Warren W. Esty

THE GEOMETRIC MODEL FOR ESTIMATING THE NUMBER OF DIES

Statistical methods for estimating the number of dies have evolved over time. When it became clear that not all dies produced equal numbers of coins, Carter proposed the "negative-binomial" model with parameter p=2 and its corresponding "simplified method" which was a great improvement on the previous "equal output" estimator.[1] The graph of the numbers of coins produced by dies looks like Figure 1 in that negative-binomial model.

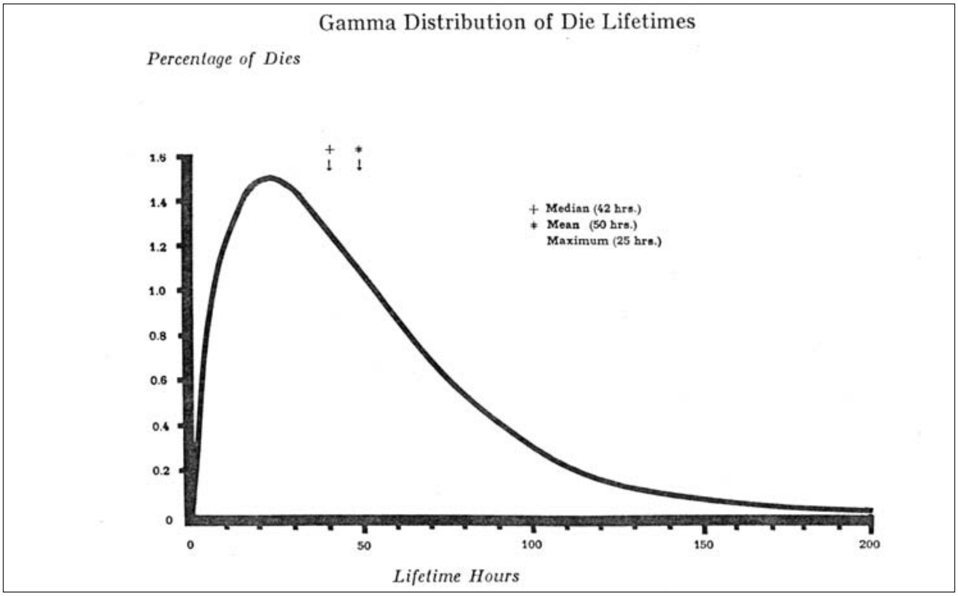

Fig. 1. - Carter's original 1983 plot of a proposal for the probability distribution of die-lifetime. The horizontal axis gives the lifetime, and the vertical axis gives the percentage of dies with the given lifetime.

The model was proposed and accepted because he showed it gave estimates quite close to the true number of dies in the case of the Crepusius denarii in the Roman Republican series where there was extensive die-study data and the true value was approximately known.[2] Since then a great deal more data from many other series of coins has become available. This paper analyzes various estimation methods using the suitable issues from over 600 sets of data assembled by de Callataÿ, as well as additional Crepusius data.

The research below shows that an overwhelming majority of data sets do not fit the $p = 2$ model. In contrast, the $p = 1$ model, which is simply the famous "exponential" or "geometric" model of failure time, fits most data sets quite well, including the Crepusius data updated in 1997.[3]

The Data

For any given issue the data consists of the sample size (the number of coins), n, the number of different dies observed, d, and the die-frequencies–the number of dies observed exactly 1, 2, 3, ... times, denoted d_1, d_2, d_3, ... Then $d = d_1 + d_2 + d_3 + ...$ and $n = 1d_1 + 2d_2 + 3d_3 + ...$. The number of dies observed exactly 1 time, d_1, is the number of singletons. The number observed 0 times is d_0. This is the number of unobserved dies, which is exactly what we want to know to estimate the original number of dies, which is $D = d + d_0$. The mean number of coins per observed die is $R = n/d$, which sometimes has been called the "characteristic index". It is a measure of how thorough the sample is.

The two curves in Figure 2 plot on the vertical scale the expected number of dies of the given frequency in the geometric case. Compare the mean 2 (plotted with triangles) and mean 5 (plotted with squares) curves.[4] If the true mean is 2 there is a higher chance of any given die being observed not at all or only once, compared to mean 5, and a lower chance of it being observed many times. Also, the chance of a die not being observed in a sample is always greater than the chance it is observed any other particular number of times (Every plot is higher on the left). However, this is not to say there will always be unobserved dies. If the mean number of coins per observed die is high enough, all the individual die-frequency probabilities are low, including the chance there remains an unobserved die.

Under the geometric model, actual data will vary randomly, but a plot of

[1] Carter 1983, 197-198. The "gamma" distribution is the continuous version of the discrete "negative-binomial" distribution.

[2] The reverse dies have sequential Roman numerals on them (1 to 519) and the obverse dies have each of various symbols combined with each of 21 letters.

[3] The Crepusius data was very kindly provided by Dr. Ted Buttrey. The "exponential" distribution is the continuous version of the discrete "geometric" distribution.

