CMPE 320

Spring 2023

Project III: The Central Limit Theorem

Cole Cavanagh

Monday/Wednesday 4:00-5:00

Dr. Chuck LaBerge

1. Introduction

In this project I will generate different sets of random variables and demonstrate the Central Limit Theorem. I will do this by summing these sets together and show that as more sets are summed, the result approaches a gaussian distribution no matter what the random variables are based on.

2. Simulation and Discussion

2.1. Sum of IID Random Variables U(0,1)

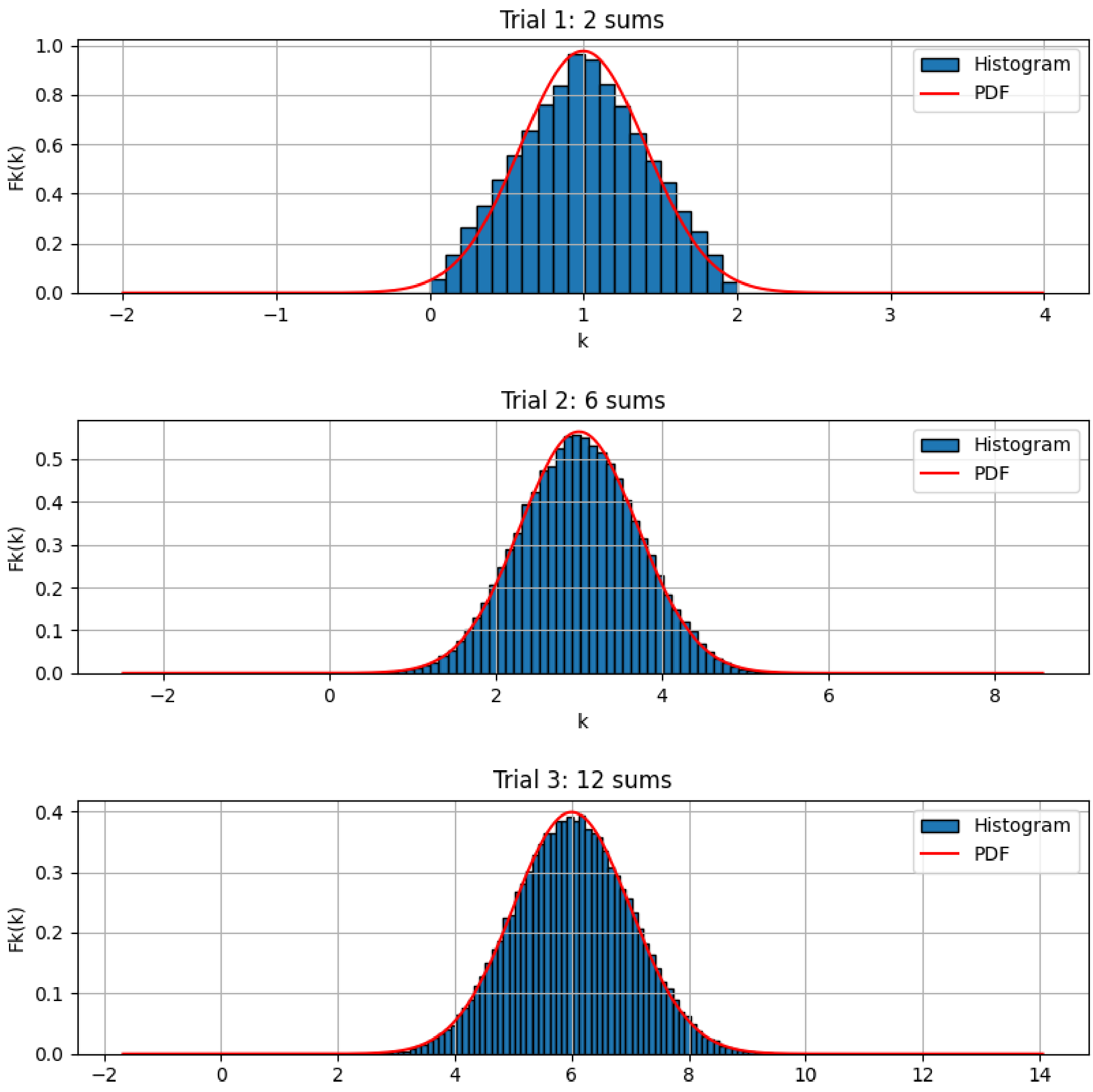


Figure 1: Random Variable and PDF of Random Variable

Table 1: Mean and Variance

	Calculated Mean	Actual Mean	Calculated Variance	Actual Variance
Trial 1	1.00	0.99	0.16	0.16
Trial 2	3.00	2.99	0.50	0.49
Trial 3	6.00	6.00	1.00	1.00

