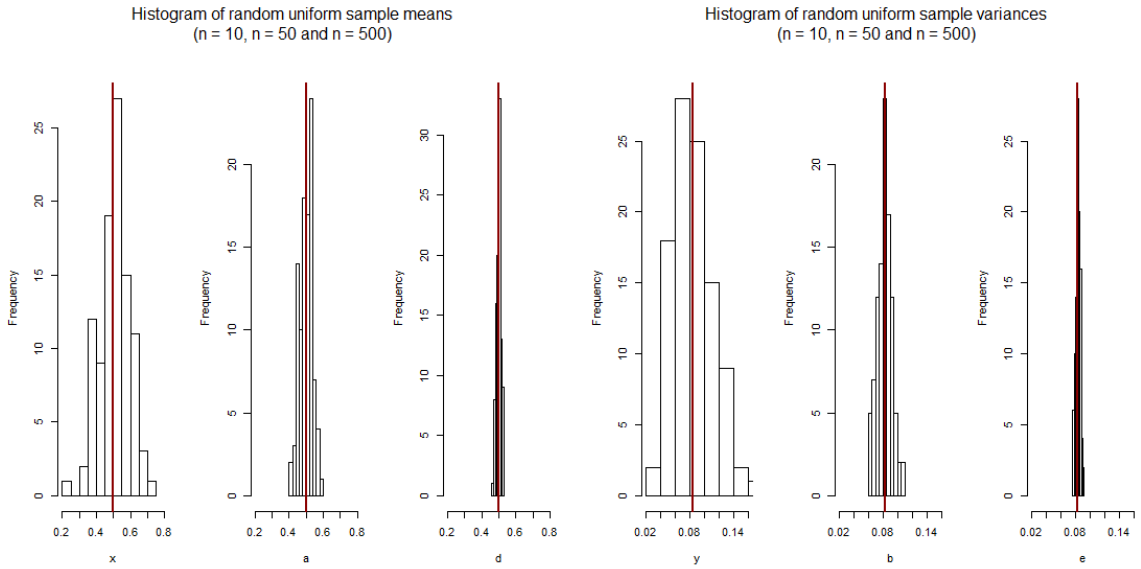
NORTHWESTERN UNIVERSITY | SCHOOL OF PROFESSIONAL STUDIES

R Lesson 7 - Solutions
MSPA 401 – Introduction to Statistical Analysis

1) Use the uniform distribution over the interval 0 to 1. Draw 100 random samples of size 10. Calculate the means for each sample. Using the 100 mean values plot a histogram. Repeat with 100 random samples of size 50. Repeat with 100 samples of size 500. Present the three histograms using par(). Calculate the variance of each histogram and compare to the original uniform distribution. What do you conclude?



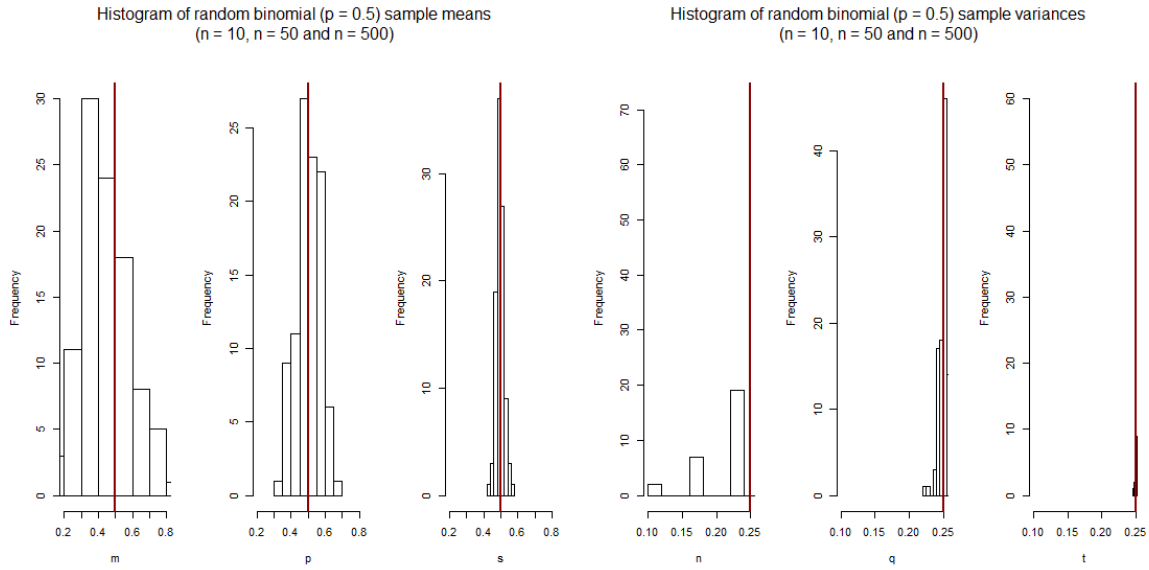
The random uniform sample means and sample variances converge to the respective "true" values as *n* is increased.

