Math 3070 § 1. **Approximating the Hypergeometric** Name: Example
Treibergs **by Binomial Distribution Example.** June 12, 2011

The hypergeometric distribution describes the probability that if a sample of size n is chosen without replacement from a population of N consisting of M successes and $L = N - M$ failures, that there be x successes in the sample. In **R**, this is given by

$$P(X = x) = \mathrm{h}(x; n, M, N) = \texttt{dhyper}(\texttt{x}, \texttt{M}, \texttt{L}, \texttt{n}).$$

Devore's rule of thumb, is that it's O.K. to approximate if the sample is small compared to the population.

> In an experiment where each trial results in S or F, but the sampling is without replacement from a population of size N, if the sample size n is at most 5% of the population size, then the experiment may be analyzed as if it were a binomial experiment.

We print and graph the hypergeometric pmf and its binomial approximation. We see how the errors compare as n/N decreases for several p's

R Session:

```
R version 2.10.1 (2009-12-14)
Copyright (C) 2009 The R Foundation for Statistical Computing
ISBN 3-900051-07-0

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

  Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[R.app GUI 1.31 (5538) powerpc-apple-darwin8.11.1]

[Workspace restored from /Users/andrejstreibergs/.RData]

> # Devore Rule of Thumb for hypergeometric
> # N = number of trials
> # N = population size
> # n = sample size
> # M = number of successes in pop
> # x = number of successes in sample
> N <- 20
> M <- 8
> n <- 7
```

```
>
> ################ SUBPROGRAM TO LIST HYP AND BINOM PMF'S ###################
> # Use fix(listHB) to write function subprogram.
> listHB <- function(N,M,n)
+               {
+                L<- N-M
+                p<- M/N
+                x <- matrix(numeric(3*(n+1)),ncol=3,
+                dimnames=list(0:n,c("dhyper", "dbinom", "difference")))
+                x[,1]<-dhyper(0:n,M,L,n)
+                x[,2]<-dbinom(0:n,n,p)
+                x[,3]<-x[,1]-x[,2]
+                error <- max(abs(x[,3]))
+                if(n/N <= .05)
+                              {answer <- "IS"}
+                else
+                              {answer<-"IS NOT"}
+                cat("Binomial Approximation to Hypergeometric\n Pop. Size N =",
+                N,"\n No. Successes in Pop., M =",M, "\n No. failures in Pop.,
+                L =",L,"\n Proportion of Successes p = M / N =",p,
+                "\n Sample Size n =",n,"\n Sample Frction of Population,
+                n / N =",n/N, "\n Devore's Rule of Thumb ",answer,
+                " satisfied.\n\n")
+                print(x)
+                cat("\n\n Maximum Error =",error,"\n\n")
+               }

> listHB(12,6,5)
Binomial Approximation to Hypergeometric
 Pop. Size N = 12
 No. Successes in Pop., M = 6
 No. failures in Pop., L = 6
 Proportion of Successes p = M / N = 0.5
 Sample Size n = 5
 Sample Frction of Population, n / N = 0.4166667
 Devore's Rule of Thumb  IS NOT  satisfied.

       dhyper  dbinom  difference
0 0.007575758 0.03125 -0.02367424
1 0.113636364 0.15625 -0.04261364
2 0.378787879 0.31250  0.06628788
3 0.378787879 0.31250  0.06628788
4 0.113636364 0.15625 -0.04261364
5 0.007575758 0.03125 -0.02367424


 Maximum Error = 0.06628788
>
>
> #  .066 error is large, but n is 42% of  N
```

```
> listHB(50,6,5)
Binomial Approximation to Hypergeometric
 Pop. Size N = 50
 No. Successes in Pop., M = 6
 No. failures in Pop., L = 44
 Proportion of Successes p = M / N = 0.12
 Sample Size n = 5
 Sample Frction of Population, n / N = 0.1
 Devore's Rule of Thumb  IS NOT  satisfied.

        dhyper       dbinom    difference
0 5.125677e-01 0.5277319168 -1.516419e-02
1 3.844258e-01 0.3598172160  2.460858e-02
2 9.376239e-02 0.0981319680 -4.369579e-03
3 8.929751e-03 0.0133816320 -4.451881e-03
4 3.115030e-04 0.0009123840 -6.008810e-04
5 2.831845e-06 0.0000248832 -2.205135e-05


 Maximum Error = 0.02460858

> listHB(50,25,5)
Binomial Approximation to Hypergeometric
 Pop. Size N = 50
 No. Successes in Pop., M = 25
 No. failures in Pop., L = 25
 Proportion of Successes p = M / N = 0.5
 Sample Size n = 5
 Sample Frction of Population, n / N = 0.1
 Devore's Rule of Thumb  IS NOT  satisfied.

