第4章　モナド

モナド（monad）は、Haskell（あるいは他の関数型言語）で破壊的代入（変数の値など状態を変更すること）や入出力のような、他の言語では“副作用”（side effect）として実現される特徴を扱うための手法である。

もともとは数学のカテゴリー理論（圏論）で使われている言葉を借用したものであるが、Haskell で使用するときには、背景となるカテゴリー理論のことを知っている必要はない。

4.1　参照透明性と副作用

純粋な関数型言語には、式は値を表すためだけのものであり、「変数の出現はその定義の右辺の式で置き換えても全体の意味は変わらない」という性質がある。このような性質を＿＿＿＿＿＿（referential transparency）と呼び、プログラムの“意味”を考察していく上でとても重要な性質である。（参照透明性のおかげで、帰納法などによる証明が容易になる。） 一方、副作用は、式がその＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿のことである。式が副作用を持ちうるということは、その言語では参照透明性が成り立たない、ということである。Haskell の式は副作用を持っていない。

C のような言語では、入出力や破壊的代入を扱う部分では、副作用を使っているため参照透明性は成り立たない。例えば、C で、

```
c = getchar();  putchar(c);  putchar(c);  // C-code ①
```

と

```
putchar(getchar()); putchar(getchar());   // C-code ②
```

とは、getchar() が副作用を持っているために異なるプログラムである。（上のプログラムでは 1 文字、下のプログラムでは 2 文字の入力が消費される。）

一見、参照透明性と入出力や破壊的代入は相容れない性質のように見える。しかし、Haskell では次のように考える。

> 入出力や、変数などの状態の変更は、何らかの“アクション”であり、単なる値とは異なる型を持っている。副作用を持つ C 言語などの関数に対応する Haskell の関数は、このアクションを含めて戻り値として返す関数として考える。（つまりアクションを“副”作用ではなく、一人前の値として扱う。）この“アクション”の型は、アクションに対応している文脈でしか使用することができない。従って、アクションと値は区別され、参照透明性が保たれる。

例えば、Haskell で getchar, putchar に対応する関数は、それぞれ次のような型を持っている。

```
getChar :: __________

putChar :: ____________________
```

この IO という型構成子がアクションの型を表す。また、() はユニット型と読み、意味のある値を持たない型である。そもそも、C 言語の putchar(getchar()) という式に対応する putChar getChar のような Haskell の式は、型エラーになってしまう。IO 型に用意されている演算子 “>>=”（バインド、と読む）を用いて、次のように書かなければいけない。

```
getChar >>= (\ c -> putChar c)
```

ここで、 (>>=) の型は IO a -> (a -> IO b) -> IO b である。（後述するように、実際にはより一般的な型を持っている。）この式では、getChar というアクションの結果得られる Char 型の値が、c という変数に束縛され、続いて putChar c というアクションが実行される。アクションの型を持つ式から値だけを抽出するには、 >>= のような演算子を介する必要がある。Char 型の変数 c に対して、c = getChar と書けるわけではないので、型の面からも c が getChar と置き換えられないことが明白である。

Q 4.1.1 C-code ①と C-code ②に対応する Haskell プログラム（の断片）を getChar, putChar, >>=を使って書け。

副作用を持つプログラムは実行する順番により意味が変わるので、並列に実行されるプログラムでは問題となることがある。参照透明性を保ちながらアクションを扱えることは理論的に有用であるだけでなく、並列コンピューティングなど実用面でも必要となってきている。

4.2 モナドとは

このような、さまざまな言語で副作用として実現される特徴（入出力・破壊的代入・例外処理・非決定性など）に対するアクションの型が、実は共通の構造を持っていて、同じ構造を持つ演算子で取り扱えることがわかっている。この共通の構造を持つ型（正確には型構成子）のことを _____(monad) という。つまり、モナドはアクションの型である。上記の IO もモナドである。ただし、モナドの中

にはアクションという言葉がふさわしくないものもあるので、以下では代わりに“計算” (computation) という言葉を使う。

具体的にはモナドとは

$$\begin{aligned} \mathit{return} &:: a \to M\ a \\ (>\!\!>\!\!=) &:: M\ a \to (a \to M\ b) \to M\ b \end{aligned}$$

