Business Management Dynamics
Vol.5, No.3, Sep 2015, pp.12-17

Account Screening Based Upon Digital Frequency Profiling in the Internal Audit Context: A Cartesian Distance Likelihood Triaging Protocol

Edward J. Lusk, PhD[1] and Michael Halperin, PhD[2]

Abstract
A central feature in the montage of executing the internal audit, in a non-forensic context, is a random sample of sufficient size to create the evidence needed to make the decision if extended investigative procedures are warranted or not. This is critical not only in maintaining a best practices profile for the department of internal auditing but also in "partnering" with the external auditors by assuring, insofar as possible, that the work of the internal audit group can be accepted as audit evidence by the external auditors and so conserving scarce organizational resources. In fact, this partnering is one of the long term goals of the PCAOB for controlling the cost of the external certification audit. However, a contentious issue in sampling is: How should the internal auditor ferret out those accounts that are likely candidates for discovery sampling? This is the point of departure of our paper where we present a simple and validated protocol based upon the digital frequency paradigm introduced by Newcomb & Benford and popularized in the audit context by Nigrini to identify accounts under audit that seem reasonable candidates for extended procedures discovery sampling testing.

Key words: Signaling Errors, Median-Conformity Triage

Available online
www.bmdynamics.com
ISSN: 2047-7031

INTRODUCTION: Identification of Account Candidates for Extended Procedures Testing through Screening

The Institute of Internal Auditors [IIA: http://www.theiia.org/guidance/standards-and-guidance/ippf/definition-of-internal-auditing/?search%C2%BCdefinition (15 Dec 2014)] defines Internal Auditing as:

An independent, objective assurance and consulting activity designed to add value and improve an organization's operations. It helps an organization accomplish its objectives by bringing a systematic, disciplined approach to evaluate and improve the effectiveness of risk management, control, and governance processes.

In this context, the internal auditor has as the principal charge: *To provide oversight of and assurance relative to the integrity and the functioning of the Accounting Information System [AIS]*. The extended implication of this charge is two-fold: (i) the internal auditor is responsible for testing the AIS to determine if management's system of Internal Control over Financial Reporting is *adequate,* and (ii) such adequacy implies that the information in the four financial statements offered by management *is a reasonable reflection of the period results of operations*. Also one recognizes these two aspects as the charge to the external auditor. This is why the external auditor could, in certain circumstances, rely on tests conducted by the internal auditor as evidence in executing the certification audit. Such reliance makes possible a beneficial cost swap—to wit, the cost of the testing effected by the internal audit group will be demonstrably lower than that of similar testing done by the external auditor as a function of the differential of the salaries of the internal audit staff compared to the hourly billing of the audit LLP. In this extended context, the internal auditor can be a *tacit-agent* of the external auditor. This agency is noted by the Public Company Accounting Oversight Board [***http://www.sec.gov/rules/pcaob/2007/34-55876.pdf*** SEC (Release No. 34-55876; File No. PCAOB-2007-02) p.147]:

As discussed in the proposing release, the Board had several objectives in proposing this standard. The first was to better integrate the financial statement audit and the audit of internal control by having only one framework for using the work of others in both audits. Additionally, the Board wanted to encourage auditors to use the work of

[1] *CPA Professor, Department of Accounting, School of Business and Economics, State University of New York, Plattsburgh, USA & Emeritus: The Wharton School, Department of Statistics: University of Pennsylvania, Philadelphia; PA USA*
E-mail: luskej@plattsburgh.edu
[2] *Former Director: Lippincott-Wharton Library: The University of Pennsylvania, Philadelphia; PA, USA*
E-mail: halperin@wharton.upenn.edu

others to a greater extent when the work is performed by sufficiently competent and objective persons. Among other things, under the proposed standard, auditors would have been able to use the work of sufficiently competent and objective company personnel – not just internal auditors –and third parties working under the direction of management or the audit committee for purposes of the financial statement audit as well as the audit of internal control.

