Simulating quantile distributions

Ben Mazzotta

Wednesday, November 19, 2014

This is a proof of concept document for simulating quantile shares of a random variable.

Problem. What share of X do the bottom two quintiles possess?

```
set.seed(1213)
t.mu <- 8
t.sigma <- 2
qlnorm(.6, t.mu, t.sigma)
```

```
## [1] 4948
```

```
t.draw <- rlnorm(1e6, t.mu, t.sigma)
str(t.draw)
```

```
##  num [1:1000000] 317.4 23914.8 20.2 33.8 28848.2 ...
```

```
t.under40 <- subset(t.draw, t.draw<qlnorm(.40, t.mu, t.sigma))
sum(t.draw)
```

```
## [1] 2.179e+10
```

```
sum(t.under40)
```

```
## [1] 266762430
```

```
sum(t.under40) / sum(t.draw)
```

```
## [1] 0.01224
```

Replicate for various sample sizes.

```
# ## Declare a list of quantiles to cutoff
# t.quant = c(.2, .4, .6, .8)

## Declare a list of sample sizes
t.sample = c(1e4, 1e5, 1e6)

## Simulate the distribution with three different sample sizes; save to list
t.draw2 <- lapply(t.sample, function(x) rlnorm(x, t.mu, t.sigma))
## Make a data frame of summary statistics
##      Column 1: sample size
t.df <- data.frame(t.sample)
##      Column 2: sum of X
t.df$sum <- sapply(t.draw2, sum)
```

```
##      Column 3: sum of X from bottom two deciles
t.df$sum40 <- sapply(t.draw2, function(x) sum(subset(x, x < qlnorm(.4, t.mu, t.sigma))))
##      Column 4: share of X belonging to bottom two deciles.
t.df$share40 <- t.df$sum40 / t.df$sum

cat("The bottom two quintiles of X are", t.df[3,4]*100, "% of the total.")
```

```
## The bottom two quintiles of X are 1.209 % of the total.
```

Ben Mazzotta works on the Cost of Cash project at the Institute for Business in the Global Context. Contact: benjamin.mazzotta@tufts.edu
+1.617.627.4561 160 Packard Ave, Medford MA 02155.