Statistics Assignment 1

HET551 – Design and Development Project 1

Michael Allwright
Haddon O'Neill

Tuesday, 24 May 2011

1 Normal Approximation to the Binomial Distribution

This section of the assignment shows how a normal curve can be used to approximate the binomial distribution. This section of the assignment was completed using a MATLAB function (shown in Listings 1) which would generate and save plots of the various Binomial Distributions after normalisation, and then calculate the errors between the standard normal curve and the binomial distribution.

The plots in Figures 1 and 2 show the binomial distribution for various n trials with probability $p = \frac{1}{3}$ and $p = \frac{1}{2}$ respectively. These binomial plots have been normalised so that they can be compared with the standard normal distribution.

From these plots it can be seen that once the binomial distribution has been normalised, the normal approximation is a good approach to estimating the binomial distribution. To determine its accuracy, the data in Table 1 shows the evaluation of $q_n = P(b_n \geq \mu_n + 2\sigma_n)$ for both the normal curve and binomial distribution.

| | $q_n = P(b_n \geq \mu_n + 2\sigma_n)$ Calculation | | | Error | |
n	$\mathcal{N}(0,1)$	$\mathcal{B}(n,\frac{1}{2})$	$\mathcal{B}(n,\frac{1}{3})$	$\mathcal{B}(n,\frac{1}{2})$	$\mathcal{B}(n,\frac{1}{3})$
1	0.0228	0.0000	0.0000	-0.02278	-0.02278
2	0.0228	0.0000	0.0000	-0.02278	-0.02278
3	0.0228	0.0000	0.0370	-0.02278	0.01426
4	0.0228	0.0000	0.0123	-0.02278	-0.01043
5	0.0228	0.0313	0.0453	0.00847	0.02249
10	0.0228	0.0107	0.0197	-0.01203	-0.00312
20	0.0228	0.0207	0.0376	-0.00208	0.01486
30	0.0228	0.0214	0.0188	-0.00139	-0.00398
40	0.0228	0.0192	0.0214	-0.00354	-0.00134
50	0.0228	0.0164	0.0222	-0.00636	-0.00059
100	0.0228	0.0176	0.0276	-0.00518	0.00479

Table 1: Calculating the error of the normal approximation to the binomial for various n and p

2 Analytical investigation of the Exponential Distribution

For this part of the assignment the density function shown in Equation 1 was given.

$$f(x) = \lambda e^{-\lambda x} \text{ for } x \geq 0 \text{ and } \lambda \geq 0 \tag{1}$$

Before any calculations were attempted, the area under graph was checked to show that $\int_{-\infty}^{\infty} f(x)\,dx = 1$. That is that the total probability of all possible values was 1.
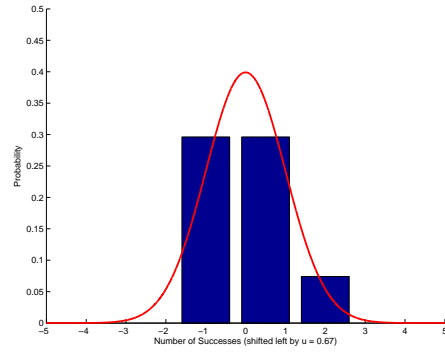
2.1 Derivation of CDF

To find the CDF of the given function, the function was integrated with 0 and x being the lower and upper bound respectively. This derivation is shown in Equations 2 to 4.
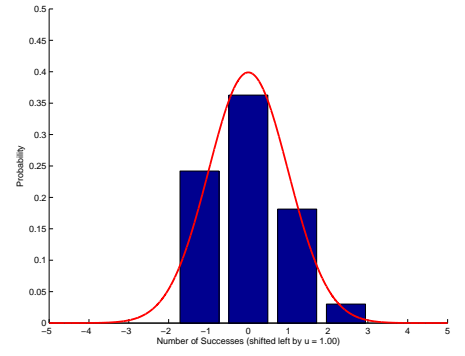
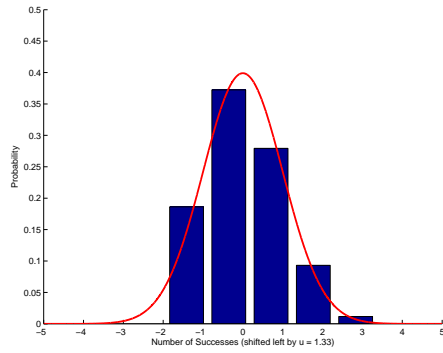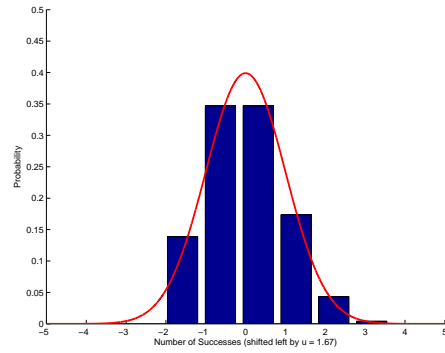$$\text{CDF} = \int_{o}^{x} f(x)\,dx = \int_{o}^{x} \lambda e^{-\lambda x}\,dx \tag{2}$$
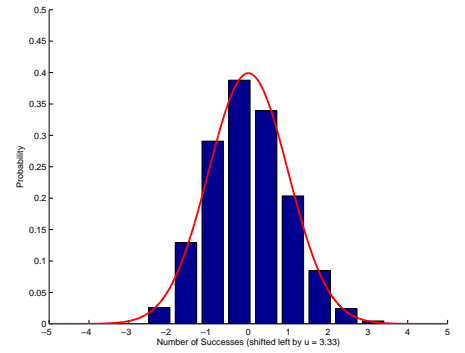
(a) $n = 1$

(b) $n = 2$

(c) $n = 3$

(d) $n = 4$

(e) $n = 5$
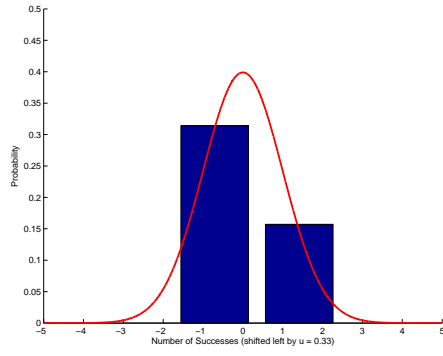
(f) $n = 10$

(g) $n = 20$

(h) $n = 30$

(i) $n = 40$

(j) $n = 50$

(k) $n = 100$

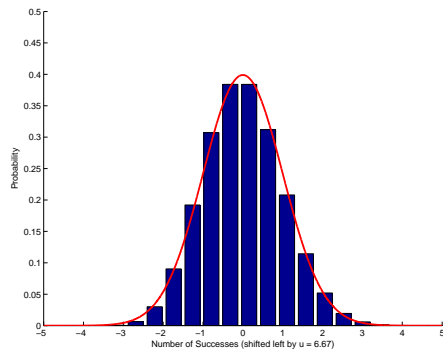Figure 1: Binomial Distribution with $p = \frac{1}{3}$
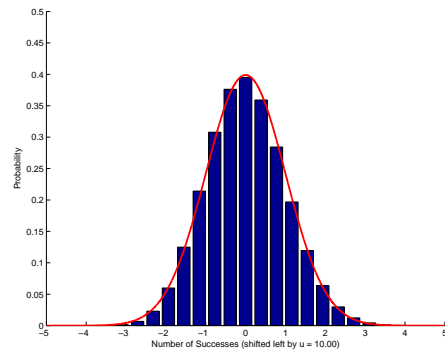
3