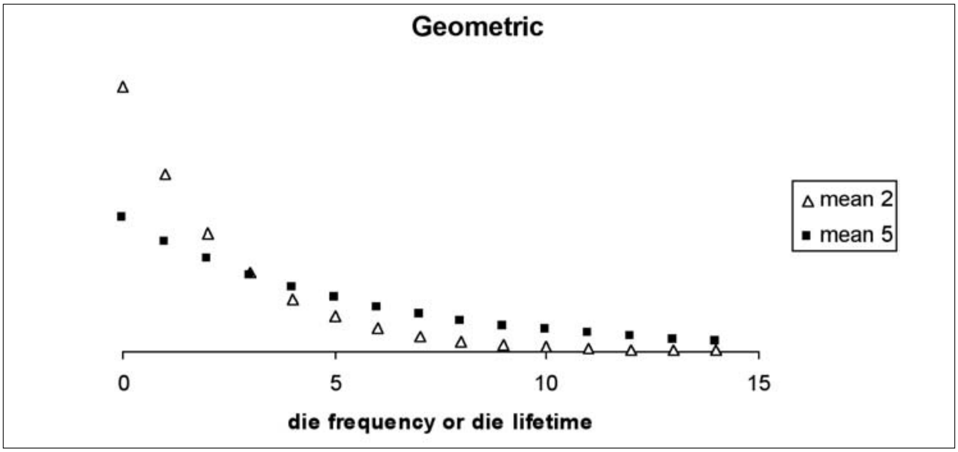

Fig. 2. - The vertical axis plots expected die frequencies in the geometric case (p = 1) if the number of coins observed per die has mean 2 or 5.

observed die frequencies should look somewhat like these curves, or a similar curve generated with a different mean.

Figure 2 also can be regarded as plotting the distribution of the original numbers of coins produced by dies. If the original distribution is geometric (perhaps with a mean in the thousands or tens of thousands), random sampling from it yields a sample which is also geometric, which was the first interpretation of Figure 2. So the same figure serves double duty. To depict the original numbers of coins produced by dies, discard the horizontal labels "5", "10" and "15" and let the mean number of coins per die replace "5" (So, the label "5" might now be replaced by, say, "10,000"). Then the "mean 5" symbols plot the probability density of the original number of coins per die. Under the geometric assumption, dies are more likely to produce few coins than the average number of coins (Each plot is higher on the left, regardless of the mean), and some produce far more than the average number of coins (the tail on the right is long). It is likely the geometric model would have been proposed and accepted long ago had the initial analysis of that one anomalous Crepusius data set not been so convincing, because the exponential distribution is the classic model of failure time in engineering and should leap to mind first when modeling the time until a die fails.[5] Now the work of de Callataÿ in assembling additional data and the work of Carter and Buttrey in extending the

[4] The $p = 1$ model corresponds to counting the trials before a single particular event happens in a sequence of independent trials. For example, count the number of rolls of a die before a "3" appears. The distribution will be the "mean 5" curve in Figure 2, with the leftmost square above 0 at height 1/6. The $p = 2$ model has much less variance.

[5] A possible explanation of the anomaly discussed below.

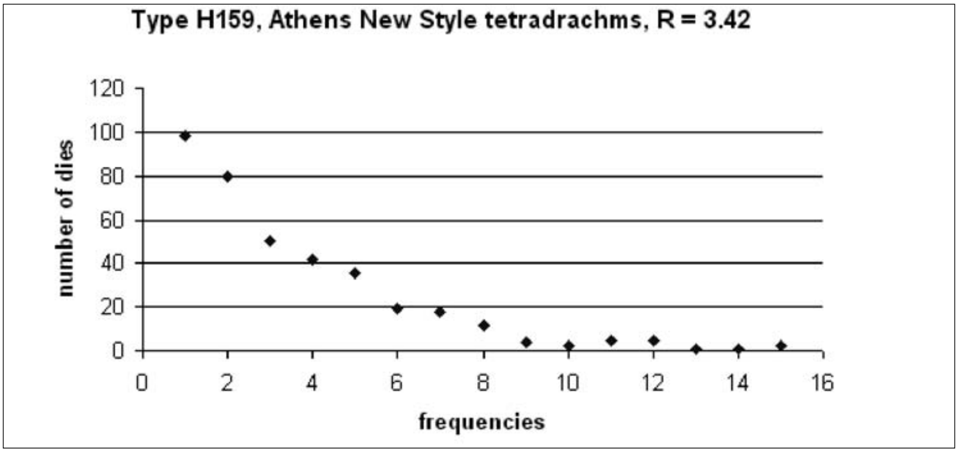

Fig. 3. - Die-frequencies from Hellenistic type 159, Athens new style tetradrachms.
n = 1320 coins from d = 375 observed dies
The horizontal "frequencies" axis gives the number of coins from a given die, and the vertical axis gives the number of dies represented by exactly that many coins.

Crepusius data set have made it possible to demonstrate that this geometric-distribution model fits the data much better. The associated estimator of the original number of dies is

$$(1) \quad e_1 = \frac{nd}{n-d}$$

where n is the sample size (number of coins) and d is the number of different dies observed.[6]

The number of singletons is also important because it is closely related to the number of unobserved dies. A plot of die-frequencies, where d_i is plotted above i, has d_0 just to the left of d_1. For example, consider de Callataÿ Hellenistic type 159, which is a typical data set of its size.

