A Few Thoughts on the Geometric Mean

CEE 202 | Project 1 | Kevin Smith

December 17, 2014

Required Libraries

```
require(plyr)
require(ggplot2)
require(gridExtra)
require(pander)
require(doMC)
```

Setup Parallel Processing

```
registerDoMC(cores = 4)
getDoParWorkers()
```

```
## [1] 4
```

Relative Efficiency as an Estimator of the Lognormal Mean

```
experiment <- function(n, mu, sigma){
  sample <- rlnorm(n, mu, sigma)
  m <- mean(sample)
  g <- exp(mean(log(sample)))
  d <- median(sample)
  result <- data.frame(m = m, g = g, d = d)
}
```

Calculating Mean Squared Error

```
mse <- function(x, X){
  x <- unlist(x)
  se <- (x - X)^2
  return(mean(se))
}
```

Setting up the Monte Carlo Simulator

```r
monte.carlo <- function(n.samples, mu, sigma){
  M <- exp(mu + sigma/2)
  experimental.results <- replicate(
    n.experiments,
    experiment(n.samples, mu, sigma),
    simplify = "data.frame")

  mean.squared.error <- (adply(experimental.results, 1, mse, M))
  t(data.frame(mean.squared.error, row.names = 1))
}
```

Setting Up the Experimental Classes

```r
n.experiments = 5000;

n.samples <- c(3, 5, 7, 11, 19)
mu <- seq(-2,2)
sigma <- seq(0.01, 2, length.out = 10)

experimental.classes <- expand.grid(
  mu = mu,
  sigma = sigma,
  n.samples = n.samples)
```

Running the Simulations

```r
set.seed(202)

rmse.values <- adply(experimental.classes, 1, mutate,
                mse = monte.carlo(n.samples, mu, sigma),
                eff = mse[,1] / mse, .parallel = TRUE)
```

Reproducing Figure 6.6

Defining A Plotting Function

```r
g.plotter <- function(this.mu){
  sub <- subset(rmse.values, mu == this.mu & sigma < 1.2)
  p <- ggplot(sub, aes(x = sigma, y = eff[,2]))
  p <- p + geom_rect( xmin = -10, xmax = 10, ymin = 1, ymax = 10, fill = "grey20")
  p <- p + geom_rect( xmin = -10, xmax = 10, ymin = -10, ymax = 1, fill = "grey40")
  p <- p + geom_line(aes(color = as.factor(n.samples)), size = 1.5)
  p <- p + geom_point(aes(color = as.factor(n.samples)), size = 4)
  p <- p + theme(legend.position = "none")
  p <- p + ggtitle(paste("Relative Efficiency of the Arithmetic Mean to the Geometric Mean
                         as Estimators of the Lognormal Population Mean
```

```
                        as a function of scale ($\\sigma$) and sample size (n)
                        Monte Carlo Simulationsm = ", n.experiments, " $\\mu$ =", this.mu))
  p
}
```

```
g.plotter(0)
```