      dhyper  dbinom   difference
0 0.02507599 0.03125 -0.006174012
1 0.14926183 0.15625 -0.006988168
2 0.32566218 0.31250  0.013162180
3 0.32566218 0.31250  0.013162180
4 0.14926183 0.15625 -0.006988168
5 0.02507599 0.03125 -0.006174012


 Maximum Error = 0.01316218
```

```
> listHB(500,250,5)
Binomial Approximation to Hypergeometric
 Pop. Size N = 500
 No. Successes in Pop., M = 250
 No. failures in Pop., L = 250
 Proportion of Successes p = M / N = 0.5
 Sample Size n = 5
 Sample Frction of Population, n / N = 0.01
 Devore's Rule of Thumb  IS  satisfied.

      dhyper  dbinom    difference
0 0.03062564 0.03125 -0.0006243624
1 0.15561808 0.15625 -0.0006319228
2 0.31375629 0.31250  0.0012562852
3 0.31375629 0.31250  0.0012562852
4 0.15561808 0.15625 -0.0006319228
5 0.03062564 0.03125 -0.0006243624


 Maximum Error = 0.001256285

> listHB(500,250,50)
Binomial Approximation to Hypergeometric
 Pop. Size N = 500
 No. Successes in Pop., M = 250
 No. failures in Pop., L = 250
 Proportion of Successes p = M / N = 0.5
 Sample Size n = 50
 Sample Frction of Population, n / N = 0.1
 Devore's Rule of Thumb  IS NOT  satisfied.

         dhyper       dbinom    difference
0  5.823448e-17 8.881784e-16 -8.299439e-16
1  3.621547e-15 4.440892e-14 -4.078737e-14
2  1.093725e-13 1.088019e-12 -9.786460e-13
3  2.137882e-12 1.740830e-11 -1.527041e-11
4  3.041505e-11 2.045475e-10 -1.741324e-10
5  3.357821e-10 1.881837e-09 -1.546055e-09
6  2.995144e-09 1.411378e-08 -1.111863e-08
7  2.219176e-08 8.871517e-08 -6.652341e-08
8  1.393520e-07 4.768440e-07 -3.374920e-07
9  7.529897e-07 2.225272e-06 -1.472282e-06
10 3.542996e-06 9.123616e-06 -5.580620e-06
11 1.465436e-05 3.317678e-05 -1.852243e-05
12 5.369232e-05 1.078246e-04 -5.413223e-05
13 1.753678e-04 3.151795e-04 -1.398117e-04
14 5.132843e-04 8.329743e-04 -3.196899e-04
15 1.352206e-03 1.999138e-03 -6.469324e-04
16 3.218140e-03 4.373115e-03 -1.154975e-03
17 6.940505e-03 8.746230e-03 -1.805724e-03
18 1.359978e-02 1.603475e-02 -2.434972e-03
19 2.426455e-02 2.700590e-02 -2.741355e-03
```

```
20 3.949055e-02 4.185915e-02 -2.368598e-03
21 5.871252e-02 5.979878e-02 -1.086264e-03
22 7.983412e-02 7.882567e-02  1.008450e-03
23 9.936850e-02 9.596169e-02  3.406813e-03
24 1.132867e-01 1.079569e-01  5.329846e-03
25 1.183418e-01 1.122752e-01  6.066676e-03
26 1.132867e-01 1.079569e-01  5.329846e-03
27 9.936850e-02 9.596169e-02  3.406813e-03
28 7.983412e-02 7.882567e-02  1.008450e-03
29 5.871252e-02 5.979878e-02 -1.086264e-03
30 3.949055e-02 4.185915e-02 -2.368598e-03
31 2.426455e-02 2.700590e-02 -2.741355e-03
32 1.359978e-02 1.603475e-02 -2.434972e-03
33 6.940505e-03 8.746230e-03 -1.805724e-03
34 3.218140e-03 4.373115e-03 -1.154975e-03
35 1.352206e-03 1.999138e-03 -6.469324e-04
36 5.132843e-04 8.329743e-04 -3.196899e-04
37 1.753678e-04 3.151795e-04 -1.398117e-04
38 5.369232e-05 1.078246e-04 -5.413223e-05
39 1.465436e-05 3.317678e-05 -1.852243e-05
40 3.542996e-06 9.123616e-06 -5.580620e-06
41 7.529897e-07 2.225272e-06 -1.472282e-06
42 1.393520e-07 4.768440e-07 -3.374920e-07
43 2.219176e-08 8.871517e-08 -6.652341e-08
44 2.995144e-09 1.411378e-08 -1.111863e-08