Section 2.1 used a random variable with a uniform distribution between 0 and 1. Trial 1 through 3 summed up 2, 6, and 12 sets of 100000 variables each respectively. Interestingly, even with summing 2 sets of random variables, the result very closely matched a gaussian distribution. All trials means and variances were extremely close to their calculated counterparts demonstrating the power of the Central Limit Theorem even at low sum counts.

2.2. Sum of IID Discrete Random Variables U(1,8)

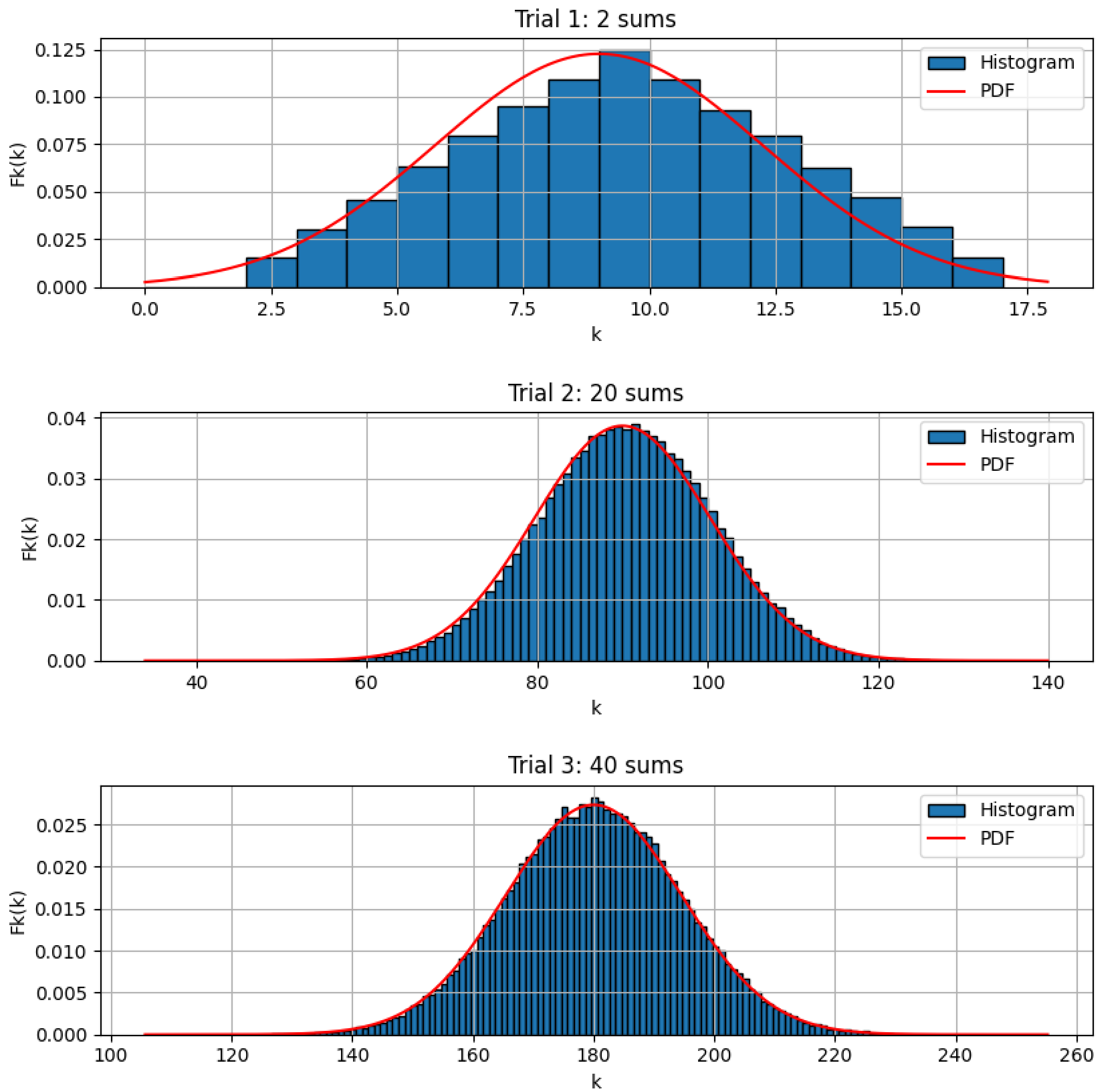


Figure 2: Random Variable and PDF of Random Variable

Table 2: Mean and Variance

	Calculated Mean	Actual Mean	Calculated Variance	Actual Variance
Trial 1	9.00	9.00	10.58	10.45
Trial 2	90.00	89.98	106.58	103.61
Trial 3	180.00	179.99	213.25	208.93

Section 2.2 used a random variable with the uniform distribution of an 8-sided die. Trial 1 through 3 summed up 2, 20, and 40 sets of 100000 variables each respectively. Again, even with summing 2 sets of random variables, the result very closely matched a gaussian distribution. Although the calculated and actual means matched, the graph of Trial 1 shows that the histogram does not perfectly match the PDF and is slightly shifted higher. This can be partially described by the discrepancy between the calculated and actual variances.

