Check Digit Schemes
MATH 171 Freshman Seminar

J. Robert Buchanan
Department of Mathematics

2009
In our increasingly computerized world, products, things, and people are identified using numbers. These numbers are often transmitted over the telephone, internet, or even through the mail. Sometimes errors are made during transmission. Sometimes unscrupulous people try to forge identification numbers. A check digit is often included in an identification number to try and catch errors and attempts at deception. The most common types of transmission errors are summarized in the table below.
<table>
<thead>
<tr>
<th>Error Type</th>
<th>Form</th>
<th>Rel. Freq.</th>
<th>Correct</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single Digit</td>
<td>$a \rightarrow b$</td>
<td>0.791</td>
<td>191433</td>
<td>191933</td>
</tr>
<tr>
<td>Adjacent trans.</td>
<td>$ab \rightarrow ba$</td>
<td>0.102</td>
<td>191433</td>
<td>191343</td>
</tr>
<tr>
<td>Jump trans.</td>
<td>$abc \rightarrow cba$</td>
<td>0.008</td>
<td>191433</td>
<td>193413</td>
</tr>
<tr>
<td>Twin</td>
<td>$aa \rightarrow bb$</td>
<td>0.005</td>
<td>191433</td>
<td>191455</td>
</tr>
<tr>
<td>Phonetic</td>
<td>$a0 \rightarrow 1a$</td>
<td>0.005</td>
<td>191433</td>
<td>194033</td>
</tr>
<tr>
<td>Jump twin</td>
<td>$aca \rightarrow bcb$</td>
<td>0.003</td>
<td>191433</td>
<td>393433</td>
</tr>
</tbody>
</table>
Avoiding these errors

Several different check digit schemes are in common use today all around us. We will examine their principles of operation, their strengths, and their weaknesses.

Example

Choose one of the lists of numbers below and determine the common characteristic of all the numbers in the list.

- \{5, 21, 37, 85, 125\}
- \{1, 13, 37, 61, 121\}
- \{8, 26, 44, 53, 98\}
- \{4, 11, 32, 67, 81\}

J. Robert Buchanan
Check Digit Schemes
Avoiding these errors

Several different check digit schemes are in common use today all around us. We will examine their principles of operation, their strengths, and their weaknesses.

Example

Choose one of the lists of numbers below and determine the common characteristic of all the numbers in the list.

- \( \{5, 21, 37, 85, 125\} \)
- \( \{1, 13, 37, 61, 121\} \)
- \( \{8, 26, 44, 53, 98\} \)
- \( \{4, 11, 32, 67, 81\} \)
Modular Arithmetic

**Definition**
We say that $x$ is **congruent to** $r$ **modulo** $n$, denoted

$$x \equiv r \pmod{n}$$

provided $n$ divides $x - r$ evenly.

**Definition**
$x \pmod{n}$ is the remainder obtained by dividing $x$ by $n$.

**Theorem**

\[
\begin{align*}
(a + b) \pmod{n} &= ((a \pmod{n}) + (b \pmod{n})) \pmod{n} \\
(a - b) \pmod{n} &= ((a \pmod{n}) - (b \pmod{n})) \pmod{n} \\
(a \cdot b) \pmod{n} &= ((a \pmod{n}) \cdot (b \pmod{n})) \pmod{n}
\end{align*}
\]
Modular Arithmetic

Definition
We say that \( x \) is congruent to \( r \) modulo \( n \), denoted

\[ x \equiv r \pmod{n} \]

provided \( n \) divides \( x - r \) evenly.

Definition
\( x \pmod{n} \) is the remainder obtained by dividing \( x \) by \( n \).

Theorem

\[
\begin{align*}
(a + b) \pmod{n} &= ((a \pmod{n}) + (b \pmod{n})) \pmod{n} \\
(a - b) \pmod{n} &= ((a \pmod{n}) - (b \pmod{n})) \pmod{n} \\
(a \cdot b) \pmod{n} &= ((a \pmod{n}) \cdot (b \pmod{n})) \pmod{n}
\end{align*}
\]
Modular Arithmetic

Definition
We say that $x$ is congruent to $r$ modulo $n$, denoted

$$x \equiv r \pmod{n}$$

provided $n$ divides $x - r$ evenly.