45 3.357821e-10 1.881837e-09 -1.546055e-09
46 3.041505e-11 2.045475e-10 -1.741324e-10
47 2.137882e-12 1.740830e-11 -1.527041e-11
48 1.093725e-13 1.088019e-12 -9.786460e-13
49 3.621547e-15 4.440892e-14 -4.078737e-14
50 5.823448e-17 8.881784e-16 -8.299439e-16


 Maximum Error = 0.006066676
```

```
> listHB(1000,500,50)
Binomial Approximation to Hypergeometric
 Pop. Size N = 1000
 No. Successes in Pop., M = 500
 No. failures in Pop., L = 500
 Proportion of Successes p = M / N = 0.5
 Sample Size n = 50
 Sample Frction of Population, n / N = 0.05
 Devore's Rule of Thumb  IS  satisfied.

          dhyper       dbinom    difference
0  2.446417e-16 8.881784e-16 -6.435368e-16
1  1.356107e-14 4.440892e-14 -3.084785e-14
2  3.667939e-13 1.088019e-12 -7.212247e-13
3  6.451686e-12 1.740830e-11 -1.095661e-11
4  8.298729e-11 2.045475e-10 -1.215602e-10
5  8.322805e-10 1.881837e-09 -1.049556e-09
6  6.775968e-09 1.411378e-08 -7.337809e-09
7  4.604015e-08 8.871517e-08 -4.267502e-08
8  2.663769e-07 4.768440e-07 -2.104671e-07
9  1.332465e-06 2.225272e-06 -8.928071e-07
10 5.831272e-06 9.123616e-06 -3.292343e-06
11 2.253854e-05 3.317678e-05 -1.063824e-05
12 7.753112e-05 1.078246e-04 -3.029343e-05
13 2.388664e-04 3.151795e-04 -7.631303e-05
14 6.625822e-04 8.329743e-04 -1.703921e-04
15 1.662013e-03 1.999138e-03 -3.371256e-04
16 3.783888e-03 4.373115e-03 -5.892274e-04
17 7.843262e-03 8.746230e-03 -9.029683e-04
18 1.484019e-02 1.603475e-02 -1.194567e-03
19 2.568680e-02 2.700590e-02 -1.319104e-03
20 4.074637e-02 4.185915e-02 -1.112782e-03
21 5.932137e-02 5.979878e-02 -4.774123e-04
22 7.935605e-02 7.882567e-02  5.303754e-04
23 9.762858e-02 9.596169e-02  1.666894e-03
24 1.105273e-01 1.079569e-01  2.570396e-03
25 1.151904e-01 1.122752e-01  2.915208e-03
26 1.105273e-01 1.079569e-01  2.570396e-03
27 9.762858e-02 9.596169e-02  1.666894e-03
28 7.935605e-02 7.882567e-02  5.303754e-04
29 5.932137e-02 5.979878e-02 -4.774123e-04
30 4.074637e-02 4.185915e-02 -1.112782e-03
31 2.568680e-02 2.700590e-02 -1.319104e-03
32 1.484019e-02 1.603475e-02 -1.194567e-03
33 7.843262e-03 8.746230e-03 -9.029683e-04
34 3.783888e-03 4.373115e-03 -5.892274e-04
35 1.662013e-03 1.999138e-03 -3.371256e-04
36 6.625822e-04 8.329743e-04 -1.703921e-04
37 2.388664e-04 3.151795e-04 -7.631303e-05
38 7.753112e-05 1.078246e-04 -3.029343e-05
39 2.253854e-05 3.317678e-05 -1.063824e-05
40 5.831272e-06 9.123616e-06 -3.292343e-06
```

```
41 1.332465e-06 2.225272e-06 -8.928071e-07
42 2.663769e-07 4.768440e-07 -2.104671e-07
43 4.604015e-08 8.871517e-08 -4.267502e-08
44 6.775968e-09 1.411378e-08 -7.337809e-09
45 8.322805e-10 1.881837e-09 -1.049556e-09
46 8.298729e-11 2.045475e-10 -1.215602e-10
47 6.451686e-12 1.740830e-11 -1.095661e-11
48 3.667939e-13 1.088019e-12 -7.212247e-13
49 1.356107e-14 4.440892e-14 -3.084785e-14
50 2.446417e-16 8.881784e-16 -6.435368e-16


 Maximum Error = 0.002915208

> listHB(100,10,6)
Binomial Approximation to Hypergeometric
 Pop. Size N = 100
 No. Successes in Pop., M = 10
 No. failures in Pop., L = 90
 Proportion of Successes p = M / N = 0.1
 Sample Size n = 6
 Sample Frction of Population, n / N = 0.06
 Devore's Rule of Thumb  IS NOT  satisfied.

