ASEN5044 Assignment 4

Solutions

4 October 2018

1. Assuming cards are only dealt to the player, observe that there are four aces, four of each of the three types of face cards, and four tens in an ordinary deck of playing cards, and that the deck contains 52 unique cards. The total number of two-card combinations consisting of one ace and one face card or ten is given by

$$(4 \text{ aces}) \times (4 \times 3 \text{ face cards} + 4 \text{ tens}) = 64$$

The total number of possible two-card combinations that the player may be dealt is given by

$$\binom{52}{2} = 1326$$

Therefore, the probability of being dealt the desired hand is

$$\frac{64}{1326} = \frac{32}{663} \approx 0.04827$$

2. (a) In order for X and Y to be independent, $P(X|Y) = P(X)$ and $P(Y|X) = P(Y)$ must hold for all values of X and Y. However,

$$P(X=1|Y=1) = \frac{\frac{1}{18}}{3 \cdot \frac{1}{18}} = \frac{1}{3} \neq P(X=1) = 3 \cdot \frac{1}{18} = \frac{1}{6}$$

Therefore, X and Y are not independent.

(b)

$$P(Y=5) = \sum_{i=1}^{3} P(Y=5, X=X_i) = \frac{1}{18} + \frac{1}{6} + \frac{1}{3} = \frac{5}{9}$$

(c)

$$\begin{aligned} P(Y=5|X=3) &= \frac{P(Y=5, X=3)}{P(X=3)} = \frac{P(Y=5, X=3)}{\sum_{i=1}^{3} P(X=3, Y=Y_i)} \\ &= \frac{\frac{1}{6}}{\frac{1}{6} + \frac{1}{18} + \frac{1}{18}} = \frac{3}{5} \end{aligned}$$

3. A valid probability density function (PDF) satisfies

$$\int_{-\infty}^{\infty} f_X(x)dx = 1$$

Thus

$$\begin{aligned}
\int_{-\infty}^{\infty} f_X(x)dx &= \int_0^1 ax(1-x)dx = a\int_0^1 (x-x^2)dx \\
&= a\left[\frac{1}{2}x^2 - \frac{1}{3}x^3\right]_0^1 = a\left(\frac{1}{2}-\frac{1}{3}\right) = \frac{a}{6} = 1
\end{aligned}$$

Solving for a gives

$$a = 6$$

4. (a) The probability distribution function, also known as the cumulative density function (CDF), is given by

$$P(X < x) = \int_{-\infty}^{x} f_X(t)dt = \int_0^x ae^{-at}dt = \left[-e^{-at}\right]_0^x = 1 - e^{-ax}$$

(b) The mean of a random variable is given by

$$\mu_X = \int_{-\infty}^{\infty} x f_X(x)dx = \int_0^{\infty} xae^{-ax}dx$$

Integrating by parts gives

$$\begin{aligned}
\int_0^{\infty} xae^{-ax}dx &= \left[-xe^{-ax}\right]_0^{\infty} - \int_0^{\infty} -e^{-ax}dx \\
&= \lim_{x\to\infty} -xe^{-ax} - \left[\frac{1}{a}e^{-ax}\right]_0^{\infty} \\
&= \lim_{x\to\infty} -xe^{-ax} + \frac{1}{a}
\end{aligned}$$

Using l'Hôpital's Rule to evaluate the limit gives

$$\lim_{x\to\infty} -\frac{x}{e^{ax}} = \lim_{x\to\infty} -\frac{1}{ae^{ax}} = 0$$

Therefore, the mean of the exponential distribution is

$$\mu_X = \frac{1}{a}$$

(c) The second moment is given by

$$\int_{-\infty}^{\infty} x^2 f_X(x)dx = \int_0^{\infty} x^2ae^{-ax}dx$$

Integrating by parts gives

$$\int_0^\infty x^2 ae^{-ax}dx = \left[-x^2e^{-ax}\right]_0^\infty - \int_0^\infty -2xe^{-ax}dx$$

Integrating by parts again

$$\begin{aligned}\int_0^\infty x^2 ae^{-ax}dx &= \left[-x^2e^{-ax}\right]_0^\infty + \left[-\frac{2xe^{-ax}}{a}\right]_0^\infty + \frac{2}{a}\int_0^\infty e^{-ax}dx \\ &= \lim_{x\to\infty} -x^2e^{-ax} + \lim_{x\to\infty} -\frac{2xe^{-ax}}{a} - \frac{2}{a^2}\left[e^{-ax}\right]_0^\infty \\ &= \lim_{x\to\infty} -x^2e^{-ax} + \lim_{x\to\infty} -\frac{2xe^{-ax}}{a} + \frac{2}{a^2}\end{aligned}$$

Using l'Hôpital's Rule

$$\begin{aligned}\lim_{x\to\infty} -\frac{x^2}{e^{ax}} &= \lim_{x\to\infty} -\frac{2x}{ae^{ax}} = \lim_{x\to\infty} -\frac{2}{a^2e^{ax}} = 0 \\ \lim_{x\to\infty} -\frac{2x}{ae^{ax}} &= \lim_{x\to\infty} -\frac{2}{a^2e^{ax}} = 0\end{aligned}$$

The second moment is therefore

$$\int_{-\infty}^\infty x^2 f_X(x)dx = \frac{2}{a^2}$$

(d) The variance is given by

$$\sigma_X^2 = \int_{-\infty}^\infty (x-\mu_X)^2 f_X(x)dx = \int_0^\infty (x-\frac{1}{a})^2 ae^{-ax}dx$$

Integrating by parts as before gives

$$\int_0^\infty (x-\frac{1}{a})^2 ae^{-ax}dx = \left[-(x-\frac{1}{a})^2e^{-ax}\right]_0^\infty - \int_0^\infty -2(x-\frac{1}{a})e^{-ax}dx$$

