EE 313 Homework 1: Central Limit Theorem

Ömer Alperen Katı
2017401069
omer.kati@boun.edu.tr

February 8, 2021

1

1.1

Let A be a random variable which can be represented as the sum of a group of random variables.

$$A = K_1 + K_2 + ... + K_n \tag{1}$$

where $K_1, K_2, ..., K_n$ are independent and identically distributed (iid). Therefore, we can write

$$E[K] = E[K_1] = E[K_2] = ... = E[K_n], \tag{2}$$

and

$$Var[K] = Var[K_1] = Var[K_2] = ... = Var[K_n]. \tag{3}$$

Using (1), Theorem 6.1, and (2),

$$E[A] = E[K_1 + K_2 + ... + K_n] = E[K_1] + E[K_2] + ... + E[K_n] = \boxed{nE[K]} \tag{4}$$

Also, Theorem 6.2 states that, the variance of $W_n = X_1 + \cdots + X_n$ is

$$Var[W_n] = \sum_{i=1}^{n} Var[X_i] + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} Cov[X_i, X_j] \tag{5}$$

Put A in the place of W, and K that of X. Then, since those Ks are independent, covariances are zero. It becomes,

$$Var[A] = \sum_{i=1}^{n} Var[K_i] = \boxed{nVar[K]}. \tag{6}$$

2 Part 2

2.1

Let $K_1, K_2, ..., K_n$ be a sequence of uniform random variables with the uniformity range of [-5, 5].

Since all Kn's are uniformly distributed in the range of [-5, 5], $E[K_n] = E[K] = 0$, and $nE[K] = 0$ for all n. Also, $Var[K_n] = Var[K]$ for each n, since they're iid. $Var[K]$ can be calculated as

$$Var[K] = E[K^2] - (E[K])^2 = \int_{-5}^{5} k^2 \frac{1}{10} dk - (\int_{-5}^{5} k \frac{1}{10} dk)^2 = \boxed{\frac{25}{3}} \tag{7}$$

$$Var[A] = nVar[K] = n\frac{25}{3} \tag{8}$$

Var[A] is calculated to be [8.3, 16.67, 25, 41.67, 83.3],
$\sigma_A = \sqrt{Var[A]} = [2.8868, 4.0825, 5.0000, 6.4550, 9.1287]$, where n = [1, 2, 3, 5, 10] respectively. The vertical lines representing these parameters can be seen in Fig. 1. As n increases, histograms converges more to a bell-shaped curve, and for n = 10, the Gaussian curves constructed by the calculated parameters almost fits to the distribution, as in Fig. (1). This is consistent with the statement of the central limit theorem.

