six times that of drawing a Z. We need to assign probabilities that are proportional to these likelihoods. We must have some number c such that, for each letter, the probability of drawing that letter should be c times the number of copies of that letter.

$$\Pr[\text{drawing letter X}] = c \cdot \text{number of copies of letter X}$$

Summing over all letters, we get

$$1 = c \cdot \text{total number of tiles}$$

Since the total number of tiles is 95, we define $c = 1/95$. The probability of drawing an E is therefore 12/95, which is about .126. The probability of drawing an A is 9/95, and so on. In Python, the probability distribution is

```
{'A':9/95, 'B':2/95, 'C':2/95, 'D':4/95, 'E':12/95, 'F':2/95,
          'G':3/95, 'H':2/95, 'I':9/95, 'J':1/95, 'K':1/95,  'L':1/95,
          'M':2/95, 'N':6/95, 'O':8/95, 'P':2/95, 'Q':1/95,  'R':6/95,
          'S':4/95, 'T':6/95, 'U':4/95, 'V':2/95, 'W':2/95,  'X':1/95,
          'Y':2/95, 'Z':1/95}
```

0.4.2 Events, and adding probabilities

In Example 0.4.4 (Page 10), what is the probability of drawing a *vowel* from the bag?

A set of outcomes is called an *event*. For example, the event of drawing a vowel is represented by the set $\{A, E, I, O, U\}$.

Principle 0.4.5 (Fundamental Principle of Probability Theory): The probability of an event is the sum of probabilities of the outcomes making up the event.

According to this principle, the probability of a vowel is

$$9/95 + 12/95 + 9/95 + 8/95 + 4/95$$

which is 42/95.

0.4.3 Applying a function to a random input

Now we think about applying a function to a random input. Since the input to the function is random, the output should also be considered random. Given the probability distribution of the input and a specification of the function, we can use probability theory to derive the probability distribution of the output.

Example 0.4.6: Define the function $f : \{1,2,3,4,5,6\} \longrightarrow \{0,1\}$ by

$$f(x) = \begin{cases} 0 & \text{if } x \text{ is even} \\ 1 & \text{if } x \text{ is odd} \end{cases}$$

Consider the experiment in which we roll a single die (as in Example 0.4.2 (Page 10)), yielding one of the numbers in $\{1,2,3,4,5,6\}$, and then we apply $f(\cdot)$ to that number, yielding either a 0 or a 1. What is the probability function for the outcome of this experiment?

The outcome of the experiment is 0 if the rolled die shows 2, 4, or 6. As discussed in Example 0.4.2 (Page 10), each of these possibilies has probability 1/6. By the Fundamental Principle of Probability Theory, therefore, the output of the function is 0 with probability $1/6 + 1/6 + 1/6$, which is 1/2. Similarly, the output of the function is 1 with probability 1/2. Thus the probability distribution of the output of the function is {0: 1/2., 1:1/2.}.

Quiz 0.4.7: Consider the flipping of a penny and a nickel, described in Example 0.4.3 (Page 10). The outcome is a pair (x, y) where each of x and y is `'H'` or `'T'` (heads or tails). Define the function

$$f : \{('H', 'H')\ ('H', 'T'),\ ('T','H'),\ ('T','T')\}$$

by

$f((x, y))$ = the number of `H`'s represented

Give the probability distribution for the output of the function.

Answer

`{0: 1/4., 1:1/2., 2:1/4.}`

Example 0.4.8 (Caesar plays Scrabble): Recall that the function f defined in Example 0.3.11 (Page 5) maps A to 0, B to 1, and so on. Consider the experiment in which f is applied to a letter selected randomly according to the probability distribution described in Example 0.4.4 (Page 10). What is the probability distribution of the output?

Because f is an invertible function, there is one and only one input for which the output is 0, namely A. Thus the probability of the output being 0 is exactly the same as the probability of the input being A, namely 9/95.. Similarly, for each of the integers 0 through 25 comprising the co-domain of f, there is exactly one letter that maps to that integer, so the probability of that integer equals the probability of that letter. The probability distribution is thus

```
{0:9/95., 1:2/95., 2:2/95., 3:4/95., 4:12/95., 5:2/95.,
          6:3/95., 7:2/95., 8:9/95., 9:1/95., 10:1/95.,  11:1/95.,
          12:2/95., 13:6/95., 14:8/95., 15:2/95., 16:1/95.,  17:6/95.,
          18:4/95., 19:6/95., 20:4/95., 21:2/95., 22:2/95.,  23:1/95.,
          24:2/95., 25:1/95.}
```

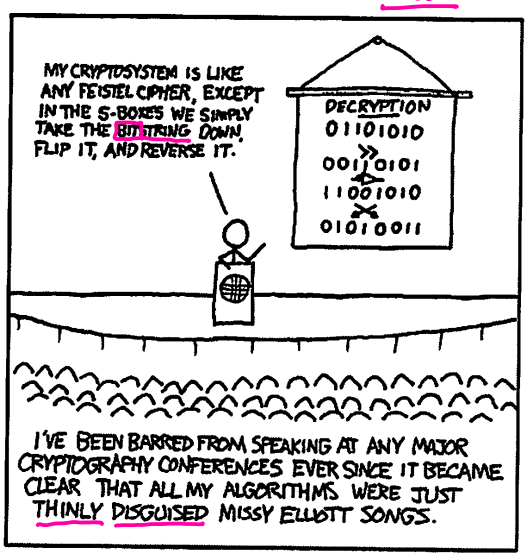
The previous example illustrates that, if the function is invertible, the probabilities are preserved: the probabilities of the various outputs match the probabilities of the inputs. It follows that, if the input is chosen according to a uniform distribution, the distribution of the output is also uniform.