2.3. Sum of IID Random Variables from $F_x(x) = 0.5e^{-0.5(x-2)}U(x-2)$

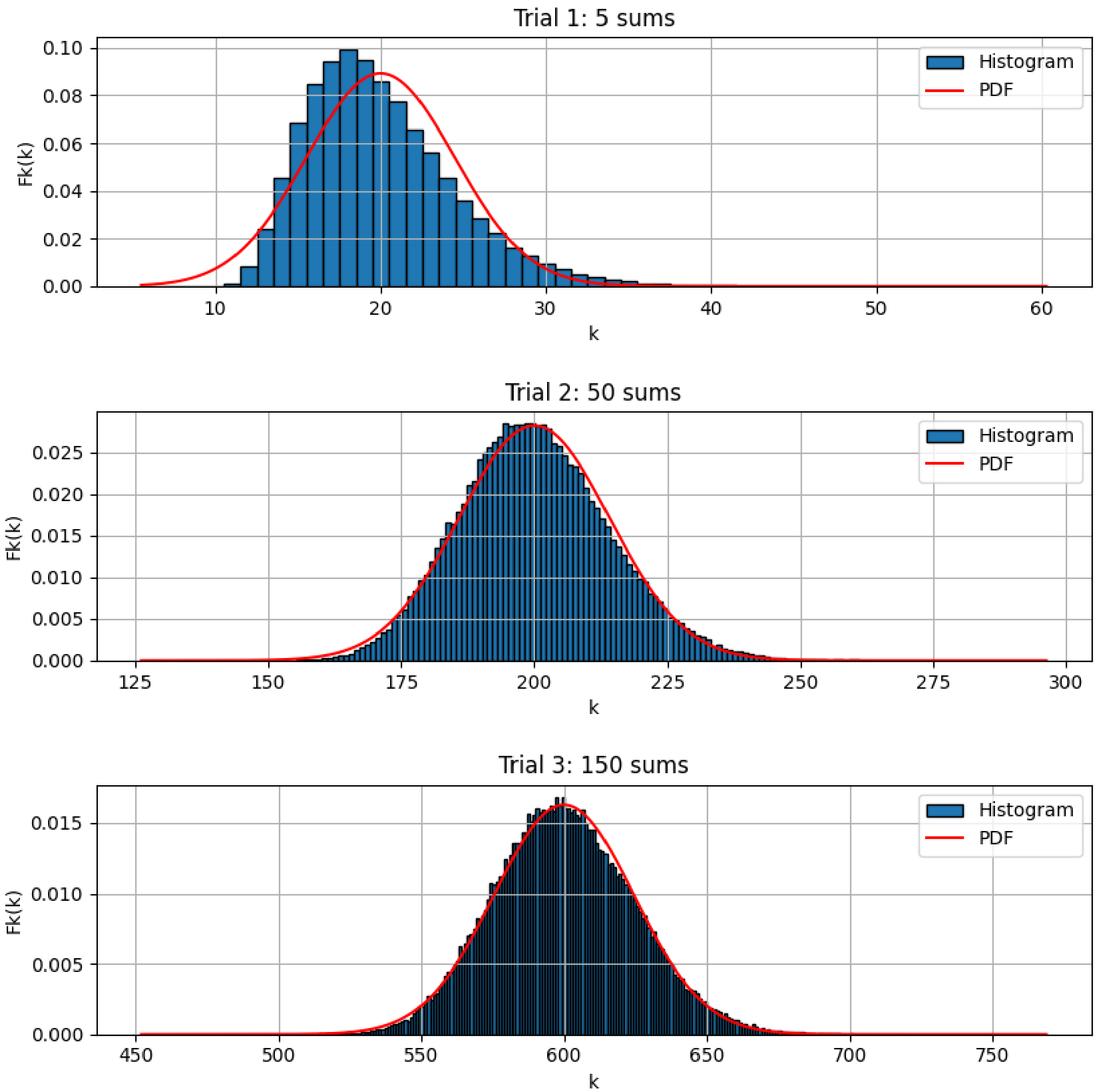