        dhyper   dbinom    difference
0 5.223047e-01 0.531441 -9.136251e-03
1 3.686857e-01 0.354294  1.439171e-02
2 9.645847e-02 0.098415 -1.956531e-03
3 1.182633e-02 0.014580 -2.753674e-03
4 7.055478e-04 0.001215 -5.094522e-04
5 1.902601e-05 0.000054 -3.497399e-05
6 1.761668e-07 0.000001 -8.238332e-07


 Maximum Error = 0.01439171
```

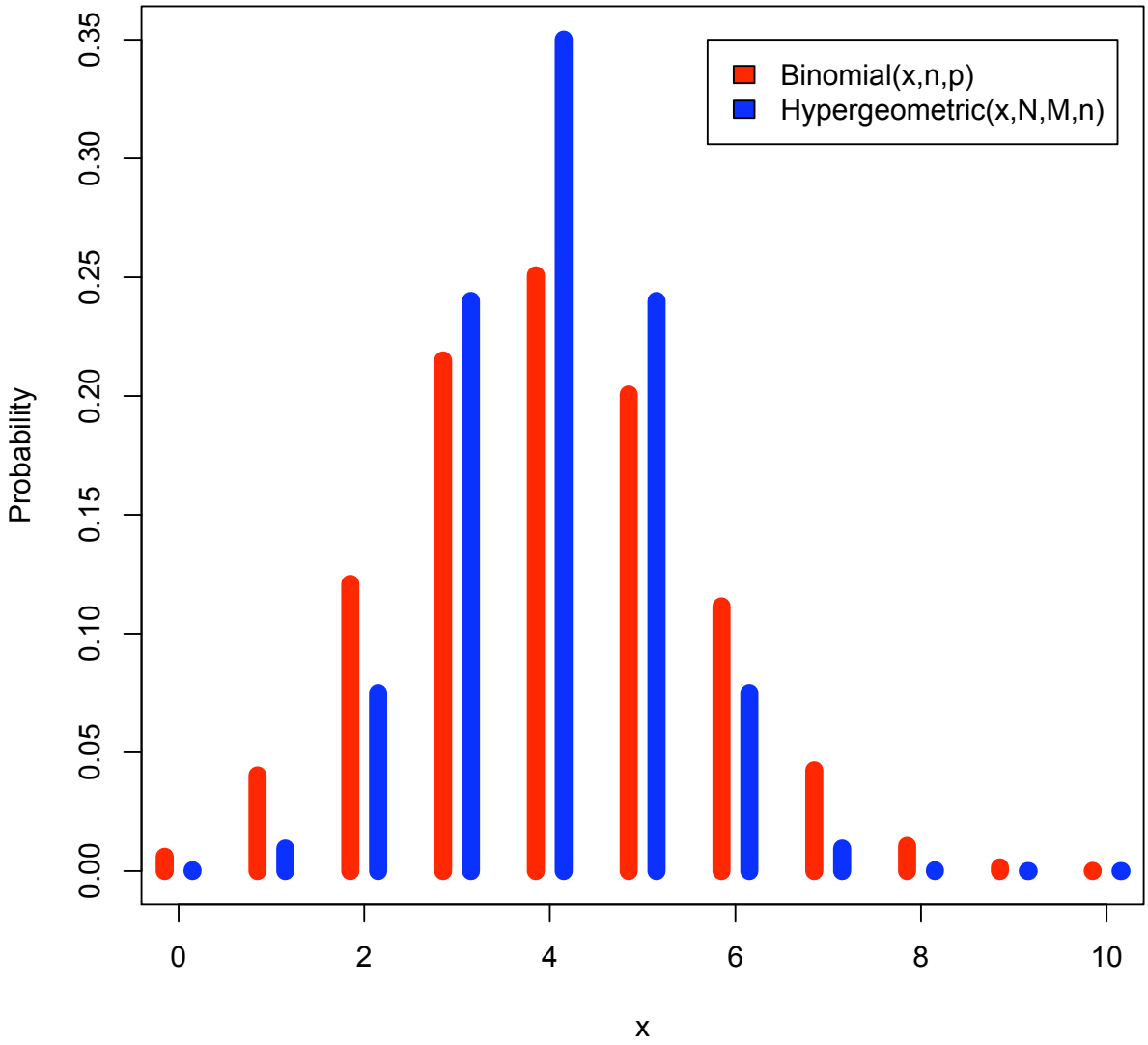
```
> listHB(200,10,6)
Binomial Approximation to Hypergeometric
 Pop. Size N = 200
 No. Successes in Pop., M = 10
 No. failures in Pop., L = 190
 Proportion of Successes p = M / N = 0.05
 Sample Size n = 6
 Sample Frction of Population, n / N = 0.03
 Devore's Rule of Thumb  IS  satisfied.

