The context now is that we will give more of the mathematical background for W. Friedman's decisive attack on the Vigenere cipher.

This attack illustrates that single-letter frequencies in natural languages combined with some maybe-not-so-intuitive mathematical manipulations can be used to break ciphers.

Kasiski has also broken Vigenere, about 1880.

Strangely, Vigenere was still believed to be unbreakable in the early 20th century.

And keep in mind that a mis-used OTP degenerates into Vigenere, so a mis-used OTP is also completely broken.

Random variables

A **random variable** X is really just a real-valued function on a probability space Ω (which, recall, is basically a set with a probability measure on its subsets).

For a real number x, the **probability that** X **takes value** x is denoted $P(X = x)$, and by definition is

$$P(X = x) = P(\{\omega \in \Omega : X(\omega) = x\})$$

For example, if $\Omega = \{H, T\}$ is the sample space for flipping a fair coin, we could define a random variable X for $\omega \in \Omega$ by

$$X(\omega) = \text{ no. heads when } \omega \text{ occurs}$$

Yes,

$$X(\mathrm{H}) = 1 \qquad X(\mathrm{T}) = 0$$

For Ω the set of outcomes ω of 4 flips of a fair coin we could similarly define

$$X(\omega) = \text{no. H's occuring in } \omega$$

In this example, the notation means

$$\begin{aligned}
P(X = 0) &= P(\text{zero Hs in 4 flips}) \\
P(X = 1) &= P(\text{exactly one H in 4 flips}) \\
P(X = 2) &= P(\text{exactly two Hs in 4 flips}) \\
P(X = 3) &= P(\text{exactly three H in 4 flips}) \\
P(X = 4) &= P(\text{exactly four H in 4 flips})
\end{aligned}$$

For any other real value x, $P(X = x) = 0$, since we can't get any other number of Hs in 4 flips.

Expected values

The **expected value** $E(X)$ or EX of a random variable X on a probability space Ω is a kind of *weighted average* of the values of X, with the weights being the probabilities of the different inputs/outputs. The precise definition is

$$\begin{aligned}\text{expected value of } X = E(X) \\ = \sum_{\omega \in \Omega} P(\omega) \cdot X(\omega)\end{aligned}$$

We can *group* the inputs according to the output value produced, so this is also equal to

$$E(X) = \sum_{\text{values } x \text{ of } X} P(X = x) \cdot x$$

where (again) the notation $P(X = x)$ means the probability that X takes value x:

$$P(X = x) = P(\{\omega \in \Omega : X(\omega) = x\})$$

About notation

Yes, the notation and terminology for random variables is different from, and in conflict with, the kind of notation used for functions and their values in calculus and differential equations.

First, and most importatly, yes, random *variables* are actually *functions*.

Yes, the random variable's name is oftne X, unlike the f or g in calculus.

Yes, usually the *input* to a function is called x, not the *output*, as in $X(\omega) = x$.

Examples of expected values

With X being the random variable counting Hs in a single flip of a fair coin,

$$E(X) = \sum_{\text{values } x \text{ of } X} P(X = x) \cdot x$$

$$= P(X = 0) \cdot 0 + P(X = 1) \cdot 1$$

$$= \frac{1}{2} \cdot 0 + \frac{1}{2} \cdot 1 = \frac{1}{2}$$

Note that we will never actually get 1/2 head in a flip of a fair coin.

But, as with many averages, the average or weighted average of integer values may be a non-integer.

That's ok.

With X being the random variable counting Hs in 3 flips of a fair coin,

$$E(X) = \sum_{\text{values } x \text{ of } X} P(X = x) \cdot x$$

$$= P(X = 0) \cdot 0 + P(X = 1) \cdot 1$$

$$+ P(X = 2) \cdot 2 + P(X = 3) \cdot 3$$

$$= \binom{3}{0} 2^{-3} \cdot 0 + \binom{3}{1} 2^{-3} \cdot 1$$

$$+ \binom{3}{2} 2^{-3} \cdot 2 + \binom{3}{3} 2^{-3} \cdot 3$$

$$= \frac{0 + 3 \cdot 1 + 3 \cdot 2 + 1 \cdot 3}{8} = \frac{3}{2}$$

This may be an intuitively appealing answer, if we imagine that we get an *average* of 1/2 head per flip in 3 flips.