We do not need to know the correct model to estimate, by eye, that the graph would cross the y-axis at about 130; the number of unobserved dies is about 130. The new model estimates that it is 149. The $p = 2$ model estimates it to be 73, which appears too low. When the number of coins per die is smaller, it is harder to pick a particular value for the number of unobserved dies. Project the curve in Figure 4 to frequency 0 and the estimate of d_0 will be very high. The number of

[6] This is derived below. It yields a point estimate. An associated confidence interval, such as formula (4) in Esty 2006, can be very wide.

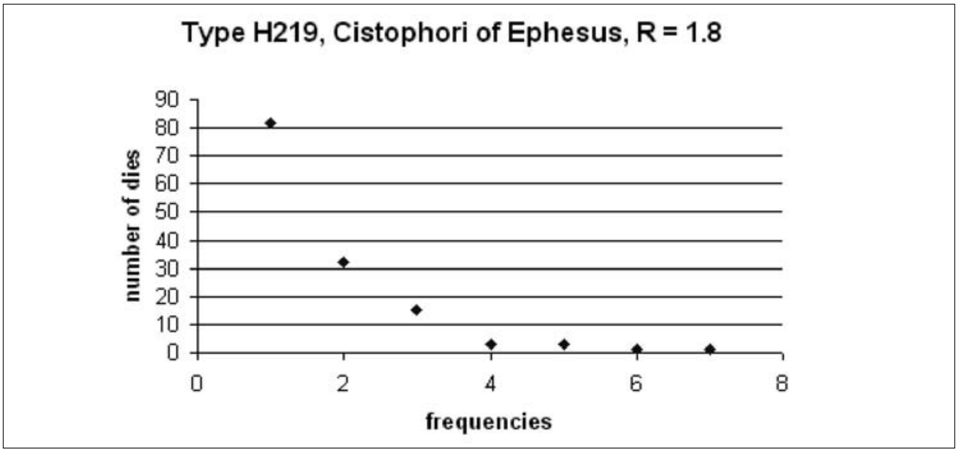

Fig. 4. - Hellenistic Type 219. n = 255 and d = 141.

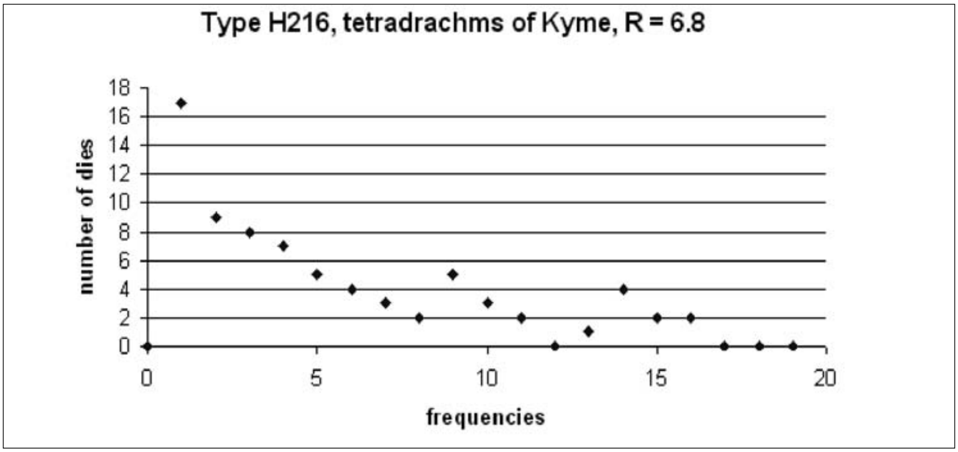

Fig. 5. - Die frequencies from Hellenistic type 216. n = 537 and d = 79.

unobserved dies is large and the deviation of the estimate from the true value could be very large.

The old model estimates d_0 to be 111. The new geometric model estimates it to be 174. Figure 5 exhibits a more-thorough sample in which there are almost seven coins per die.

The old model estimates d_0 to be 5. The new model estimates d_0 to be 14. Even with $R = 6.8$ coins per observed die, there are still many singletons – more than any other frequency. It is typical to see more singletons than any other frequency. This is a feature of the geometric model that the $p = 2$ model does not have.

All "negative binomial" models (including Figures 1 and 2) have the convenient statistical property that the expected die-frequencies in a sample have the same type of distribution as the original die-lifetime distribution, just rescaled. Looking at Figure 1, this means that we relabel the *x*-axis with frequencies (like Figures 3-5) then the area above the numbers would resemble the frequencies. If R is large, in Figure 1 the number of singletons would be plotted just to the right of the *y*-axis where the curve is low; there would be fewer singletons than dies with higher frequencies. However, this result is rarely observed in hundreds of sets of data. Figure 1 must be wrong. In fact, when the number of dies is large, as more and more coins are entered into a sample, even as old singletons turn into doubletons or tripletons, new singletons keep appearing, as expected in the geometric model.