        dhyper       dbinom    difference
0 7.321404e-01 7.350919e-01 -2.951507e-03
1 2.374509e-01 2.321343e-01  5.316654e-03
2 2.872390e-02 3.054398e-02 -1.820081e-03
3 1.638440e-03 2.143438e-03 -5.049974e-04
4 4.575431e-05 8.460937e-05 -3.885506e-05
5 5.810071e-07 1.781250e-06 -1.200243e-06
6 2.548277e-09 1.562500e-08 -1.307672e-08


 Maximum Error = 0.005316654


> listHB(200,20,6)
Binomial Approximation to Hypergeometric
 Pop. Size N = 200
 No. Successes in Pop., M = 20
 No. failures in Pop., L = 180
 Proportion of Successes p = M / N = 0.1
 Sample Size n = 6
 Sample Frction of Population, n / N = 0.03
 Devore's Rule of Thumb  IS  satisfied.

        dhyper   dbinom    difference
0 5.269439e-01 0.531441 -4.497138e-03
1 3.613329e-01 0.354294  7.038934e-03
2 9.751883e-02 0.098415 -8.961684e-04
3 1.322289e-02 0.014580 -1.357108e-03
4 9.471454e-04 0.001215 -2.678546e-04
5 3.386442e-05 0.000054 -2.013558e-05
6 4.703391e-07 0.000001 -5.296609e-07


 Maximum Error = 0.007038934
```

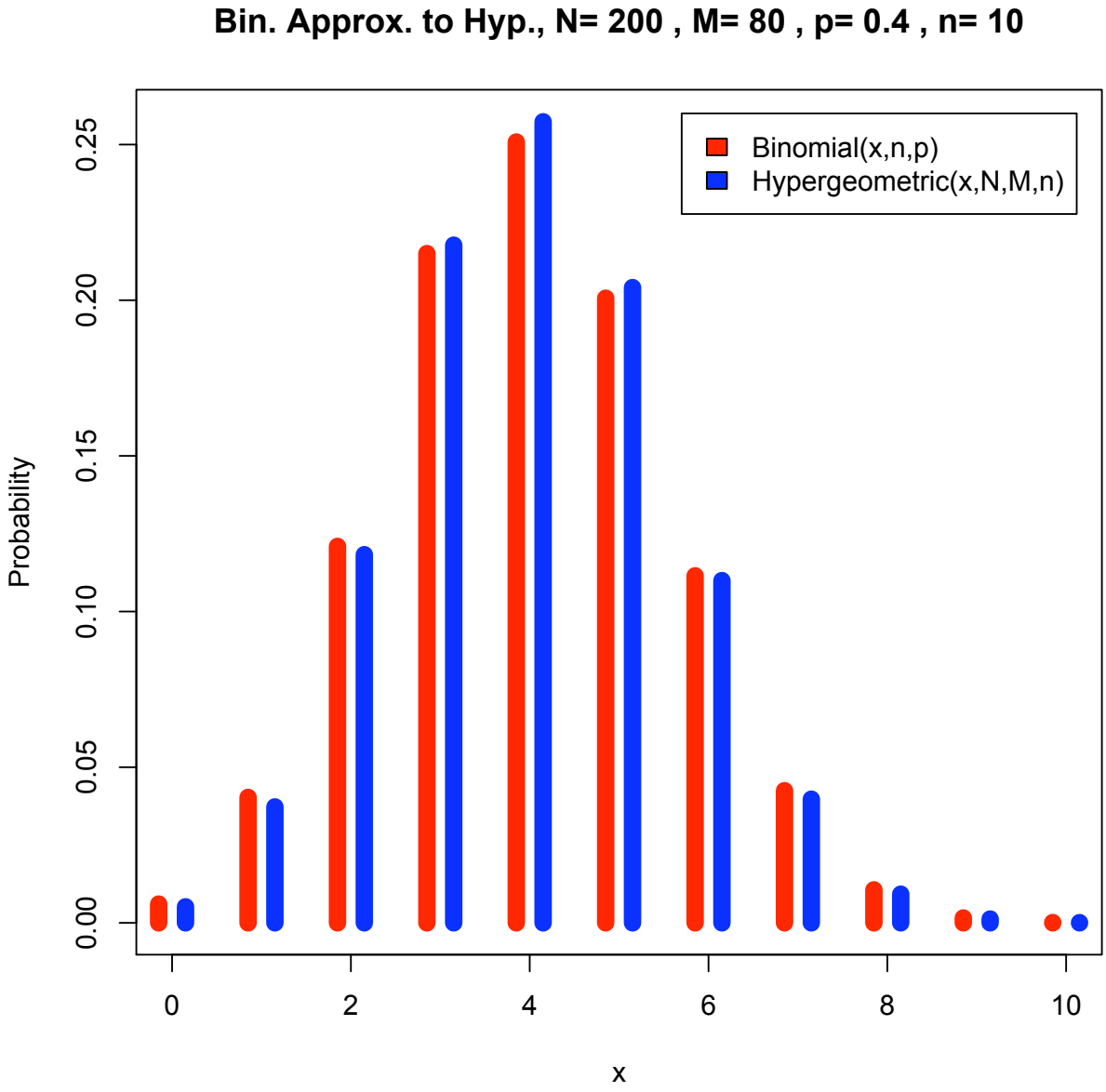
```
> listHB(200,100,6)
Binomial Approximation to Hypergeometric
 Pop. Size N = 200
 No. Successes in Pop., M = 100
 No. failures in Pop., L = 100
 Proportion of Successes p = M / N = 0.5
 Sample Size n = 6
 Sample Frction of Population, n / N = 0.03
 Devore's Rule of Thumb  IS  satisfied.

      dhyper   dbinom   difference
0 0.01446514 0.015625 -0.001159859
1 0.09135879 0.093750 -0.002391215
2 0.23553437 0.234375  0.001159368
3 0.31728341 0.312500  0.004783410
4 0.23553437 0.234375  0.001159368
