Testing

Let’s say I want to merge two sorted lists.

```
-- | Assuming the input lists are sorted, combine the lists into a
-- sorted output.
merge :: Ord a => [a] -> [a] -> [a]
merge (x:xs) (y:ys)
  | x < y     = x : merge xs ys
  | otherwise = y : merge xs ys
merge _      _      = []
```

Does this function work as expected? I could run a few tries in GHCi, but that’s a little unsatisfactory: I have to do the work to think up a test, but I get to use it only once. Instead, it’s much better to write the test in my file, and that way I can re-run it every time I update my merge function.

The technique described above is often referred to as *unit testing* and is used exensively in the real world. But is unit testing even that great? Sure you can re-run all your tests whenever you want, but that doesn’t get around the issue that you actually have to *write* all of the tests in the first place. Coming up with specific test cases is often tedious, repetitive, and *arbitrary*.

Can we do better?

Property-based testing

Writing test cases is boring. And, it’s easy to miss out on unexpected behavior. Much better (and, more along the lines of *wholemeal programming*) is to define *properties* we wish our function to have. Then, we can get the computer to generate the test cases for us.

QuickCheck is the standard Haskell library for property-based testing. The idea is that you define a so-called *property*, which is then tested using pseudo-random data.

For example:

```
prop_numElements :: [Integer] -> [Integer] -> Bool
prop_numElements xs ys
  = length xs + length ys == length (merge xs ys)
```

This property is saying that the sum of the lengths of the input lists should be the same as the length of the output list. (It is customary to begin property names with `prop_`.) Let’s try it!

```
*Main> quickCheck prop_numElements
*** Failed! Falsifiable (after 5 tests and 4 shrinks):
[]
[0]
```

(Your results may differ slightly. Remember: it’s using randomness.)

The first thing we notice is that our function is clearly wrong, with lots of stars and even an exclamation point! We then see that QuickCheck got through 5 tests before discovering the failing test case, so our function isn’t terrible. QuickCheck tells us what the failing arguments are: `[]` and `[0]`. Indeed GHCi tells us that `merge [] [0]` is `[]`, which is wrong.

What’s so nice here is that QuickCheck found us such a nice, small test case to show that our function is wrong. The way it can do so is that it uses a technique called *shrinking*. After QuickCheck finds a test case that causes failure, it tries successively smaller and smaller arguments (according to a customizable definition of “smaller”) that keep failing. QuickCheck then reports only the smallest failing test case. This is wonderful, because otherwise the test cases QuickCheck produces would be unwieldy and hard to reason about.

(If you want to observe QuickCheck doing that, use `verboseCheck` instead of `quickCheck`.)

A final note about this property is that the type signature tells us that the property takes lists of integers, not any type `a`. This is so that GHC doesn't choose a silly type to test on, such as `()`. We must always be careful about this when writing properties of polymorphic functions. Numbers are almost always a good choice.

Implications

Let us take another stab at our function:

```
merge :: MergeFun
merge (x:xs) (y:ys)
  | x < y     = x : merge2 xs (y:ys)
  | otherwise = y : merge2 (x:xs) ys
merge [] ys = ys
merge xs [] = xs
```

```
*Main> quickCheck prop_numElements
+++ OK, passed 100 tests.
```

Huzzah!

Is that it? Are we done? Not quite. Let's try another property:

```
prop_sorted :: [Integer] -> [Integer] -> Bool
prop_sorted xs ys
  = merge xs ys == sort (xs ++ ys)
```

```
*Main> quickCheck prop_sorted
*** Failed! Falsifiable (after 4 tests and 3 shrinks):
[]
[1,0]
```

Drat. QuickCheck quite reasonably tried the list `[1,0]` as an input to our function. Of course, this isn't going to work because it's not already sorted. We need to specify an implication property:

```
prop_sorted :: [Integer] -> [Integer] -> Property
prop_sorted xs ys
  = isSorted xs && isSorted ys ==> merge xs ys == sort (xs ++ ys)

isSorted :: Ord a => [a] -> Bool
isSorted (a:b:rest) = a <= b && isSorted (b:rest)
isSorted _          = True    -- must be fewer than 2 elements
```

In `prop_sorted`, we see the use of the operator `(==>)`. Its type is `Testable prop => Bool -> prop -> Property`. It takes a `Bool` and a `Testable` thing and produces a `Property`. Note how `prop_sorted` returns a `Property`, not a `Bool`. We'll sort these types out fully later, but I wanted to draw your attention to the appearance of `Property` there.

Let's see how this works:

```
*Main> quickCheck prop_sorted
*** Gave up! Passed only 21 tests.
```

There aren't any failures, but there aren't a lot of successes either. We can see the problem if we use `verboseCheck` instead of `quickCheck`: QuickCheck will run the test only when both randomly-generated lists are in sorted order. The odds that a randomly-generated list of length `n` is sorted is `1/n!`, which is generally quite small. And we need *two* sorted lists. This isn't going to work out well.