Example 0.4.9: In Caesar's Cyphersystem, one encrypts a letter by advancing it three positions. Of course, the number k of positions by which to advance need not be three; it can be any integer from 0 to 25. We refer to k as the *key*. Suppose we select the key k according to the uniform distribution on $\{0, 1, \ldots, 25\}$, and use it to encrypt the letter P. Let $w : \{0, 1, \ldots, 25\} \longrightarrow \{A, B, \ldots, Z\}$ be the the function mapping the key to the cyphertext:

$$\begin{aligned} w(k) &= h(f(P) + k \bmod 26) \\ &= h(15 + k \bmod 26) \end{aligned}$$

The function $w(\cdot)$ is invertible. The input is chosen according to the uniform distribution, so the distribution of the output is also uniform. Thus when the key is chosen randomly, the cyphertext is equally likely to be any of the twenty-six letters.

0.4.4 Perfect secrecy



Cryptography (`http://xkcd.com/153/`)

We apply the idea of Example 0.4.9 (Page 12) to some even simpler cryptosystems. A cryptosystem must satisfy two obvious requirements:

- the intended recipient of an encrypted message must be able to decrypt it, and
- someone for whom the message was not intended should *not* be able to decrypt it.

The first requirement is straightforward. As for the second, we must dispense with a misconception about security of cryptosystems. The idea that one can keep information secure by not revealing the *method* by which it was secured is often called, disparagingly, *security through obscurity*. This approach was critiqued in 1881 by a professor of German, Jean-Guillame-Hubert-Victor-François-Alexandre-August Kerckhoffs von Niewenhof, known as August Kerckhoffs. The *Kerckhoffs Doctrine* is that *the security of a cryptosystem should depend only on the secrecy of the key used, not on the secrecy of the system itself.*

There is an encryption method that meets Kerchoffs' stringent requirement. It is utterly unbreakable if used correctly.[1] Suppose Alice and Bob work for the British military. Bob is the commander of some troops stationed in Boston harbor. Alice is the admiral, stationed several miles away. At a certain moment, Alice must convey a one-bit message p (the plaintext) to Bob: whether to attack by land or by sea (0=land, 1=sea). Their plan, agreed upon in advance, is that Alice will encrypt the message, obtaining a one-bit *cyphertext* c, and send the cyphertext c to Bob by hanging one or two lanterns (say, one lantern = 0, two lanterns = 1). They are aware that the fate of a colony might depend on the secrecy of their communication. (As it happens, a rebel, Eve, knows of the plan and will be observing.)

Let's go back in time. Alice and Bob are consulting with their cryptography expert, who suggests the following scheme:

[1] For an historically significant occurence of the former Soviet Union failing to use it correctly, look up VENONA.

Bad Scheme: Alice and Bob randomly choose k from $\{\clubsuit, \heartsuit, \spadesuit\}$ according to the uniform probability function $(\text{pr}(\clubsuit) = 1/3, \text{pr}(\heartsuit) = 1/3, \text{pr}(\spadesuit) = 1/3)$. Alice and Bob must both know k but must keep it secret. It is the *key*.
When it is time for Alice to use the key to encrypt her plaintext message p, obtaining the cyphertext c, she refers to the following table:

p	k	c
0	♣	0
0	♡	1
0	♠	1
1	♣	1
1	♡	0
1	♠	0

The good news is that this cryptosystem satisfies the first requirement of cryptosystems: it will enable Bob, who knows the key k and receives the cyphertext c, to determine the plaintext p. No two rows of the table have the same k-value and c-value.

The bad news is that this scheme leaks information to Eve. Suppose the message turns out to be 0. In this case, $c = 0$ if $k = \clubsuit$ (which happens with probability 1/3), and $c = 1$ if $k = \heartsuit$ or $k = \spadesuit$ (which, by the Fundamental Principle of Probability Theory, happens with probability 2/3). Thus in this case $c = 1$ is twice as likely as $c = 0$. Now suppose the message turns out to be 1. In this case, a similar analysis shows that $c = 0$ is twice as likely as $c = 1$.