5 0.09135879 0.093750 -0.002391215
6 0.01446514 0.015625 -0.001159859


 Maximum Error = 0.00478341
```

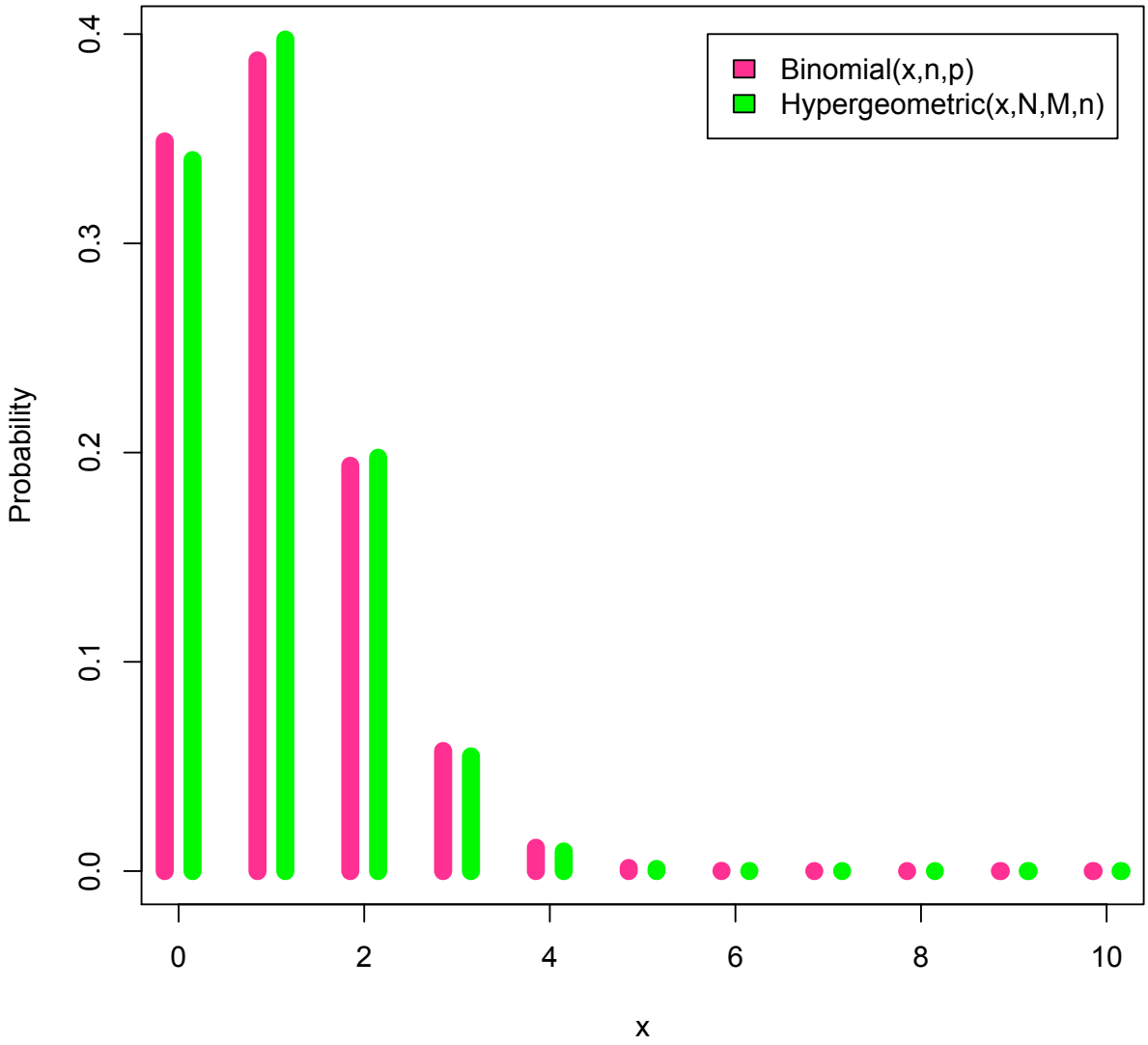
```
> ############## FUNCTION TO COMPUT ERROR OVER ENTIRE LIST ###################
> # To tabulate errors over various choices of N, M, n
> maxd <- function(N,M,n){}
> fix(maxd)
>
> maxd <- function(N,M,n)
+                        {
+                         L <- N-M
+                         p <- M/N
+                         max(abs(dhyper(0:n,M,L,n)-dbinom(0:n,n,p)))
+                        }

> maxd(200,10,6)
[1] 0.005316654


> Ns <- c(200,400,1000,2000)
> Ms <-.2*Ns
> ns <- c(5,10,20,50,100,200)
> Ms
[1]  40  80 200 400
> Ns
[1]  200  400 1000 2000
> ns
[1]   5  10  20  50 100 250

> y <- matrix(numeric(24),ncol=4,dimnames=list(ns,Ns))
> for(i in 1:6){
+               for(j in 1:4){
+                             y[i,j]<- maxd(Ns[j],Ms[j],ns[i])
+                            }
+              }
> y
>
> # n\N  Max Error Table for  p = .2
>
            200         400        1000         2000
5   0.005194780 0.002578546 0.001026953 0.0005127372
10  0.007809532 0.003838583 0.001520028 0.0007574849
20  0.011747070 0.005655574 0.002213311 0.0010987637
50  0.021512405 0.009630446 0.003628964 0.0017804114
100 0.040824680 0.015320001 0.005366314 0.0025784830
200 0.929630401 0.029039258 0.008297400 0.0038046684
```

```
> Ms <-.1*Ns
> for(i in 1:6){
+ for(j in 1:4){
+ y[i,j]<- maxd(Ns[j],Ms[j],ns[i])}}
> y
>
> # n\N  Max Error Table for  p = .1
>

            200         400        1000         2000
