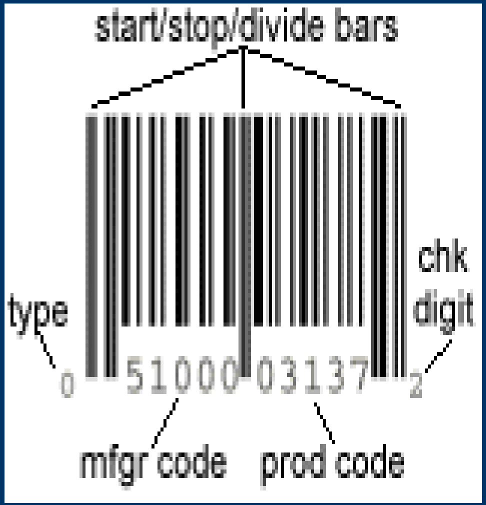
Check Digit Schemes and Error Detecting Codes

By Breanne Oldham
Union University

What are Check Digit Schemes?

Check digit schemes are numbers appended to an identification number that allow the accuracy of information stored to be checked by an algorithm.



What are error detecting codes?

Error detecting codes are algorithms that utilize check digits to detect, but not to correct, errors in entering identification numbers.

Where can error detecting schemes and check digits be found?

- International Standard Book Numbers (ISBN)
- Driver's License Numbers
- Universal Product Codes (UPC)
- Personal Checks
- Library Cards
- Shipping Labels

* W E L C O M E *
Copyright 1999 by Nick Johnson

Types of Common Errors

- Single digit (79.1%) 19456 → 19486
- Transposition of adjacent digits (10.2%) 19456→19546
- Jump transposition (0.8%) 19456→19654

Types of Common Errors (cont)

- Twin (0.5%) — 194<u>55</u>→194<u>66</u>
- Phonetic (0.5%) — 19<u>40</u>6→19<u>14</u>6
- Jump twin (0.3%) — <u>191</u>56→<u>494</u>56

3 Types of Error Detecting Codes

- Modular Arithmetic
- Permutations
- Noncommutative Schemes

ISBN Numbers

- A 10-digit number assigned to books published in most industrialized nations.
- Uses a modulo 11 arithmetic error detecting scheme with the last digit being the check digit.
- Detects 100% of single digit errors and adjacent transposition errors

ISBN with a Single Digit Error

Take the ISBN number 0-669-19496-4

Let it be incorrectly transmitted as 0-669-1<u>6</u>496-4

The ISBN check digit is determined by

$$10d_{10}+9d_9+8d_8+7d_7+6d_6+5d_5+4d_4+3d_3+2d_2+d_1 \equiv 0 \pmod{11}$$

So, for our example,

$$10(0)+9(6)+8(6)+7(9)+6(1)+5(6)+4(4)+3(9)+2(6)+1(4) \overset{?}{\equiv} 0 \pmod{11}$$

$260 \equiv 7 \pmod{11}$

Therefore, you know that there is a mistake.

ISBN with a Transposition of Adjacent Digits Error

Take the ISBN number 0-669-19496-4

Let it be incorrectly transmitted as

0-669-<u>91</u>496-4

So, in our example,

$10(0)+9(6)+8(6)+7(9)+6(9)+5(1)+4(4)+3(9)+2(6)+1(4) \stackrel{?}{\equiv} 0 \pmod{11}$

$283 \stackrel{?}{\equiv} 0 \pmod{11}$

$283 \equiv 8 \pmod{11}$

There is an error.

Proof of the ISBN Error Detecting Code (adjacent transposition)

- Let the ISBN number be denoted by

$$d_{10}d_9d_8d_7d_6d_5d_4d_3d_2d_1$$

- Then, the following represents the incorrect ISBN with transposition of adjacent digits

$$d_{10}d_9d_8\underline{d_6}\underline{d_7}d_5d_4d_3d_2d_1$$

- Let T denote the check sum for the ISBN number and let S denote the check sum for the incorrect ISBN number.

Proof of the ISBN Error Detecting Code (cont.)

$T = 10d_{10}+9d_9+8d_8+7d_7+6d_6+5d_5+4d_4+3d_3+2d_2+d_1$

and

$S = 10d_{10}+9d_9+8d_8+7d_6+6d_7+5d_5+4d_4+3d_3+2d_2+d_1$

Assume that T is a multiple of 11 and we show that S is not a multiple of 11.

Consider T-S. After cancellations, we obtain

$$T-S = 7d_7-7d_6+6d_6-6d_7 = d_7-d_6$$

Since d_7-d_6 is a one-digit number, then

$$-10 < d_7-d_6 < 10$$

Proof of the ISBN Error Detecting Code (cont.)

The only multiple of 11 between -10 and 10 is zero. But if $d_7 - d_6 = 0$, then $d_7 = d_6$ and no error has occurred. This is a contradiction.