Therefore, when Eve sees the cyphertext c, she learns something about the plaintext p. Learning c doesn't allow Eve to determine the value of p with certainty, but she can revise her estimate of the chance that $p = 0$. For example, suppose that, before seeing c, Eve believed $p = 0$ and $p = 1$ were equally likely. If she sees $c = 1$ then she can infer that $p = 0$ is twice as likely as $p = 1$. The exact calculation depends on Bayes' Rule, which is beyond the scope of this analysis but is quite simple.

Confronted with this argument, the cryptographer changes the scheme simply by removing ♠ as a possible value for p.

Good Scheme: Alice and Bob randomly choose k from $\{\clubsuit, \heartsuit\}$ according to the uniform probability function $(\text{pr}(\clubsuit) = 1/2, \text{pr}(\heartsuit) = 1/2)$
When it is time for Alice to encrypt her plaintext message p, obtaining the cyphertext c, she uses the following table:

p	k	c
0	♣	0
0	♡	1
1	♣	1
1	♡	0

0.4.5 Perfect secrecy and invertible functions

Consider the functions

$$f_0 : \{\clubsuit, \heartsuit\} \longrightarrow \{0, 1\}$$

and

$$f_1 : \{\clubsuit, \heartsuit\} \longrightarrow \{0, 1\}$$

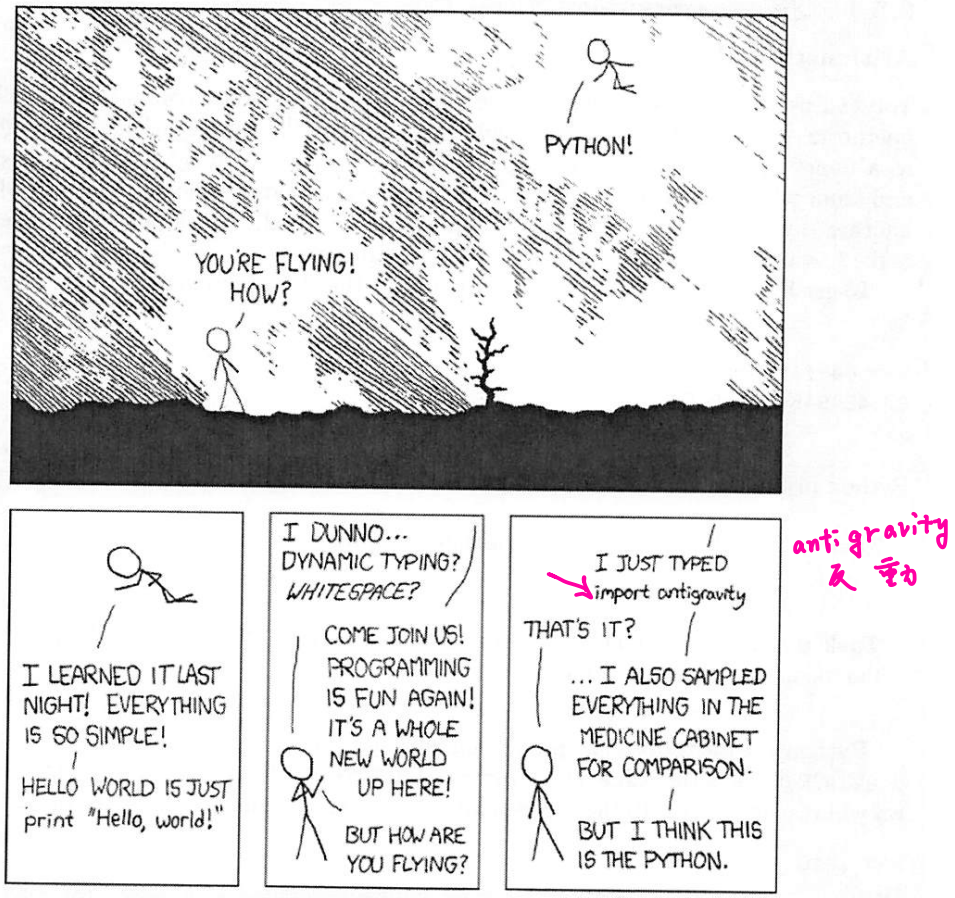
defined by

$$f_0(x) = \text{encryption of 0 when the key is } x$$

$$f_1(x) = \text{encryption of 1 when the key is } x$$

Each of these functions is invertible. Consequently, for each function, if the input x is chosen uniformly at random, the output will also be distributed according to the uniform distribution. This in turn means that the probability distribution of the output does not depend on whether 0 or 1 is being encrypted, so knowing the output gives Eve no information about which is being encrypted. We say the scheme achieves *perfect secrecy*.

0.5 *Lab: Introduction to Python—sets, lists, dictionaries, and comprehensions*



Python http://xkcd.com/353/

We will be writing all our code in Python (Version 3.x). In writing Python code, we emphasize the use of *comprehensions*, which allow one to express computations over the elements of a set, list, or dictionary without a traditional for-loop. Use of comprehensions leads to more compact and more readable code, code that more clearly expresses the mathematical idea behind the computation being expressed. Comprehensions might be new to even some readers who are familiar with Python, and we encourage those readers to at least skim the material on this topic.