(a) $n = 1$

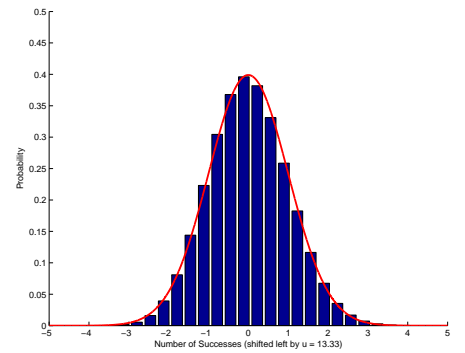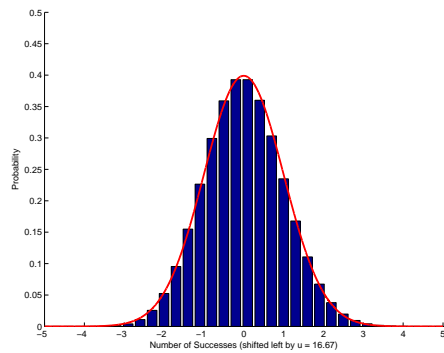(b) $n = 2$

(c) $n = 3$

(d) $n = 4$
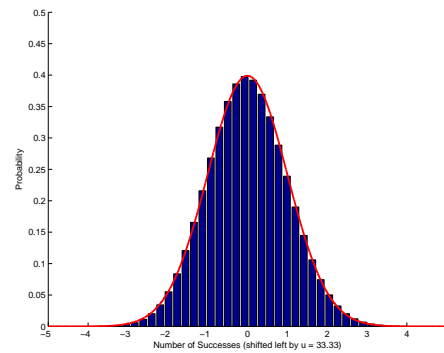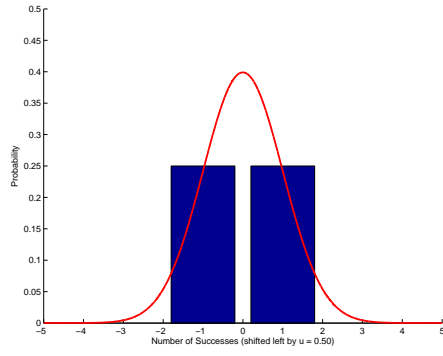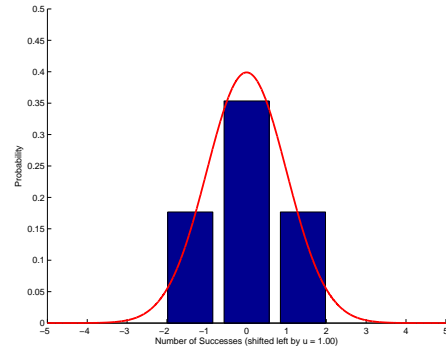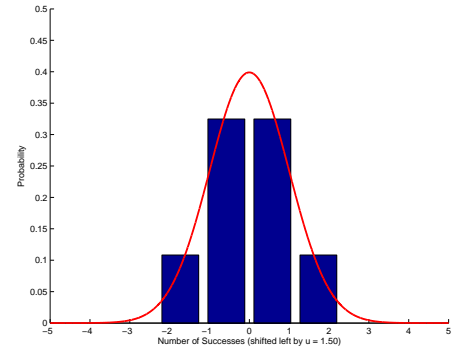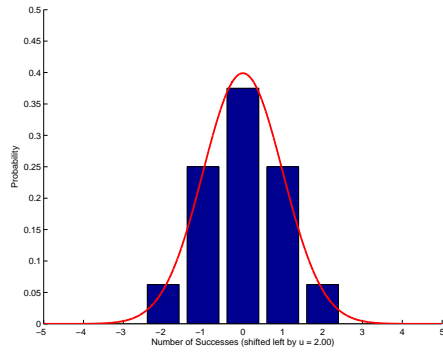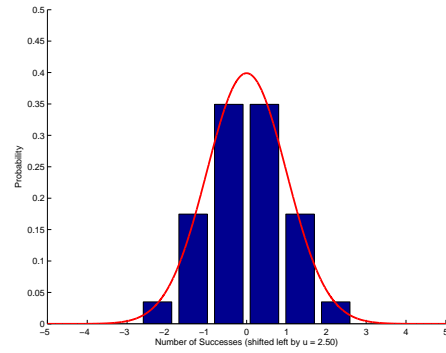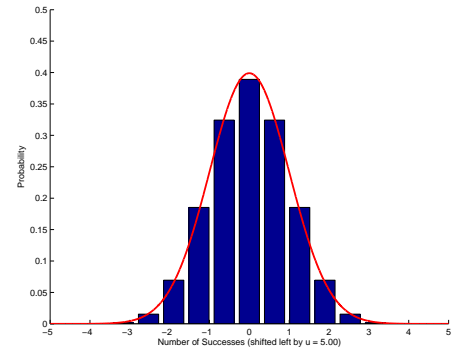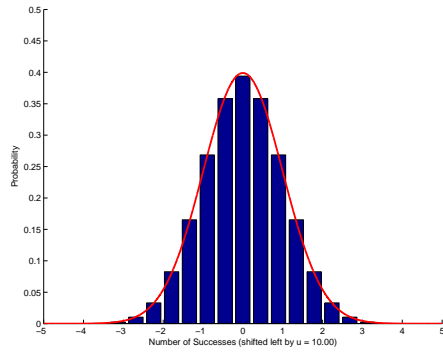
(e) $n = 5$