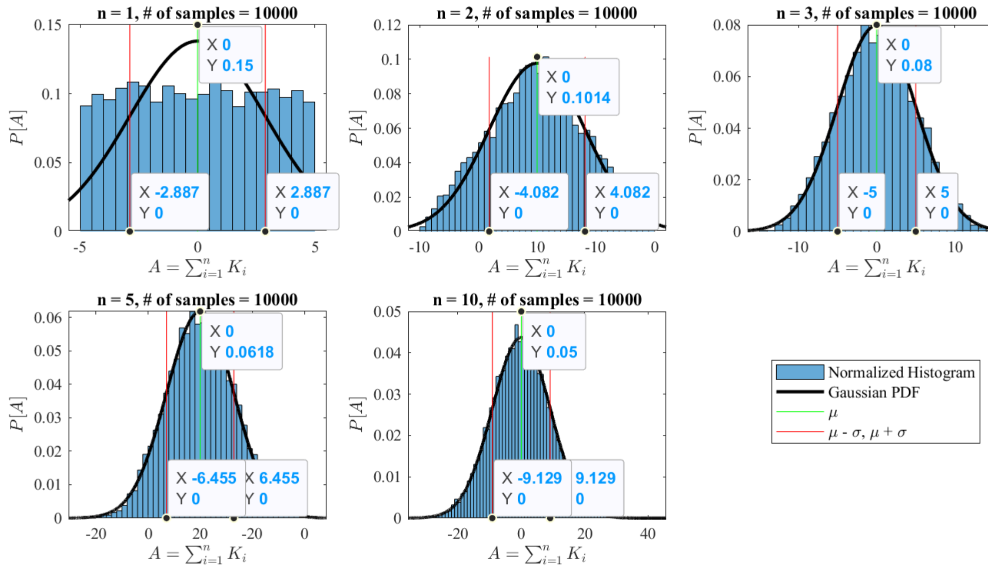

Figure 1: Normalized histograms of each sample vector and Gaussian PDFs whose parameters are calculated for each case with the help of the expected value and variance expressions found earlier for n = {1, 2, 3, 5, 10}. Note that A is represented as the sum of n uniform random variables. Number of samples is 10000 for each case. K_i has a uniform distribution in the range of [-5, 5] for each i.

2.2

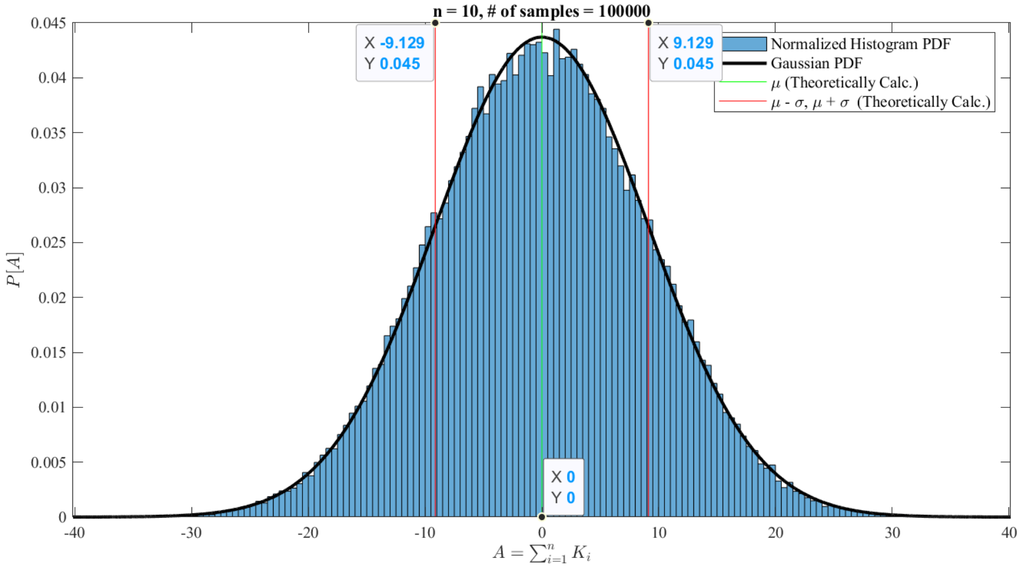

Figure 2: Normalized histogram of a vector consisting of 100000 samples and corresponding Gaussian PDF whose parameters are calculated with the help of the expected value and variance expressions found earlier for n = 10. Note that A is represented as the sum of n uniform random variables. K_i has a uniform distribution in the range of [-5, 5] for each i.

In Fig. (2), it can be seen that for more amount of samples Gaussian PDFs fit to the histograms. The thing is, as the number of samples increase, they fit more.

2.3

When the uniformity range is extended from [-5, 5] to [-10, 10], by symmetry, E[K] stays as 0 but the Var[K] changes as the following,

$$Var[K] = E[K^2] - (E[K])^2 = \int_{-10}^{10} k^2 \frac{1}{20} dk - (\int_{-10}^{10} k \frac{1}{20} dk)^2 = \boxed{\frac{100}{3}} \tag{9}$$

Then $Var[A] = nVar[K] = 1000/3$. For the range of [-10, 10], the Gaussian PDF and histogram are plotted in Fig. (3). Since the range doubles, standard deviation also doubles as expected. But the expected value doesn't change, and stays at 0.