To start Python, simply open a console (also called a shell or a terminal or, under Windows, a "Command Prompt" or "MS-DOS Prompt"), and type `python3` (or perhaps just `python`) to the console (or shell or terminal or Command Prompt) and hit the `Enter` key. After a few lines telling you what version you are using (e.g., Python 3.4.1), you should see `>>>` followed by a space. This is the *prompt*; it indicates that Python is waiting for you to type something. When you type an expression and hit the `Enter` key, Python evaluates the expression and prints the result, and then prints another prompt. To get out of this environment, type `quit()` and `Enter`, or `Control-D`. To interrupt Python when it is running too long, type `Control-C`.

This environment is sometimes called a *REPL*, an acronym for "read-eval-print loop." It reads what you type, evaluates it, and prints the result if any. In this assignment, you will interact with Python primarily through the REPL. In each task, you are asked to come up with an expression of a certain form.

There are two other ways to run Python code. You can import a *module* from within the REPL, and you can run a Python script from the command line (outside the REPL). We will discuss modules and importing in the next lab assignment. This will be an important part of your interaction with Python.

0.5.1 *Simple expressions*

Arithmetic and numbers

You can use Python as a calculator for carrying out arithmetic computations. The binary operators +, *, -, / work as you would expect. To take the negative of a number, use - as a unary operator (as in -9). Exponentiation is represented by the binary operator **, and truncating integer division is //. Finding the remainder when one integer is divided by another (modulo) is done using the % operator. As usual, ** has precedence over * and / and //, which have precedence over + and -, and parentheses can be used for grouping.

To get Python to carry out a calculation, type the expression and press the Enter/Return key:

```
>>> 44+11*4-6/11.
87.454545454545454
>>>
```

Python prints the answer and then prints the prompt again.

Task 0.5.1: Use Python to find the number of minutes in a week.

Task 0.5.2: Use Python to find the remainder of 2304811 divided by 47 without using the modulo operator %. (Hint: Use //.)

Python uses a traditional programming notation for scientific notation. The notation `6.022e23` denotes the value 6.02×10^{23}, and `6.626e-34` denotes the value 6.626×10^{-34}. As we will discover, since Python uses limited-precision arithmetic, there are round-off errors:

```
>>> 1e16 + 1
1e16
```

Strings

A string is a series of characters that starts and ends with a single-quote mark. Enter a string, and Python will repeat it back to you:

```
>>> 'This sentence is false.'
'This sentence is false.'
```

You can also use double-quote marks; this is useful if your string itself contains a single-quote mark:

```
>>> "So's this one."
"So's this one."
```

Python is doing what it usually does: it *evaluates* (finds the value of) the expression it is given and prints the value. The value of a string is just the string itself.

Comparisons and conditions and Booleans

You can compare values (strings and numbers, for example) using the operators ==, < , >, <=, >=, and !=. (The operator != is inequality.)

```
>>> 5 == 4
False
>>> 4 == 4
True
```

The value of such a comparison is a Boolean value (True or False). An expression whose value is a boolean is called a Boolean expression.

Boolean operators such as and and or and not can be used to form more complicated Boolean expressions.

```
>> True and False
False
>>> True and not (5 == 4)
True
```

Task 0.5.3: Enter a Boolean expression to test whether the sum of 673 and 909 is divisible by 3.

0.5.2 *Assignment statements*

The following is a *statement*, not an expression. Python executes it but produces neither an error message nor a value.

```
>>> mynum = 4+1
```

The result is that henceforth the variable mynum is bound to the value 5. Consequently, when Python evaluates the expression consisting solely of mynum, the resulting value is 5. We say therefore that the value of mynum is 5.

A bit of terminology: the variable being assigned to is called the *left-hand side* of an assignment, and the expression whose value is assigned is called the *right-hand side*.

A variable name must start with a letter and must exclude certain special symbols such as the dot (period). The underscore _ is allowed in a variable name. A variable can be bound to a value of any type. You can rebind mynum to a string:

```
>>> mynum = 'Brown'
```

This binding lasts until you assign some other value to mynum or until you end your Python session. It is called a *top-level* binding. We will encounter cases of binding variables to values where the bindings are temporary.

It is important to remember (and second nature to most experienced programmers) that an assignment statement binds a variable to the *value* of an expression, not to the expression itself. Python first evaluates the right-hand side and only then assigns the resulting value to the left-hand side. This is the behavior of most programming languages.

Consider the following assignments.

```
>>> x = 5+4
>>> y = 2 * x
>>> y
18
>>> x = 12
>>> y
18
```

In the second assignment, y is assigned the value of the expression 2 * x. The value of that expression is 9, so y is bound to 18. In the third assignment, x is bound to 12. This does not change the fact that y is bound to 18.

0.5.3 *Conditional expressions*

There is a syntax for conditional expressions:

⟨*expression*⟩ `if` ⟨*condition*⟩ `else` ⟨*expression*⟩

The *condition* should be a Boolean expression. Python evaluates the condition; depending on whether it is `True` or `False`, Python then evaluates either the first or second *expression*, and uses the result as the result of the entire conditional expression.