Definition
$x \pmod{n}$ is the remainder obtained by dividing $x$ by $n$.

Theorem

\[
\begin{align*}
(a + b) \pmod{n} & = ((a \pmod{n}) + (b \pmod{n})) \pmod{n} \\
(a - b) \pmod{n} & = ((a \pmod{n}) - (b \pmod{n})) \pmod{n} \\
(a \cdot b) \pmod{n} & = ((a \pmod{n}) \cdot (b \pmod{n})) \pmod{n}
\end{align*}
\]
Example

Calculate each of the following expressions:

- $23 \pmod{7}$
- $106 \pmod{13}$
- $(25 - 9) \pmod{7}$
- $(12 \cdot 18) \pmod{6}$
- $13^4 \pmod{6}$
The identification code on a United States Postal Service money order is an 11-digit number. The first ten digits are the actual identification code. The 11th digit is the check digit.
The check digit is calculated using the following algorithm:

- Let \( a_1 a_2 a_3 a_4 a_5 a_6 a_7 a_8 a_9 a_{10} a_{11} \) be a valid USPS money order identification number,
- then

\[
a_{11} \equiv (a_1 + a_2 + a_3 + a_4 + a_5 + a_6 + a_7 + a_8 + a_9 + a_{10}) \pmod{9}
\]
Example

Determine if the following USPS money order identification numbers are valid.

- 67021200988
- 90123212125
- 31059112852

Determine the correct check digit for the money order identifier 3980062110.
Discussion Questions

1. Which type of single digit error will this scheme fail to catch?

2. Which of our common transmission errors will this check digit scheme miss?

3. Which type of transposition error will this scheme detect?
   You may wish to experiment/conjecture with the following valid numbers.

   Correct number: 27914009534 27914009534
   Incorrect number: 27194009534 27914009543
The identification number on an airline ticket is a 15-digit number. The first fourteen digits are the actual identification number and the 15th digit is the check digit.
The check digit is calculated from the identification number using the following algorithm.

- Let \( a_1 a_2 a_3 a_4 a_5 a_6 a_7 a_8 a_9 a_{10} a_{11} a_{12} a_{13} a_{14} a_{15} \) be a valid airline ticket identification number,
- then

\[
a_{15} \equiv (a_1 a_2 a_3 a_4 a_5 a_6 a_7 a_8 a_9 a_{10} a_{11} a_{12} a_{13} a_{14}) \pmod{7}.
\]
Example

Consider the fifteen digit number 001479769423786. The check digit is 6. We can verify that

\[
00147976942378 = 147976942378 = (7)(21139563196) + 6
\]
Example

Determine if the following airline ticket identification numbers are valid.

- 0-037-1222494405-0
- 0-037-1222495505-0

Determine the correct check digit for the airline ticket number 0-004-2871911233.
Discussion Questions

1. What are all the possible numbers which could be check digits for airline ticket identifiers?

2. This scheme will not detect all single digit errors. Which will it miss? You may wish to experiment/conjecture with the following numbers some of which are valid, others are invalid but the check digit scheme fails to detect the error.

<table>
<thead>
<tr>
<th>Correct</th>
<th>Incorrect</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-004-2871911233-4</td>
<td>0-004-2871511233-4</td>
</tr>
<tr>
<td>0-004-2871911233-4</td>
<td>0-004-2171911233-4</td>
</tr>
</tbody>
</table>

3. This scheme will not detect all adjacent transposition errors. Which will it miss? You may want to consider the following two examples.

<table>
<thead>
<tr>
<th>Correct</th>
<th>Incorrect</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-004-2870911233-3</td>
<td>0-004-2807911233-3</td>
</tr>
<tr>
<td>0-004-2870911233-3</td>
<td>0-004-2870912133-3</td>
</tr>
</tbody>
</table>
There are actually five versions of the Universal Product Code (UPC) in use in the United States. Every other country uses the European Article Numbering (EAN) code. We will explore here the 12-digit UPC code often called version A.
The first number of the twelve identifies the type of product. The type codes are listed in the following table.