QuickCheck's types

How does QuickCheck generate the arbitrary test cases, anyway? It uses the `Arbitrary` class:

```
class Arbitrary a where
  arbitrary :: Gen a
  shrink    :: a -> [a]
```

We'll leave `shrink` to the online documentation and focus on `arbitrary`. The `arbitrary` method gives us a `Gen a` – a generator for the type `a`. Of course, the `arbitrary` method for lists doesn't care about ordering (indeed, it can't, due to parametricity), but we do. Luckily, this is a common problem, and QuickCheck offers a solution in the form of `OrderedList`, a wrapper around lists that have the right `Arbitrary` instance for our needs:

```
newtype OrderedList a = Ordered { getOrdered :: [a] }
instance (Ord a, Arbitrary a) => Arbitrary (OrderedList a) where ...
```

Now, let's rewrite our property:

```
prop_sorted :: OrderedList Integer -> OrderedList Integer -> Bool
prop_sorted (Ordered xs) (Ordered ys)
  = merge xs ys == sort (xs ++ ys)
```

```
*Main> quickCheck prop_sorted
+++ OK, passed 100 tests.
```

Huzzah! Just by changing the types a bit, we can affect instance selection to get what we want.

Yet, this all seems like black magic. How does QuickCheck do it? Let's look more in depth at the types.

```
quickCheck :: Testable prop => prop -> IO ()

class Testable prop where ...
instance Testable Bool where ...
instance Testable Property where ...
instance (Arbitrary a, Show a, Testable prop) => Testable (a -> prop) where ...
```

We can `quickCheck` anything that's `Testable`. Boolean values are `Testable`, as are the somewhat mysterious `Property`s. But it's the last instance listed here of `Testable` that piques our curiosity. It says that a *function* is `Testable` as long as its argument has an `arbitrary` method, the argument can be printed (in case of failure), and the result is `Testable`.

Is `[Integer] -> [Integer] -> Bool` `Testable`? Sure it is. Recall that `[Integer] -> [Integer] -> Bool` is equivalent to `[Integer] -> ([Integer] -> Bool)`. Because `[Integer]` has both an `Arbitrary` instance and a `Show` instance, we can use the last instance above as long as `[Integer] -> Bool` is `Testable`. And that's `Testable` because we (still) have an `Arbitrary` and a `Show` instance for `[Integer]`, and `Bool` is `Testable`. So, that's how `quickCheck` works – it uses the `Arbitrary` instances for the argument types. And, that's how changing the argument types to `OrderedList Integer` got us the result we wanted.

Generating arbitrary data

When you want to use QuickCheck over your own datatypes, it is necessary to write an `Arbitrary` instance for them. Here, we'll learn how to do so.

Let's have a look at our own list type

```
data List a = Empty | Entry a (List a)

instance Show a => List a where
  show = show . toList
```

```haskell
toList :: List a -> [a]
toList Empty           = []
toList (a `Entry` as) = a : toList as

fromList :: [a] -> List a
fromList []     = Empty
fromList (a:as) = a `Entry` fromList as
```

If we want an `Arbitrary` instance, we must define the `arbitrary` method, of type `Gen (List a)`. Luckily for us, `Gen` is a monad (did you see that coming?), so some of its details are already familiar. We also realize that if we want arbitrary lists of `a`, we'll need to make arbitrary `as`. So, our instance looks like

```haskell
instance Arbitrary a => Arbitrary (List a) where
  arbitrary = genList
```

At this point, check out the combinators available in the "Generator combinators" section of the **QuickCheck documentation**.

It's helpful to think about how you, as a human, would generate an arbitrary list. One way to do it is to choose an arbitrary length (say, between 0 and 10), and then choose each element arbitrarily. Here is an implementation:

```haskell
genList :: Arbitrary a => Gen (List a)
genList = do
  len <- choose (0, 10)
  vals <- replicateM len arbitrary
  return $ fromList vals
```

Let's try it out:

```
*Main> sample genList
[(),(),(),(),(),()]
[]
[(),(),(),(),(),(),(),(),()]
[(),(),(),(),(),()]
[(),(),(),(),(),(),(),(),(),()]
[()]
[(),(),(),(),(),(),(),()]
[(),(),(),(),(),(),(),(),()]
[(),(),()]
[(),(),(),(),(),()]
[(),(),(),(),(),(),(),(),(),()]
```

The arbitrary lengths are working, but the element generation sure is boring. Let's use a type annotation to spruce things up (and override GHC's default choice of `()`)!

```
*Main> sample (genList :: Gen (List Integer))
[0,0,0,0,0,0,0,0,0,0]
[]
[-2,3,1,0,4,-1]
[-5,0,2,1,-1,-3]
[-5,-6,-7,-2,-8,7,-3,4,-6]
[4,-3,-3,2,-9,9]
[]
[10,-1]
[9,-7,-16,3,15]
[0,14,-1,0]
[3,18,-13,-17,-20,-8]
```

That's better.