> Hellenistic type 134 (Tetradrachm of questeur Aesillas): $n = 506$, $d = 92$. $R = 5.5$.
> Frequencies: 37, 14, 8, 6, 3, 1, 6, 1, and others on out to one die with 24 coins, one with 32 and one with 50.
>
> Hellenistic type 171 (Dichalcon of Hermione, Peloponesus): $n = 47$, $d = 33$. $R = 1.42$.
> Frequencies: 25, 3, 4, 1.
>
> Hellenistic type 309 (Tigranes of Armenia tetradrachm): $m = 241$, $d = 49$. $R = 4.92$.
> Frequencies: 17, 8, 6, 0, 4, 2, 0, 1, 2, 0, 2, 0, 1, 0, 2, 2, 1, 1

These three examples are typical and illustrate that singletons are common. There are more singletons than fit the $p = 2$ model, but, in the past we did not know what to do about it because the Crepusius case seemed conclusive. By the way, the Type H171 data is not really out of line. Even with $p = 2$, d_1 is estimated by 22.8 (for 25) and with the geometric model the estimate is 23.2, as close as could be expected. It is just that the number of coins per die is small, so singletons are frequent.

So, were we just noticing a few exceptional cases? Already in 1992 Esty and Carter noted that the data sets they had were fit better using a value of p less than 2, but it took the publication of the voluminous de Callataÿ data in 1993 and 2003 to allow certainty that there are regularly more singletons than other frequencies. De Callataÿ noted the problem and proposed a non-specific solution: There is some unknown fraction of dies that broke very early (Figure 6).[7]

The theory was that, in issues with many dies, as samples became more thorough, the numerous low-output dies (on the left) would occasionally appear, yielding singletons, even if many of the dies had already appeared numerous times.

Often the number of dies is used as a proxy for the size of the issue. More dies suggest a larger issue. Unfortunately, the true number of dies would include the

[7] de Callataÿ 1993, 45.

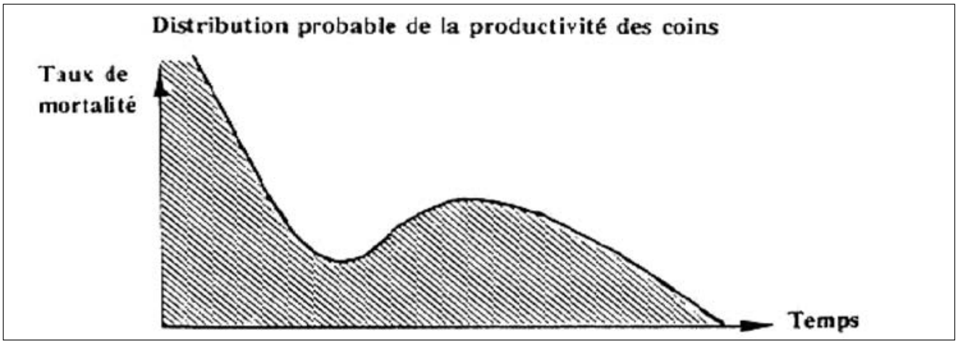

Fig. 6. - de Callataÿ 's 1993 proposed plot of the numbers of dies with various die lifetimes.

number of low-output dies, even though they made a relatively small contribution to the overall output of coins and to the coverage of a sample. If there are many dies, with this model there is no way to know if there are many more low-output dies that remain unseen in the sample. Should we estimate the number of dies to include some fraction of low-output dies, or should we base our estimate on relatively few low-output dies and be content with a die-number that reflects the number of dies that did not break almost immediately?

Dealing with the coverage instead of the original number of dies provided a partial solution. The coverage of a sample, C, is:

(2) $$\frac{\text{the number of coins in the original population struck by dies in the sample}}{\text{the total number of coins struck by all the dies}}$$

The coverage measures how thorough the sample is. It is a property of the sample, not of the population, and is not observable. However, it can be estimated by this formula:

(3) $$C_{est} = 1 - \frac{d_1}{n}$$

This estimator is non-parametric and remarkably robust, which strongly supports its use.[8] Furthermore, even if we make the unwarranted assumption that we know the true model and its value of p, that extra knowledge would not make much difference to the accuracy of the estimates. Figure 7, with hypothetical data,

[8] Esty 1986 showed it is remarkably efficient compared to even the most restrictive parametric model, the "equally likely" case (the rightmost column in Figure 7). The efficiency is usually in the 98% range and never less than 85%, which occurs only when the number of coins per die is high. It is even somewhat robust, that is, it is relatively insensitive to some non-randomness. This estimator of the coverage is excellent and only non-randomness could introduce unexpected errors.

negative binomial	shape parameter *p*	0.5	*p = 1 geometric*	*p = 2*	3	10	∞
R = n/d = 1.43 e.g. *n* = 100 and *d* = 70	Estimate of *C*	0.504	*0.510*	*0.517*	0.521	0.529	0.533
	Estimate of *D*	336	*233*	*182*	165	141	131
R = n/d = 2.5 e.g. *n* = 100 and *d* = 40	Estimate of *C*	0.826	*0.840*	*0.855*	0.864	0.882	0.892
	Estimate of *D*	90.5	*66.7*	*55.2*	51.5	46.7	44.8

Fig. 7. - The parametric negative-binomial estimates of C and D based on n and d alone in two typical cases. The parameter value ∞ corresponds to the (incorrect) case when all dies produce equal numbers of coins.

shows that different values of *p* can make a huge difference in the estimate of the number of dies, *D*, while making relatively little difference in the estimate of the coverage.