(f) $n = 10$

(g) $n = 20$
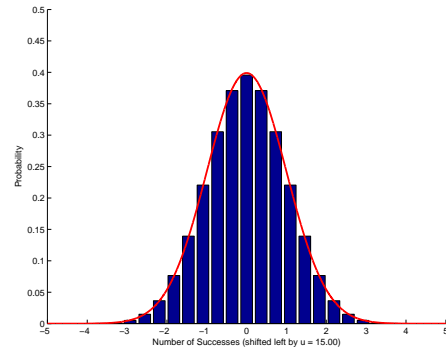
(h) $n = 30$
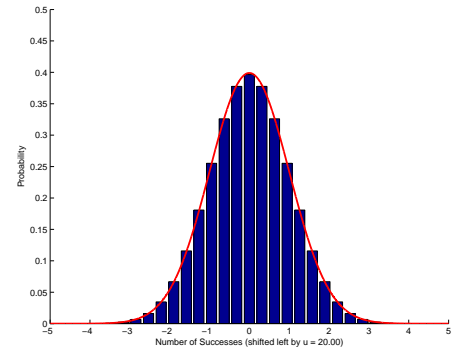
(i) $n = 40$
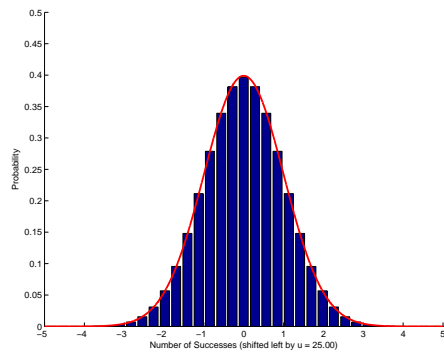
(j) $n = 50$
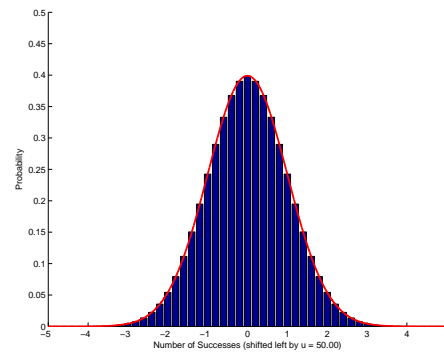
(k) $n = 100$

Figure 2: Binomial Distribution with $p = \frac{1}{2}$

$$\int_{o}^{x} \lambda e^{-\lambda x}\, dx = \left[-e^{-\lambda x} dx\right]_{0}^{x} \tag{3}$$

$$\int_{o}^{x} \lambda e^{-\lambda x}\, dx = 1 - e^{-\lambda x} \tag{4}$$

2.2   Mean Calculation

$$E[X] = \int_{-\infty}^{\infty} x\, f(x)\, dx$$

$$E[X] = \lambda \int_{0}^{\infty} x\, e^{-\lambda x}\, dx$$

To evaluate the integral on the right hand side, integration by parts must be applied as follows

$$\int f(x)g'(x)dx = f(x)g(x) - \int f'(x)g(x)dx$$

$$f(x) = x \qquad\qquad f'(x) = 1$$

$$g'(x) = e^{-\lambda x} \qquad\quad g(x) = \tfrac{-1}{\lambda}e^{-\lambda x}$$

$$E[X] = \lambda \int_{0}^{\infty} x\, e^{-\lambda x}\, dx = \left[\frac{-x}{\lambda}e^{-\lambda x}\right]_{0}^{\infty} + \frac{1}{\lambda}\int_{0}^{\infty} e^{-\lambda x}dx$$

$$E[X] = \left[\frac{-x}{\lambda}e^{-\lambda x}\right]_{0}^{\infty} + \left[\frac{-1}{\lambda^2}e^{-\lambda x}\right]_{0}^{\infty}$$

$$E[X] = \lambda\left[\left(-\frac{x}{\lambda} - \frac{1}{\lambda^2}\right)e^{-\lambda x}\right]_{0}^{\infty} \tag{5}$$

$$E[X] = \lambda\left(\lim_{x\to\infty}\left[\left(-\frac{x}{\lambda} - \frac{1}{\lambda^2}\right)e^{-\lambda x}\right] - \lim_{x\to 0}\left[\left(-\frac{x}{\lambda} - \frac{1}{\lambda^2}\right)e^{-\lambda x}\right]\right)$$

as $x \to \infty$ the $e^{-\lambda x}$ of first part decays faster than $\frac{x}{\lambda}$ grows. This allows the first term of the expression to be considered zero.

$$E[X] = (-\lambda)(-\frac{1}{\lambda^2}) = \frac{1}{\lambda} \tag{6}$$

5

2.3 Variance Calculation

$$\text{Var}(X) = \int (x - E[x])^2 f(x)\, dx \tag{7}$$

Substituting Equation 1 and 6 into the expression in Equation 7 yields and can be expanded to the following.

$$\text{Var}(X) = \int_0^\infty (x^2 - \frac{2x}{\lambda} + \frac{1}{\lambda^2}) \lambda e^{-\lambda x}\, dx$$

(Note: to simplify working, the upper and lower bounds have been temporarily omitted)

$$\text{Var}(X) = \lambda \int x^2 e^{-\lambda x}\, dx - 2 \int x e^{-\lambda x}\, dx + \frac{1}{\lambda} \int e^{-\lambda x}\, dx \tag{8}$$

To solve this expression, integration by parts was utilised to simplify the first term

$$\int x^2 e^{-\lambda x} = \int f(x) g'(x) dx = f(x) g(x) - \int f'(x) g(x) dx$$

where:

$$f(x) = x^2 \qquad f'(x) = 2x$$

$$g'(x) = e^{-\lambda x} \qquad g(x) = \tfrac{-1}{\lambda} e^{-\lambda x}$$

$$\int x^2 e^{-\lambda x} = \frac{-x^2}{\lambda} e^{-\lambda x} + \frac{2}{\lambda} \int x e^{-\lambda x} dx + C_1 \tag{9}$$

Substituting Equation 9 into the expression in Equation 8.

$$\text{Var}(X) = \lambda \left(\frac{-x^2}{\lambda} e^{-\lambda x} + \frac{2}{\lambda} \int x e^{-\lambda x} dx \right) - 2 \int x e^{-\lambda x}\, dx + \frac{1}{\lambda} \int e^{-\lambda x}\, dx$$

$$\text{Var}(X) = -x^2 e^{-\lambda x} + 2 \int x e^{-\lambda x} dx - 2 \int x e^{-\lambda x}\, dx + \frac{1}{\lambda} \int e^{-\lambda x}\, dx$$

$$\text{Var}(X) = -x^2 e^{-\lambda x} + \frac{1}{\lambda} \int e^{-\lambda x}\, dx$$

$$\text{Var}(X) = -x^2 e^{-\lambda x} - \frac{1}{\lambda^2} e^{-\lambda x} + C_2$$

Removing the constant of integration and applying the upper and lower limits of the integral allows this expression to be factorised.

$$\text{Var}(X) = \left[-e^{-\lambda x} \left(x^2 + \frac{1}{\lambda^2} \right) \right]_0^\infty$$