Therefore, the Policies, Procedures and Protocols that underlie internal audit testing are critically important factors in the decision made by the external auditor to rely on the internal audit test results as direct evidential support of the two opinions required by the PCAOB: (i) The adequacy of *Internal Control over Financial Reporting* and the assurance that the *Financial Statements are Reasonably Free from Material Error*. This is the point of departure of our paper. Following we wish to offer an Account Screening Protocol that is formed on the, *en vogue*, Digital Frequency Profile [DFP] introduced in the audit context by Nigrini (1996). The nature of the screening protocol, and therefore the focus of this paper, is:

*The development of a DFP-**Screening Protocol** to facilitate the identification of a set of **Accounts** that may be argued as reasonable candidates for extended procedures discovery sample testing in executing the charge of the internal auditor in a manner that would create audit evidence for the external auditor.*

ACCOUNT SCREENING AND DISCOVERY SAMPLING USING DIGITAL FREQUENCY PROFILES

Screening protocols are often employed in tandem with Discovery Sampling Models. Simply put:

Screening Protocols look for "anomalous profiles" in the account under audit scrutiny that suggest that there is reason to effect an extended procedures examination using Discovery Sampling Models. In this mode: Screening is a Pre-curser to Discovery Sampling.

The specific measure that we will use in developing the screening information is the Digital Frequency Profile. Consider this measure next.

The historically available information on the Digit Frequency Profile [DFP] starts in 1881 when, Simon Newcomb (1835-1909) notices a curious pattern of *Wear & Tear* of his logarithmic tables—his Decision Support System of the 19th century. He offers (1881, p.39):

That the ten digits do not occur with equal frequency must be evident to any one making much use of logarithmic tables, and noticing how much faster the first pages wear out than the last ones. The first significant figure is oftener 1than any other digit, and the frequency diminishes up to 9.

Some fifty years later Frank Benford (1938), an electrical engineer with *General Electric Inc.* with many patents to his credit, makes and records the same observation.

Newcomb and Benford both arrived at a simple formula to characterize the likely distribution of the nine first digits. To wit the [N-B Profile]:

EQ1 $\quad$ Frequency$[d_i]$ = $\mathrm{LOG}_{10}\,(1 + 1/d_i)$ for i= 1, 2, - - -, 9

Important questions which is begged by the "non-intuitive" observations of Newcomb and Benford is:

Why should EQ1 work as a general DFP-estimator of generating processes the output of which is subsequently measured, and further under what conditions can an auditor reasonably expect to use EQ1 as a profiling-screen for the use of extended procedures?

The first research to address the theoretic underpinning of the Log_{10} formula, EQ1, as a reasonable and appropriate surrogate for data generating processes only starts to appear some 25 years after Benford's paper. The ground breaking work is offered by Pinkham (1961), Adhikari and Sarkar (1968), Duncan (1969), and Raimi (1969). However, Hill (1995a, b, 1996, 1998) is usually credited with providing the conclusive theoretical support for the *Why*, *How*, and *When* questions posed above.

THE MEASURE: THE DISTANCE DIFFERENTIAL IN CARTESIAN SPACE

We will be using the non-directional Cartesian distance measure that has been suggested by Cho and Gaines (2007) and also detailed by Reddy and Sebastin (2012) both of which are highly recommended readings that cover the expansive nature of various metrics formed around Digital Frequency Profiling. We will use the simplest measure which is the non-directional difference in Cartesian Coordinates line-space. Additionally, in the service of robustness, we will be using a measure formed from two independent line-scoring measures: The Mean Absolute Difference in Distance [MAD] and the Median Absolute Difference in Distance [MdAD].

Computationally, we will be forming the MAD and MdAD as:

EQ2 $$MAD = [\sum_1^9 ABS[BPP_i - O_i]]/9$$

EQ3 $$MdAD = ABS[BPP_5 - O_5]$$

Where: BPP_i is the Benford Practical Profile developed by Lusk & Halperin (2014) at the i[th] frequency, O_i is the Observed frequency, and, ABS is the absolute value operator.

The final distance measure that we will use for a particular dataset will be the simple average of the MAD and MdAD, or:

EQ4 $$d = [MAD + MdAD]/2$$

EQ4 gives the measure of the non-directional difference relative to the first digit difference frequencies in Cartesian distance for the particular dataset under audit examination. Using this metric, consider now the overview of the triage testing that we will use to make the decision if the dataset under audit is likely to be *Conforming* in nature or is likely to be *Non-Conforming* in nature. This is the crucial decision in the following way:

If the dataset is likely to be Conforming then extended procedures are not likely to be warranted; if on the other hand, the dataset is likely to be Non-Conforming then extended procedures are likely to be warranted.

All of these elements of the distance-signaling functionality are programed in an Excel™-VBA® Decision Support System called Distance:DSS. It is available from the authors as a download at no cost and with no restriction on it use.