This generation still isn't great, though, because perhaps a function written over `Lists` fails only for lists longer than 10. We'd like unbounded lengths. Here's one way to do it:

```haskell
genList :: Arbitrary a => Gen (List a)
genList = do
  stop_now <- arbitrary
  if stop_now
     then return Empty
     else do
       x <- arbitrary
       xs <- genList
       return (x `Entry` xs)
```

```
*Main> sample (genList :: Gen (List Integer))
[0,0,0,0,0,0]
[]
[3,-3]
[]
[]
[-1,-1]
[-10]
[]
[]
[11]
[-20,-14]
```

The lengths are unbounded, but we're getting a *lot* of empty lists. This is because at every entry in the list, there's a 50% chance of producing `Empty`. That means that a list of length n will appear only one out of 2^n times. So, lengths are unbounded, but very unlikely.

The way to make progress here is to use the `sized` combinator. QuickCheck is set up to try "simple" arbitrary things before "complex" arbitrary things. The way it does this is using a size parameter, internal to the `Gen` monad. The more generating QuickCheck does, the higher this parameter gets. We want to use the size parameter to do our generation.

Let's look at the type of `sized`:

```haskell
sized :: (Int -> Gen a) -> Gen a
```

An example is the best way of explaining how this works:

```haskell
genList :: Arbitrary a => Gen (List a)
genList = sized $ \size -> do
  len <- choose (0, size)
  vals <- replicateM len arbitrary
  return $ fromList vals
```

```
*Main> sample (genList :: Gen (List Integer))
[]
[-2]
[-1,3,4]
[-4,-2,1,-1]
[]
[]
[12,3,11,0,3,-12,10,5,11,12]
[-4,-8,-9,2,14,5,8,11,-1,7,11,-8,2,-6]
[6,10,-5,15,6]
[-3,-18,-4]
[9,19,13,-19]
```

That worked nicely – the lists tend to get longer the later they appear. The idea is that `sized` takes a *continuation*: the thing to do with the size parameter. We just use a lambda function as the one argument to `sized`, where the lambda binds the `size` parameter, and then we can use it internally. If that's too painful (say we just want to produce the size parameter, without using a continuation), you could always do something like this:

```haskell
getSize :: Gen Int
getSize = sized return
```

I'll leave it to you to figure out how that works. Follow the types!

As one last example, we can also choose arbitrary generators from a list based on frequency. Although the length method of `genList` works well for lists, the following technique is much better for trees:

```haskell
genList :: Arbitrary a => Gen (List a)
genList = sized $ \size -> do
  frequency [ (1, return Empty)
            , (size, do x <- arbitrary
                        xs <- resize (size - 1) genList
                        return (x `Entry` xs) )]
```

```
*Main> sample (genList :: Gen (List Integer))
[]
[2,0]
[4,0,-1]
[-6,-2,3,-1,-1]
[6,6]
[2,2,2,6,6]
[-6,-9,5,-5]
[8,7,5,7,7,-1,-2,-1,-5,-3]
[15,-12,14,13,-5,-10,-9,-8,-2]
[12,-11,-8,6,-6,-4,11,11]
[-7,1,-3,4,-3,-9,4,6,-2,10,-9,-7,5,7,1]
```

Let's look at the type of `frequency`:

```haskell
frequency :: [(Int, Gen a)] -> Gen a
```

It takes a list of `(Int, Gen a)` pairs and produces a `Gen a`. The numbers in the list give the likelihood of choosing that element. Above, we fixed the frequency of `Empty` at 1, but let the likelihood of `Entry` vary according to the desired size. Then, in the recursive call to `genList`, we used `resize` to lower the `size` parameter. Otherwise, it's too likely to get a runaway list that goes on toward infinity.

Conclusion

Property based testing is an amazingly effective and satisfying approach to testing. It works especially well in Haskell because of the combination of purity, which means that one only needs to vary the arguments to a function, and the type system, which allows us to use type classes to very generically and composable create code that comes up with random input. QuickCheck came out in 1999 has since then inspired re-implementations in other languages as well.

Practical hints

Very likely, your project has some code that can be tested this way, and you should do that. Here are some pointers to look at.

- **QuickCheck** itself.
- QuickCheck is random, which is nice. An alternative approach is to test *all* input, up to a given size. This way your test suite is deterministic, and at least no small corner cases remain. A library implementing that approach is **SmallCheck**.
- Unit tests are still useful. The default library to go for is **HUnit**.
- You do not want to run your tests in GHCi only. There are libraries that allow you to name and group your tests and create an executable that runs your tests and has nice shiny colored output. Check out **tasty** for one of those. It has support for QuickCheck, SmallCheck, HUnit and more.