Finally evaluating the integral yields:

$$\text{Var}(X) = \lim_{x \to \infty} \left[-e^{-\lambda x} \left(x^2 + \frac{1}{\lambda^2} \right) \right] - \lim_{x \to 0} \left[-e^{-\lambda x} \left(x^2 + \frac{1}{\lambda^2} \right) \right]$$

$$\text{Var}(X) = \frac{1}{\lambda^2}$$

2.4 K$^{\text{th}}$ Moment Expression

Starting with the expression $M_k = E[X^k] = \int_0^\infty x^k e^{-\lambda x}\, dx$, integration by parts is used to develop an expression in terms of M_{k-1}, where $M_{k-1} = \int_0^\infty x^{k-1} e^{-\lambda x}\, dx$.

$$M_k = \int_0^\infty x^k\, f(x)\, dx$$

$$\int f(x)g'(x)dx = f(x)g(x) - \int f'(x)g(x)dx$$

$$f(x) = x^k \qquad\qquad f'(x) = kx^{k-1}$$

$$g'(x) = e^{-\lambda x} \qquad\qquad g(x) = \tfrac{-1}{\lambda}e^{-\lambda x}$$

$$M_k = \left[-\frac{x^k}{\lambda}e^{-\lambda x}\right]_0^\infty + \frac{k}{\lambda}\int_0^\infty x^{k-1}e^{-\lambda x}dx$$

Evaluating the limits of the first term and substituting second term for M_{k-1} then yields:

$$M_k = \lim_{x\to\infty}\left[-\frac{x^k}{\lambda}e^{-\lambda x}\right] - \lim_{x\to 0}\left[-\frac{x^k}{\lambda}e^{-\lambda x}\right] + \frac{k}{\lambda}M_{k-1}$$

$$M_k = \frac{k}{\lambda}M_{k-1}$$

Hence the recursive expression for the K$^{\text{th}}$ moment of the exponential distribution.

3 Properties of Variance Estimators

3.1 $\sigma^2 = \text{Var}(X_i)$ Calculation

For a uniform random variable in the range [0,1], the distribution is represented by:

$$F(x) = \begin{cases} 0 & x < 0 \\ x & 0 < x < 1 \\ 1 & 1 < x \end{cases}$$

From this, $f(x)$ can be found given that $f(x) = \frac{d}{dx}F(x)$:

$$f(x) = \begin{cases} 0 & x < 0 \\ 1 & 0 < x < 1 \\ 0 & 1 < x \end{cases}$$

or, for simplification

$$f(x) = 1,\ 0 < x < 1 \text{ otherwise } 0$$
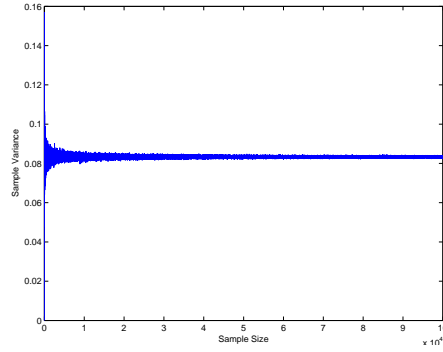
Figure 3: Variance across 100,000 samples

The expectation of X is defined as:

$$E[X] = \int x f(x) dx$$

Substituting in the derived density function:

$$E[X] = \int\limits_0^1 x \, dx = \left[\frac{1}{2}x^2\right]_0^1 = \frac{1}{2}$$

Using this result, the Variance can now be defined such that:

$$Var(X_i) = \int_0^1 (x - \frac{1}{2})^2 dx = \int_0^1 x^2 - x - \frac{1}{4} dx = \left[\frac{1}{3}x^3 - \frac{1}{2}x^2 + \frac{1}{4}x\right]_0^1 = \frac{1}{3} - \frac{1}{2} + \frac{1}{4}$$

$$Var(X_i) = \frac{1}{12} = 0.08\bar{3}\bar{3}$$

3.2 Estimation of σ^2 using Sample Variance

As stated before, the sample variance is defined as:

$$Var(X) = \int (x - E[X])^2 f(x) dx$$

To simplify the simulation, the MATLAB function Var(X) to calculate the sample variance.

The MATLAB code estimating σ^2 by generating a plot of variance of a independent uniform random variables against the number of samples is in Listing 2

MATLAB calculated the mean as $\mu = 0.0833$.

Comparing this to what the generated plot shows from the 10^5 samples that were generated, it is fair assumption for it to be approximately around this figure.

8

Figure 4: Distribution of Sample Variance over 100000 sequences where $n = 5$

3.3 Deriving the Mean of the distribution of the Sample Variance from simulation.

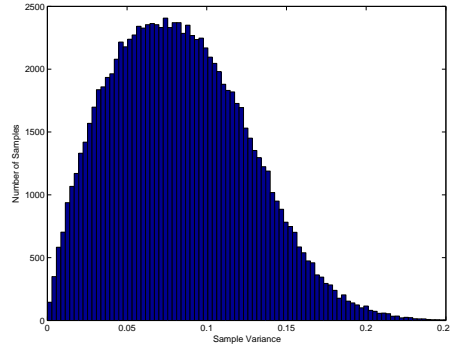We have been asked to plot the distribution of the sample variance of a number of sequences of $n = 5$. To do this, an array of sequences of length 5 were generated and the variance of each sequence was calculated.

The MATLAB code to first create the data sets and then plot the distribution is in Listing 3. The result of this simulation is shown in Figure 4.

The mean of `SamVarSet` was calculated out:

$$\sigma^2 = 0.0833 \approx \frac{1}{12}$$

$$\sigma^2 \approx Var(X_i)$$

Using the generated Histogram of the distribution, the mean is observed to be approximately the calculated value.

3.4 Estimator Bias Considerations

For this section we are asked to prove that $\tilde{S}^2 = \frac{\sum_{i=1}^{n}(X_i - \frac{1}{2})^2}{n}$ is an unbiased estimator of the variance, which is defined as $Var(X) = \int(x - E[X])^2 f(x) dx$.

To prove that it is unbiased, we must compare it against both the calculated variance and the sample variance and ensure that the results match.

To do this, we used the result from Part 3.3 using $\tilde{S}^2 = \frac{\sum_{i=1}^{n}(X_i - \frac{1}{2})^2}{n}$ instead of $Var(x)$ in the MATLAB code in Listing 3. The same number of samples were generated. This MATLAB code to generate the sample sequences and plot the distribution is shown in Listing 4.