Inspection of Figure 7 shows that the estimate of the number of dies, *D*, depends strongly on the choice of the parameter *p*, whereas the parametric estimate of the coverage, *C*, does not. In fact, we might as well use the non-parametric estimate if *C*, (3), regardless of *p*. If we do, we can derive another estimator for the original number of dies, *D*.[9]

$$(4) \qquad e_p = \left(\frac{d}{C_{est}}\right)\left(1 + \frac{d_1}{pd}\right)$$

The first factor alone would be an "equal output" estimate, so the second factor makes it clear how much the estimate is increased because of using a non-equal-output model. Smaller values of *p* make larger estimates. Esty (1996 and 2006) accepted the $p = 2$ model and recommended using (4) instead of the derived formula in because it is more robust, simpler, and by incorporating the number of singletons it usually adjusted the die-number estimate somewhat upward as seemed desirable to fit the number of singletons.[10]

In the geometric case $p = 1$ and (4) is algebraically equivalent to:

$$(5) \qquad e_{1C} = \frac{n(d + d_1)}{n - d_1}$$

[9] The derivation of (4) is much like the derivation below of (1). Fix *p*. Compute E(*C*) as a function of *n* and *e*. Use the observed value of *n* and replace E(*C*) by its estimate from (3). An equation results which can be solved for *e* (even when *p* is not 1 or 2). This is a "Method of Moments" estimator.

[10] The complicated derived formula, which corresponded to Carter's simplified method, is formula (3) in the 2006 article. By using d_1 as a component of the estimate of *D* rather than relying entirely on *n* and *d*, the number of singletons fit better using (4). This paper shows that using $p = 1$ is better than using $p = 2$.

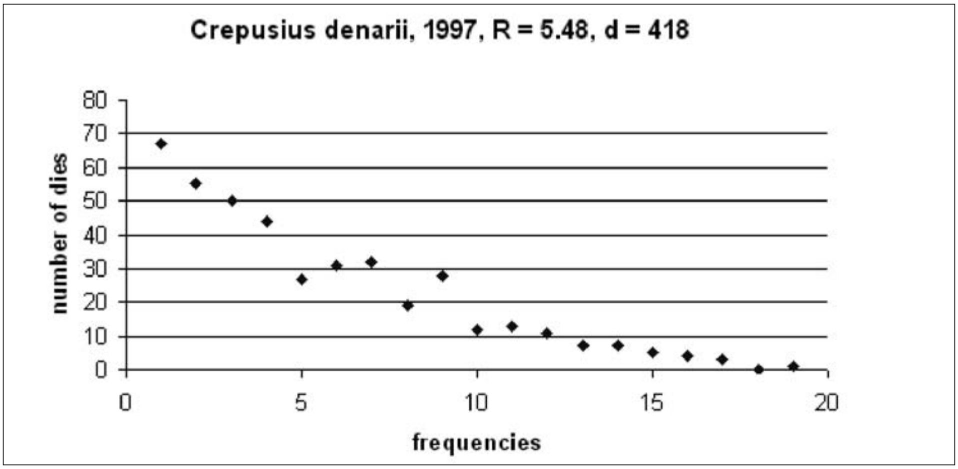

Fig. 8. - Crepusius reverse-die data from 1997 (n = 2290 and d = 418).

This estimator yields estimates that are similar to the estimates from formula (1).

Notice how the data plots resemble a theoretical geometric plot in Figure 2. There are hundreds of plots that look like this, but we do not know the true value of d_0 so we cannot simply check to see that the resulting estimator really works well. However, the newer Crepusius data can be used. Figure 8 has Crepusius data as of October 1997 for the dies with Roman numerals from 1 to 519, for graciously provided by Dr. Buttrey.

Again, the plot looks geometric. The actual fit of this data is discussed below.

The de Callataÿ Data

De Callataÿ summarized 329 sets of die-frequency data from Hellenistic (H) types and 279 from Archaic and Classical (A&C) types. This is a gold mine of data, without which this paper could not have been written. Unfortunately, we do not have *a-priori* knowledge of the true number of dies for any of them. However, it has more than enough data sets of sufficient size to address whether the negative-binomial model fits data well.

Using a model of output we can use n and d to estimate d_0, which is unobserved and what we want to know. However, we can also use it to estimate d_1, which is observed. If the actual values of d_1 do not resemble their estimated values, something is wrong. However, if they do resemble their estimated values, the model is supported.

First we must acknowledge that there are some problems with the data. Some of the recorded data is wrong because dies were identified incorrectly.[11] We ignore this, expecting it to be a minor issue. Worse is that die-studies use samples that are certainly not random.

