More counting

Without listing them, count the pairs of disjoint 3-element and 5-element subsets of a 12-element set.

There are 12 choices for the first element of the first set, $12-1$ for the second, $12-2$ for the third, so there are $12(12-1)(12-2)$ choices for an **ordered** subset of 3 elements. But this style of choosing artificially orders the chosen elements. To take this into account, divide by 3!, the number of ways to order a set with 3 elements. (As earlier) there are

$$12(12-1)(12-2)/3! = \binom{12}{3}$$

choices for a 3-element subset of a 12-element set.

From the remaining $(12-3)$-element subset, there are $(12 - 3)$ choices for the first element of the second set, $(12 - 3) - 1$ choices for the second element of the second set, and so on. Divide by $5!$ to discount the artificial ordering. So for each choice of the first set there are

$$(12-3)(12-3-1)\ldots(12-3-5+1)/5!$$

Thus, altogether there are

$$\frac{12!}{(12-3)!\,3!}\cdot\frac{(12-3)!}{(12-3-5)!\,5!}$$

$$=\frac{12!}{(12-3-5)!\,3!\,5!}$$

choices of 3-element and 5-element subsets of a 12-element set.

Note that we get the same answer if the roles of 3 and 5 are reversed in the derivation.

One more counting problem

Count the number of sets of 3 disjoint 2-element subsets of a 12-element set.

As above, there are $\binom{12}{2}$ choices for the first (?!) subset, $\binom{12-2}{2}$ choices for the second, and $\binom{12-2-2}{2}$ choices for the third. *But* there is no ordering on the set of 2-element subsets, so our choice procedure will choose the same thing several times (unlike the case where the disjoint subsets are different sizes)! For example,

$$\{\{1,2\},\{7,8\},\{3,4\}\}$$

would be chosen separately as

$$\{\{7,8\},\{3,4\},\{1,2\}\}$$

and altogether 3! ways. Thus, we must divide by 3!, the number of ways to order 3 things, getting the final count

$$\binom{12}{2}\binom{12-2}{2}\binom{12-2-2}{2}/3!$$

More Bad Old Ciphers: Affine Cipher

Shift ciphers use *addition* and reduction modulo 26, but the keyspace is too small.

An obvious attempt to make a more complex (**bigger keyspace**) version of a shift cipher, using natural mathematical manipulation of letters `abc...xyz` encoded as numbers $0, 1, 2, \ldots, 23, 24, 25$ is to use *multiplication* as well as *addition*.

An **affine cipher** has keys (a, b) where a is an *odd integer* not divisible by 13 and b is any integer. The encryption step is

$$E_{a,b}(x) = (a \cdot x + b) \,\%\, 26$$

When $a = 1$ this is just a shift cipher. For example, with key $(5, 11)$, the character `t` encrypts as

$$\texttt{t} \to 19 \to (5 \cdot 19 + 11) \,\%\, 26 = 106 \,\%\, 26$$

$$= 2 \to \texttt{c}$$

It may not be apparent, but decryption is of the same form. Let a^{-1} be the **multiplicative inverse** of a **modulo** 26. Then for a key (a, b) the decryption step is

$$x \to ((a^{-1} \cdot x) - (a^{-1} \cdot b)) \,\%\, 26$$

Of we pretend that the reduction modulo 26 is simply not there, then verification that this does really decrypt is easy:

$$a^{-1}(ax + b) - a^{-1}b = x + a^{-1}b - a^{-1}b = x$$

The fact that reduction modulo m allows an essentially idential computational style is not obvious, but is critical.

Reduction versus arithmetic

It is not obvious, but is true, and will be proven later, that addition and multiplication interact well with reduction modulo m:

$$((x \,\%\, m) + (y \,\%\, m)) \,\%\, m$$

$$= ((x \,\%\, m) + y) \,\%\, m = (x + y) \,\%\, m$$

and

$$((x \,\%\, m) \cdot (y \,\%\, m)) = ((x \,\%\, m) \cdot y) \,\%\, m$$

$$= ((x \,\%\, m) \cdot y) \,\%\, m = (x \cdot y) \,\%\, m$$

In other words, we can reduce *or not* modulo m at any point in a computation, as long as we reduce modulo m at the end.