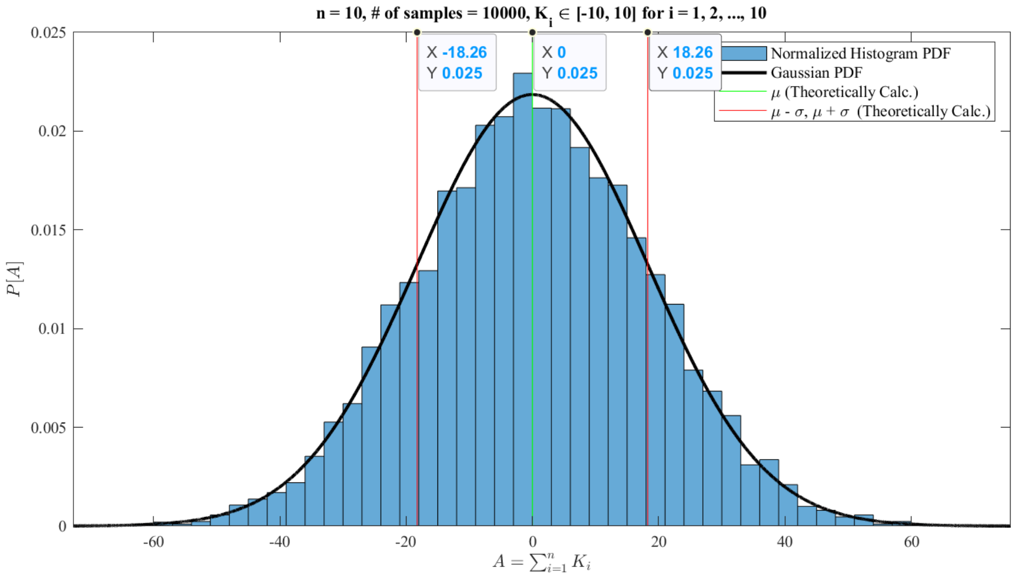

Figure 3: Normalized histogram of a vector consisting of 10000 samples and corresponding Gaussian PDF whose parameters are calculated with the help of the expected value and variance expressions found earlier for n = 10. Note that A is represented as the sum of n uniform random variables. K_i has a uniform distribution in the range of [-10, 10] for each i.

3

3.1

Let $K_1, K_2, ..., K_n$ be a sequence of exponential random variables where $\lambda = 1$. As Theorem 3.8 states, $E[K] = \frac{1}{\lambda} = 1$ and $Var[K] = \frac{1}{\lambda^2} = 1$. Since all K_n's are iid, $E[A] = nE[K] = [1, 2, 3, 5, 10]$ and $Var[A] = nVar[K] = [1, 2, 3, 5, 10]$ where $n = [1, 2, 3, 5, 10]$.

In Fig. 4, it is observed that as n increases, the distribution converges to the Gaussian as the central limit theorem states. For the cases given, it can be said that they converged to the Gaussian but for higher values of n, it will almost perfectly converge.

3.2

In Fig. (5), it can be seen that for both amount of samples Gaussian PDFs fit to the histograms. The thing is, as the number of samples increase, they fit more.

3.3

λ becomes 2, and n is kept constant as 10. As Theorem 3.8 states, $E[K] = \frac{1}{\lambda} = 0.5$ and $Var[K] = \frac{1}{\lambda^2} = 0.25$. Since all K_n's are iid, $E[A] = nE[K] = 5$ and $Var[A] = nVar[K] = 2.5$ where $n = 10$.

