Modular Arithmetic in the Marketplace

JOSEPH A. GALLIAN AND STEVEN WINTERS
Department of Mathematics and Statistics, University of Minnesota, Duluth, MN 55812

1. Introduction. It is not surprising that there are many schemes that utilize modular arithmetic to append a check digit to product identification numbers for error detection. Two of the better known schemes—the ZIP code bar code and the International Standard Book Number (ISBN)—have been the subjects of articles in the *UMAP Journal* [2], [7]. The schemes used for the Universal Product Code (UPC) and passports are described in [6] and [1].

What is surprising to us is the diversity of the methods in use and the fact that some of them are poorly conceived! In this note we examine several of these schemes.

2. Check digits: from money orders to library books. We begin with the least effective of the methods we have found and work our way up to the best one. The Postal Service's money-order identification number consists of ten digits and a check digit. The check digit is the remainder modulo 9 of the 10-digit number. In contrast to the schemes mentioned in the introduction, this method does not detect all single errors! Indeed, excluding the check digit, a substitution of a 9 for a 0 or vice versa goes undetected. All single-digit errors involving the check digit are detectable. Thus the single-digit error-detection rate for this method is 970/990 or 98.0%. (We assume all errors are equally likely.) Moreover, the only transposition errors detected by this method are those involving the check digit. That is, an error resulting from the transposition of two consecutive digits such as $\dots 53 \dots$ instead of $\dots 35 \dots$ is undetected while $\dots 53$ instead of $\dots 35$ is detected. Because the digit 9 can never occur as a check digit, this method detects $9 \cdot 10^9 - 8$ of a total of $90 \cdot 10^9 - 8$ possible transposition errors for a rate of 10.0%.

A similar and equally ineffective method is employed on VISA traveler's checks. There the check digit is the additive inverse modulo 9 of the remainder upon division by 9.

Federal Express, airline companies, and the United Parcel Service use a method that is slightly less effective for detecting single errors but fairly effective for detecting transposition errors in their identification numbers. The U.P.S. identification number, for instance, consists of nine digits plus a check digit. The check digit is the remainder modulo 7 of the 9-digit number. Of course, any substitution of b for a in the first nine digits where $|a - b| = 7$ will go undetected. The single-error detection rate for this method is 846/900 or 94.0%. This method detects transposition errors at the rate of $(762 \cdot 10^7 - 5)/(810 \cdot 10^7 - 5)$ or 94.1%.

The Chemical Abstract Service assigns chemicals a registry number together with a check digit calculated in the following way. The number $a_1 a_2 \cdots a_k$ ($k \leqslant 7$) has appended the check digit $(a_1, a_2, \dots, a_k) \cdot (k, k-1, \dots, 2, 1) \bmod 10$. All single errors in positions with weighting factors 1, 3, or 7 as well as the check digit position are detected; errors of the form $a \to b$ where $|a - b| = 5$ go undetected in positions with weighting factors 2, 4, or 6; errors of the form $a \to b$ where $|a - b|$ is even go undetected in the position with a weighting factor of 5. For $k = 7$, this yields a single-error detection rate of 65/72 or 90.3%. On the other hand, *all* transposition errors not involving the check digit are detected. The errors of the type $\ldots ab \to \ldots ba$ go undetected when $(a_1, a_2, a_3, a_4, a_5, a_6) \cdot (7,6,5,4,3,2) \equiv 5 \bmod 10$. Since the probability that this dot product is any particular digit is .1, this yields a transposition error detection rate of 62/63 or 98.4%.

Identification numbers for banks (as appearing on checks, for instance) have eight digits, $a_1 a_2 \cdots a_8$, and a check digit

$$c = (a_1, a_2, \ldots, a_8) \cdot (7,3,9,7,3,9,7,3) \bmod 10.$$

For example, the bank number 09190204 gives $c = 0 + 27 + 9 + 63 + 0 + 18 + 0 + 12 = 129 \equiv 9 \bmod 10$. This method detects 100% of all single errors and 80/90 or 88.9% of all transposition errors. In particular, the only undetectable transpositions are those of the form $ab \to ba$ where $|a - b| = 5$. The advantage this weighting scheme has over one involving just two distinct weighting factors such as $(7,3,7,3,7,3,7,3)$ is that the former will detect most errors of the form $\ldots abc \ldots \to \ldots cba \ldots$ while the latter will detect no errors of this form except those involving the check digit.