LIKELIHOOD TRIAGING: JUDGING THE LIKELY CONFORMITY FOR AN ACCOUNT UNDER AUDIT EXAMINATION

What underlies our developmental validity of the screening DFP distance model, is the misclassification jeopardy of any signaling system. This is the relationship between the *True* state of nature and the signal created by the screening model. There are only four case contingencies. In our context: IF the ***True*** state of nature is *Non-Conforming* and the Screening model signals: *Non-Conforming* OR IF the ***True*** state of nature is *Conforming* and the Screening models signals: *Conforming*, then these are the correct signaling assessments. In like manner, the signal can be wrong in the two cases remaining—these are the misclassifications. We will use these four classification cases to develop and also to evaluate the signaling system to be proposed. Consider now the screening model.

a The elements of the Functionality of the DFT Likelihood Protocol as programmed in the Distance DSS

The classification measurement montage uses:

1. ***Conforming* and *Non-Conforming* Datasets** We have collected from the literature 56 datasets: 33 of which have been argued as *Conforming* and 23 that have been offered as *Non-Conforming*. These State of Nature datasets are all included in the Distance:DSS.
2. **An Inferential Metric** We will use the value of *d* as scripted in EQ4.
3. **The Triaging Partition** The screening metric used to triage a particular audit dataset will use the Median value of *d* over all [∀] 56 datasets to form the triage model. We note the value of this median triage point as ***d****. Simply: It is certainly expected that:
$$\{d_\forall^c\} < \sim d^* \sim < \{d_\forall^{nc}\}, \text{ or}$$
that the collection of *d*-values of *Conforming* datasets: $\{d_\forall^c\}$ will be mostly less [< ~] than ***d**** and that the collection of *d*-values of *Non-Conforming* datasets: $\{d_\forall^{nc}\}$ will be mostly greater than [> ~] than ***d****. If this logical expectation is realized, then ***d****, over all the State of Nature datasets in the audit context, should be an effective triage point. We will, of course, evaluate this in the four classifications cases as a validation of the triage cut-point.

b Testing the Functionality Profile of d*

Using the elements of this screening montage, we calculated the Median of *d*, for the two collected components and also for the overall data sample. These profiles of *d* are presented in Table A.

Table A Distance Metric Profile for the *Non-Conforming* and the *Conforming* Datasets

Statistical Profile of the value of *d*	***Non-Conforming* Datasets, n = 23**	***Conforming* Datasets, n = 33**	**Both Datasets, n =56**
Mean/Median	0.0292/0.0210	0.0127/0.0082	0.0195/**0.0170**+
Range/StDev	0.1278/0.0269	0.0431/0.0093	0.1401/0.0201
95% CI	[0.0176 – 0.0409]	[0.0094 – 0.0160]	[0.0141-0.0249]

+***0.0169996278*** is the Median value of *d* value to 10 decimal units. This is the specific cut-point value that we recommend using as the triage partition. In the application of the triage we used ***d*= 0.0169996278***
To be clear, the expectation is that the *Conforming* Data *d*-profile will have a lower Median distance between the BPP profile and that of the *Conforming* sample compared to the Median distance between the BPP profile and that of the *Non-Conforming* sample. These results are clearly realized as presented in Table A as:

$$\{d_{\forall}^{c}: 0.0082\} < \sim d^{*} = 0.0170 \sim < \{d_{\forall}^{nc}: 0.0210\}$$

However, to inferentially validate the separation, as we will use these two datasets to form the ***d****-likelihood partition, we conducted the usual inferential tests. Using the parametric [*Welsh Adjusted t-test version*] and non-parametric [*Wilcoxon: Kruskal-Wallis Rank Sum Test*], the two tailed p-values were both < 0.0001 suggesting a strong rejection of the Null that there are likely no *d*-differences in the Mean or Median between the *Non-Conforming* and *Conforming* datasets.

c The Likelihood Triaged Partition

Given the evident and tested separation between the two reference datasets, we decided to use the most conservative triage point of separation—that is the Median of the overall dataset, n = 56 or in this case: d*=0.0169996278. Therefore, the exhaustive Likelihood triage point will be:

For a particular audit dataset under examination,

If the value of d is <= to d, then the likely set membership of this Audit Dataset is the Conforming Dataset and therefore:* ***Extended Procedures may not be warranted****.*

If the value of d is > to d, then the likely set membership of this Audit Dataset is the Non-Conforming Dataset and therefore:* ***Extended Procedures may be warranted****.*