For example, the value of the expression `x if x>0 else -x` is the absolute value of `x`.

Task 0.5.4: Assign the value -9 to `x` and 1/2 to `y`. Predict the value of the following expression, then enter it to check your prediction:

```
2**(y+1/2) if x+10<0 else 2**(y-1/2)
```

0.5.4 *Sets*

Python provides some simple data structures for grouping together multiple values, and integrates them with the rest of the language. These data structures are called *collections*. We start with sets.

A set is an unordered collection in which each value occurs at most once. You can use curly braces to give an expression whose value is a set. Python prints sets using curly braces.

```
>>> {1+2, 3, "a"}
{'a', 3}
>>> {2, 1, 3}
{1, 2, 3}
```

Note that duplicates are eliminated and that the order in which the elements of the output are printed does not necessarily match the order of the input elements.

The *cardinality* of a set S is the number of elements in the set. In Mathese we write $|S|$ for the cardinality of set S. In Python, the cardinality of a set is obtained using the procedure `len(·)`.

```
>>> len({'a', 'b', 'c', 'a', 'a'})
3
```

Summing

The sum of elements of collection of values is obtained using the procedure `sum(·)`.

```
>>> sum({1,2,3})
6
```

If for some reason (we'll see one later) you want to start the sum not at zero but at some other value, supply that value as a second argument to `sum(·)`:

```
>>> sum({1,2,3}, 10)
16
```

Testing set membership

Membership in a set can be tested using the `in` operator and the `not in` operator. If S is a set, x `in` S is a Boolean expression that evaluates to `True` if the value of x is a member of the set S, and `False` otherwise. The value of a `not in` expression is just the opposite

```
>>> S={1,2,3}
>>> 2 in S
True
>>> 4 in S
False
>>> 4 not in S
True
```

Set union and intersection

The *union* of two sets S and T is a new set that contains every value that is a member of S or a member of T (or both). Python uses the vertical bar | as the *union* operator:

```
>>> {1,2,3} | {2,3,4}
{1, 2, 3, 4}
```

The *intersection* of S and T is a new set that contains every value that is a member of both S and T. Python uses the ampersand & as the *intersection* operator:

```
>>> {1,2,3} & {2,3,4}
{2, 3}
```

Mutating a set

A value that can be altered is a *mutable* value. Sets are mutable; elements can be added and removed using the `add` and `remove` methods:

```
>>> S={1,2,3}
>>> S.add(4)
>>> S.remove(2)
>>> S
{1, 3, 4}
```

The syntax using the dot should be familiar to students of object-oriented programming languages such as Java and C++. The operations `add(·)` and `remove(·)` are *methods*. You can think of a method as a procedure that takes an extra argument, the value of the expression to the left of the dot.

Python provides a method `update(...)` to add to a set all the elements of another collection (e.g. a set or a list):

```
>>> S.update({4, 5, 6})
>>> S
{1, 3, 4, 5, 6}
```

Similarly, one can intersect a set with another collection, removing from the set all elements not in the other collection:

```
>>> S.intersection_update({5,6,7,8,9})
>>> S
{5, 6}
```

Suppose two variables are bound to the same value. A mutation to the value made through one variable is seen by the other variable.

```
>>> T=S
>>> T.remove(5)
>>> S
{6}
```

This behavior reflects the fact that Python stores only one copy of the underlying data structure. After Python executes the assignment statement T=S, both T and S point to the same data structure. This aspect of Python will be important to us: many different variables can point to the same huge set without causing a blow-up of storage requirements.

Python provides a method for copying a collection such as a set:

```
>>> U=S.copy()
>>> U.add(5)
>>> S
{6}
```

The assignment statement binds U not to the value of S but to a copy of that value, so mutations to the value of U don't affect the value of S.

Set comprehensions

Python provides for expressions called *comprehensions* that let you build collections out of other collections. We will be using comprehensions a lot because they are useful in constructing an expression whose value is a collection, and they mimic traditional mathematical notation. Here's an example:

```
>>> {2*x for x in {1,2,3}}
{2, 4, 6}
```

This is said to be a set comprehension *over* the set {1,2,3}. It is called a set comprehension because its value is a set. The notation is similar to the traditional mathematical notation for expressing sets in terms of other sets, in this case $\{2x : x \in \{1,2,3\}\}$. To compute the value, Python iterates over the elements of the set {1,2,3}, temporarily binding the control variable x to each element in turn and evaluating the expression 2*x in the context of that binding. Each of the values obtained is an element of the final set. (The bindings of x during the evaluation of the comprehension do not persist after the evaluation completes.)

Task 0.5.5: Write a comprehension over $\{1,2,3,4,5\}$ whose value is the set consisting of the squares of the first five positive integers.

Task 0.5.6: Write a comprehension over $\{0,1,2,3,4\}$ whose value is the set consisting of the first five powers of two, starting with 2^0.