2) Using the histogram determined above for samples of size 50, find the quartiles. Using the normal distribution with the true mean and variance for a uniform distribution over the interval 0 to 1, determine the theoretical quartiles for a sample mean from 50 observations. Compare the two sets of quartiles. What do you conclude?

```
> quantile(a)
       0%       25%       50%       75%      100%
0.4021239 0.4674217 0.5029306 0.5290173 0.5946016
> quantile(j)
       0%       25%       50%       75%      100%
0.4158362 0.4810670 0.5022179 0.5252440 0.6535995
```

The two sets of quartiles are similar up through the third quartile and may converge as sample size increases. Largest difference is observed in the fourth quartile.

3) Use the binomial distribution with p= 0.5. Draw 100 random samples of size 10. Calculate the means for each sample. Using the 100 mean values plot a histogram. Repeat with 100 random samples of size 50. Repeat with 100 samples of size 500. Present the three histograms using par(). Calculate the variance of each histogram and compare to the original mean and variance for the binomial. What do you conclude?



The binomial sample means and sample variances converge to the respective "true" values as *n* is increased.

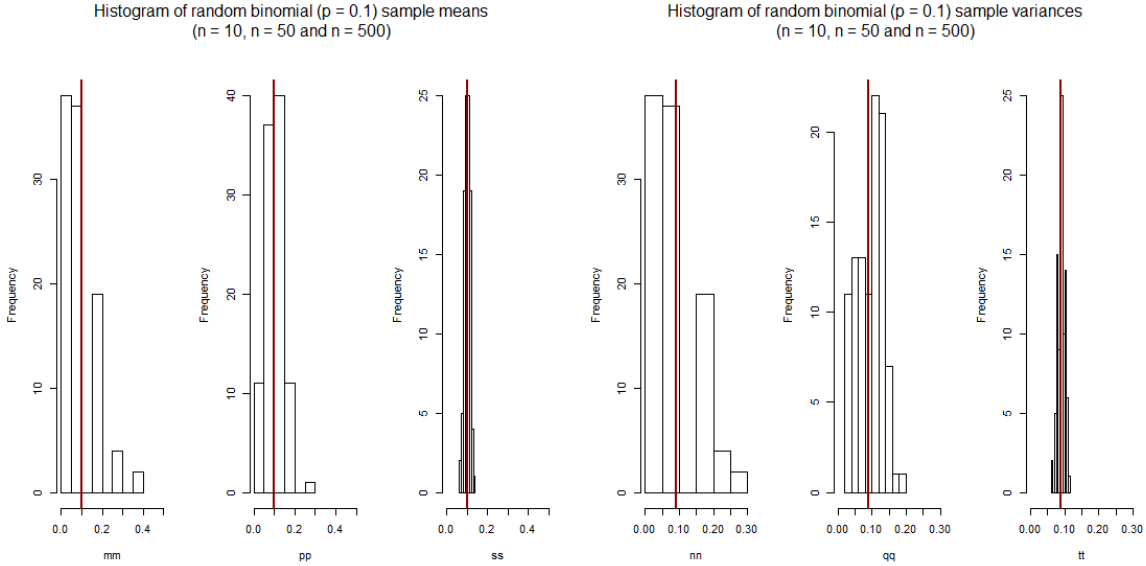
4) Using the histogram determined above for samples of size 50, find the quartiles. Using the normal distribution with the true mean and variance for a binomial distribution with p = 0.5, determine the theoretical quartiles for a sample mean from 50 observations. Compare the two sets of quartiles. What do you conclude?

```
> quantile(p)
  0%  25%  50%  75% 100%
0.32 0.46 0.52 0.56 0.66
> quantile(v)
       0%       25%       50%       75%      100%
0.3328496 0.4511897 0.4955020 0.5505300 0.6580152
```

The two sets of quartiles are very similar, likely to converge as sample size is increased.

5) Use the binomial distribution with p= 0.1. Draw 100 random samples of size 10. Calculate the means for each sample. Using the 100 mean values plot a histogram. Repeat with 100 random samples of size 50. Repeat with 100 samples of size 500. Present the three histograms using

par(). Calculate the variance of each histogram and compare to the original mean and variance for the binomial. What do you conclude?



The binomial sample means and sample variances converge to the respective "true" values as *n* is increased.

6) Using the histogram determined above for samples of size 50, find the quartiles. Using the normal distribution with the true mean and variance for a binomial distribution with p = 0.1, determine the theoretical quartiles for a sample mean from 50 observations. Compare the two sets of quartiles. What do you conclude?

```
> quantile(pp)
  0%  25%  50%  75% 100%
0.02 0.08 0.12 0.14 0.26
> quantile(vv)
          0%          25%          50%          75%         100%
-0.02883012  0.08130462  0.10593952  0.13540200  0.19509431
```

The quartiles appear dissimilar, especially the first and fourth. A normal distribution, with a mean of 0.1, is not left-bound at zero as is a binomial distribution. However, with increased sample size, it is likely the distributions will further converge.