But notice that the *definition* hands us an expression whose value is not obviously the answer what we expect, though it turns out to be so.

Sums and products of random variables

The **sum random variable** $X + Y$ made from two random variables X, Y defined on the *same* probability space Ω is defined, reasonably enough, to be the function whose values are the sum of the values of X and Y. That is, for $\omega \in \Omega$

$$(X + Y)(\omega) = X(\omega) + Y(\omega)$$

Similarly, the **product random variable** $X \cdot Y$ is

$$(X \cdot Y)(\omega) = X(\omega) \cdot Y(\omega)$$

The basic theorem on expected values

Our intuition about certain examples (like flipping a coin several times) is justified by the basic theorem about expected values:

Theorem: Let $X_1, \ldots, X_n$ be random variables on a common probability space Ω. Then

$$E(X_1 + \ldots + X_n) = E(X_1) + \ldots + E(X_n)$$

That is, the expected-value function E is **additive** (or *linear*).

Most functions do not have the additive property, though naive presumption of additivity (or linearity) is common. For example, despite many errors by novices, *generally*

$$\sin(a+b) \neq \sin a + \sin b$$
$$\sqrt{a+b} \neq \sqrt{a} + \sqrt{b}$$
$$(a+b)^2 \neq a^2 + b^2$$

For example, to compute the expected number of Hs in 10 flips of a fair coin, let X be the random variable on the probability space of all possible outcomes of 10 flips. The *definition* of expected value of X is what we want, namely

$$\text{expected no. Hs in 10 flips} = E(X)$$

$$= \sum_{k=0}^{10} P(X=k) \cdot k = \sum_{k=0}^{10} \binom{10}{k} \cdot 2^{-10} \cdot k$$

$$[1 \cdot 0 + 10 \cdot 1 + 45 \cdot 2 + 120 \cdot 3$$

$$+ 210 \cdot 4 + 252 \cdot 5 + 210 \cdot 6 + 120 \cdot 7$$

$$+ 45 \cdot 8 + 10 \cdot 9 + 1 \cdot 10] \, /1024$$

$$= (amazingly)5$$

It is completely *not* obvious that this big computation will yield the intuitively suggested answer

$$10 \cdot \frac{1}{2} = 5 \text{ expected Hs in 10 flips}$$

Invocation of the *Theorem* allows us to legitimize our intuition here. Define random variables $X_1, \ldots, X_{10}$ by

$$X_i = \text{ no. Hs on the } i^{\text{th}} \text{ flip of 10}$$

Note that these are all defined on the same probability space. Then

$$X = X_1 + \ldots + X_{10}$$

By the theorem,

$$E(X) = E(X_1) + \ldots + E(X_{10})$$

We evaluate each $E(X_i)$ via the definition

$$E(X_i) = \sum \text{values } k \, P(X_i = k) \cdot k$$

$$= P(X_i = 0) \cdot 0 + P(X_i = 1) \cdot 1$$

Since the flips are independent and the coin is fair, for any index i the probability that H appears on the i^{th} flip is $1/2$, so this is

$$E(X_i) = \frac{1}{2} \cdot 0 + \frac{1}{2} \cdot 1 = \frac{1}{2}$$

Then

$$E(X) = E(X_1) + \ldots + E(X_{10})$$

$$= \underbrace{\frac{1}{2} + \ldots + \frac{1}{2}}_{10} = 10 \cdot \frac{1}{2} = 5$$

It bears repeating that this is *not* the definition of expected value, is *not* obviously correct. Happily, it *is* intuitively correct and in the end our intuition (in this case) is vindicated by the Theorem.

Beware, though, that not all functions are additive or linear.

Evaluation by generating functions

But, even though it turns out that we do not need it in the above example, we might also want to be able to evaluate expressions such as

$$\sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \cdot k$$

directly. This is possible, and the methodology has many applications.