Estimators might not give good estimates if the data is not random. Some sets of data seem to show non-random groups of coins which stayed together since they were minted. This causes n to be greater than it should be in a random sample and that causes estimates of the number of dies and singletons to be lower than they should be, making singletons look unexpectedly common in samples.

> Archaic and Classical Type 10 (drachm of Velia): $n = 157$, $d = 52$, $R = 3.02$
> Frequencies: 28, 12, 3, 0, 3, ..., 1 die with 16 coins and 1 die with 35
>
> Archaic and Classical Type 209 (drachm of Sinope): $n = 594$, $d = 163$, $R = 3.64$
> Frequencies: 110, 13, 10, ..., and dies with 20, 28, 60, and 97 coins (from a single hoard!)

We can be certain those 97 coins from one die were not assembled randomly. They were probably all almost uncirculated. That non-random group increases n without increasing d, making the sample seem more thorough than it is and the estimate lower than it should be (Type A&C10 almost certainly has this problem with the 35 coins from one die, and Hellenistic Type 134 mentioned above probably has the same problem with 50 coins from one die). That makes any estimate of d_1 based on n and d too low because n is too high. So, this type of non-randomness makes the observed number of singletons higher than the model would fit. A partial fix to this problem is discussed below.

Nevertheless, we can consider the numerous data sets and see if the estimator based on n and d alone yields an estimate of d_1 that resembles the observed value of d_1.

The values of n, d, d_1, d_2, and d_3 were entered into a spreadsheet for all 449 types with sample size $n \geq 30$. The other data sets have too few coins for statistical purposes. 117 Hellenistic data sets and 124 Archaic and Classical data sets had $n \geq 100$. For $p = 2$ and $p = 1$ (and many other values of p) the estimates of D were computed and the corresponding estimates of d_1 were computed and compared to the actual values of d_1. Some cases are extreme, for example when R is so large that all dies are seen multiple times and $d_1 = 0$ (Type A&C17, with 712 coins from 28 dies, which has $R = 25.4$ coins per observed die). On the other hand, some have R very low (Type A&C122, with $n = 108$ and $d = 93$ has $R = 1.16$ and $d_1 = 80$) where

[11] I have personally noted that it can be extremely difficult to determine die identities. The possibilities of die-deterioration, recut dies, and hubbing complicate the issue. Numerous authors have disputed the die-identifications of other authors. For a recent example, see Hoover 2008, 67.

we cannot expect much accuracy in the estimate. To avoid the difficulties at the extremes, the summaries for the 117 and 124 data sets are given with medians instead of means because medians are far less sensitive to erroneous and extreme results.

Also, another 41 Hellenistic and 49 Archaic and Classical data sets had between 50 and 99 coins from at least 10 different dies, and they were also analyzed, but separately because of the smaller sample size.

Checking the model using d_1

Given the negative binomial family of models and parameter p, we can derive an estimator of D based on n and d alone. It effectively estimates d_0 the number of unobserved dies. We can check the model by seeing if it does a good job of estimating d_1, which is observed and closely related to d_0. Consider the ratio of the actual number of singletons observed to the number of singletons predicted by the model. The ratio is:

$$\frac{d_1}{E(d_1|n,d,p)}$$

If all the assumptions hold, this ratio should tend to be near 1.0.

The ratio of the actual value of d_1 to its estimated value helps show whether the data fits the model. If this ratio is greater than 1, the model underestimates the number of singletons and therefore probably underestimates the number of unobserved dies. If this ratio is near 1, the model does a good job of estimating the number of singletons and therefore probably does a good job of estimating the number of unobserved dies.

Here is a summary of the analysis of the Greek coin data. Because it is possible for a model to be incorrect in general yet fortuitously fit fairly well for particular ranges of R, each group of data sets, Hellenistic or Archaic and Classical, was categorized as a whole and again by R value. The Archaic and Classical data sets had many types with very high R and $d_1 = 0$. When d_1 is 0, the expected value of d_1 is usually close to 0 and that category would not be recommended for statistical usage because numbers less than 5 are usually considered to be too small. But, those data sets have very important data. For example, if the expected number of singletons is 0.1 and yet singletons appear, that is strong evidence that something is wrong. We do not want to discard that evidence. To include as much evidence as possible, the data sets with very high values of R were included whenever their expected number of singletons was at least 1.0 (row 7 in the table).

	1	2	3	4	5	6	7	8
	Data Series	sample criteria	# data sets	median $R = n/d$	$p = 2$ median d_1/d_1 est*	$p = 1$ median d_1/d_1 est*	$p = 2$ normalized**	$p = 1$ normalized**
1	H	$n \geq 100$	117	4.20	1.31	1.053	1.12	0.26
2	A&C	$n \geq 100$	124	4.54	1.32	1.049	1.16	0.26
3	H	$50 \leq n < 100$ and $d \geq 10$	41	3.17	1.39	1.125	0.78	0.253
4	A&C	$50 \leq n < 100$ and $d \geq 10$	49	2.39	1.096	1.014	0.375	0.054
5	H	$n \geq 100$ and $R \geq 5$	43	6.80	1.79	0.98	1.30	-0.017
6	H	$n \geq 100$ and $R < 5$	74	3.21	1.217	1.058	1.07	0.31
7	A&C	$n \geq 100$ and $R \geq 10$ and $E(d_1) \geq 1$	9‡ 21†	14.4	2.69‡	0.91†	1.81‡	-0.09†
8	A&C	$n \geq 100$ and $5 \leq R < 10$.	22	6.95	2.14	1.24	1.73	0.509
9	A&C	$n \geq 100$ and $R < 5$.	62	2.98	1.28	1.106	1.33	0.598