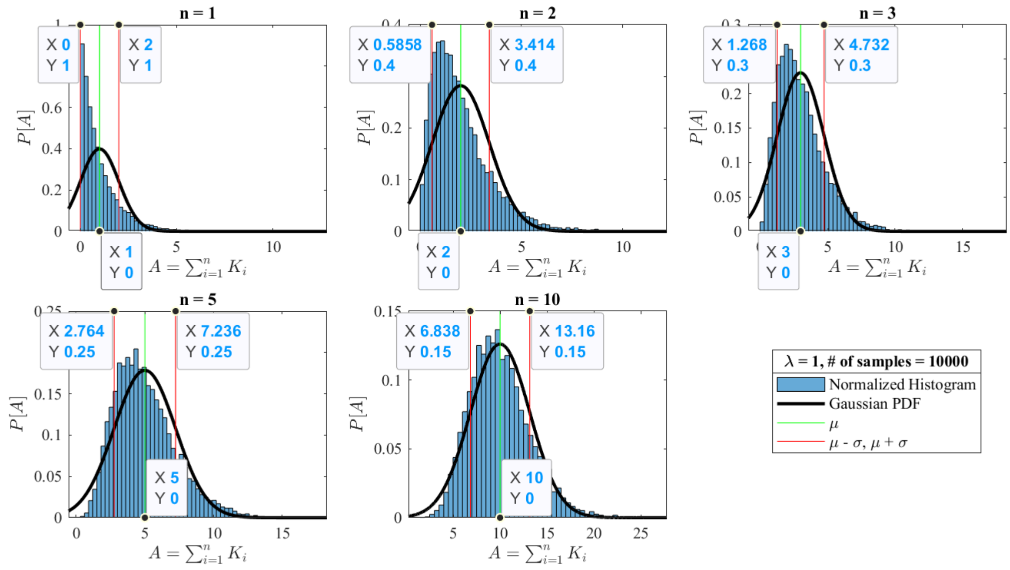

Figure 4: Normalized histograms of each sample vector and Gaussian PDFs whose parameters are calculated for each case with the help of the expected value and variance expressions found earlier for n $= \{1, 2, 3, 5, 10\}$. Note that A is represented as the sum of n exponential random variables. Number of samples is 10000 and $\lambda = 1$ for each case.

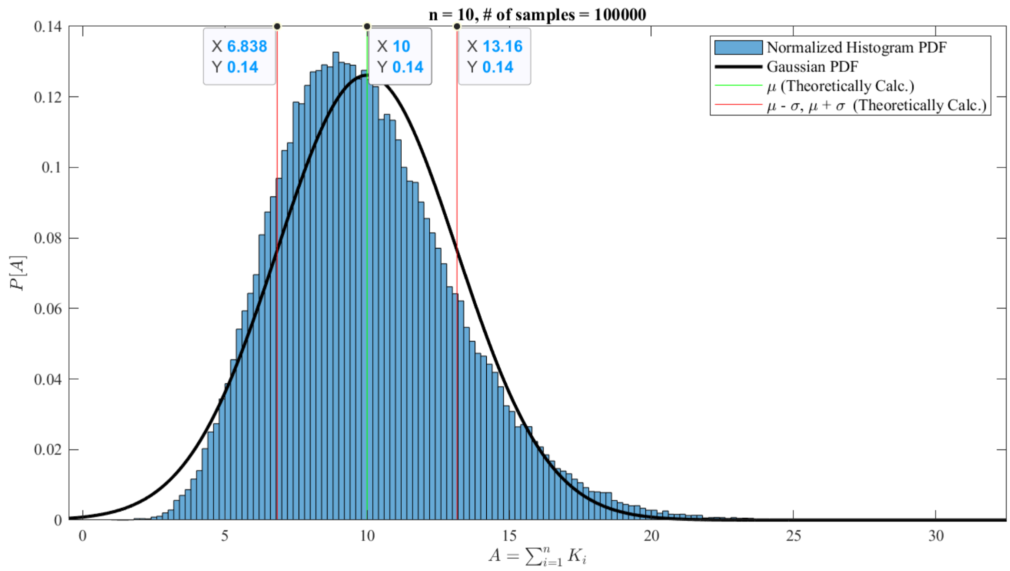

Figure 5: Normalized histogram of a vector consisting of 100000 samples and corresponding Gaussian PDF whose parameters are calculated with the help of the expected value and variance expressions found earlier for n $= 10$ and $\lambda = 1$. Note that A is represented as the sum of n exponential random variables.

These parameters are consistent with the experimental results, as can be seen in Fig. (6).

In Fig. (6), it is observed that as λ doubles, the expected value halves, and standard variance decreases to its quarter.