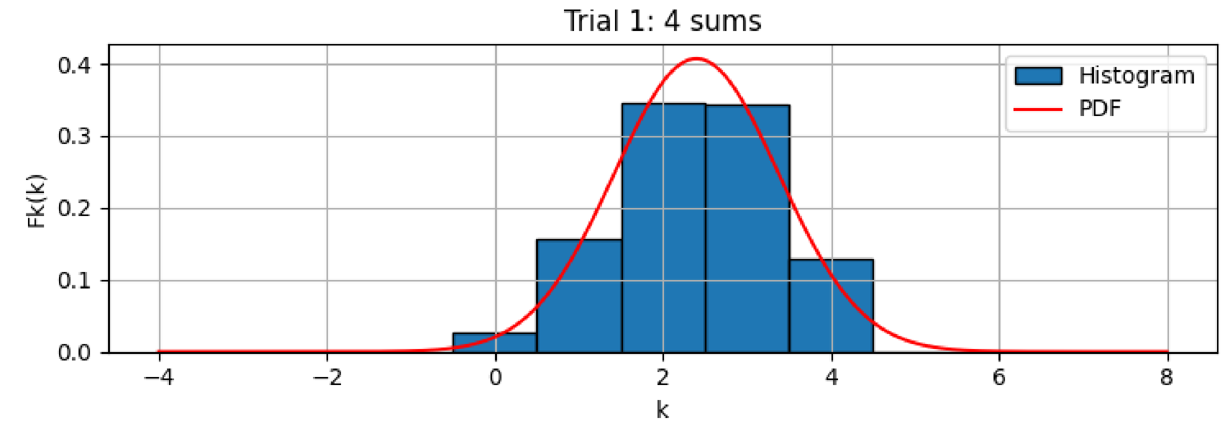
Figure 3: Random Variable and PDF of Random Variable

