601 HW 3

Vinny Paris

2/11/2018

```
#Simple Comparison
# 1/M * sum(indicator(y>x))
#Above ran 2000 times

mc_approx_up <- function(x){
  holding <- matrix(ncol=11, nrow = 2000)
  holding <- sapply(1:2000, function(y){sum(rnorm(x)<rnorm(x))/x})
  return(holding)
}



#Double Comparison
#1/M*sum_i(1/M*sum_j(y_i>x_j))
#Above ran 2000 times
#simplification of 1/M^2 * sum(expanded(y) > x)
#x cycled

mc_approx_double_up <- function(x){
  holding <- matrix(ncol = 11, nrow = 2000)
  fake_function <- function(y){sum(kronecker(rnorm(x), rep(1,x)) <rnorm(x))/x^2}
  holding <- sapply(1:2000, fake_function)
  return(holding)
}

mm <- c(100,200,300,400,500,600,700,800,900,1000,2000)
```

```
double_mc <- read.csv('ki.csv')[,-1]
mc <- read.csv('ka.csv')[,-1]
```

Relative Precision of Single MC

```
#sample variance vs believed variance for single mc
sample_est <- apply(mc, 2, var)
j <- apply(mc, 2, mean)
bernoulli_est <- j*(1-j)/mm
ratio <- bernoulli_est/sample_est
data <- data.frame(sample_est, bernoulli_est, ratio)
data <- t(data)
colnames(data) <- c("M100", "M200", "M300", "M400", "M500", "M600",
                    "M700", "M800", "M900", "M1000", "M2000")
round(data, 6)
```

```
##                   M100     M200     M300     M400     M500     M600
## sample_est    0.002504 0.001232 0.000828 0.000656 0.000497 0.000413
## bernoulli_est 0.002500 0.001250 0.000833 0.000625 0.000500 0.000417
```

```
## ratio         0.998368 1.014526 1.006371 0.952971 1.005585 1.008340
##                    M700     M800     M900    M1000    M2000
## sample_est    0.000355 0.000312 0.000286 0.000254 0.000126
## bernoulli_est 0.000357 0.000312 0.000278 0.000250 0.000125
## ratio         1.005059 1.001698 0.971735 0.983407 0.995613
```

It appears that the Bernoulli estimator is extremely close and has no advantage/disadvantage over the sample variance. This is fully expected since in effect we have 2000*M bernoulli trials with expected value .5. We merely broke up the trials into two groups (the first one being the sample size M's and the second group being the 2000 N's). The full variance of this then could be written as, for I being total number of 1's in the data frame, I(2000M - I)/(2000M)^2 which is equilavent to p(1-p)/M where p is the grand mean proportion.

Relative Precision of Double MC

```
#sample variance vs believed variance for double mc
sample_est_double <- apply(double_mc, 2, var)
j_double <- apply(double_mc, 2, mean)
bernoulli_est_double <- j_double*(1-j_double)/mm
ratio <- bernoulli_est_double/sample_est_double
data <- data.frame(sample_est_double, bernoulli_est_double, ratio)
data <- t(data)
colnames(data) <- c("M100", "M200", "M300", "M400", "M500", "M600",
                    "M700", "M800", "M900", "M1000", "M2000")
round(data, 6)
```

```
##                          M100     M200     M300     M400     M500     M600
## sample_est_double    0.001700 0.000885 0.000571 0.000447 0.000326 0.000285
## bernoulli_est_double 0.002500 0.001250 0.000833 0.000625 0.000500 0.000417
## ratio                1.470362 1.412272 1.460453 1.399649 1.532206 1.462739
##                          M700     M800     M900    M1000    M2000
## sample_est_double    0.000232 0.000212 0.000189 0.000165 0.000082
## bernoulli_est_double 0.000357 0.000312 0.000278 0.000250 0.000125
## ratio                1.538105 1.475211 1.467331 1.512372 1.517562
```

So here using the expected variance does not work well at all. I believe the main issue for the excess variation in them is that using this formula is still assuming that each bernoulli trial is independent. This is not true though. Take X to be the 'outer' monte carlo sample and Y to be the 'inner'. If, for X_1 and Y_i for i between 1-20 (say), Indicator(X_1 > Y_i) is 1 for all Y_i then I'll put money down on what the indicator will be for X > Y_21. This positive covariance structure reduces the variance we observe compared to what we naively expected. I believe the best way to model this variance would be using the sum of non-identical but independent binomials which according to a quick google search does not appear to be a simple task and is beyond the scope of this assignment (Actually looks like my undergrad advisor might have wrote a paper on this. Kind of cool).

Relative Precision For Double vs Single MC

```
#Final Solution
sample_ratio <- sample_est/sample_est_double
bernoulli_ratio <- bernoulli_est/bernoulli_est_double

data <- t(data.frame(sample_ratio, bernoulli_ratio))
colnames(data) <- c("M100", "M200", "M300", "M400", "M500", "M600",
```

```
                    "M700", "M800", "M900", "M1000", "M2000")
round(data, 6)
```

```
##                      M100     M200     M300     M400     M500     M600
## sample_ratio    1.472768 1.392051 1.451204 1.468721 1.523696 1.450641
## bernoulli_ratio 1.000002 1.000000 0.999998 1.000000 1.000000 1.000000
##                      M700     M800     M900    M1000    M2000
## sample_ratio    1.530360 1.472709 1.510011 1.537891 1.524249
## bernoulli_ratio 0.999998 0.999999 1.000000 1.000000 1.000000
```

So it appears that the double monte carlo procedure, using the sample variances, is about 2/3 more precise than using the the single monte carlo procedure. Trying to estimate the variance the same way we would for iid bernoulli trials does not produce an advantage to either procedure but again, I think it would be a poor decision to use this method. Also worth noting is that it appears to be irrelevant the M size for choosing which method is preferred.