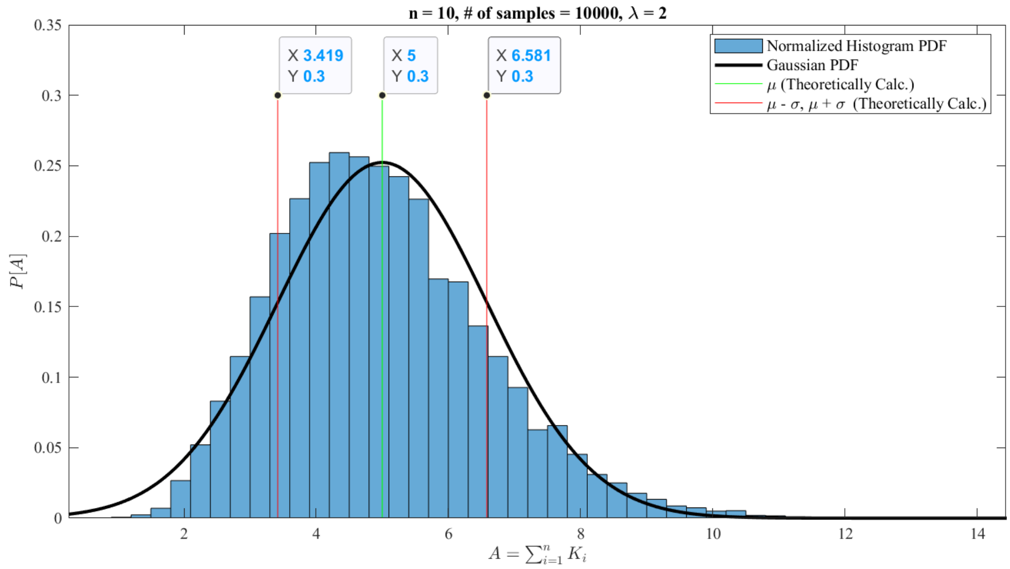

Figure 6: Normalized histogram of a vector consisting of 10000 samples and corresponding Gaussian PDF whose parameters are calculated with the help of the expected value and variance expressions found earlier for n = 10 and $\lambda = 2$. Note that A is represented as the sum of n exponential random variables.

4

4.1

Let $K_1, K_2, ..., K_n$ be a sequence of Bernoulli random variables where p = 0.5. As in Theorem 2.15 and Theorem 2.4, $E[A] = nE[K] = np = [0.5, 1, 1.5, 2.5, 5]$ and $Var[A] = nVar[K] = np(1-p) = [0.25, 0.5, 0.75, 1.25, 2.5]$ where n = 1, 2, 3, 5, 10.

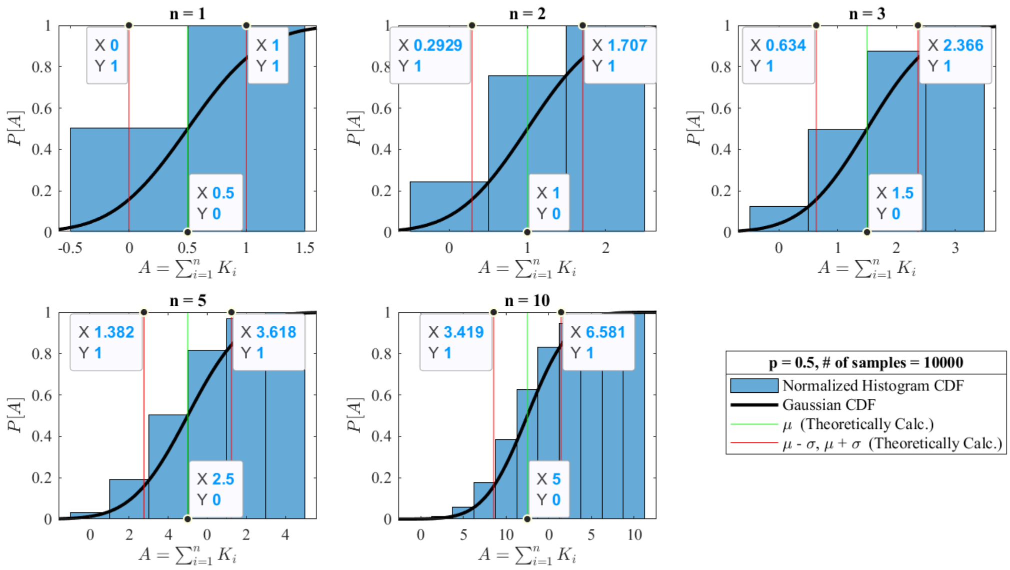

Figure 7: CDFs of the normalized histograms of each sample vector and Gaussian CDFs whose parameters are calculated for each case with the help of the expected value and variance expressions found earlier for n = {1, 2, 3, 5, 10}. Note that A is represented as the sum of n Bernoulli random variables where p = 0.5. Number of samples is 10000 for each case.