In this case, as is clear from Table A the d-value of the *Non-Conforming* data is larger than the d-value of the *Conforming* dataset by a factor of approximately 10. The test of this likelihood cut-off of *d**is, of course, how it performs in the re-classification of the 56 datasets. In this regard, we will use the Distance:DSS to classify the 56 series.
As an illustration of this triage partition consider the Nigrini (1996) and the Hill (1998) datasets that are in the sample of 56 datasets that we collected. The Hill dataset is taken from an experiment where he asked 743 students to suggest a six-digit random number. Hill (1998, p.363 Figure 5) offers this dataset as a clear illustration of *Non-Conforming* data. Nigrini (1996) collected income tax information which is argued by Hill (1998, p.363: Figure 5) as *Conforming*. Hill indicates: "- - - the first significant digits of true tax data taken by Mark Nigrini from the lines of 169,622 IRS model files follow Benford's Law closely." The *d*-measures of this data are presented in Table F.

Table B Hill *Non-Conforming* and Nigrini *Conforming* Dataset profiled against d*

Digits	**BPP Expectation**	**Hill Actual**	**Nigrini Actual**	**Hill ABS differences**	**Nigrini ABS differences**
1	0.289189	0.147	0.3059	0.142189	0.016711
2	0.194770	0.100	0.1779	0.094770	0.016870
3	0.126650	0.104	0.1270	0.022650	0.000350
4	0.090612	0.133	0.0948	0.042388	0.004188
5	0.075436	0.097	0.0778	0.021564	0.002364
6	0.064314	0.157	0.0650	0.092686	0.000686
7	0.054081	0.120	0.0563	0.065919	0.002219

8	0.054872	0.084	0.0503	0.029128	0.004572
9	0.050522	0.058	0.0450	0.007478	0.005522

Using EQ2, EQ3 & EQ4 applied to the data in Table B, produces the following results:

Hill *Non-Conforming*: Dataset: MAD = 0.057641, MdAD = 0.042388 and therefore d= 0.050015

Nigrini *Conforming* Dataset: MAD = 0.005942, MdAD = 0.004188 and therefore d = 0.005065

The results in Table B are the anecdotal demonstration of the rational for a Likelihood partition of d*. As expected,

$$\{0.005065\} < \sim 0.0170 \sim < \{0.050015\}$$

In this case, the d-value of the *Non-Conforming* data is larger than the d-value of the *Conforming* dataset by a factor of 10. The test of this likelihood cut-off of *d**is, of course, how it performs in the re-classification of the 56 datasets. In this regard, we will use the Distance:DSS to classify the 56 series. Using the usual [2x2] Classification/Misclassification matrix we will examine the functional profile of the d*-Triage Protocol.

EVALUATION OF THE FUNCTIONALITY OF THE TRIAGING PROTOCOL

FPIE and FNIE Evaluation

The next issue that we wish to discuss is the performance of the *d**-likelihood cut-off in terms of the False Positive Investigation Error [FPIE] and the False Negative Investigation Error [FNIE]. For the FPIE, investigating when it is NOT likely warranted, we tested the *d**-classification protocol using *Back-Casting*. Using the Back-Cast evaluation for the dataset used in the creation of the information in Table A, for the *Conforming* Data, in 10 of the 33 cases the *d-computed* was > than *d** suggesting *Non-Conformity* which incorrectly suggests that extended procedures are warranted. In this case, this is incorrect as the dataset was *a priori* judged to be *Conforming* and so the internal auditor commits a FPIE 30.3% [10/33] of the time. As for the FNIE—failing to investigate when warranted—we ran the 23 *Non-Conforming* datasets using the Distance:DSS and five times of the 23 the d-computed was <= *d** indicating that the dataset is likely to be *Conforming* and so extended procedures are not warranted. In this case, this is a FNIE—i.e., failing to investigate when it may be warranted which occurs 21.7% [5/23] of the time. The statistical profile of classification matrix test is that the Likelihood Ratio is 13.1, with a p-value = 0.0003. Also, using the Fisher's Exact test for a conservative test of the classification, i.e., the 2-tailed version, the p-value is = 0.0009. Both indications provide strong support for (1) rejecting the Null that the classifications made by the DFT Likelihood *d**-Protocol as programmed in the Distance:DSS are random and that this particular realization is the exception, and (2) provides support for the efficacy of the d* triage as the FPIE and the FNIE are certainly within reasonable bounds for the internal audit investigation of accounts that may be in need of extended procedures.

