

## Another Technique for Finding the Length of the Keyword

From the discussion that resulted in the method for calculating  $I$ , we notice that Friedman's determination of whether a cipher is polyalphabetic or monoalphabetic depends upon the likelihood of "drawing" two of the same letters from a ciphertext. The probability that the two letters will be the same if the cipher is polyalphabetic is approximately 0.038, and the probability that the two letters will be the same if the cipher is monoalphabetic is approximately 0.065. So, it is more likely that the two letters will be the same if they are drawn from the same cipher alphabet.

We can use this information to develop another method for determining the length of the keyword.

Consider the ciphertext that we used as an example in the discussion of the index of coincidence:

DBZMG	AOIYS	OPVFH	OWKBW	XZPJL	VVRFG
NBKIX	DVUIM	OPFQL	VVPUD	KPRVW	OARLW
DVLMW	AWINZ	DAKBW	MMRLW	QIICG	PAKYU
CVZKM	ZARPS	DTRVD	ZWEYG	ABYYE	YMGYF
YAFHL	CMWLW	LCVHL	MMGYL	DBZIF	JNCYL
OMIAJ	JCGMA	IBVRL	OPVFW	OBVLK	OPVUJ
ZDVLQ	XWDGG	IQEYF	BTZMZ	DVRMM	ANZWA
ZVKFQ	GWEAL	ZFKNZ	ZZVCK	VDVLQ	BWFXU
CIEWW	OPRMU	JZIIK	KWEXA	IOIYH	ZIKYV
GMKNW	MOIIM	KADUQ	WMWIM	ILZHL	CMTCH
CMINW	SBRHV	OPVSO	DTCMG	HMKCE	ZASYD
JKRNW	YIKCF	OMIPS	GAFZK	JUVGM	GBZJD
ZWWNZ	ZVLGT	ZZFZS	GXYUT	ZBJCF	PAVNZ
ZAVWS	IJVZG	PVUVQ	NKRHF	DVXNZ	ZKZJZ
ZZKYP	OIEXX	MWDNZ	ZQIMH	VKZHY	DVKYD
GQXYF	OOLYK	NMJGS	YMRML	JBYYF	PUSYJ
JNRFH	CISYL	N			

Let us think of the ciphertext message as one (very long) string of letters. In the example we will only use a portion of the ciphertext message

DBZMGAOIYSOPVFHOWKBWXPJLVVRFGNBKIXDVUIMOPFQLVVPUDKPRVWOARLWDVLMW

Make two copies of this string and slide one along the other. For each shift, count the number of vertical pairs of letters that are duplicate letters. We view each vertical pair of letters as a “draw” – the top letter is from one cipher alphabet and the bottom letter is from another. Friedman’s test tells us that we are mostly likely to draw pairs of duplicate letters if we are drawing from the same cipher alphabet – when both the top letter and the bottom letter come from the same cipher alphabet. The first time this happens is when the shift corresponds to the length of the keyword.

Here is the example; instances of duplicate letters are in bold.

#### Shift of one

DBZMGAOIYSOPVFHOWKBWXPJLV**V**RFGNBKIXDVUIMOPFQLV**V**PUDKPRVWOARLWDVLMW  
DBZMGAOIYSOPVFHOWKBWXPJLVVRFGNBKIXDVUIMOPFQLVVPUDKPRVWOARLW

Two pairs of duplicate letters

#### Shift of two

DBZMGAOIYSOPVFHOWKBWXPJLVVRFGNBKIXDVUIMOPFQLVVPUDKPRVWOARLWDVLMW  
DBZMGAOIYSOPVFHOWKBWXPJLVVRFGNBKIXDVUIMOPFQLVVPUDKPRVWOARLW

No duplicate letters

#### Shift of three

DBZMGAOIYSOPVFHOWKB**W**XPJLVVRFGNBKIXDVUIMOPFQLVVPUDKPRVWOARLWDVLMW  
DBZMGAOIYSOPVFHOW**K**BWXPJLVVRFGNBKIXDVUIMOPFQLVVPUDKPRVWOARLW

One pair of duplicate letters

#### Shift of four

DBZMGAOIYSOPVFHOWKBWXPJLVVRFGNBKIXDVUIMOPFQLVVPUDK**P**RVWOARLWDVLMW  
DBZMGAOIYSOPVFHOWKBWXPJLVVRFGNBKIXDVUIMOPFQLVVPUDKPRV**W**OARLW

Two pairs of duplicate letters

#### Shift of five

DBZMGAOIYSOPV**H**OWKBWXPJLVVRFGNBKIXDVUIMOPFQLVVPUDKPRVWOARL**W**DVLMW  
DBZMGAOIYS**O**PVFHOWKBWXPJLVVRFGNBKIXDVUIMOPFQLVVPUDKPRV**W**OARLW

Four pairs of duplicate letters

Etc.

Recall that the length of the keyword is 5.