The most sophisticated method we have found in use is the "Code-a-bar" system used by many libraries. Here each thirteen digit identification number $a_1 a_2 \cdots a_{13}$ is assigned the check digit

$$-(a_1, a_2, \ldots, a_{13}) \cdot (2,1,2,1,2,1,2,1,2,1,2,1,2) - r \bmod 10,$$

where r is the number of the digits among $a_1, a_3, a_5, a_7, a_9, a_{11}, a_{13}$ greater than or equal to 5. For example, the identification number 3125600196431 yields the check digit

$$-(6 + 1 + 4 + 5 + 12 + 0 + 0 + 1 + 18 + 6 + 8 + 3 + 2) - 2 = -68 \equiv 2 \bmod 10.$$

This method detects all single errors and all transposition errors except $09 \leftrightarrow 90$. Thus the detection rate for transposition errors is 88/90 or 97.8%. As shown by Gumm [5], it is not possible to improve upon these rates with any system that uses conventional techniques based on addition modulo 10.

3. A foolproof method. The highly effective scheme used by libraries raises the interesting question of whether it is possible to devise a method that will detect all single errors and all transposition errors with a single check digit. Actually, the ISBN method achieves this [7]. But it does so in an artificial way by using the character X to represent the possible check number of 10, which consists of *two* characters. (The method involves modulo 11 arithmetic.)

Recently Gumm [3], [4] has discovered a group theoretical method that uses a single check digit and is 100% effective in detecting single errors and transposition erros. To describe this method we need the dihedral group of order 10, D_5, represented in the form shown in TABLE 1, and the permutation $\sigma = (0)(14)(23)(58697)$. To append a check digit to any string of digits we "weight" the digits with powers of σ and, using TABLE 1, multiply them and take the inverse of the product. For example, consider 1793. The check digit is

$$(\sigma^4(1) * \sigma^3(7) * \sigma^2(9) * \sigma(3))^{-1} = (1 * 6 * 5 * 2)^{-1} = 4^{-1} = 1.$$

For 17326, we obtain the check digit

$$\left(\sigma^5(1)*\sigma^4(7)*\sigma^3(3)*\sigma^2(2)*\sigma(6)\right)^{-1}=(4*9*2*2*9)^{-1}=0^{-1}=0.$$

To see that this scheme detects all single-digit errors we observe that an error-free number $a_n a_{n-1} \cdots a_1 a_0$ (where a_0 is the check digit) has the property that

$$\sigma^n(a_n)*\sigma^{n-1}(a_{n-1})*\cdots*\sigma(a_1)*a_0=0$$

and, therefore, any particular factor in this product is uniquely determined by all of the others. Thus a single-digit error does not result in a product of 0. That all transposition errors are detected can be verified by showing that for all distinct a and b, $\sigma(a)*b \neq \sigma(b)*a$. For then, for all i,

$$\sigma^{i+1}(a)*\sigma^i(b) \neq \sigma^{i+1}(b)*\sigma^i(a)$$

and consequently a transposition will not result in a product of 0.

In addition to being foolproof in detecting single errors and transpositions, Gumm's method will detect approximately 90% of all other types of errors.

TABLE 1. The multiplication table of D_5.

*	0	1	2	3	4	5	6	7	8	9
0	0	1	2	3	4	5	6	7	8	9
1	1	2	3	4	0	6	7	8	9	5
2	2	3	4	0	1	7	8	9	5	6
3	3	4	0	1	2	8	9	5	6	7
4	4	0	1	2	3	9	5	6	7	8
5	5	9	8	7	6	0	4	3	2	1
6	6	5	9	8	7	1	0	4	3	2
7	7	6	5	9	8	2	1	0	4	3
8	8	7	6	5	9	3	2	1	0	4
9	9	8	7	6	5	4	3	2	1	0

For the purpose of comparison, we summarize the calculations given above in TABLE 2 and average the two rates. However, this average does not represent the detection rate for either one of a single-digit error or a transposition error unless these two types of errors are equally likely to occur.