Recall the **Binomial Theorem**

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k \, y^{n-k}$$

Partial differentiation with respect to x gives

$$n(x+y)^{n-1} = \sum_{k=0}^{n} \binom{n}{k} k x^{k-1} \, y^{n-k}$$

Anticipating that we'll let $x = p$ and $y = 1 - p$ eventually, we see we're missing a factor of x on the right in that equality

$$n(x+y)^{n-1} = \sum_{k=0}^{n} \binom{n}{k} k x^{k-1} y^{n-k}$$

so multiply through by x:

$$nx(x+y)^{n-1} = \sum_{k=0}^{n} \binom{n}{k} k x^{k} y^{n-k}$$

Letting $x = p$ and $y = 1 - p$ gives

$$p \cdot n = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} k$$

In the simple case $p = 1/2$ we get the same conclusion as we got earlier via the Theorem.

Just fooling around, differentiating

$$n(x+y)^{n-1} = \sum_{k=0}^{n} \binom{n}{k} k x^{k-1} y^{n-k}$$

once more gives

$$n(n-1)(x+y)^{n-2} = \sum_{k=0}^{n} \binom{n}{k} k(k-1) x^{k-2} y^{n-k}$$

and letting $x = y = 1$ gives

$$n(n-1)2^{n-2} = \sum_{k=0}^{n} \binom{n}{k} k(k-1)$$

As a variation, if we multiply through

$$n(x+y)^{n-1} = \sum_{k=0}^{n} \binom{n}{k} k x^{k-1} y^{n-k}$$

by x before differentiating again, we get

$$\frac{d}{dx}\left(x\, n(x+y)^{n-1}\right)$$

$$= \frac{d}{dx} \sum_{k=0}^{n} \binom{n}{k} k x^{k} y^{n-k}$$

$$= \sum_{k=0}^{n} \binom{n}{k} k^2 x^{k-1} y^{n-k}$$

and letting $x = y = 1$ again

$$n(n+1)2^{n-2} = \sum_{k=0}^{n} \binom{n}{k} k^2$$

As another example, consider the problem of *how long we should expect to wait in flipping a fair coin until we get an H.*

That is, let X be the random variable which counts the number of flips up to and including the first flip which gives a H. Then

$$E(X) = \sum_{k=0}^{\infty} P(X = k) \cdot k$$

$$= P(\mathtt{H}) \cdot 1 + P(\mathtt{TH}) \cdot 2 + P(\mathtt{TTH}) \cdot 3$$

$$+ P(\mathtt{TTTH}) \cdot 4 + P(\mathtt{TTTTH}) \cdot 5 + \ldots$$

$$= P(\mathtt{H}) \cdot 1 + P(\mathtt{T})P(\mathtt{H}) \cdot 2 + P(\mathtt{T})^2 P(\mathtt{H}) \cdot 3$$

$$+ P(\mathtt{T})^3 P(\mathtt{H}) \cdot 4 + P(\mathtt{T})^4 P(\mathtt{H}) \cdot 5 + \ldots$$

by *independence* of flips. Without even thinking about the fairness, let $P(\mathtt{H}) = p$ and $P(\mathtt{T}) = q$, where $p + q = 1$.

Then we're wanting to evaluate

$$\sum_{k=0}^{\infty} p\, q^{k-1} \cdot k$$

The infinite series we know how to evaluate is the **geometric series**

$$\sum_{k=0}^{\infty} q^k = \frac{1}{1-q}$$

for $|q| < 1$. Differentiating both sides of this with respect to q gives

$$\sum_{k=0}^{\infty} q^{k-1} \cdot k = \frac{1}{(1-q)^2}$$

This is missing a factor of p, so multiply both sides by pq and using $p + q = 1$

$$\sum_{k=0}^{\infty} pq^{k-1} \cdot k = \frac{p}{(1-q)^2} = \frac{1}{p}$$

In the identity

$$\sum_{k=0}^{\infty} pq^{k-1} \cdot k = \frac{1}{p}$$

let $p = q = \frac{1}{2}$ to obtain

expected flips of fair coin to get a H

$$= \sum_{k=0}^{\infty} \frac{1}{2} \left(\frac{1}{2}\right)^{k-1} \cdot k = \frac{1}{1/2} 2$$

This might suggest that we should *expect* to get a `H` on the second flip, so that we get a `T` on the first flip? But the same discussion would say that the expected number of flips to get a `T` is also 2.

