1) Variation between randomizations will decrease as you increase the sample size, and increase as you reduce the sample size.

With sample sizes of 100, the means move +/- .25 around 2 (pop1) and 5 (pop2); standard deviations vary from 1.1 to 1.4 for pop1 and 1.2 to 1.7 for pop2. With sample sizes of 1000, changes of the sample mean between randomizations drop to less than .1, and changes of standard deviation drop to less than .2. Sample sizes of 20 massively increase variability between randomizations, with mean and standard deviation values reaching significantly higher and lower than in the sample size 100 runs.

2) I expect the p value to rise and t-value to fall, as the two samples get closer in mean and standard deviation due to the convergence of p in the binomial function.

Sample1 p = .3: t = 10.67183824060483, pvalue = 2.705154846456434e-21

Sample1 p = .4: t = 4.555020581273329, pvalue = 9.146047524394654e-06

The closer together the inputs to the set generation get, the more that the resulting samples will resemble each other, on average. Thus, the t-value drops, and p-value increases.

3) Using the laplace distribution, I made two pairs of populations, each with one of its variables kept the same and the other multiplied by 10 in the second population. The results were illustrative, though not particularly surprising, given the nature of this distribution.

The first pair kept the location of the mean the same, but multiplied the scale by 10; this, naturally, made the spread of the bell curve produced by sample2’s histogram *much* broader compared to the narrow peak of sample1, causing the p-value between the two to be over 0.5, and the t-value to bounce around between 0 +/- 2, depending on the randomization.

The second pair multiplied the location of the mean by 10, and kept the scale the same. This merely moves the center of the distribution over, without changing the shape or scale of it in any meaningful or consistent way. Thus, the p-value of the two samples was so low as to effectively be zero, and the T-value was in the 40s.

One question I have on terminology: what is the proper term for what I’ve been calling randomizations, i.e. a particular run of the random-number-generator function?