^2 = \alpha\,\beta^2$

When β the scale parameter is 1, $m = s^2$ as in PD.

As described in Appendix A1, the Negative Binomial Distribution (NBD) is a mixture distribution MD which is a convolution of Poisson and Gamma distributions.

$$NBD:\ P(x) = \{\Gamma(k+x)/[\ \Gamma(k)\ x!]\}\ p^k\ q^x \qquad (k > 0,\ x = 0,1,2......) \qquad \text{(A3)}$$

BD can be rewritten using factorials as

$$BD:\ P(x,n) = \{\ \Gamma(n)/[\ \Gamma(n\text{-}x)x]\}!\ u^x\ v^{(n-x)} \qquad (x = 0.1.2,3.....n) \qquad \text{(A5)}$$

The correspondence between BD and NBD is:

$$n \rightarrow k + x \qquad u \rightarrow q \qquad v \rightarrow p$$

for k a positive integer.

The BD can be generalized to negative values of n since the Binomial theorem from which BD is derived is valid for $n < 0$. It can be shown that NBD behaves like a BD with a negative power index. The coefficient in BD $\binom{n}{x}$ has the index n a positive integer. For a negative index k

$$\binom{-k}{x} = -k\ (-k-1)(-k-2)................(-k-x+1) / x!$$

$$= (-1)^x\ (k+x-1).............(k+2)(k+1)k / x!$$

$$= (-1)^x \binom{k+x-1}{x}$$

$$= (-1)^x\{\ \Gamma\ (k+x)/[\ \Gamma\ (k)\ x!)]\}$$

or

$$\{\ \Gamma\ (k+x)/[\ \Gamma\ (k)\ x!)]\} = (-1)^x \binom{-k}{x}$$

Substituting the above in NBD equation (A3)

$$NBD:\ P(x) = p^k\ \binom{-k}{x}(-q)^x \qquad (x = 0,1,2....)$$

The above can be viewed as p^k times the x^{th} coefficient in the binomial expansion of

$$(1 - q)^{-k}$$

with a negative power index $-k$. The interrelationship of BD, PD, GD and NBD is shown in Box 1.

When the sample size n the power index of the BD (equation A5) is very large the BD becomes the normal distribution (ND).

$$ND: P(x) = (1/\sigma\sqrt{2\pi})\ \exp [-(x-\mu)^2/2\ \sigma^2] \qquad -\infty < x < +\infty \qquad (4)$$

Here the mean $\mu = n\ u$ and the variance $\sigma^2 = n\ u(1-u)$. As n increases, σ which is proportional to $\sqrt{n}$ increases; so the distribution becomes broader and more symmetric. Skewness, a measure of the asymmetry is $(1 - 2u)/\ \sigma$. It is exactly 0 only when $u = 0.5$, but tends to 0 as n increases.

The ND variously called 'normal', 'Gaussian' or the 'bell curve' is also called – very appropriately for our purposes – the 'curve of error'. It is perhaps the most important of all distributions in statistics (Keeping 1964). An important variant of ND is the lognormal distribution LND1. In LND1, instead of x it is $\ln x$ which is the variate.

$$LND1: P(x) = (1/\sigma\sqrt{2\pi})(1/x)\ \exp [-(\ln x-\mu)^2/2\ \sigma^2] \qquad x > 0 \qquad (5)$$

Here $\ln x$ is the natural logarithm of x. Whereas ND is symmetric, LND1 is asymmetric with a long tail just like the NBD, with two parameters μ and σ.

In another variant of ND - the LND2 - $\ln (x+X_o)$ is the variate.

$$LND2: P(x) = (1/\sigma\sqrt{2\pi})[1(/x+ X_0)]\ \exp \{-[\ln(x+ X_0)-\mu])^2/2\ \sigma^2]\} \qquad x > 0 \qquad (6)$$

X_0 is a constant, positive or negative (Cramer 1953). When $X_0 = 0$, LND2 reduces to LND1.

The distribution of errors in crossword puzzles (x) is equally well described by NBD and LND2. This dichotomy may exist only in the 'rare error' regime, i.e. when the mean number of errors is small. We have not established a direct link between NBD and LND2, although both are traceable to the BD (Box 1). LND2 requires three free adjustable parameters and is the less preferred compared to NBD which has only two parameters.

CROSSWORD PUZZLES - "ERROR" (FAILURE) ANALYSIS

COMPARISON OF BINOMIAL (BD), POISSON (PD) AND NEGATIVE BINOMIAL (NBD) DISTRIBUTIONS			
p(x) = probability of x failures. (x=0,1,2....)			
	BD	**PD**	**NBD**
p(x)	**nCx. u^x.v^(n-x)**	**exp(-λ).λ^x/x!**	**[Γ(k+x)/([Γ(k) x!)] p^k.q^x**
Parameters	n= #of clues in grid	λ (>0)	k (>0), [Γ(k)=(k-1)[Γ(k-1), k>1
	u=prob of failure (<<1)		[Γ(k) =(k-1)! (k integer)
	v=1-u		p (<1), q=1-p
Mean **m**	nu	λ	k(1-p)/p
Variance (**s*s**)	nu(1-u)	λ	k(1-p)/(p*p)
D = m/(s*s*)	1/(1-u)	1	p
Recurrence Rel of p(x)	**p(0) = (1-u)^n**	**p(0) = exp (-λ)**	**p(0) = p^k**
R(x)=p(x+1)/p(x)	**[1/(1+x)] (n-x)u/(1-u)**	**[1/(1+x)].λ**	**[1/(1+x)] (k+x)(1-p)**
Given p(x*) is max	u=(1+x*)/(1+n)	λ=1+x*	k=(1+px*)/(1-p)
Parameters	u=1-(s*s)/m =1-(1/D)	λ=m=s*s	p=m/(s*s) = D
in terms of m,s*s			k=mp/(1-p)