Using the union operator | or the intersection operator &, you can write set expressions for the union or intersection of two sets, and use such expressions in a comprehension:

```
>>> {x*x for x in S | {5, 7}}
{1, 25, 49, 9}
```

By adding the phrase `if` ⟨*condition*⟩ at the end of the comprehension (before the closing brace "}"), you can skip some of the values in the set being iterated over:

```
>>> {x*x for x in S | {5, 7}  if x > 2}
{9, 49, 25}
```

I call the conditional clause a *filter*.

You can write a comprehension that iterates over the Cartesian product of two sets:

```
>>>{x*y for x in {1,2,3} for y in {2,3,4}}
{2, 3, 4, 6, 8, 9, 12}
```

This comprehension constructs the set of the products of every combination of x and y. I call this a *double comprehension*.

Task 0.5.7: The value of the previous comprehension,

```
{x*y for x in {1,2,3} for y in {2,3,4}}
```

is a seven-element set. Replace `{1,2,3}` and `{2,3,4}` with two other three-element sets so that the value becomes a nine-element set.

Here is an example of a double comprehension with a filter:

```
>>> {x*y for x in {1,2,3} for y in {2,3,4} if x != y}
{2, 3, 4, 6, 8, 12}
```

Task 0.5.8: Replace `{1,2,3}` and `{2,3,4}` in the previous comprehension with two disjoint (i.e. non-overlapping) three-element sets so that the value becomes a five-element set.

Task 0.5.9: Assume that `S` and `T` are assigned sets. Without using the *intersection operator* `&`, write a comprehension over `S` whose value is the intersection of `S` and `T`. Hint: Use a membership test in a filter at the end of the comprehension.

Try out your comprehension with `S = {1,2,3,4}` and `T = {3,4,5,6}`.

Remarks

The empty set is represented by `set()`. You would think that `{}` would work but, as we will see, that notation is used for something else.

You cannot make a set that has a set as element. This has nothing to do with Cantor's Paradox—Python imposes the restriction that the elements of a set must not be mutable, and sets are mutable. The reason for this restriction will be clear to a student of data structures from the error message in the following example:

```
>>> {{1,2},3}
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
TypeError: unhashable type: 'set'
```

There is a nonmutable version of set called *frozenset*. Frozensets can be elements of sets. However, we won't be using them.

0.5.5 *Lists*

Python represents sequences of values using *lists*. In a list, order is significant and repeated elements are allowed. The notation for lists uses square brackets instead of curly braces. The empy list is represented by `[]`.

```
>>> [1,1+1,3,2,3]
[1, 2, 3, 2, 3]
```

There are no restrictions on the elements of lists. A list can contain a set or another list.

```
>>> [[1,1+1,4-1],{2*2,5,6}, "yo"]
[[1, 2, 3], {4, 5, 6}, 'yo']
```

However, a set cannot contain a list since lists are mutable.

The *length* of a list, obtained using the procedure `len(·)`, is the number of elements in the list, even though some of those elements may themselves be lists, and even though some elements might have the same value:

```
>>> len([[1,1+1,4-1],{2*2,5,6}, "yo", "yo"])
4
```

As we saw in the section on sets, the sum of elements of a collection can be computed using `sum(·)`

```
>>> sum([1,1,0,1,0,1,0])
4
>>> sum([1,1,0,1,0,1,0], -9)
-5
```

In the second example, the second argument to `sum(·)` is the value to start with.

Task 0.5.10: Write an expression whose value is the average of the elements of the list `[20, 10, 15, 75]`.

List concatenation

You can combine the elements in one list with the elements in another list to form a new list (without changing the original lists) using the `+` operator.

```
>>> [1,2,3]+["my", "word"]
[1, 2, 3, 'my', 'word']
>>> mylist = [4,8,12]
>>> mylist + ["my", "word"]
[4, 8, 12, 'my', 'word']
>>> mylist
[4, 8, 12]
```

You can use `sum(·)` on a collection of lists, obtaining the concatenation of all the lists, by providing `[]` as the second argument.

```
>>> sum([ [1,2,3], [4,5,6], [7,8,9] ])
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
TypeError: unsupported operand type(s) for +: 'int' and 'list'
>>> sum([ [1,2,3], [4,5,6], [7,8,9] ], [])
[1, 2, 3, 4, 5, 6, 7, 8, 9]
```

List comprehensions

Next we discuss how to write a list comprehension (a comprehension whose value is a list). In the following example, a list is constructed by iterating over the elements in a set.

```
>>> [2*x for x in {2,1,3,4,5} ]
[2, 4, 6, 8, 10]
```

Note that the order of elements in the resulting list might not correspond to the order of elements in the set since the latter order is not significant.

You can also use a comprehension that constructs a list by iterating over the elements in a list:

```
>>> [ 2*x for x in [2,1,3,4,5] ]
[4, 2, 6, 8, 10]
```

Note that the list `[2,1,3,4,5]` specifies the order among its elements. In evaluating the comprehension Python iterates through them in that order. Therefore the order of elements in the resulting list corresponds to the order in the list iterated over.