5   0.005567055 0.002758405 0.001097423 0.0005477289
10  0.009976442 0.004914302 0.001948439 0.0009713762
20  0.015285460 0.007376067 0.002890347 0.0014354635
50  0.028308379 0.012708635 0.004795623 0.0023537993
100 0.053836860 0.020292772 0.007119921 0.0034226446
200 0.906363689 0.038506820 0.011030217 0.0050604070
> Ms
[1]  20  40 100 200
> Ns
[1]  200  400 1000 2000
> Ms <-.05*Ns
> Ms
[1]  10  20  50 100
> for(i in 1:6){
+ for(j in 1:4){
+ y[i,j]<- maxd(Ns[j],Ms[j],ns[i])}}
> y
>
> # n\N  Max Error Table for  p = .05
>

            200         400         1000         2000
5   0.003823851 0.001893539 0.0007530712 0.0003758172
10  0.011644430 0.005707603 0.0022564285 0.0011238505
20  0.020043329 0.009717930 0.0038181004 0.0018978370
50  0.036886535 0.016458374 0.0061925967 0.0030368332
100 0.072452423 0.027559691 0.0097023964 0.0046684448
200 0.871642627 0.052411041 0.0150905960 0.0069306180
>
>
>
```

```
> ################### PLOT HYPERGEOMETRIC AND BINOMIAL PMF'S ##############
> N <- 20
> M <- 8
> L <- N-M
> p <- M/N; p
[1] 0.4
> n <- 10
> plot(c(0:n,0:n),c(dbinom(0:n,n,p), dhyper(0:n,M,L,n)), type = "n",
+      main = paste("Bin. Approx. to Hyp., N=", N, ", M=", M, ", p=", p,
+      ", n=",n), ylab = "Probability",xlab="x")
> points((0:n)-.15,dbinom(0:n,n,p), type = "h", col = "red", lwd=10)
> points((0:10)+.15,dhyper(0:n,M,L,n), type = "h", col = "blue", lwd=10)
> legend(5.7,.35,legend=c("Binomial(x,n,p)","Hypergeometric(x,N,M,n)"),
+        fill=c("red","blue"),bg="white")
> # M3074ApproxHyp1.pdf