In Fig. (7), it can be seen that, as n increases, CDF converges to the Gaussian CDF. For all the cases, it seems there is a match between CDFs and histograms but as n increases the match become more obvious

4.2

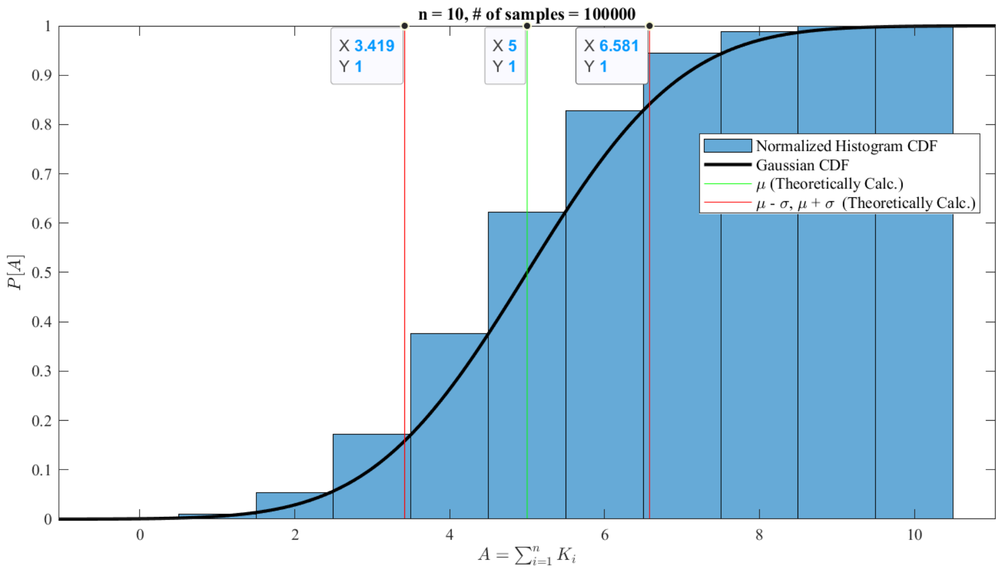

Figure 8: Normalized histogram of a vector consisting of 100000 samples and corresponding Gaussian PDF whose parameters are calculated with the help of the expected value and variance expressions found earlier for n = 10. Note that A is represented as the sum of n Bernoulli random variables.

In Fig. 8, since there is already enough samples for a Bernoulli random variable, CDF and the Gaussian parameters doesn't change much but it theoretically gets better.