Fig. 9. - H = Hellenistic. A = Archaic and Classical. Entries in the four columns on the right are medians for the given number of data sets.
* Entries give the ratio of d_1 to its estimated value under the model. If the model and estimation were perfect, this would be 1.00.
** Entries give $((d_1/(d_1 \text{ est}) - 1)\sqrt{(d_1 \text{ est})}$, which puts larger and smaller estimates on the same scale of variability. If the model and estimation were perfect, this would be 0.0.
‡ Based on the 9 sets with $E(d_1)$ at least 1.0 with the $p = 2$ model when R is very high.
† Based on the 21 sets with $E(d_1)$ at least 1.0 with the $p = 1$ model when R is very high.

In Figure 9, rows 1 through 4 use four disjoint sets of data. Rows 1 and 2 have samples sizes at least 100 and rows 3 and 4 have sample sizes from 50 through 99. Rows 5 and 6 subdivide the Hellenistic data in row 1 into two cases, high numbers of coins per observed die ($R \geq 5$) and lower number of coins per die. Rows 7 through 9 subdivide the Archaic and Classical data, only this time there are more sets of data with very high R value ($R \geq 10$) and they are considered separately because so many have no singletons and possibly no missing dies.

Columns 5 and 6 would ideally have entries near 1.0. The fact that column 5 deviates from 1.0 by at least 3 times as much as column 6 shows that the geometric model in column 6 fits the number of singletons much better.

Columns 7 and 8 use the same raw data, but normalized. Because larger samples should tend to produce more accurate results, deviations based on a larger sample should carry more weight than deviations based on smaller samples. The appropriate weighting factor is the square root of the expectation. Columns 7 and 8 include these weights and the positive entries show that the number of singletons tends to exceed the number predicted, but not by much in the geometric case and by 3 or more times as much in the $p = 2$ case. The geometric model fits well.

Crepusius Data

Rev.dies	n	d	d_1	d_2	d_3	True value?	p=2 estimate	p=1 estimate
1-100	551	85	12	6	11	100+1	90.5	100.5
101-200	551	83	15	8	8	100+1	88.2	97.7
201-300	430	81	12	14	7	100+1	88.5	99.8
301-400	362	78	10	19	6	100	87.2	99.4
401-519	396	91	20	8	17	119	103.1	118.2
All 519	2290	418	69	55	49	519+3	457.6	515.6

Fig. 10. - 1997 Crepusius data subdivided into groups of 100.

The first four rows consider groups of 100 Roman numerals. Rarely, the same numeral was on two different dies, inflating the true number in the group by one. The highest numeral is 519, so numerals 401-519 were put into one class. The bottom row includes all 519. The true values are dramatically underestimated by the $p = 2$ model and remarkably well estimated by the estimator (1) which results from the geometric model. Why does the geometric model work so well now with this augmented data when the 1983 data set supported the selection of $p = 2$? The answer is probably that much of the initial data came from museums and large collections that did not collect randomly, but sometimes sought new die numerals and avoided additional duplicates. It is easy to imagine a collector or a museum seeking a new die number and avoiding duplicates. The 1983 data had $n = 1082$ and $d = 384$. The new estimate would be 595, far too high, probably because n was less than it would have been in a random sample. If 375 more coins of the type had been seen, but not recorded or not acquired, the geometric estimate with the increased $n = 1082+375$ would be the desired value of 522 and the $p = 2$ estimate would be only 449. For the original data set, the choice of the $p = 2$ model, which tends to underestimate the original number of dies, worked to compensate for the inflation of the number caused by the extra singletons and missing duplicates. Now that more extensive data has been assembled, the sampling bias has been diminished.

The model, singletons and non-randomness

The $p = 2$ model regularly yields a ratio much higher than the expected 1.0, often 2 or 3, which proves that some assumption is wrong.

Theoretically, the error could be that: (1) the data are incorrectly recorded; (2) the samples are not random samples, or (3) the model is wrong. It seems unlikely that incorrect data could make this happen. Certainly the samples are not random, but we must deal with what we have, so non-randomness is not a pleasing excuse.

We already discussed one type of non-randomness and there is a second which influences the results in the opposite direction. Collections may avoid duplicates, but hoards may have non-random groups of extra duplicates which cause n to be inflated without inflating d, which decreases the estimate of d_1, which would make the observed ratio greater than it should be.[12] How much difference does this type of non-randomness make?

Consider the geometric model. In Figure 9, row 1, the median ratio is only about 1.053, quite close to 1.00, and when normalized using $(d_1/d_1 \text{ est})\sqrt{(d_1 \text{ est})}$ the median is 0.259, not too far from 0.00. This amount of deviation from the best possible result could be easily be explained by non-randomness inflating n.