You can also write list comprehensions that iterate over multiple collections using two control variables. As I mentioned in the context of sets, I call these "double comprehensions". Here is an example of a list comprehension over two lists.

```
>>> [ x*y for x in [1,2,3] for y in [10,20,30] ]
[10, 20, 30, 20, 40, 60, 30, 60, 90]
```

The resulting list has an element for every combination of an element of `[1,2,3]` with an element of `[10,20,30]`.

We can use a comprehension over two sets to form the Cartesian product.

Task 0.5.11: Write a double list comprehension over the lists `['A','B','C']` and `[1,2,3]` whose value is the list of all possible two-element lists [letter, number]. That is, the value is

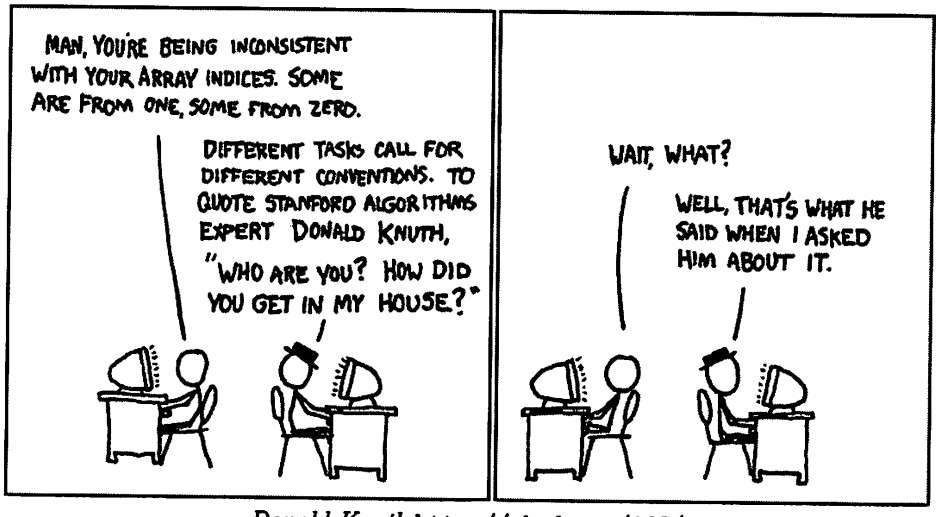
```
[['A', 1], ['A', 2], ['A', 3], ['B', 1], ['B', 2],['B', 3],
['C', 1], ['C', 2], ['C', 3]]
```

Task 0.5.12: Suppose `LofL` has been assigned a list whose elements are themselves lists of numbers. Write an expression that evaluates to the sum of all the numbers in all the lists. The expression has the form

`sum([sum(...`

and includes one comprehension. Test your expression after assigning `[[.25, .75, .1], [-1, 0], [4, 4, 4, 4]]` to `LofL`. Note that your expression should work for a list of any length.

Obtaining elements of a list by indexing



Donald Knuth `http://xkcd.com/163/`

There are two ways to obtain an individual element of a list. The first is by indexing. As in some other languages (Java and C++, for example) indexing is done using square brackets around the index. Here is an example. Note that the first element of the list has index 0.

```
>>> mylist[0]
4
>>> ['in','the','CIT'][1]
'the'
```

Slices: A *slice* of a list is a new list consisting of a consecutive subsequence of elements of the old list, namely those indexed by a range of integers. The range is specified by a colon-separated pair $i:j$ consisting of the index i as the first element and j as one past the index of the last element. Thus `mylist[1:3]` is the list consisting of elements 1 and 2 of `mylist`.

Prefixes: If the first element i of the pair is 0, it can be omitted, so `mylist[:2]` consists of the first 2 elements of `mylist`. This notation is useful for obtaining a prefix of a list.

Suffixes: If the second element j of the pair is the length of the list, it can be omitted, so `mylist[1:]` consists of all elements of `mylist` except element 0.

```
>>> L = [0,10,20,30,40,50,60,70,80,90]
>>> L[:5]
[0, 10, 20, 30, 40]
>>> L[5:]
[50, 60, 70, 80, 90]
```

Slices that skip You can use a colon-separated *triple* $a:b:c$ if you want the slice to include every c^{th} element. For example, here is how you can extract from `L` the list consisting of even-indexed elements and the list consisting of odd-indexed elements:

```
>>> L[::2]
[0, 20, 40, 60, 80]
>>> L[1::2]
[10, 30, 50, 70, 90]
```

Obtaining elements of a list by unpacking

The second way to obtain individual elements is by *unpacking*. Instead of assigning a list to a single variable as in `mylist =[4,8,12]`, one can assign to a list of variables:

```
>>> [x,y,z] = [4*1, 4*2, 4*3]
>>> x
4
>>> y
8
```

I called the left-hand side of the assignment a "list of variables," but beware: this is a notational fiction. Python does not allow you to create a value that is a list of variables. The assignment is simply a convenient way to assign to each of the variables appearing in the left-hand side.