The mean of SamVarSetEstimator was calculated out:

$$\tilde{S}^2 = 0.0836 \approx \frac{1}{12}$$

$$\tilde{S}^2 \approx Var(X_i) \approx \sigma^2$$

From Figure 5 we can say that $\tilde{S}^2$ is an unbiased estimator of the variance, as proven by it's close approximation to to calculated Variance found in Part 3.1 and the sample variance generated by MATLAB, Parts 3.2 and 3.3.
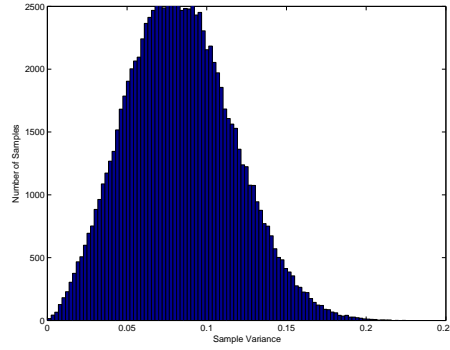
9

Figure 5: Distribution of the Estimated Variance over 100000 sequences of length 5

4 The Gamma Distribution

4.1 Determining the Normalising constant

It is required to find the normalizing constant C contained withing the following function:

$$f(x) = Cx^{\alpha-1}e^{-\lambda x}, \ x \geq 0, \ C > 0$$

From research, the standard notation of the probability density function for gamma random variables is as follows:

$$f(x) = \frac{\lambda^{\alpha}}{\Gamma(\alpha)}x^{\alpha-1}e^{-\lambda x}, \ x > 0$$

$$\text{Following } \Gamma(\alpha) = \int_{0}^{\infty} x^{a-1}e^{-x}dx$$

$$\therefore \ C = \frac{\lambda^{\alpha}}{\Gamma(\alpha)}$$

4.2 Visualisation of the Gamma distribution with respect to α and λ

To generate plots of gamma densities for differing values of α and λ, the MATLAB code in Listing 5 was generated to show how varying these parameters affects the distribution. These variations are shown in Figure 6.

Figure 6 shows the Gamma distributions generated by varying α with respect to x and the probability density at x, for 6 distinct values of λ.

4.3 Relationship between the Gamma Distribution and the Exponential Distribution

The Exponential Distribution is a special case of the Gamma Distribution that occurs when $\alpha = 1$. This is shown by the following:

$$f(x) = \frac{\lambda^{\alpha}}{\Gamma(\alpha)}x^{\alpha-1}e^{-\lambda x}, \ x > 0$$

10
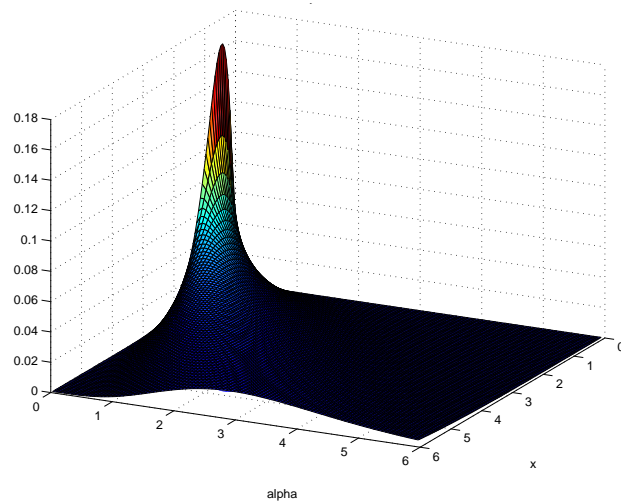
(a) $h = 0.100$
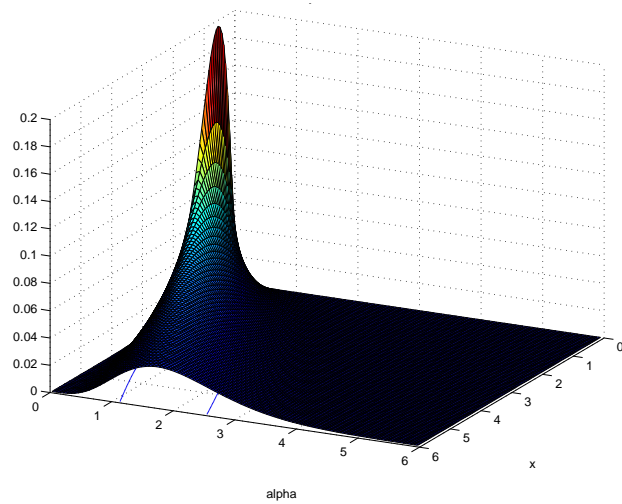
(b) $h = 0.167$
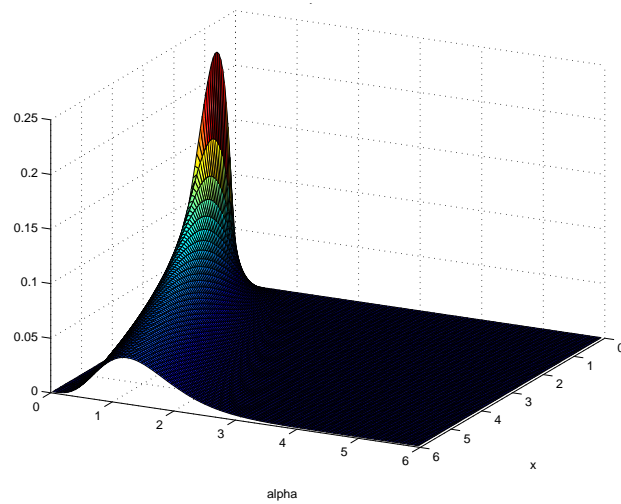
(c) $h = 0.233$
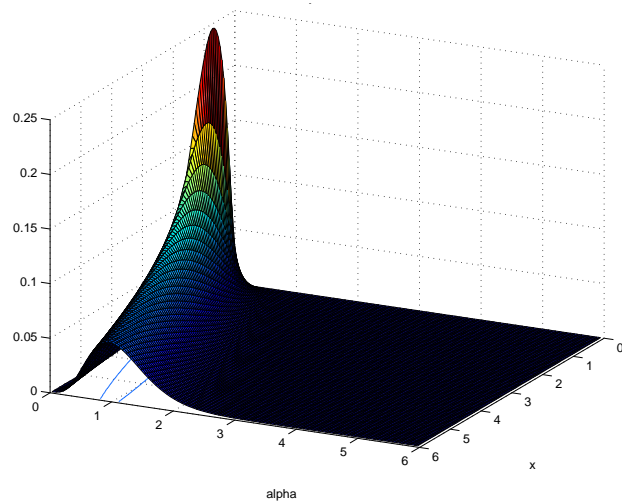
(d) $h = 0.300$
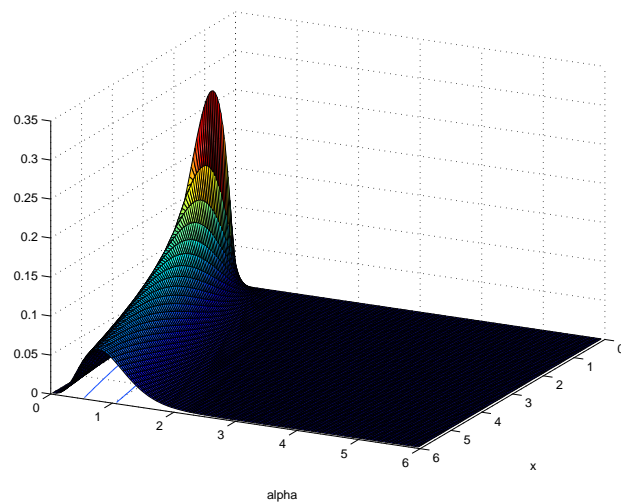
(e) $h = 0.367$
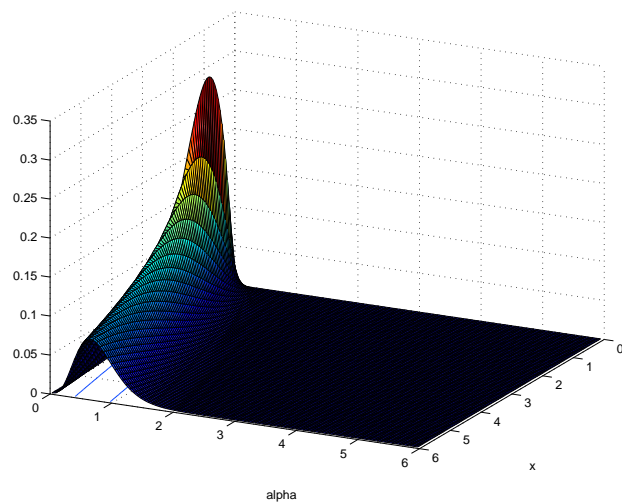
(f) $h = 0.433$