Variable **x** is positive integer. p(x*)=p(x*+1).

All possible values of D are covered: D >1 (BD), D=1 (PD), D < 1 (NBD)

Crossword puzzle data favor D < 1 (NBD). Simulations done for BD, PD too

Approx: BD tends to PD when n is large, u is small and nu =λ(small)

NBD tends to PD when k is large, q is small and kq =λ(small)

Table 4

Finally we mention the χ^2- distribution that is used for testing hypothesis. The χ^2-distribution is actually a Gamma distribution with its variate related to the standardized normal variate $z = (x - \mu)/ \sigma$.

In Table 4 is given a comparison of the properties of three discrete distributions, Poisson, Binomial and Negative Binomial Distributions.

ACKNOWLEDGEMENT.

I thank Daniel Peake for his ready consent to my request to use data from his "2500 clue challenge". Shuchismita Upadhyay noted the significance of Peake's results for my work. She extracted the relevant histograms (of failures and successes) which served as the basic inputs for my analysis. I am grateful to her for providing a vital link in my tests of NBD for crosswords error distribution.

DEDICATION.

This paper is dedicated to the memory of Visalakshi Naranan. As in life, in after-life too she is a source of encouragement in my academic pursuits.

Chennai S. Naranan

20 June 2015 e-mail: snaranan@gmail.com

REFERENCES

Aitchison, J. And Brown, J.A.C. (1957). *The Lognormal Distribution*. Cambridge . University Press, Cambridge.

Cramer, H. (1955). *The Elements of Probability Theory and some of its applications*. John Wiley & Sons, Inc. New York, London, Sydney.

Feller, W. (1972). *An Introduction to Probability Theory and its applications.* Vol.1. Wiley Eastern Private Ltd. New Delhi, India.

Keeping, E.S. (1962). *Introduction to Statistical Inference*. D. Van Nostrand Co. Inc. Published as East-West Student Edition, New Delhi (1964).

Naranan, S. (2010). *A Statistical Study of Failures in Solving Crossword Puzzles*. Journal of Quantitative Linguistics. 17, 191-211. See also http://vindhiya.com/snaranan/Xword-puzzle/xwordarticle2012-1.pdf

(This paper is commented upon in
http://www.crosswordunclued.com/2010/09/failures-in-solving-crosswords.html
http://www.openthemagazine.com/article/nation/the-high-of-hidden-words
https://books.google.co.in/books/about/The_Centenary_of_the_Crossword.html?id=Ob0FmwEACAAJ&hl=en)

Naranan, S.(2011)
http://vindhiya.com/snaranan/wordpuzzlelottery/wordpuzzlelottery-2ss.pdf

Peake, D. http://danielpeake.com/blog1/2500-clue-challenge/

Putnam, Don. (1974). *Crosswords for the Devotee*. Paperfronts, Elliot Right Way Books, U.K).

Upadhyay, Shuchismita. http://www.crosswordunclued.com.

CRYPTIC CROSSWORD PUZZLES: A STATISTICAL ANALYSIS OF ERRORS IN SOLUTION

CONTENTS

1. Introduction. 2
2. Cryptic Crosswords: Data on failures in solution. 4
   - 2.1. Negative Binomial Distribution (NBD). 6
   - 2.2. Lognormal Distribution. 9
3. Crossword Puzzles: Daniel Peake's data. 12
4. Complexity of a composer's puzzle. 15
5. Simulations of NBD: Different levels of solvers' skill. 16
6. Predictive power of NBD. 16
7. One puzzle, many solvers: Word Puzzle Lottery. 19
8. Concluding Remarks 22

Appendix A1. Superposition of NBD's. 26

Appendix A2. Interrelationship of some statistical distributions. 27

References. 33

Contents 34.

TABLE AND FIGURE CAPTIONS

Table 1. Crossword failure data update.

Table 2. Comparison of NBD and LND2 parameters.

Table 3. Word Puzzle Lottery.

Table 4. Comparison of BD, PD and NBD.

Figure 1. Crossword failure data update: NBD fits to data.

Figure 2. Crossword failure data update: LND2 fits to data.

Figure 3. Daniel Peake's data: unsuccessful clues.

Figure 4. Daniel Peake's data: successful clues.

Figure 5. Explore NBD with wide range of complexity of puzzles.

Figure 6. Binomial distribution: Distribution of errors.

Box 1. Interrelationship of some statistical distributions.