Extra exercise 5

The results of this run of the PSLS "Probability"-applet were typed into Minitab (the first two columns below, labeled “trials” and “count”), and displayed together with the observed proportion and the deviations from the expected numbers for both proportions and counts:

```
MTB > name c1 'trials'
MTB > name c2 'count'
MTB > Name C3 'proportion'
MTB > Let 'proportion' = 'count'/'trials'
MTB > Name C4 'diffprop'
MTB > Let 'diffprop' = 'proportion'-0.5
MTB > Name C5 'diffcount'
MTB > Let 'diffcount' = 'count'-0.5*'trials'
MTB > Print 'trials'-'diffcount'.
```

Data Display

Data

Row	trials	count	proportion	diffprop	diffcount
1	50	23	0.460000	-0.0400000	-2
2	100	50	0.500000	0.0000000	0
3	200	93	0.465000	-0.0350000	-7
4	400	183	0.457500	-0.0425000	-17
5	1000	473	0.473000	-0.0270000	-27
6	2000	979	0.489500	-0.0105000	-21
7	4000	1963	0.490750	-0.0092500	-37
8	6000	2975	0.495833	-0.0041667	-25
9	8000	3958	0.494750	-0.0052500	-42
10	10000	4967	0.496700	-0.0033000	-33

Comments:
From the discussion in the lecture and the text(s) we would expect the observed proportions to get closer to 0.5 as the number of trials (n) increases. This does indeed appear to be the case, but perhaps not as rapidly as one would expect. The deviation from 0.5 after 10,000 trials above is well within the statistical uncertainty (we will later in the course see how to assess this).

It is less clear from the results above whether the deviations between the observed and expected counts will stabilize as well. It can be shown mathematically that a stabilization will not happen, and in fact the difference between the observed and expected counts will not remain within any bounds around zero, as n increases. For example, the probability that these two numbers differ by at most 100 will tend to zero as n increases to very large numbers. It is a mathematical result and in a sense of less practical use when one does not know how large n needs to be for this to happen.

From a more practical perspective we can just note that the difference between observed and expected counts is “unpredictable” in the sense that it can be large in either direction. The limitation to how large it can be, follows from when dividing the difference by n, the value will get successively smaller (towards zero) as n increases.