4.3

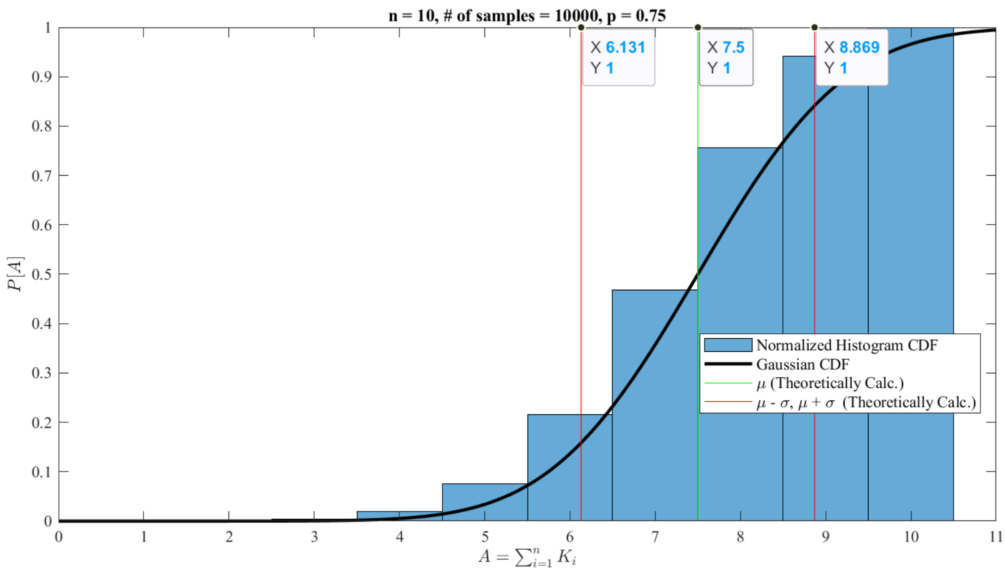

Figure 9: Normalized histogram of a vector consisting of 10000 samples and corresponding Gaussian PDF whose parameters are calculated with the help of the expected value and variance expressions found earlier for n = 10 and p = 0.75. Note that A is represented as the sum of n Bernoulli random variables.

In Fig. (9), as p increases, the mean increases but the standard deviation decreases.

5

5.1

Let $K_1, K_2, ..., K_n$ be a sequence of Poisson random variables where $\lambda = 1$. As in Theorem 2.15 and Theorem 2.6, $E[A] = nE[K] = n\lambda = [1, 2, 3, 5, 10]$ and $Var[A] = nVar[K] = n\lambda = [1, 2, 3, 5, 10]$ where n = 1, 2, 3, 5, 10.

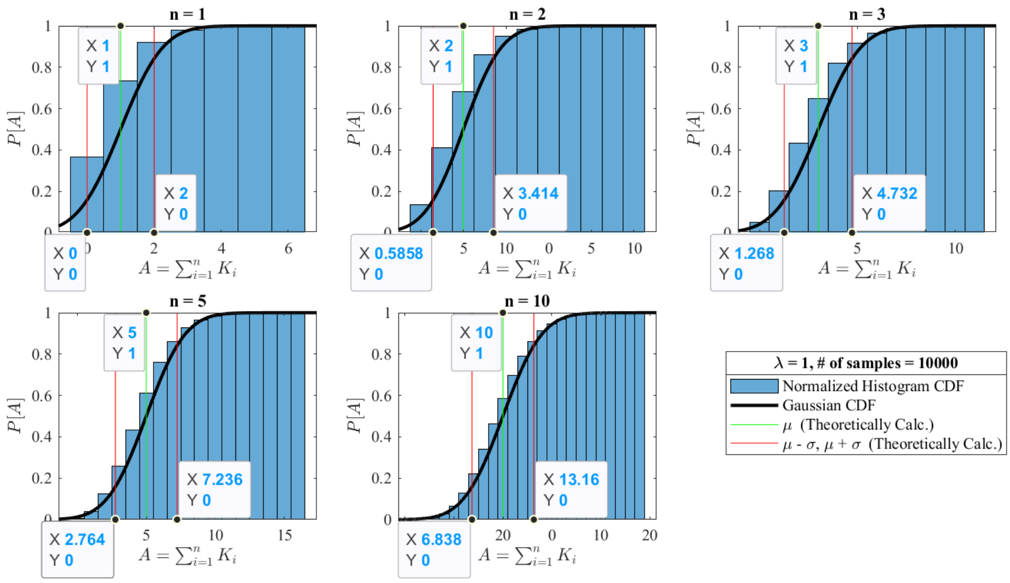

Figure 10: CDFs of the normalized histograms of each sample vector and Gaussian CDFs whose parameters are calculated for each case with the help of the expected value and variance expressions found earlier for n = {1, 2, 3, 5, 10}. Note that A is represented as the sum of n Poisson random variables where $\lambda = 1$. Number of samples is 10000 for each case.

In Fig. (10), Histograms and Gaussian CDFs match more as n increases.

5.2

As in Fig. (11), if number of samples increases to 100000, the histogram distribution fits technically more to the Gaussian CDF.

5.3

Since for the sum of Poisson random variables, both expected value and variance equal to $n\lambda$, both parameters double, standard variance is multiplied with $\sqrt{2}$ as expected in Fig. (12).