Figure 6: Gamma Distributions with respect to parameters α and λ

$$\text{where } \Gamma(\alpha) = \int_0^\infty x^{a-1} e^{-x} dx$$

$$\text{when } \alpha = 1$$

$$f(x) = \frac{\lambda^1}{\Gamma(1)} x^{1-1} e^{-\lambda x}, \ x > 0$$

$$\Gamma(1) = \int_0^\infty x^{1-1} e^{-x} dx = \int_0^\infty e^{-x} dx = [-xe^{-x}]_0^\infty = 1$$

$$f(x) = \lambda e^{-\lambda x}, \ x > 0$$

Hence when $\alpha = 1$, the expression for the Gamma Distribution simplifies to $f(x) = \lambda e^{-\lambda x}, \quad x > 0$ which is the Exponential Distribution.

4.4 Derivation of the Mean and Variance

4.4.1 Mean

The following shows the calculation of $E[X]$ for the Gamma Distribution in Equation 10 is derived:

$$f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} \tag{10}$$

$$E[X] = \int_0^\infty x \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} dx$$

$$E[X] = \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{\alpha-1} e^{-\lambda x} dx$$

$$E[X] = \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty \frac{\lambda^{\alpha+1}}{\Gamma(\alpha+1)} x^\alpha e^{-\lambda x} dx \frac{\Gamma(\alpha+1)}{\lambda^{\alpha+1}}$$

$$E[X] = \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty \frac{\lambda^{\alpha+1}}{\Gamma(\alpha+1)} x^\alpha e^{-\lambda x} dx \frac{\Gamma(\alpha+1)}{\lambda^{\alpha+1}}$$

$$E[X] = \frac{\lambda^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha+1)}{\lambda^{\alpha+1}} \int_0^\infty \frac{\lambda^{\alpha+1}}{\Gamma(\alpha+1)} x^\alpha e^{-\lambda x} dx$$

Where $\int \frac{\lambda^{\alpha+1}}{\Gamma(\alpha+1)} x^\alpha e^{-\lambda x} dx = \int \frac{\lambda^{\alpha+1}}{\Gamma(\alpha+1)} x^{(\alpha+1)-1} e^{-\lambda x} dx$ is the definition of the PDF:

$$\Gamma(\alpha+1, \lambda) = 1$$

$$E[X] = \frac{\lambda^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha+1)}{\lambda^{\alpha+1}}$$

$$E[X] = \frac{\Gamma(\alpha + 1)}{\lambda\Gamma(\alpha)}$$

Since $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$

$$E[X] = \frac{\alpha\Gamma(\alpha)}{\lambda\Gamma(\alpha)}$$

$$E[X] = \frac{\alpha}{\lambda}$$

4.4.2   Variance

$$Var(X) = \int_0^\infty x^2 f(x)dx - (E[X])^2$$

$$Var(X) = \int_0^\infty x^2 \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} dx - (E[X])^2$$

$$Var(X) = \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{\alpha-1} e^{-\lambda x} dx - (E[X])^2$$

$$Var(X) = \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty \frac{\lambda^{\alpha+2}}{\Gamma(\alpha+2)} x^{((\alpha+2)-1)} e^{-\lambda x} dx \lambda^{\alpha+2} - (E[X])^2$$

Where $\int \frac{\lambda^{\alpha+1}}{\Gamma(\alpha+1)} x^\alpha e^{-\lambda x} dx = \int \frac{\lambda^{\alpha+1}}{\Gamma(\alpha+1)} x^{(\alpha+1)-1} e^{-\lambda x} dx$ is the definition of the PDF:

$$\Gamma(\alpha + 1, \lambda) = 1$$

$$Var(X) = \frac{\lambda^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha+2)}{\lambda^{\alpha+2}} - (E[X])^2$$

$$Var(X) = \frac{\lambda^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha+2)}{\lambda^{\alpha+2}} - \frac{\alpha^2}{\lambda^2} \tag{11}$$

$$\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1) \tag{12}$$

$$\Gamma(\alpha + 2) = ((\alpha + 2) - 1)\Gamma((\alpha + 2) - 1)$$

$$\Gamma(\alpha + 2) = (\alpha + 1)\Gamma(\alpha + 1) \tag{13}$$

$$\Gamma(\alpha + 1) = \alpha\Gamma(\alpha) \tag{14}$$

Substituting Equation 14 into Equation 13

$$\Gamma(\alpha + 2) = (\alpha + 1)\alpha\Gamma(\alpha)$$

13

$$\Gamma(\alpha + 2) = (\alpha^2 + \alpha)\Gamma(\alpha) \tag{15}$$

Substituting Equation 15 into Equation 11

$$Var(X) = \frac{(\alpha^2 + \alpha)\Gamma(\alpha)}{\lambda^2 \Gamma(\alpha)} - \frac{\alpha^2}{\lambda^2}$$

$$Var(X) = \frac{\alpha^2 + \alpha - \alpha^2}{\lambda^2}$$

$$Var(X) = \frac{\alpha}{\lambda^2}$$

4.5 Derivation of the Sample α and λ in terms of the random sample.

From the Sections 4.4.1 and 4.4.2, the actual values of m_1 and m_2 are defined in as follows:

$$\frac{\alpha}{\lambda} = m_1$$

$$\frac{\alpha}{\lambda^2} = m_2$$

Rearranging these Equations to give an expression in terms of α and λ.