<table>
<thead>
<tr>
<th>$a_1$</th>
<th>Specific Use</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>General groceries</td>
</tr>
<tr>
<td>2</td>
<td>Meat and produce</td>
</tr>
<tr>
<td>3</td>
<td>Drugs and health products</td>
</tr>
<tr>
<td>4</td>
<td>Non-food items</td>
</tr>
<tr>
<td>5</td>
<td>Coupons</td>
</tr>
<tr>
<td>6, 7</td>
<td>Other items</td>
</tr>
<tr>
<td>1, 8, 9</td>
<td>Reserved for future use</td>
</tr>
</tbody>
</table>
The 12th digit is the check digit assigned to the first 11 digits. It is calculated in the following way.

- If the Version A UPC is $a_1 a_2 a_3 a_4 a_5 a_6 a_7 a_8 a_9 a_{10} a_{11} a_{12}$,

then

$$3(a_1 + a_3 + a_5 + a_7 + a_9 + a_{11}) + (a_2 + a_4 + a_6 + a_8 + a_{10} + a_{12}) \pmod{10} = 0.$$
### Example

<table>
<thead>
<tr>
<th></th>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Is the UPC code is 038000312106 valid?</td>
</tr>
<tr>
<td>2</td>
<td>Is the UPC code is 070501091104 valid?</td>
</tr>
<tr>
<td>3</td>
<td>Is the UPC code is 123004244107 valid?</td>
</tr>
<tr>
<td>4</td>
<td>Given the first 11 digits of the UPC code 50200391562, find the check digit.</td>
</tr>
</tbody>
</table>
1. What are all the possible numbers which could be check digits for UPC Version A identifiers?

2. The UPC scheme will catch all single-digit errors. Suppose the UPC $a_1 \cdots a_i \cdots a_{12}$ where $1 \leq i \leq 12$ is mis-transmitted as $a_1 \cdots b_i \cdots a_{12}$ where $a_i \neq b_i$. A single-digit error has occurred where $a_i$ was replaced by $b_i$. Explain how the error will be caught by the check digit scheme. There are two cases to consider.
All books published since 1972 have a 10-digit ISBN number associated with them. The first nine digits identify the language, publisher, and book. The tenth digit is the check digit.
If the first nine digits are $a_1a_2a_3a_4a_5a_6a_7a_8a_9$ then the check digit $a_{10}$ is the solution to the equation:

$$10a_1 + 9a_2 + 8a_3 + 7a_4 + 6a_5 + 5a_6 + 4a_7 + 3a_8 + 2a_9 + a_{10} = 0 \pmod{11}$$

where the letter "X" is used if the check digit is “10”.
Example

Consider the ISBN number 0883857200. Note that the check digit is “0”. We can verify that

\[(10)(0) + (9)(8) + (8)(8) + (7)(3) + (6)(8) + (5)(5) + (4)(7) + (3)(2) + (2)(0) + 0 \]
\[= 264 = (24)(11) \equiv 0 \pmod{11}\]
Example

Given the first 9 digits of the book identifier 354096576, we can let the check digit be $C$. Then $C$ must satisfy the equation,

\[
(10)(3) + (9)(5) + (8)(4) + (7)(0) + (6)(9) + (5)(6) \\
+ (4)(5) + (3)(7) + (2)(6) + C \\
\equiv 0 \pmod{11}
\]

\[
244 + C \equiv 0 \pmod{11}
\]

Thus the correct check digit is 9.
1. What all the possible numbers which could be check digits for ISBN identifiers?

2. The ISBN scheme will catch all single-digit and adjacent transposition errors. Suppose the ISBN $a_1 \cdots a_i \cdots a_{10}$ where $1 \leq i \leq 10$ is mis-transmitted as $a_1 \cdots b_i \cdots a_{10}$ where $a_i \neq b_i$. A single-digit error has occurred where $a_i$ was replaced by $b_i$. Explain how the error will be caught by the check digit scheme.
Since January 1, 2007 ISBN numbers have 13 digits. In a paragraph:

- explain the difference between the ISBN-10 number and the ISBN-13 number,
- explain the check digit calculation for an ISBN-13 identifier and give a valid example for an actual book (include the title and author of the book), and
- explain whether the ISBN-13 check digit scheme will catch all adjacent transposition errors.