No, it's just that *we should not expect to get the expected value*, since it's just an average.

gcd's and lcm's

An integer d **divides** an integer n if $n \,\%\, d \;=\; 0$. And in that situation n is a **multiple** of d. The notation is

$$d|n$$

In this notation the line is *vertical*, not slanted. For example

$$5|10 \qquad 35|105 \qquad 2 \not| 5$$

where the last illustrates the slash to denote *does not divide*.

Thus, to say d divides n is to say in more colloquial terms that d divides n *evenly*, but in mathematics that qualification is always implied.

A **proper divisor** d of n is a divisor of n in the range

$$1 < d < n$$

The **greatest common divisor** $\gcd(x, y)$ of two integers x, y is the largest positive integer d which divides both x, y, that is, $d|x$ and $d|y$. For example,

$$\gcd(3, 5) = 1 \quad \gcd(12, 18) = 6$$

$$\gcd(49, 56) = 7 \quad \gcd(105, 49) = 7$$

The **least common multiple** $\text{lcm}(x, y)$ of two integers is the smallest positive integer m which is a multiple of both x, y. For example,

$$\text{lcm}(3, 5) = 15 \quad \text{lcm}(12, 18) = 36$$

$$\text{lcm}(49, 56) = 392 \quad \text{lcm}(105, 49) = 735$$

Especially with the larger numbers, we should admit that we cannot claim to directly intuit the answer. *We want a systematic procedure.*

The intuitively fairly obvious approach to computing *lcm*s and *gcd*s uses *prime factorization*.

We grant for now the **unique factorization of integers into primes**, meaning that for a given positive integer n there is an expression

$$n = p_1^{e_1} p_2^{e_2} \ \ldots \ p_t^{e_t}$$

where the p_i are *distinct* primes, and the exponents e_i are positive integers.

For example,

$$\begin{aligned} 6 &= 2 \cdot 3 \\ 8 &= 2^3 \\ 12 &= 2^2 \cdot 3 \\ 18 &= 2 \cdot 3^2 \\ 96 &= 2^4 \cdot 3 \\ 735 &= 3 \cdot 5 \cdot 7^2 \\ 10205150 &= 2 \cdot 5^2 \cdot 53 \cdot 3851 \end{aligned}$$

If we have the prime factorization of both x and y, then

The prime factorization of $\gcd(x, y)$
has prime factors that occur in *both* factorizations, with corresponding exponents equal to the *minimum* of the exponents in the two.

The prime factorization of $\text{lcm}(x, y)$
has prime factors that occur in *either* factorization, with coresponding exponents equal to the *maximum* of the exponents in the two.

For example, with

$$x = 1001 = 7 \cdot 11 \cdot 13$$

$$y = 735 = 3 \cdot 5 \cdot 7^2$$

we have

$$\gcd(1001, 735) =$$

$$= 3^{\min(0,1)}\, 5^{\min(0,1)}\, 7^{\min(1,2)}\, 13^{\min(0,1)}$$

$$= 3^0\, 5^0\, 7^1\, 13^0 = 7$$

And still with

$$x = 1001 = 7 \cdot 11 \cdot 13$$

$$y = 735 = 3 \cdot 5 \cdot 7^2$$

we have

$$\mathrm{lcm}(1001, 735) =$$

$$= 3^{\max(0,1)}\, 5^{\max(0,1)}\, 7^{\max(1,2)}\, 13^{\max(0,1)}$$

$$= 3^1\, 5^1\, 7^2\, 13^1 = 9555$$

This approach is acceptable for relatively small number, or in any case if we have *prime factorizations* or can obtain them readily.

Trial division

The basic method to obtain the factorization of smallish integers into primes is **trial division**.