$$\lambda = \frac{m_1}{m_2}$$

$$\alpha = \frac{m_1^2}{m_2}$$

Expressing the moments α and λ in terms $\hat{m}_1$, $\hat{m}_2$ and the random sample

$$\hat{\lambda} = \frac{\left(\sum_{i=1}^n x_i\right)}{\left(\sum_{i=1}^n x_i^2\right)}$$

$$\hat{\alpha} = \frac{\left(\sum_{i=1}^n x_i\right)^2}{n\left(\sum_{i=1}^n x_i^2\right)}$$

4.6 Validation of Estimator Bias though Simulation

To show that the estimators from Part 4.5 are unbiased, a set of Gamma distributively random variables for three sets of α and λ are generated.

The MATLAB code in Listing 6 generates the random variables following the Gamma distribution, compares the Estimators and plots a distribution of the sample mean and sample variance for each of the set of generated data as shown in Figures 7, 8 and 9.

A comparison of the distributions to both the Estimated mean and Sample mean is shown in Table 2.

(a) Sample Mean
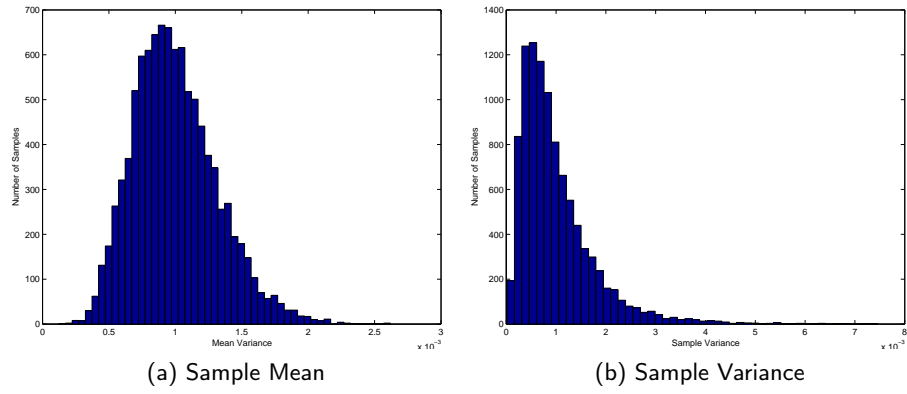
(b) Sample Variance

Figure 7: Distributions of Sample Mean and Variance for $\alpha = 0.001$ and $\lambda = 1$

(a) Sample Mean
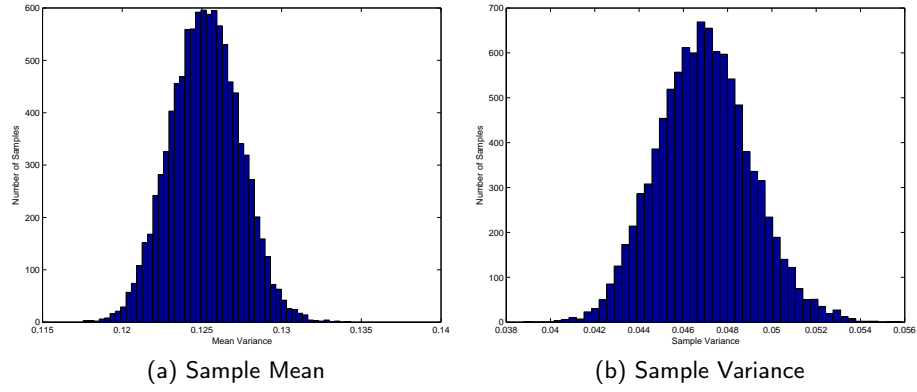
(b) Sample Variance

Figure 8: Distributions of Sample Mean and Variance for $\alpha = 0.334$ and $\lambda = 2.667$

(a) Sample Mean
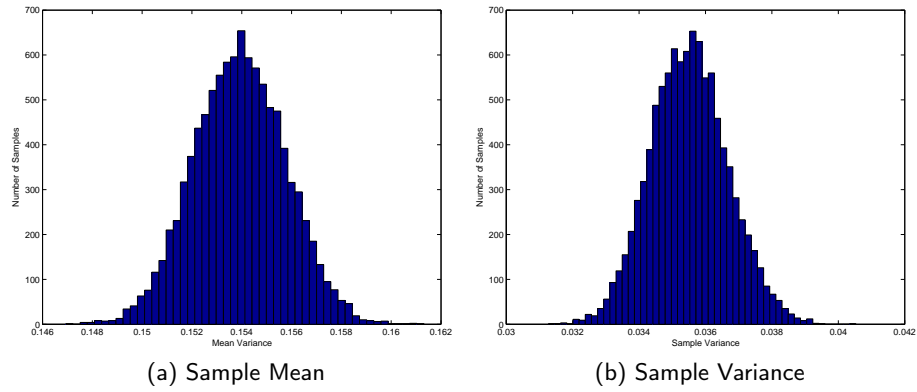
(b) Sample Variance

Figure 9: Distributions of Sample Mean and Variance for $\alpha = 0.667$ and $\lambda = 4.333$

	Set 1 ($\alpha = 0.001, \lambda = 1$)	Set 2 ($\alpha = 0.334, \lambda = 2.667$)	Set 3 ($\alpha = 0.667, \lambda = 4.333$)
Estimated Mean	0.0010	0.1253	0.1539
Simulated Mean	0.0012	0.1248	0.1537
Estimated Variance	0.0010	0.0470	0.0355
Simulated Variance	0.0013	0.0463	0.0348