という型の関数の存在する型構成子 M のことである。より厳密には、*return*, *(>>=)* が

$$\begin{aligned} (\mathit{return}\ a) >\!\!>\!\!= k &= k\ a \\ m >\!\!>\!\!= (\lambda a \to \mathit{return}\ a) &= m \\ (m_1 >\!\!>\!\!= k_1) >\!\!>\!\!= k_2 &= m_1 >\!\!>\!\!= (\lambda a \to (k_1\ a >\!\!>\!\!= k_2)) \end{aligned}$$

の 3 つの等式（monad law）を満たす必要がある。これらの等式は、直感的には *return* が *(>>=)* の“単位元”であること、*(>>=)* が結合則を満たすこと、を示している。つまり、数に対する、加算 (+) と加算の単位元 0 の次の等式に相当する。

$$\begin{aligned} 0 + x &= x \\ x + 0 &= x \\ (x + y) + z &= x + (y + z) \end{aligned}$$

直観的には $M\ a$ という型は、_____ の型を意味する。また、*return* と *(>>=)* の直観的な意味は次の通りである。

- *return a* $\cdots$ 値 *a* をそのまま返す何もしない（アクションが実質ない）計算を表す。
- `m >>= k` $\cdots$ まず、$m :: M\ a$ を計算し、その結果の値の部分を関数 $k :: a \to M\ b$ に渡して、続けて計算することを表す。また、アクションは *m*, *k* のアクションを続けて実行するアクションに相当する。

4.3 モナドと型クラス

モナド M の定義は模倣したい副作用により異なるし、付随する関数 *return*, *(>>=)* の定義ももちろん異なる。しかし、*return*, *(>>=)* は M をパラメーターとして、決まった型を持つので、型クラスを用いてアドホック多相な関数として定義することができる。

```
1  class Monad m where
2     return :: a -> m a
3     (>>=)  :: m a -> (a -> m b) ->  m b
```

実際は、`Monad` は型ではなく、型構成子に対する型クラス（正確には型構成子クラス）になっている。

4.4 IO モナド

IO は Haskell の Prelude（最初から読み込まれているライブラリ）に入っているモナドである。型クラス Monad のインスタンスである。その定義は一般のプログラマからは見えない組込みの型となっている。ただし、直観的には次のような定義を持つ型だと考えることができる。

```
-- type IO a ≒ ____________________________________

-- instance Monad IO where
-- -- 注:  実際には type で定義された型名を instance には使えない。
--   return a = \ w -> (a,w)
--   m >>= k  = \ w -> let (a,w1) = m w in k a w1
```

ただし RealWorld は、コンピューター全体のファイルなどの状態を表す型である。IO は関数の型だが、引数の RealWorld 型のデータは隠されていて、他の計算で使われないことが保証できるため、破壊的に書き換えて、戻り値の RealWorld 型のデータのために使っても良い、ということである。

IO は Haskell で入出力や状態を効率的に扱うために、処理系で特別な扱いを受ける。主なプリミティブとして、putChar, getChar の他にも、以下のような関数がある。

```
putStr      :: String -> IO () -- 文字列を出力する
putStrLn    :: String -> IO () -- 文字列を出力して、改行する

getLine     :: IO String       -- 一行を読み込む
getContents :: IO String       -- EOF まで読み込む

print  :: Show a => a -> IO () -- 文字列に変換して出力し、改行する
readLn :: Read a => IO a       -- 一行を読み込んでパースする
```

これらの他にファイルに対して入出力するための関数なども用意されている。

特に getContents は遅延評価で標準入力の内容を読み込む。つまり次のようなプログラムでは、標準入力のすべてを読み込むことはなく、最初の 1 行のみを出力する。

```
main :: IO ()
main = getContents >>=
       (\ all -> let ls = lines all
                 in putStrLn (head ls))
```

ただし、lines :: String -> [String] は文字列を改行文字で分割する関数、head :: [a] -> a はリストの先頭要素を返す関数である。