Task 0.5.13: Find out what happens if the length of the left-hand side list does not match the length of the right-hand side list.

Unpacking can similarly be used in comprehensions:

```
>>> listoflists = [[1,1],[2,4],[3, 9]]
>>> [y for [x,y] in listoflists]
[1, 4, 9]
```

Here the two-element list `[x,y]` iterates over all elements of `listoflists`. This would result in an error message if some element of `listoflists` were not a two-element list.

Mutating a list: indexing on the left-hand side of =

You can mutate a list, replacing its i^{th} element, using indexing on the left-hand side of the =, analogous to an assignment statement:

```
>>> mylist = [30, 20, 10]
>>> mylist[1] = 0
>>> mylist
[30, 0, 10]
```

Slices can also be used on the left-hand side but we will not use this.

0.5.6 *Tuples*

Like a list, a tuple is an ordered sequence of elements. However, tuples are immutable so they can be elements of sets. The notation for tuples is the same as that for lists except that ordinary parentheses are used instead of square brackets.

```
>>> (1,1+1,3)
(1, 2, 3)
>>> {0, (1,2)} | {(3,4,5)}
{(1, 2), 0, (3, 4, 5)}
```

Obtaining elements of a tuple by indexing and unpacking

You can use indexing to obtain an element of a tuple.

```
>>> mytuple = ("all", "my", "books")
>>> mytuple[1]
'my'
>>> (1, {"A", "B"}, 3.14)[2]
3.14
```

You can also use unpacking with tuples. Here is an example of top-level variable assignment:

```
>>> (a,b) = (1,5-3)
>>> a
1
```

In some contexts, you can get away without the parentheses, e.g.

```
>>> a,b = (1,5-3)
```

or even

```
>>> a,b = 1,5-3
```

You can use unpacking in a comprehension:

```
>>> [y for (x,y) in [(1,'A'),(2,'B'),(3,'C')] ]
['A', 'B', 'C']
```

Task 0.5.14: Suppose S is a set of integers, e.g. $\{-4,-2,1,2,5,0\}$. Write a triple comprehension whose value is a list of all three-element tuples (i,j,k) such that i,j,k are elements of S whose sum is zero.

Task 0.5.15: Modify the comprehension of the previous task so that the resulting list does not include $(0,0,0)$. Hint: add a filter.

Task 0.5.16: Further modify the expression so that its value is not the list of all such tuples but is the first such tuple.

The previous task provided a way to compute three elements i, j, k of S whose sum is zero—if there exist three such elements. Suppose you wanted to determine if there were a hundred elements of S whose sum is zero. What would go wrong if you used the approach used in the previous task? Can you think of a clever way to quickly and reliably solve the problem, even if the integers making up S are very large? (If so, see me immediately to collect your Ph.D.)

Obtaining a list or set from another collection

Python can compute a set from another collection (e.g. a list) using the constructor `set(·)`. Similarly, the constructor `list(·)` computes a list, and the constructor `tuple(·)` computes a tuple

```
>>> set([0,1,2,3,4,5,6,7,8,9])
{0, 1, 2, 3, 4, 5, 6, 7, 8, 9}
>>> set([1,2,3])
{1, 2, 3}
>>> list({1,2,3})
[1, 2, 3]
>>> set((1,2,3))
{1, 2, 3}
```

23/11/24

Task 0.5.17: Find an example of a list `L` such that `len(L)` and `len(list(set(L)))` are different.

0.5.7 Other things to iterate over

Tuple comprehensions—not! Generators

One would expect to be able to create a tuple using the usual comprehension syntax, e.g. `(i for i in [1,2,3])` but the value of this expression is not a tuple. It is a *generator*. Generators are a very powerful feature of Python but we don't study them here. Note, however, that one can write a comprehension over a generator instead of over a list or set or tuple. Alternatively, one can use `set(·)` or `list(·)` or `tuple(·)` to transform a generator into a set or list or tuple.

Ranges

A range plays the role of a list consisting of the elements of an arithmetic progression. For any integer n, `range(`n`)` represents the sequence of integers from 0 through $n-1$. For example, `range(10)` represents the integers from 0 through 9. Therefore, the value of the following comprehension is the sum of the squares of these integers: `sum({i*i for i in range(10)})`.

Even though a range represents a sequence, it is not a list. Generally we will either iterate through the elements of the range or use `set(·)` or `list(·)` to turn the range into a set or list.

```
>>> list(range(10))
[0, 1, 2, 3, 4, 5, 6, 7, 8, 9]
```

Task 0.5.18: Write a comprehension over a range of the form `range(`n`)` such that the value of the comprehension is the set of odd numbers from 1 to 99.

You can form a range with one, two, or three arguments. The expression `range(`a`,`b`)` represents the sequence of integers $a, a+1, a+2, \ldots, b-1$. The expression `range(`a`,`b`,`c`)` represents $a, a+c, a+2c, \ldots$ (stopping just before b).