randPort vignette

Mike Flynn, Angel Zhou, Dave Kane

July 10, 2013

A simple case to consider is one with only 3 stocks in the universe, say GE, IBM and Coca-Cola (Ko). Assume that there is some data available about these stocks:

```
set.seed(40)
stocks <- data.frame(ticker = c("GE", "IBM", "KO"), sizerank = c(1, 0, -1))
stocks

##   ticker sizerank
## 1     GE        1
## 2    IBM        0
## 3     KO       -1
```

The most basic constraint is the long-only constraint i.e. the weights of the stock must add up to one and be positive. This can be represented by a linear equation: let the weights of "GE","IBM", and "KO", be x, y, and z, respectively. Then $x + y + z = 1$, $x, y, z \geq 0$ are the only constraints. The `randPort` package is designed to sample underdetermined equations of the form $Ax = b$, with $x \geq 0$ and so it is ideal for this set of constraints. In this case $A = [1, 1, 1]$ and $b = 1$. With `hitandrun`, we can easily obtain a set of weights that are randomly distributed and fulfill the constraints.

```
A <- matrix(c(1, 1, 1), nrow = 1, ncol = 3)
A

##      [,1] [,2] [,3]
## [1,]    1    1    1

b <- 1

w <- hitandrun(A, b, n = 1000)
points <- as.data.frame(t(w))
colnames(points) <- stocks$ticker
head(points)
```

```
##        GE     IBM     KO
## 1 0.10506 0.03877 0.8562
## 2 0.07990 0.83320 0.0869
## 3 0.09234 0.80191 0.1057
## 4 0.14397 0.44635 0.4097
## 5 0.43460 0.26308 0.3023
## 6 0.13072 0.19263 0.6766
```

The pairwise scatters display the fact that any pair of variables has a sum bounded by 1, but can be anywhere between that line and the axes.

Graphs go here

Perhaps now we want to match a portfolio that has an exposure to size rank of 0.5. This is the same thing as adding a new row to A and b, corresponding to the column "sizerank" of our stock data and 0.5 respectively.

```r
A <- rbind(A, stocks$sizerank)
b <- c(b, 0.5)

w2 <- hitandrun(A, b, n = 1000)
points2 <- as.data.frame(t(w))
colnames(points2) = colnames(points)
head(points2)
```

```
##        GE     IBM     KO
## 1 0.10506 0.03877 0.8562
## 2 0.07990 0.83320 0.0869
## 3 0.09234 0.80191 0.1057
## 4 0.14397 0.44635 0.4097
## 5 0.43460 0.26308 0.3023
## 6 0.13072 0.19263 0.6766
```

Running this code is equivalent to sampling from the positive solution space of $x + y + z = 1$ and $x - z = .5$, which corresponds to the line segment parameterized by $r(t) = [t, 1.5 - 2t, t - .5]$ with $.5 \leq t \leq 0.75$. This is very clear when looking at pairwise plots of the variables:

Geometric Nature of the Problem

It is easier to picture the space of each problem when looking at things geometrically. Each row of Ax=b corresponds to a linear equation $c_1x_1 + c_2x_2 + \cdots + c_nx_n = b_m$. Geometrically, this is an (n-1)-plane in $\mathbb{R}^n$ which our solution must lie in. When there are m rows, our solution must lie in the intersection of those

m planes which is itself an $(n-m)$ plane. A concrete example is the 3 stock case with only the long-only constraint. This is the peice of the $(3-1=2)$-plane $x+y+z=1$ that is in the positive quadrant. It forms a triangle. If another constraint is added, we can see this as the intersection of two 2-planes, also known as a line, as you can see via these graphs. When we sample with our function, the random points are distributed uniformly along these shapes.