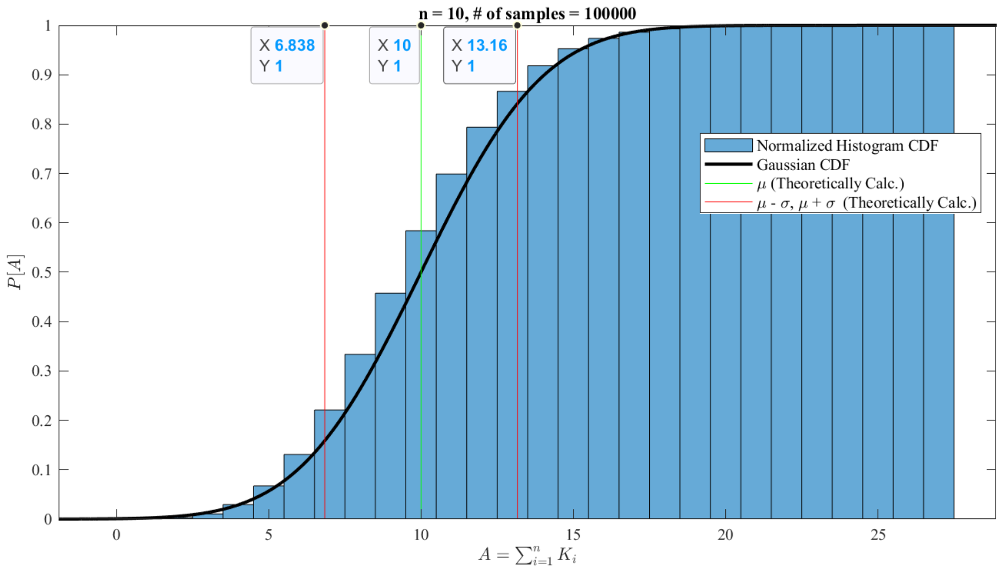

Figure 11: Normalized histogram of a vector consisting of 100000 samples and corresponding Gaussian PDF whose parameters are calculated with the help of the expected value and variance expressions found earlier for n = 10. Note that A is represented as the sum of n Poisson random variables.

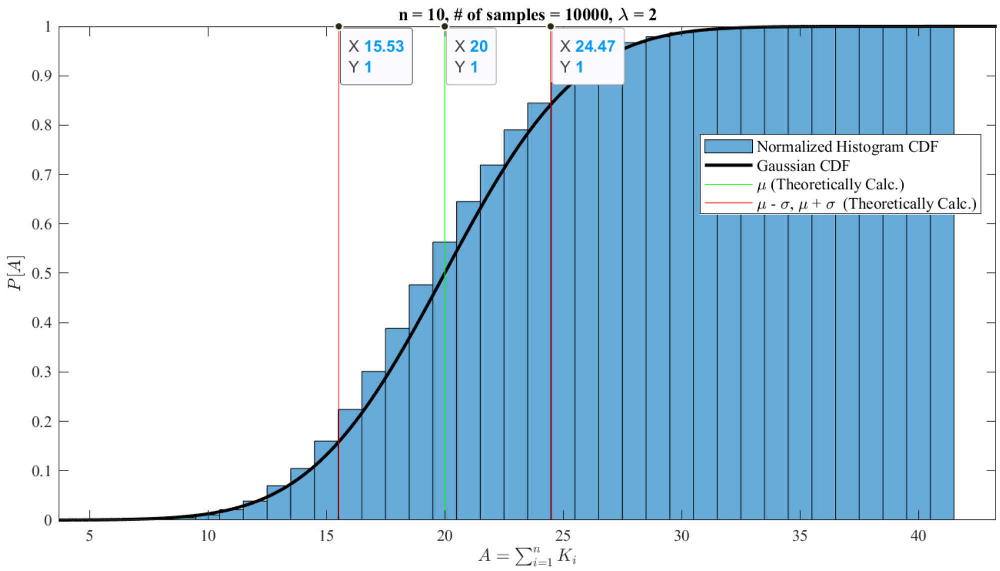

Figure 12: Histogram CDFs and corresponding Gaussian PDF fitting, $\lambda = 1$ samples = 10000 and 100000 for n = 10.