さらに、Data.IORef というモジュールを import することで、破壊的代入が可能な参照型（IORef）を扱う関数も用意される。

```
-- import Data.IORef が必要

newIORef    :: a -> IO (IORef a)     -- 新しい参照の作成
readIORef   :: IORef a -> IO a       -- 参照の読出し
```

```
writeIORef  :: IORef a -> a -> IO () -- 参照への書込み
```

次に、IORef を利用したプログラムの例をあげる。

```
import Data.IORef

foo r = readIORef r >>= \ i ->
        if i <= 0 then return ()
        else putStrLn (show i)    >>= \ _ ->
             writeIORef r (i - 1) >>= \ _ ->
             foo r

main = newIORef 10 >>= \ r ->
       foo r
```

ただし、これは無理矢理 IORef を使った例であり、IORef を使わずに同じ動作をする次のようなプログラムのほうが自然である。

```
{- foo の自然な定義は以下の通り -}
--  foo i = if i <= 0 then return ()
--          else putStrLn (show i) >>= \ _ ->
--               foo (i - 1)
--
--  main = foo 10
```

ところで、Haskell のプログラムを、GHCi のような対話環境ではなく、ghc コマンドで実行可能ファイルにコンパイルして実行するとき、C と同じように main という名前の関数から実行が開始される約束になっている。そして main 関数は、IO *t* という形の型（*t* は任意の型）を持たなければいけないことになっている。

たとえば、標準入力から読み込んだ文字列の大文字を小文字に変換したものと小文字を大文字に変換したものを出力するプログラムは以下のようになる。

```
import Data.Char   -- Data.Char モジュールを import する

-- toLower, toUpper ::  Char -> Char は大文字・小文字の変換関数
main :: IO ()
main = getContents >>= (\ s -> putStr (map toLower s ++ map toUpper s)))
```

ラムダ抽象（\... -> ... ）は、どんな中置演算子よりも優先順位が低いので、上記の定義は括弧を省略し、さらにレイアウトを整えて、次のように書くことが多い。

```
main = getContents            >>= \ s ->
       putStr (map toLower s ++ map toUpper s)
```

以降のプログラムでは、このように括弧を省略する。

実は、Haskell では Monad クラスに対して、**do** 記法という糖衣構文を用意していて、この関数は次のように書くこともできる。

```
main = do { s <- getContents;
            putStr (map toLower s ++ map toUpper s) }
```

この **do** 式も第 A 章で紹介したレイアウトルールの対象であり、適切にレイアウトすればブレースやセミコロンを省略することができる。

```
1  main = do s <- getContents
2            putStr (map toLower s ++ map toUpper s)
```

そして **do** 式は次のルールで翻訳される。（下線部に翻訳規則を再帰的に適用していく。）

$$\begin{aligned}
\texttt{do } \{e\} &\Rightarrow e \\
\texttt{do } \{e;\ stmts\} &\Rightarrow e \texttt{ >>= \textbackslash _ -> do } \{stmts\} \\
\texttt{do } \{x \texttt{ <- } e;\ stmts\} &\Rightarrow e \texttt{ >>= \textbackslash } x \texttt{ -> do } \{stmts\} \\
\texttt{do } \{\texttt{let } decls;\ stmts\} &\Rightarrow \texttt{let } decls \texttt{ in do } \{stmts\}
\end{aligned}$$

Q 4.4.1 つぎのようなプログラムを上記の IO に関する関数を用いて定義せよ。

1. 一行だけ行を読み込んで、それをオウム返しするプログラム
2. 一行だけ行を読み込んで、それを 2 回オウム返しするプログラム

なお、**for** 文などの繰返しのための構文はないので、繰り返しが必要なときは再帰関数を定義する必要がある。

問 4.4.2 IO モナドを利用して次のようなプログラムを作成せよ。（次節で紹介している関数を使用するかもしれない。）

1. 入力文字列を行毎に大文字と小文字に変換して出力する。（例えば、「Hello,↙World↙」 が 「hello,↙HELLO,↙world↙WORLD」’ になる。）
2. 入力文字列中の数字（'0'〜'9'）の出現回数をカウントして出力する。
3. 入力文字列中の'@' が出現する最初の 10 行だけを出力する。