Attacks on affine ciphers

From a *human* viewpoint, the **ciphertext-only** attack, meaning getting a message without knowing either the *key* or anything about the *message*, is harder than for the shift cipher, because of the bulk of trial decryptions to look through: even if we only trial-decrypt the first few characters... Running an outer loop over a and an inner loop over b, with $a = 1, 3, 5, 7$ and $0 \leq b \leq 25$, from ciphertext

`CRIITFTEIWCTFLERVTP`

we get nine-character strings

`criitftei, dsjjugufj, etkkvhvgk,`
`fullwiwhl, gvmmxjxim, hwnnykyjn,`
`ixoozlzko, jyppamalp, kzqqbnbmq,`
`larrcocnr, mbssdpdos, nctteqept,`
`oduufrfqu, pevvgsgrv, qfwwhthsw,`
`rgxxiuitx, shyyjvjuy, tizzkwkvz,`
`ujaalxlwa, vkbbmymxb, wlccnznyc,`

xmddoaozd, yneepbpae, zoffqcqbf,
apggrdrcg, bqhhsesdh, gzyyfpfmy,
hazzgqgnz, ibaahrhoa, jcbbisipb,
kdccjtjqc, leddkukrd, mfeelvlse,
ngffmwmtf, ohggnxnug, pihhoyovh,
qjiipzpwi, rkjjqaqxj, slkkrbryk,
tmllscszl, unmmtdtam, vonnueubn,
wpoovfvco, xqppwgwdp, yrqqxhxeq,
zsrryiyfr, atsszjzgs, buttakaht,
cvuublbiu, dwvvcmcjv, exwwdndkw,
fyxxeoelx, khoorzruo, lippsasvp,
mjqqtbtwq, nkrrucuxr, olssvdvys,
pmttwewzt, qnuuxfxau, rovvygybv,
spwwzhzcw, tqxxaiadx, uryybjbey,
vszzckcfz, wtaadldga, xubbemehb,
yvccfnfic, zwddgogjd, axeehphke,
byffiqilf, czggjrjmg, dahhksknh,
ebiiltloi, fcjjmumpj, gdkknvnqk,
helloworl, ifmmpxpsm, jgnnqyqtn,
opeedjdce, pqffekedf, qrggflfeg,
rshhgmgfh, stiihnhgi, tujjioihj,
uvkkjpjik, vwllkqkjl, wxmmlrlkm,

`xynnmsmln, yzoontnmo, zappouonp, abqqpvpoq, bcrrqwqpr, cdssrxrqs, dettsysrt, efuutztsu, fgvvuautv, ghwwvbvuw, hixxwcwvx, ijyyxdxwy, jkzzyeyxz, klaazfzya, lmbbagazb, mnccbhbac, noddcicbd`

Oops! Looking back a little bit there is `helloworl` which is, plausibly English.

This also illustrates human limitations.

How can we **automate** decryption, rather than relying upon human recognition of a natural language?

Review of linear systems of equations

Using ordinary numbers, a system of equations

$$\begin{cases} ax_1 + b = y_1 \\ ax_2 + b = y_2 \end{cases}$$

can be solved for unknowns a, b given x_1, y_1, x_2, y_2 fairly easily. For example, to solve

$$\{a \cdot 5 + b = 5a \cdot 2 + b = 11$$

subtract the second equation from the first to *eliminate* b

$$3 \cdot a = -6$$

and

$$a = -2$$

Substitute $a = -2$ back into the first equation $a \cdot 5 + b = 5$

$$-2 \cdot 5 + b = 5$$

so $b = 15$. Solution is $(a, b) = (-2, 15)$.

Non-semantic attack on affine ciphers

Example: An affine cipher with unknown key (a, b) encrypts $E_{a,b}(\texttt{m}) = \texttt{j}$ and $E_{a,b}(\texttt{t}) = \texttt{m}$. Determine the key.

Of course, encode as numbers rather than characters: `a` through `z` become 0 through 25, respectively, so `m` becomes `12`, t becomes `19`, `j` becomes `9`, and `m` becomes `12`. Thus, the given relations are

$$\begin{cases} (a \cdot 12 + b) \,\%\, 26 = 9 \\ (a \cdot 19 + b) \,\%\, 26 = 12 \end{cases}$$

Except for the reduction, it's a system of two linear equations in two unknowns.

Subtract the second equation from the first, and use the fact that addition, subtraction, and multiplication interact well with reduction modulo 26:

$$(-7 \cdot a) \,\%\, 26 = -3 \,\%\, 26$$

By brute force at worst, a multiplicative inverse to -7 mod 26 is 11. Thus,

$$(11 \cdot (-7 \cdot a)) \,\%\, 26 = (11 \cdot -3) \,\%\, 26$$

Simplifying the left-hand side using the good interaction of reduction with addition and multiplication,

$$((11 \cdot -7) \,\%\, 26) \cdot a \,\%\, 26 = (1 \cdot a) \,\%\, 26 = a \,\%\, 26$$

Thus

$$a = (11 \cdot -3) \,\%\, 26 = -33 \,\%\, 26 = 19$$

From the first equation, substituting back and solving for b, $(19 \cdot 12 + b) \,\%\, 26 = 9$ yields

$$b = (9 - 19 \cdot 12) \,\%\, 26 = 15$$

Thus, the key is $(19, 15)$.