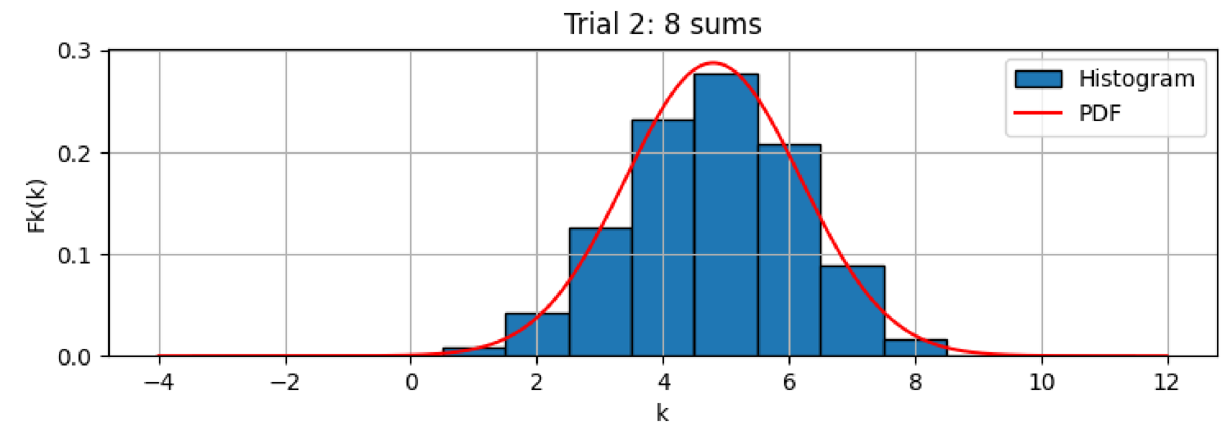
Table 3: Mean and Variance

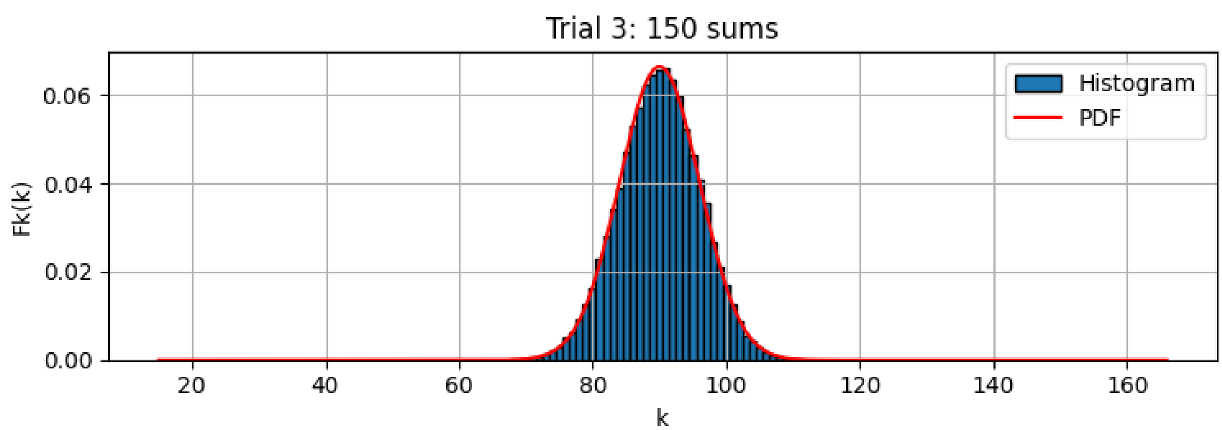
	Calculated Mean	Actual Mean	Calculated Variance	Actual Variance
Trial 1	20.00	19.99	20.00	20.08
Trial 2	200.00	199.97	200.00	198.63
Trial 3	600.00	599.94	600.00	606.02

Section 2.3 used a random variable with the distribution $F_x(x) = 0.5e^{-0.5(x-2)}$. Trial 1 through 3 summed up 5, 50, and 150 sets of 100000 variables each respectively. This distribution creates an impulse at 2 causing the histogram to not match the PDF as well in Trial 1. Even with this, after only 5 sums in Trial 1 the means and variances were very close. Despite the distribution, Trials 2 and 3 prove the Central Limit Theorem showing that at a large enough number of sets the histogram approaches Gaussian distribution.