>
> N <- 200
> M <- 80
> L <- N-M
> p <- M/N; p
[1] 0.4
> n <- 10
> plot(c(0:n,0:n),c(dbinom(0:n,n,p),dhyper(0:n,M,L,n)), type = "n",
+      main=paste("Bin. Approx. to Hyp., N=",N,", M=",M,", p=",p,",
+      n=",n),ylab="Probability",xlab="x")
> points((0:n)-.15,dbinom(0:n,n,p), type = "h", col = "red", lwd=10)
>
> points((0:10)+.15,dhyper(0:n,M,L,n), type = "h", col = "blue", lwd=10)
> legend(5.7,.26,legend=c("Binomial(x,n,p)","Hypergeometric(x,N,M,n)"),
+        fill=c("red","blue"),bg="white")
> # M3074ApproxHyp2.pdf
>
> N <- 200
> M <- 20
> L <- N-M
> p <- M/N; p
[1] 0.1
> n <- 10
> plot(c(0:n,0:n), c(dbinom(0:n,n,p), dhyper(0:n,M,L,n)), type = "n",
+      main = paste("Bin. Approx. to Hyp., N=", N, ", M=", M, ", p=", p,
+      ", n=", n), ylab = "Probability", xlab="x")
> points((0:n)-.15,dbinom(0:n,n,p), type = "h", col = clr[12], lwd=10)
> points((0:10)+.15,dhyper(0:n,M,L,n), type = "h", col = clr[5], lwd=10)
> legend(5.7,.4, legend = c("Binomial(x,n,p)", "Hypergeometric(x,N,M,n)"),
+        fill = clr[c(12,5)], bg="white")
> #M3074ApproxHyp3.pdf
```


Bin. Approx. to Hyp., N= 20 , M= 8 , p= 0.4 , n= 10
Binomial(x,n,p)
Hypergeometric(x,N,M,n)
Probability
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0
2
4
6
8
10
x


Bin. Approx. to Hyp., N= 200 , M= 80 , p= 0.4 , n= 10
Binomial(x,n,p)
Hypergeometric(x,N,M,n)
Probability
0.00
0.05
0.10
0.15
0.20
0.25
0
2
4
6
8
10
x


Bin. Approx. to Hyp., N= 200 , M= 20 , p= 0.1 , n= 10
Binomial(x,n,p)
Hypergeometric(x,N,M,n)
Probability
0.0
0.1
0.2
0.3
0.4
0
2
4
6
8
10
x