This is basically a brute force search for proper divisors, but knowing when we can stop. Note that, if $d < N$ and $d|N$ and $d > \sqrt{N}$, then $\frac{N}{d}$ is *also* a divisor of N and $1 < \frac{N}{d} \leq \sqrt{N}$. Thus, in looking for *proper* divisors it suffices to stop looking at $\sqrt{N}$ if we haven't found any by that point!

Recall that N is **prime** if N has no proper divisor and if $N > 1$. That is, N is prime if there is no $d|N$ with $1 < d < N$ *and* $N > 1$. (It is a good convention that 1 is *not* prime.)

Non-prime numbers bigger than 1 are called **composite**. The number 1 is neither prime nor composite, evidently.

Thus, for example, to test whether N is *prime*

Compute $N \,\%\, 2$
If $N \,\%\, 2 = 0$, stop, N composite
Else if $N \,\%\, 2 \neq 0$, continue
Initialize $d = 3$.
While $d \leq \sqrt{N}$:
    Compute $N \,\%\, d$
    If $N \,\%\, d = 0$, **stop**, N composite
    Else if $N \,\%\, d \neq 0$,
        Replace d by $d + 2$, continue
If reach $d > \sqrt{N}$ without termination,
    N is prime

This takes at worst $\sqrt{N}/2$ steps to confirm or deny the primality of N.

For example, to test $N = 53$ for primality:

Compute $53 \,\%\, 2 = 1$
Since $53 \,\%\, 2 \neq 0$, continue
Initialize $d = 3$.
While $d \leq \sqrt{53}$:
    Compute $53 \,\%\, d$
    Compute $53 \,\%\, 3 = 2$
    Since $53 \,\%\, 3 \neq 0$,
        replace $d = 3$ by $d + 2 = 5$, continue
    Still $d = 5 \leq \sqrt{53}$, so continue
    Compute $53 \,\%\, 5 = 3$
    Since $53 \,\%\, 5 \neq 0$,
        replace $d = 5$ by $d + 2 = 7$, continue
    Still $d = 7 \leq \sqrt{53}$, so continue
    Compute $53 \,\%\, 7 = 4$
    Since $53 \,\%\, 7 \neq 0$,
        replace $d = 7$ by $d + 2 = 9$, continue
But $9 > \sqrt{53}$, so
    53 is prime

This approach is infeasible for integers $\sim 10^{30}$ and larger.

To achieve **factorization into primes** of an integer N:

Initialize $n = N$
While $2|n$, add 2 to list of prime factors
    and replace n by $n/2$
Initialize $d = 3$
While $d \leq \sqrt{n}$:
    While $d|n$, add d to list
        and replace n by n/d
    When d does not divide n
        replace d by $d + 2$
When $d > \sqrt{n}$
    If $n = 1$ the list of prime factors
        of the original N is complete
    If $n > 1$ then add n to the list

Note that the nature of the process assures that the ds obtained are primes.

For example, to factor 24750

Initialize $n = 24750$
$2|n$, so
    put 2 on the list (just (2) so far)
    replace n by $n = 24750/2 = 12375$
Now 2 does not divide $n = 12375$
Initialize $d = 3$
$3|12375$, so
    put 3 on the list (now $(2, 3)$)
    replace n by $n = 12375/3 = 4125$
$3|4125$, so
    put 3 on the list (now $(2, 3, 3)$)
    replace n by $n = 4125/3 = 1375$
Now 3 does not divide $n = 1375$, so
    replace $d = 3$ by $d = 3 + 2 = 5$

...

$5|1375$, so

put 5 on the list (now $(2, 3, 3, 5)$)

replace n by $n = 1375/5 = 275$

$5|275$, so

put 5 on the list (now $(2, 3, 3, 5, 5)$)

replace n by $n = 275/5 = 55$

$5|55$, so

put 5 on the list (now $(2, 3, 3, 5, 5, 5)$)

replace n by $n = 55/5 = 11$

Now 5 does not divide 11, so

replace $d = 5$ by $d = 5 + 2 = 7$

Now $11 \geq \sqrt{11}$, so 11 is prime

The product of prime factors, counting how many times they appear, is

$$24750 = 2 \cdot 3 \cdot 3 \cdot 5 \cdot 5 \cdot 5 \cdot 11$$

The above process is **trial division**.