TABLE 2. Summary of error-detection rates.

scheme	single error rate	transposition error rate	average
U.S. Postal Service, VISA	98.0%	10.0%	54.0%
Airlines, U.P.S.	94.0%	94.1%	94.1%
Chemical	90.3%	98.6%	94.5%
Bank	100%	88.9%	94.5%
Library	100%	97.8%	98.9%
Gumm	100%	100%	100%

Acknowledgment. The authors are indebted to David Witte and Douglas Jungreis for their assistance. Joseph Gallian was supported by the NSF (Grant Number DMS 8407498). Steven Winters was supported by a University of Minnesota, Duluth, College of Science and Engineering Undergraduate Honors Research stipend.

REFERENCES

1. Steve Connor, The invisible border guard, *New Scientist* (January 5, 1984) 9–14.
2. Joseph A. Gallian, The ZIP code bar code, *The UMAP Journal*, 7(1986) 191–195.

3. H. Peter Gumm, A new class of check-digit methods for arbitrary number systems, *IEEE Transactions on Information Theory*, 31(1985) 102–105.
4. H. Peter Gumm, Data security through check digits, *Statistical Software Newsletter*, 11(1985) 124–127.
5. H. Peter Gumm, Encoding of numbers to detect typing errors, *International Journal of Applied Engineering Education*, 2 (1986) 61-65.
6. Ian D. Rae, Machine readable codes, *New Zealand Mathematics Magazine*, 21(1984) 109–113.
7. Philip M. Tuchinsky, International Standard Book Numbers, *The UMAP Journal*, 5(1985) 41–54.

On a Property of $x^n e^{-x}$

GABRIEL KLAMBAUER
Department of Mathematics, University of Ottawa, Ottawa, Ontario, Canada, K1N 6N5

Let $f(x) = x^n e^{-x}$, where $x \geqslant 0$ and $n \geqslant 2$. The function f is increasing in the interval $(0, n)$ and decreasing in (n, ∞) and has points of inflection for $x = n \pm \sqrt{n}$. My colleague Ian Iscoe of the University of Ottawa asked me to find a simple proof that $f(n + \sqrt{n}) > f(n - \sqrt{n})$. I wish to show here that, more generally, $f(n + h) > f(n - h)$, where $0 < h < n$.

Indeed, the inequalities

$$\frac{(n+h)^n e^{-n-h}}{(n-h)^n e^{-n+h}} > 1 \quad \text{and} \quad \ln\frac{n+h}{n-h} > \frac{2h}{n}$$

are equivalent. The latter inequality can be proved as follows. Note that

$$\ln\{(n+h)/(n-h)\}$$

is the area of the region H below the hyperbola $y = 1/x$ and above the interval $[n-h, n+h]$ on the x-axis. On the other hand, $2h/n$ can be interpreted as the area of the trapezoidal region T below the tangent line to $y = 1/x$ at $x = n$, the midpoint of the interval $[n-h, n+h]$, and above the interval $[n-h, n+h]$ on the x-axis. Observing that $y = 1/x$ is concave upward for $x > 0$, it is apparent that the region T is contained in the region H and so the area of T is smaller than the area of H.

Noncentral Difference Quotients and the Derivative

P. P. B. EGGERMONT
Department of Mathematical Sciences, University of Delaware, Newark, DE 19716

We all have proved at one time or another the result that if a function f, defined on a neighborhood of the origin, is differentiable at the origin, then the central difference quotients converge, i.e.,

$$\lim_{h\to 0} \frac{f(h) - f(-h)}{2h}$$

exists and equals $f'(0)$, but that the converse is not true in general, the counterexam-