Table 2: Comparing the simulated and estimated means for 3 sets of different α and λ values

```matlab
function [err_tbl] = napprox()

% Values for 'number of trials' to be investigated
n_arr = [1 2 3 4 5 10 20 30 40 50 100];
p_arr = [1/2 1/3];

% Function to calculate the standard normal distribution
g = inline('exp(-(x.^2 ./ 2))/sqrt(2*pi)');
g_range = -5:0.001:5;
g_range_b = 2:0.001:10;

err_tbl = zeros(size(n_arr,2),6);
err_tbl(1:size(n_arr,2)) = n_arr;

% calculate q_n for normal
err_tbl(1:size(n_arr,2),2) = sum(g(g_range_b) * 0.001);

for j = 1:1:size(p_arr,2)
    for i = 1:1:size(n_arr,2)
        % generate raw PDF
        x = 0:1:n_arr(i);
        dat = binopdf(x,n_arr(i),p_arr(j));

        % normalise
        [m v] = binostat(n_arr(i),p_arr(j));
        xn = (x - m) / sqrt(v);
        datn = dat * sqrt(v);

        % plot and convert to PDF image
        figure(1), clf(1), hold on, axis([-5 5 0 0.5]);
        bar(xn, datn);
        plot(g_range,g(g_range),'r','LineWidth',2);
        xlabel('Number of Successes');
        ylabel('Probability');
        saveas(gcf,sprintf('figures\\binpdf(p=%.2f,n=%d).pdf', p_arr(j), n_arr(i)));

        % calculate q_n for binomial
        q_n = 0;
        for k = 1:1:(n_arr(i) + 1)
            if(xn(k) > 2)
                q_n = q_n + (datn(k) / sqrt(v));
            end
        end
        err_tbl(i,j+2) = q_n;
    end
end

% calculate errors
err_tbl(1:size(n_arr,2),5) = err_tbl(1:size(n_arr,2),3) - err_tbl(1:size(n_arr,2),2);
err_tbl(1:size(n_arr,2),6) = err_tbl(1:size(n_arr,2),4) - err_tbl(1:size(n_arr,2),2);

xlswrite('err_tbl.xls',err_tbl);
```
Listing 1: MATLAB Source Code to Generate the PDF's of the Binomial Variables

```
%Number of random samples
n=100000;

%For comparison, Matlab's Var(X) is calculated using sample variance
X=rand([1 n]);
xVar=var(X)
%using sample variance, plot variance against number of samples
xVarn = [];
for i=1:1:n
    X=rand([1 i]); %new set of random variables generated for each pass, ensures no bias
    xVarn = [xVarn Var(X)];
end
x:1:1:100000;
figure(1), clf(1), plot(xVarn);
ylabel('Sample Size');
xlabel('Sample Variance');
saveas(gcf,'part b - variance x sample no.pdf');
```

Listing 2: MATLAB Source Code to generate a plot of the Variance of a data set contain n samples

```
%for code simplifacation's sake, we will use Var(X)
n=5;
sampleSetSize=100000;
SamVarSet = [];
for i=1:1:sampleSetSize
a=0;
X=rand([1 n]);
SamVarSet = [SamVarSet Var(X)];
end
figure, clf, hist(SamVarSet,100);
axis([0 0.25 0 2500]);
xlabel('Sample Variance');
ylabel('Number of Samples');
saveas(gcf,sprintf('part c - variance v samples.pdf'));
mean(SamVarSet)
```

Listing 3: MATLAB Source Code to generate a plot of the Variance Distribution over 100000 samples

```matlab
n=5;
sampleSetSize=100000;
SamVarSetEstimator = [];
for i=1:1:sampleSetSize
a=0;
X=rand([1 n]);
for i=1:1:n
    a=a+(X(i)-1/2)^2;
end
    a=a/n;
SamVarSetEstimator = [SamVarSetEstimator a];
end
figure, clf, hist(SamVarSetEstimator,100);
axis([0 0.25 0 2500]);
xlabel('Sample Variance');
ylabel('Number of Samples');
saveas(gcf,sprintf('part d - variance estimator v samples.pdf'));
mean(SamVarSetEstimator)
```
Listing 4: MATLAB Source Code to generate a plot of the estimated Variance Distribution over 100000 samples

```matlab
%f = inline('((h^a)/(factorial(a-1)))*(x^(a-1))*exp(-h*x)','x','a','h');
%Part B
% Using x from -100 to 100 and a =3, h=5

f = inline('((h^a)/(gamma(a)))*(x^(a-1))*exp(-h*x)','x','a','h');
min = 0;
steps = 100;
max = 6;
step=((max-min)/steps);
hmin=0.1;
hsteps=6;
hmax=0.5;
hstep=((hmax-hmin)/hsteps);
x=min:step:max-step;
a=min:step:max-step;
h=hmin:hstep:hmax-hstep;
aArray = zeros(steps,steps);
for i=1:1:numel(h)
    for j=1:1:steps
        for k=1:1:steps
            aArray(k,j)=f(k,a(j),h(i));
        end
    end
surfc(x,a,aArray);
view([1,0.5,0.5])
ylabel('alpha')
xlabel('x')
title(sprintf('Gamma density for Lambda = %.3f',h(i)))
saveas(gcf,sprintf('Q4b3D h=%.3f.pdf',h(i)));
end
```
Listing 5: MATLAB Source Code to generate a plot of the estimated Variance Distribution over 100000 samples

```matlab
n=10000; %Number of random samples per set of variables
sets=3; %number of sets of variables

amin=0.001;
amax=1;
hmin=1;
hmax=6;

astep=((amax-amin)/sets);
hstep=((hmax-hmin)/sets);

a=amin:astep:amax-astep;
h=hmin:hstep:hmax-hstep;

grand=zeros(3,n);

for i=1:1:steps
    grand(i,:)=gamrnd(a(i),(1/(h(i))),[1 n]);
end
%Creating the set of mean and var from calculation
meanCalc=[];
varCalc=[];
for i=1:1:steps
    meanCalc(i)=a(i)/h(i);
    varCalc(i)=a(i)/(h(i)^2);
end

%Creating a set of mean and var from simulation
meanSim=[];
varSim=[];
for i=1:1:steps
    meanSim(i)=mean(grand(i,:));
    varSim(i)=var(grand(i,:));
end

%Code for generating a distribution plot of sample mean and var
GrandMean = zeros(3,n);
GrandVar = zeros(3,n);
for j=1:1:n
    for i=1:1:sets
        grand(i,:)=gamrnd(a(i),(1/(h(i))),[1 n]);
        GrandMean(i,j)=mean(grand(i,:));
        GrandVar(i,j)=var(grand(i,:));
    end
end

%Plotting the distribution of
for i=1:1:sets
    hist(GrandMean(i,:),50)
    xlabel('Mean Variance');
    ylabel('Number of Samples');
    saveas(gcf,sprintf('Q4F - Mean a=%.3f h=%.3f.pdf',a(i),h(i)))
    hist(GrandVar(i,:),50)
    xlabel('Sample Variance');
    ylabel('Number of Samples');
    saveas(gcf,sprintf('Q4F - Var a=%.3f h=%.3f.pdf',a(i),h(i)))
end

%outputs Calculated and Simulated mean and var for comparison
meanCalc
meanSim
```