Problems with the ISBN Error Detecting Code

- Using mod 11 arithmetic, the check digit can be 10.
- Since 10 is not a single digit, it must be denoted by “X”
- Does not detect all double errors

The IBM Scheme

- Can be used with any length identification number
- Used by credit card companies, libraries, blood banks, DMVs, and German banks
- Catches <u>ALL</u> Single Digit Errors
- Uses the permutation

  $\sigma = (0)(1,2,4,8,7,5)(3,6)(9)$

Problems with the IBM Scheme

- Does not catch all transposition of adjacent digits errors
- Specifically does not catch errors involving 0 and 9

Definition of the IBM Check Digit Scheme

Let $a_1a_2a_3 \ldots a_{n-1}$ represent an identifying number.

The check digit a_n is appended to the number $a_1a_2a_3 \ldots a_{n-1}a_n$ by using the permutation

$$\sigma = (0)(1,2,4,8,7,5)(3,6)(9)$$

in one of the following two ways.

Definition of the IBM Check Digit Scheme (cont.)

1. If n is even, the check digit a_n is assigned such that
   $\sigma(a_1)+a_2+\sigma(a_3)+a_4+ \ldots +\sigma(a_{n-1})+a_n \equiv 0 (\mathrm{mod} 10)$
2. If n is odd, the check digit a_n is assigned such that
   $a_1+ \sigma(a_2) + a_3+ \sigma(a_4)+ \ldots +a_{n-1}+ \sigma(a_n) \equiv 0 (\mathrm{mod} 10)$

Proof of the IBM Error Detecting Code

- Let $a_1 \ldots a_i \ldots a_n$ be an identification number with n even and $1 \leq i \leq n$.
- Let a single digit error occur transmitting $a_1 \ldots a_i \ldots a_n$ as $a_1 \ldots b_i \ldots a_n$ with $a_i \neq b_i$
- Assume that the error is not caught.

Proof of the IBM Error Detecting Code (cont)

Case I:

Since both errors are not caught,

$$\sigma(a_1)+a_2+\cdots\sigma(a_i)+\cdots+\sigma(a_{n-1})+a_n \equiv 0(\text{mod}10)$$

and

$$\sigma(a_1)+a_2+\cdots\sigma(b_i)+\cdots+\sigma(a_{n-1})+a_n \equiv 0(\text{mod}10)$$

This can also be written

$$(\sigma(a_1)+a_2+\cdots\sigma(a_i)+\cdots+\sigma(a_{n-1})+a_n)-(\sigma(a_1)+a_2+\cdots\sigma(b_i)+\cdots+\sigma(a_{n-1})+a_n) \equiv 0(\text{mod}10)$$

Proof of the IBM Error Detecting Code (cont)

This results in:

$0 = (\sigma(a_1)+a_2+\cdots\sigma(a_i)+\cdots+\sigma(a_{n-1})+a_n) - (\sigma(a_1)+a_2+\cdots+\sigma(b_i)+\cdots+\sigma(a_{n-1})+a_n) \pmod{10}$

$= \sigma(a_1)+a_2+\cdots\sigma(a_i)+\cdots+\sigma(a_{n-1})+a_n - \sigma(a_1) - a_2 - \cdots\sigma(b_i) - \cdots - \sigma(a_{n-1}) - a_n \pmod{10}$

$= \sigma(a_i) - \sigma(b_i) \pmod{10}$

Proof of the IBM Error Detecting Code (cont)

Thus $\sigma(a_i) - \sigma(b_i) \equiv 0 \pmod{10}$. Since $a_i \neq b_i$ and σ is a permutation, $\sigma(a_i) - \sigma(b_i) = 0$. Adding $\sigma(b_i)$ to both sides yields the result $\sigma(a_i) = \sigma(b_i)$. Since σ is a permutation, this means that $a_i = b_i$. This is a contradiction. Thus the assumption that the error is not caught is false, and the error is caught.

Proof of the IBM Error Detecting Code (cont)

Case II:

Since both errors are not caught,

$$\sigma(a_1)+a_2+\cdots a_i+\cdots+\sigma(a_{n-1})+a_n \equiv 0(\text{mod}10)$$

and

$$\sigma(a_1)+a_2+\cdots b_i+\cdots+\sigma(a_{n-1})+a_n \equiv 0(\text{mod}10)$$

This can also be written

$$(\sigma(a_1)+a_2+\cdots a_i+\cdots+\sigma(a_{n-1})+a_n)-(\sigma(a_1)+a_2+\cdots b_i+\cdots+\sigma(a_{n-1})+a_n) \equiv 0(\text{mod}10)$$

Proof of the IBM Error Detecting Code (cont)

This results in:

$0 = (\sigma(a_1)+a_2+\cdots a_i+\cdots+\sigma(a_{n-1})+a_n) - (\sigma(a_1)+a_2+\cdots+b_i + \cdots+\sigma(a_{n-1})+a_n) \pmod{10}$

$= \sigma(a_1)+a_2+\cdots a_i+\cdots+\sigma(a_{n-1})+a_n - \sigma(a_1) - a_2 - \cdots b_i - \cdots - \sigma(a_{n-1}) - a_n \pmod{10}$

$= a_i - b_i \pmod{10}$

Proof of the IBM Error Detecting Code (cont)

Thus $a_i - b_i \equiv 0 \pmod{10}$. Since $a_i \neq b_i$ and both a_i and b_i are integers between 0 and 10, $a_i - b_i = 0$. Adding b_i to both sides, the calculation results in $a_i = b_i$. This is a contradiction. Thus the assumption that the error is not caught is false, and the error is caught.

Checking a Library Number for Errors Using the IBM Scheme

Let a library book number be 21005620917 9 where 9 is the check digit.

Suppose the number is entered as 21005260917 9

The calculation shows that

$\sigma(2)+1+\sigma(0)+0+\sigma(5)+2+\sigma(6)+0+\sigma(9)+1+\sigma(7)+9 \overset{?}{\equiv} 0 \mod(10)$

$4+1+0+0+1+2+3+0+9+0+1+5+9 \overset{?}{\equiv} 0 \mod(10)$

$35 \overset{?}{\equiv} 0 \mod(10)$

$35 \equiv 5 \mod(10)$

There is an error.

The Verhoeff Check Digit Scheme

- A noncommutative scheme
- Catches all errors previously mentioned
- Based on the Cayley Table for D_{10}

*	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

Definition of the Verhoeff Check Digit Scheme

Let $a_1a_2 \ldots a_{n-1}\, a_n$ be an identification number with check digit a_n. The check digit a_n is appended to the number $a_1a_2 \ldots a_{n-1}$ such that the following equation is satisfied:

$$\sigma^{n-1}(a_1) * \sigma^{n-2}(a_2) * \sigma^{n-3}(a_3) * \cdots * \sigma(a_{n-1}) = 0$$

Where $\sigma = (0)(1,4)(2,3)(5,6,7,8,9)$ and $*$ is the group operation from D_{10} as previously presented.

Checking an Identification Number Using the Verhoeff Scheme

- Let 386018429278 be an identification number that is incorrectly transmitted as 386015429278. (Single Digit Error)
- Compute the check scheme using the Cayley Table for D_{10}.
- $\sigma^{11}(3)*\sigma^{10}(8)*\sigma^{9}(6)*\sigma^{8}(0)*\sigma^{7}(1)*\sigma^{6}(5)*\sigma^{5}(4)*\sigma^{4}(2)*\sigma^{3}(9)*\sigma^{2}(2)*\sigma(7)*8 \stackrel{?}{=} 0$

Checking an Identification Number Using the Verhoeff Scheme

- This leads to

  2*8*5*0*4*6*1*2*7*2*8*8 $\stackrel{?}{=}$ 0

  $2 \neq 0$

- Thus, there must be an error.
- This error would be caught.

*	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

Checking an Identification Number Using the Verhoeff Scheme

- Let 3860184<u>29</u>278 be an identification number that is incorrectly transmitted as 3860148<u>92</u>278. (transpostion of adjacent digits error)
- Compute the check scheme using the Cayley Table for D_{10}.
- $\sigma^{11}(3)*\sigma^{10}(8)*\sigma^{9}(6)*\sigma^{8}(0)*\sigma^{7}(1)*\sigma^{6}(8)*\sigma^{5}(4)*\sigma^{4}(9)*\sigma^{3}(2)*\sigma^{2}(2)*\sigma(7)*8 \stackrel{?}{=} 0$

Checking an Identification Number Using the Verhoeff Scheme

- This leads to

$2*8*5*0*4*9*1*8*3*2*8*8 \stackrel{?}{=} 0$

$4 \neq 0$

- Thus, there must be an error.
- This error would be caught.

*	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

What's Next?

- Multiple Check Digits in a single Identification Number
- Error Correcting Codes
  - ✓ Similar to Error Detecting Codes
  - ✓ Both detect and correct the errors that are found.

Conclusion

- Check digits are an integral part of error detecting codes.
- There are three different types of error detecting codes which are progressively more sophisticated and detect more errors

Bibliography

Dixon, Emily. *Take a Break.* http://pass.maths.org.uk/issue12/features/codes/

Gallian, Joseph A. "The Mathematics of Identification Numbers." *The College Mathematics Journal.* May 1991. 194-202.

Kirtland, Joseph. *Identification Numbers and Check Digit Schemes.* Mathematical Assoc. of America. 2001.

Wheeler, Mary L. "Check Digit Schemes." *The Mathematics Teacher.* April 1994. 228-230.