Holdback Validation

As a validation test of the DFT Likelihood Protocol effectiveness, we collected a 25% holdback sample: Seven datasets that were offered in the literature or as judged as part of a certified audit as *Conforming* and seven that were reported as *Non-Conforming*. We then used the Distance:DSS to classify these Hold-Back datasets. The classification results are: The FPIE was 14.3% and the FNIE was 0%. In this case, the Likelihood Ratio is 6.7 with a p-value of 0.0003; the directional Fisher's p-value is 0.0047. This strongly supports the classification effectiveness of the DFT Likelihood Protocol and is consistent with the Back-Cast results.

SUMMARY AND OUTLOOK

Summary

Using the DFT *d*-*Likelihood Protocol the internal auditor will decide upon the use of extended procedures of discovery sampling using the following DSS messaging system:

If the d-computed is <= *d** then the Distance:DSS produces a Cell Message: ***This dataset is Likely to be a CONFORMING Dataset and so Extended Procedures Are NOT Indicated***.

If the d-computed is > *d** then the Distance:DSS displays the Cell message: ***This dataset is Likely to be a NON-CONFORMING Dataset and so Extended Procedures Are Indicated***.

This information may then be used by the internal auditor to make the decision regarding the use of extended procedures; and further this information can be presented to the external auditor as testing information for both the adequacy of Internal Control over Financial Reporting as well as the test of the reasonability of the information provided in the Financial Statements.

Outlook

It is clear that the various parameters of the *d*-Likelihood triage are dependent on the two dataset profiles; *Conforming* and *Non-Conforming*. To enrich the relevance and utility of the Distance:DSS we would hope that *Conforming* and *Non-Conforming* datasets would be made available so as to up-date the Likelihood-Triaging protocol. We would be happy to be a posting-source for such information that we will post on our "Commons" site and so be available for all who wish to use this information.

REFERENCES

Adhikari, A.K., Sarkar, B.P. (1968). "Distributions of Most Significant Digit in Certain Functions Whose Arguments Are Random Variables", *Sankhya-The Indian Journal of Statistics Series B*, Vol. 30, pp.47–58.

Benford, F. (1938). "The Law of Anomalous Numbers", *Proceedings of the American Philosophical Society*, Vol. 78, pp.551-572.

Cho, W.K.T., Gaines, B.J. (2007). "Breaking the (Benford) Law: Statistical Fraud Detection in Campaign Finance", *American Statistician*, Vol. 61, pp.218-223. http://dx.doi.org/10.1198/000313007X223496

Duncan, R. (1969). "Note on the Initial Digit", *Fibonacci Quarterly*, Vol. 7, pp.474–475.

Hill, T. (1995a). "The Significant-digit Phenomenon", *American Mathematical Monthly,* Vol. 102, pp.322–327. http://dx.doi.org/10.2307/2974952

Hill, T. (1995b). "Base-invariance Implies Benford's Law", *Proceedings of the American Mathematical Society*, Vol. 123, pp.887-895. http://dx.doi.org/10.1090/S0002-9939-1995-1233974-8

Hill, T. (1996**)**. "A Statistical Derivation of the Significant-digit Law", *Statistical Science*, Vol. 10, pp.354-363.

Hill, T. (1998). "The First Digit Phenomenon: A Century-old Observation About an Unexpected Pattern in Many Numerical Tables Applies to the Stock Market, Census Statistics and Accounting Data", *American Scientist*, Vol. 86, pp.358-363. http://dx.doi.org/10.1511/1998.4.358T. P.

Lusk, E.J., Halperin, M. (2014). "Using the Benford Datasets and the Reddy & Sebastin Results to Form an Audit Alert Screening Heuristic: A Note", *IUP Journal of Accounting Research and Audit Practices*, Vol. 8, pp.56-69.

Newcomb, S. (1881). "Note on the Frequency of Use of the Different Digits in Natural Numbers", *American Journal of Mathematics*, Vol. 4, pp.39-40.

Nigrini, M. (1996). "A Taxpayer Compliance Application of Benford's Law", *Journal of American Taxation Association,* Vol. 18, pp.72–91.

Pinkham R. (1961). "On the Distribution of First Significant Digits", *Annals of Mathematical Statistics,* Vol. *32*, pp.1223 -1230. http://dx.doi.org/10.1214/aoms/1177704862

Raimi, R. (1969). "The Peculiar Distribution of First Digits", *Scientific American,* Vol. 221, pp.109–120. http://dx.doi.org/10.1038/scientificamerican1269-109

Reddy Y.V., Sebastin, A. (2012). "Entropic Analysis in Financial Forensics", *The IUP Journal of Accounting Research and Audit Practices*, Vol. 11, pp.42-57.