4.5 続・有用なリスト処理関数

モナドとは直接関係ないが、以前紹介した以外に文字列処理プログラムで有用と思われるリスト処理関数を、以下でいくつか紹介する。これらの関数は Prelude（標準ライブラリ）に定義済みである（sort を除く）。

```
-- lines は文字列を改行文字のところで分割して、文字列のリストにする。
-- 結果の文字列には改行文字は含まれない。
lines           :: String   -> [String]

-- words は文字列を空白文字で分割して、文字列のリストにする。
words           :: String   -> [String]

-- unlines は lines の逆操作である。
-- 各文字列の末尾に改行文字を追加し、一つに結合する。
unlines         :: [String] -> String

-- unwords は words の逆操作である。
-- 各文字列を空白で区切って結合する。
unwords         :: [String] -> String

-- take n xs は xs の長さ n の先頭部分を返す。
-- n > length xs のときは xs 自身を返す。
take            :: Int -> [a] -> [a]

-- drop n xs は xs の最初の n 個の要素を除いた末尾部分を返す。
-- n > length xs のときは [] を返す
drop            :: Int -> [a] -> [a]

-- takeWhile は述語 p とリスト xs を受け取り, p を満たす要素
-- だけからなる xs の最も長い先頭部分（空の場合もある）を返す。
--
-- > takeWhile (< 3) [1,2,3,4,1,2,3,4] == [1,2]
-- > takeWhile (< 9) [1,2,3] == [1,2,3]
-- > takeWhile (< 0) [1,2,3] == []
takeWhile       :: (a -> Bool) -> [a] -> [a]

-- dropWhile p xs は、takeWhile p xs の残りの末尾部分を返す。
--
-- > dropWhile (< 3) [1,2,3,4,5,1,2,3] == [3,4,5,1,2,3]
-- > dropWhile (< 9) [1,2,3] == []
-- > dropWhile (< 0) [1,2,3] == [1,2,3]
dropWhile       :: (a -> Bool) -> [a] -> [a]

-- any は述語とリストを受け取り、リストのなかのどれかの要素が述語
-- を満たすかどうか判定する。
any             :: (a -> Bool) -> [a] -> Bool

-- all は述語とリストを受け取り、リストのなかのすべての要素が述語
-- を満たすかどうか判定する。
all             :: (a -> Bool) -> [a] -> Bool

-- リスト（空ではいけない）の先頭要素を取り出す。
head            :: [a] -> a

-- （非空かつ有限な）リストの最後の要素を取り出す。
last            :: [a] -> a
```

```
52
53  -- リスト（空ではいけない）の先頭要素を除いた末尾部分を取り出す。
54  tail            :: [a] -> [a]
55
56  -- リストの要素として含まれているか否かを判定する。
57  -- 通常 x `elem` xs のように中置記法で用いることが多い。
58  elem            :: (Eq a) => a -> [a] -> Bool
59
60  -- import Data.List が必要
61  -- sort 関数は安定なソートアルゴリズムの実装である。
62  -- 比較関数を引数にとる sortBy 関数もある
63  sort            :: (Ord a) => [a] -> [a]
64
65  -- sequence 関数はリスト中のアクションを順に実行する。
66  sequence        :: Monad m => [m a] -> m [a]
67  sequence []      = return []
68  sequence (m:ms) = m           >>= \ a ->
69                    sequence ms >>= \ as ->
70                    return (a:as)
71
72  mapM            :: Monad m => (a -> m b) -> [a] -> m [b]
73  mapM f as       = sequence (map f as)
```

Q 4.5.1 次の式の値は何か?

1. `drop 2 [1,4,5]`
2. `takeWhile (< 0) [-1,-3,2,-2,7]`
3. `any (> 0) [-1,-3,-5]`
4. `3 `elem` [2,5,3,1]`

4.6 さらに詳しく知りたい人のために ...

文献 [1] は、モナドとリストの内包表記の関係について解説している。

この章の参考文献

[1] Philip Wadler, “Comprehending Monads”
ACM Conference on Lisp and Functional Programming, Nice (France), 1990 年 6 月