Non-semantic analysis/attacks

An insight of William Friedman (and others) circa 1920 was that natural languages have a probabilistic/statistical nature that

(1) Can be used to *automate* attacks, not requiring constant human supervision.

(2) Can be used to *create* attacks that do not correspond to direct human intuition.

Even more amazing is that apparently non-semantic but merely statistical characterizations *suffice* for most purposes.

Ironically, current research efforts to describe the structure of language seem to do no better than just crude statistical explanations. That is, semantic descriptions are not as good as statistical ones!

Basic ideas of probability

What *is* probability???

To say that the *probability* of something happening is the *chance* or it happening, or the *likelihood*, or any other synonym, does not address the issue.

Can we *measure* it? This would be a problem of applied statistics, and is a profound and confusing issue in itself. We will ignore it.

Can we make *inferences* about probability?

Yes, and without knowing what it *truly* is and without worry about *measuring* it.

(By the way, the conversion between *chances-of* and *probability* is that chances-of is a percentage, while probability is a number between 0 and 1. So 34% chance is probability of 0.34)

Example: fairness. A **fair coin** is a coin with *heads* and *tails* **equally likely**. That is,

$$P(heads) = P(tails)$$

It is merely a *normalization* that the sum of the probabilities of all the possible outcomes is 1, so

$$P(heads) + P(tails) = 1$$

That is, we have a system of the form

$$\begin{cases} x & = y \\ x + y & = 1 \end{cases}$$

which we solve (without knowing what probability is)

$$P(heads) = P(tails) = \frac{1}{2}$$

We have completed a numerical computation without being able to answer any philosophical or other deeper questions.

Example: urns.

Suppose there are 3 red balls and 9 green balls in an *urn*, otherwise indistinguishable.

As with the coin, we infer that the probabilities of drawing the $12 = 3+9$ balls are all the same, and add up to 1, so are all $1/12$.

It is a small leap to infer that, since drawing one of the balls precludes drawing any other, that the probability of drawing a *red* ball is

$$P(\mathit{red}) = P(\mathit{red}_1) + P(\mathit{red}_2) + P(\mathit{red}_3)$$

$$= \frac{1}{12} + \frac{1}{12} + \frac{1}{12} = 3 \cdot \frac{1}{12} = \frac{1}{4}$$

and that of drawing a *green* ball is, similarly,

$$P(\mathit{green}) = 9 \cdot \frac{1}{12} = \frac{3}{4}$$

Example: independence. The apparent fact that the outcome of *one* flip of a coin has no effect on the outcome of *another* flip of a coin is the *independence* of the two events.

(This has nothing to do with the *fairness* or not of the coin.)

That is, neither the coin nor the universe remember prior flips, and do not try to compensate or make up for too many heads in the past, etc.

In a different world this could have been otherwise.

(In contrast, a sequence of events in which the outcome of the next event *can* be affected by the previous one is (roughly) a **Markov process**.)

Not the definition of probability

It *turns out* **not** wise to define the probability of an outcome of an event as

$$P(\text{outcome}) = \lim_{\textit{trials} \to \infty} \frac{\textit{no. times}\ \text{outcome}\ \textit{occurs}}{\textit{total no. trials}}$$

(Yet this statement *is true*, and is a *theorem*, the Law of Large Numbers.)

Cannot do infinitely many tests.

Do not know how rapidly the result of a finite number of tests approaches the limit.

Do not know that the limit exists in any sense.

Might evaluate the limit on different days and get different answers?

Different people might get different approximations?

Monty Hall Paradox

Or, in case elementary probability seems all too easy, here is a popular example that may seem less obvious.

In a game show *Let's Make a Deal* in which players were faced with 3 doors, behind one of which was a prize. The player chose a door, but the door was not opened. The host *Monty Hall* (who knew where the prize was) opened *another* door than the one guessed by the player, but *not* the one with the prize. The player was offered the chance to change their guess. *Should the player change their guess?*

Thus, the player was faced with one open door with no prize, and two closed doors, one of which was their original guess, and behind one of which is the prize.

They should always change their guess.

This may be counter-intuitive.

One way to explain this in colloquial terms is to say that the probability of originally guessing the correct door is 1/3, and that does not change. Thus, the probability is

$$1 - \frac{1}{3} = \frac{2}{3}$$

that you're *wrong*, and should change your guess.

Among many *incorrect* arguments there is the one that says that, not knowing what else is going on, since there are two doors, the probability is 1/2. This approach, in which *ignorance of facts is interpreted as equal probability*, was already disdained by Laplace 300 years ago, and we should not use it now.