For example, the median number of coins per die in the Hellenistic data is 4.2. R would be 4.2 in a hypothetical case in which $n = 420$ and $d = 100$. So, for each observed die, there would be on average 3.2 additional duplicate coins. If one in each ten of the 320 duplicate coins entered the sample in a non-random manner, that would be 32 extra duplicate coins. Then n should have been 420-32 = 388 instead of 420. If we recompute the estimates with $n = 388$ instead of 420, we find the estimate increases from 131.2 to 134.7, an increase of 2.7%. However, the estimate of d_1 increases from 23.8 to 25.8, an increase of 8.4%, more than enough to explain why the median ratio observed is 1.053, only 5.3% above the estimated value. It does not take very many coins entering a sample in groups to bias the estimates of d_1 (and d_0) so that they are 5 or 10 percent too low.

If the frequencies strongly suggest non-randomness, it would be appropriate to compute estimates with n decreased by the number of obvious non-random duplicates.

Mathematics for Statistics

Here is an outline of the derivation of the new estimator in (1). It is derived by the statistical method known as the method of moments. The original number of dies, e, is unknown, but we could compute the expected values of various observable statistics such as d and n/d if e were given. By setting the expected

[12] Lack of circulation suggests lack of mixing. If there are a large number of nearly uncirculated die duplicates, it is likely they did not enter the sample randomly. The sample size n should be decreased to compensate for this.

values of these two statistics equal to their observed values, we can solve for the unknown e. The solution is then the method of moments estimator for e.

The geometric distribution has a parameter, r, which we do not need to know *a priori*. The probability of observing exactly k is $P(X = k) = r(1 - r)^k$. The probability a given die is observed is one minus the probability it is not observed: $1 - P(X = 0) = 1 - r$. Then the expected number of dies observed is $e(1 - r)$ and the actual number of dies observed is d. Setting these two equal yields one equation, $d = e(1 - r)$. The expected number of coins per original die is $(1 - r)/r$. The expected number of coins per observed die is $1/r$. The actual number of coins per observed die is n/d. Setting these two equal yields a second equation, $n/d = 1/r$. We can solve the two equations for e. The result is formula (1).

Conclusion

The evidence of hundreds of die-count data sets and the augmented 1997 Crepusius data is overwhelming that the negative-binomial model with $p = 2$ is incorrect, and that the geometric model comes reasonably close to fitting the data. Plots 3, 4, 5 and 8 above are typical of many large data sets and strongly resemble theoretical geometric plots in Figure 2. Therefore, the geometric point estimator for the original number of dies, $nd/(n - d)$, is likely to be a good estimator. A formula for approximate confidence intervals, which are often wide, was given by Esty.[13]

The estimator assumes the sample is random, or at least not biased. Non-random groups of coins from the same die inflate n without a corresponding increase in d, so the estimates tend to underestimate the original number of dies if the sample includes non-random groups of coins. Such groups can sometimes be identified by noting die-frequencies which are clearly outliers. In those cases, it would be appropriate to do the calculations with n decreased by the apparent number of non-random duplicates.

On the other hand, samples might be non-random in the other direction because they have coins from collections that selected for different identities. This type of selection inflates d more than random samples would, and therefore estimates from such samples tend to overestimate the true number of dies. When researchers assemble samples sequentially they should include additional duplicates as they appear, even if several or many have already been recorded, or else the sample and estimates will have a significant bias.

The revised 1997 Crepusius data also yields strong evidence that estimates using (1) from the geometric model work well. Die-identity researchers should

[13] Esty 2006, 360, formula (4).

include all die-frequencies, not just n and d, so readers can judge the fit of the model. Plots like Figures 3, 4, 5 and 7 are as useful as the estimate itself and provide additional information about the fit of the data to the model. If the number of dies is large, we can expect a plot somewhat like the geometric plot in Figure 2 for which the estimator in formula (1) is appropriate.

Bibliography

G.F. Carter 1983, *A simplified method for calculating the original number of dies from die-link statistics*, ANSMN, 28, 195-206.

Fr. de Callataÿ 1993, *L'estimation du nombre de coins : en augmentant l'échantillon…*, Acta Numismatica, 21-23 (*Homenatage al Dr. Leandre Villaronga*), 31-48.

Fr. de Callataÿ 1997, *Recueil Quantitatif des Émissions Monétaires Hellénistiques*, Wetteren.

Fr. de Callataÿ 2003, *Recueil Quantitatif des Émissions Monétaires Archaïques et Classiques*, Wetteren.

W.W. Esty 1986, *The efficiency of Good's nonparametric coverage estimator*, The Annals of Statistics, 14, 1257-1260.

W.W. Esty 2006, *How to estimate the original number of dies and the coverage of a sample*, NC, 166, 359-364.

W.W. Esty and G. F. Carter 1992, *The distribution of the numbers of coins struck by dies*, AJN, sec. ser. 3-4 (1991-2), 165-186.

O.H. Hoover 2008, *Review of Silver Coinage of the Artaxiad Dynasty of Armenia by Y. T. Nercessian*, ANS Bulletin, 7 (1), 67-72.