2.4. Sum of IID Bernoulli Trials







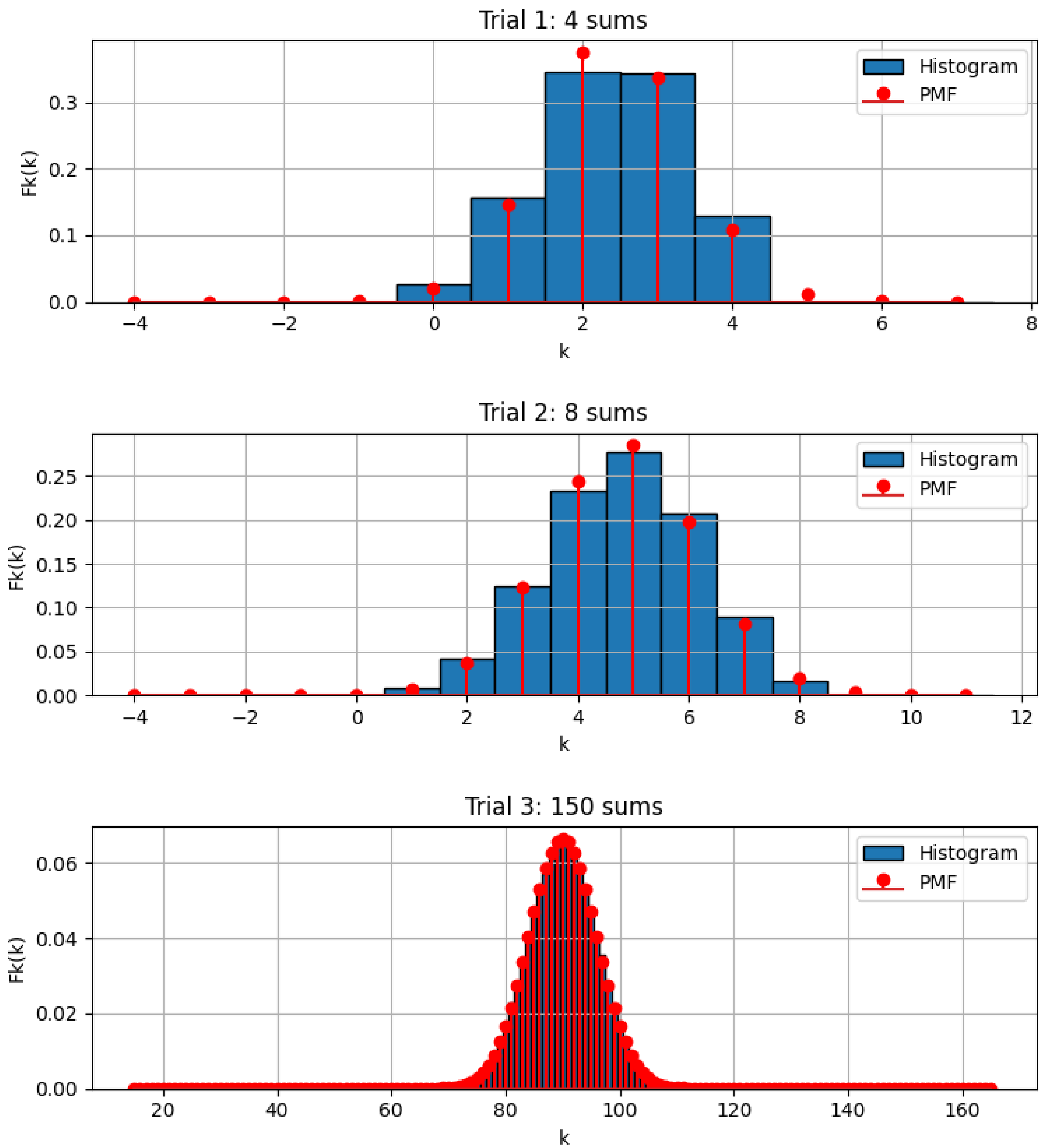


Figure 4: Random Variable and PDF of Random Variable

Table 4: Mean and Variance

	Calculated Mean	Actual Mean	Calculated Variance	Actual Variance
Trial 1	2.40	2.39	.96	.96
Trial 2	4.80	4.80	1.92	1.91
Trial 3	90.00	89.99	36.00	35.79

Section 2.4 used a random variable with the Bernoulli distribution with Pr[k = 1] = 0.6. Trial 1 through 3 summed up 4, 8, and 150 sets of 100000 variables each respectively. Figure 4 shows the histogram plotted against both the PDF and PMF. The histogram for Trial 1 does not seem to closely match the PDF due to the large bin sizes in comparison to the variance. This is required due to the sum's whole number values. The closeness can be seen instead when looking at the histogram plotted against the PMF and the closeness of the calculated and actual mean and variance values.

3. What I learned

Project 3 helped me fully understand the power of the Central Limit Theorem. Even with random variables having extreme differences in distribution, with a sufficient number of sums the histograms always approached Gaussian distributions. This was supported by both the graphs and calculated vs actual means and variance.