Integrating by parts again

$$\begin{aligned}\int_0^\infty (x-\frac{1}{a})^2 ae^{-ax}dx &= \left[-(x-\frac{1}{a})^2e^{-ax}\right]_0^\infty + \left[-\frac{2}{a}(x-\frac{1}{a})e^{-ax}\right]_0^\infty + \frac{2}{a}\int_0^\infty e^{-ax}dx \\ &= \lim_{x\to\infty} -(x-\frac{1}{a})^2e^{-ax} + \frac{1}{a^2} + \lim_{x\to\infty} -\frac{2}{a}(x-\frac{1}{a})e^{-ax} - \frac{2}{a^2} \\ &\quad + \frac{2}{a}\int_0^\infty e^{-ax}dx\end{aligned}$$

Using l'Hôpital's Rule, it can be shown that the two infinite limits go to zero. The remaining integral term is identical to that of the second moment term. Thus, the variance is given by

$$\sigma_X^2 = \frac{1}{a^2} - \frac{2}{a^2} + \frac{2}{a^2} = \frac{1}{a^2}$$

(e) The standard deviation of a random variable is the square root of its variance. Thus,

$$\sigma_X = \frac{1}{a}$$

The probability of the variable falling within one standard deviation of its mean is then

$$\begin{aligned} P(\mu_X - \sigma_X < x < \mu_X + \sigma_X) &= P(x < \mu_X + \sigma_X) - P(x < \mu_X - \sigma_X) \\ &= P(x < \frac{2}{a}) - P(x < 0) = P(x < \frac{2}{a}) \\ &= 1 - e^{-2} \approx 0.8647 \end{aligned}$$

5. The histograms of the uniformly-distributed random samples at each sample size are shown in Figures 1 to 3. The means and standard deviations of the histograms are shown in Table 1.

Table 1: Means and Standard Deviations from Monte-Carlo Approximation of Standard Uniform Random Distribution

Sample Size N	Mean μ_X	Standard Deviation σ_X
50	0.520414	0.287771
500	0.488484	0.286149
5000	0.504221	0.287722

To find the analytical mean and standard deviation, first observe that the PDF of the uniform random distribution over the range $[0, 1]$ is given by

$$f_X(x) = \begin{cases} 1, & 0 \le x \le 1 \\ 0, & \text{otherwise} \end{cases}$$

The mean is then

$$\mu_X = \int_{-\infty}^{\infty} x f_X(x) dx = \int_0^1 x dx = \frac{1}{2} \left[x^2\right]_0^1 = \frac{1}{2} \approx 0.5$$

and the standard deviation is given by

$$\sigma_X = \left[\int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x) dx\right]^{\frac{1}{2}} = \left[\int_0^1 (x - \frac{1}{2})^2 dx\right]^{\frac{1}{2}} = \sqrt{\frac{1}{3}\left[x'^3\right]_{-\frac{1}{2}}^{\frac{1}{2}}} = \frac{1}{\sqrt{12}} \approx 0.2887$$

where a change of variables of $x' = x - \mu_X$ was made to evaluate the integral.

As the sample size increases, the histograms more closely approach the shape of the true PDF, and the sample means and standard deviations more closely approach the analytical mean and standard deviation.

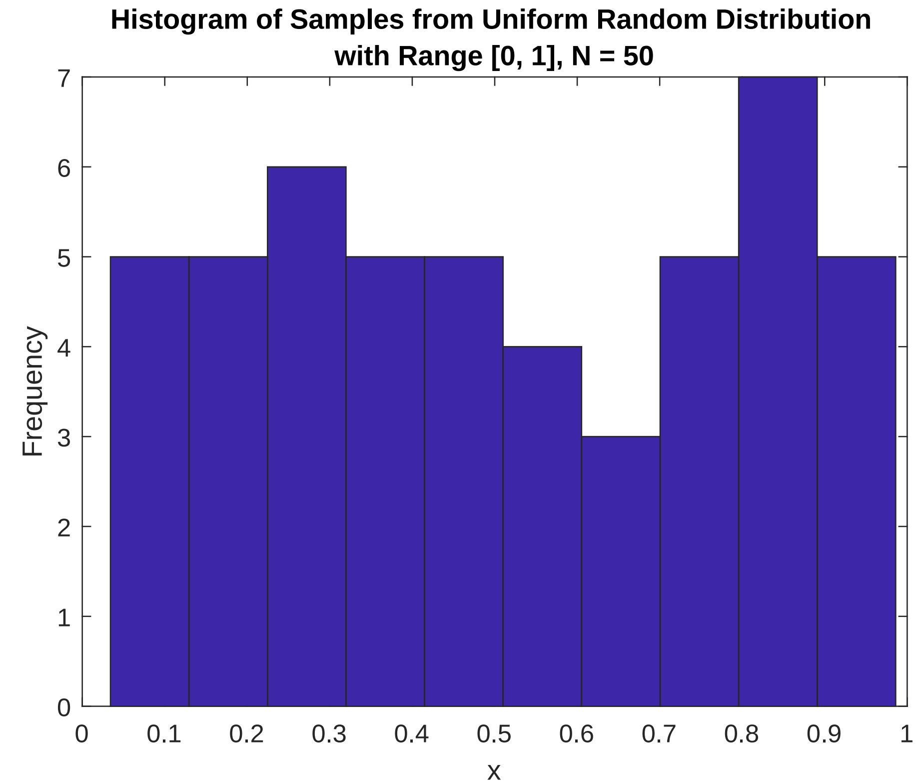

Figure 1: Histogram of Uniform Random Samples, $N = 50$

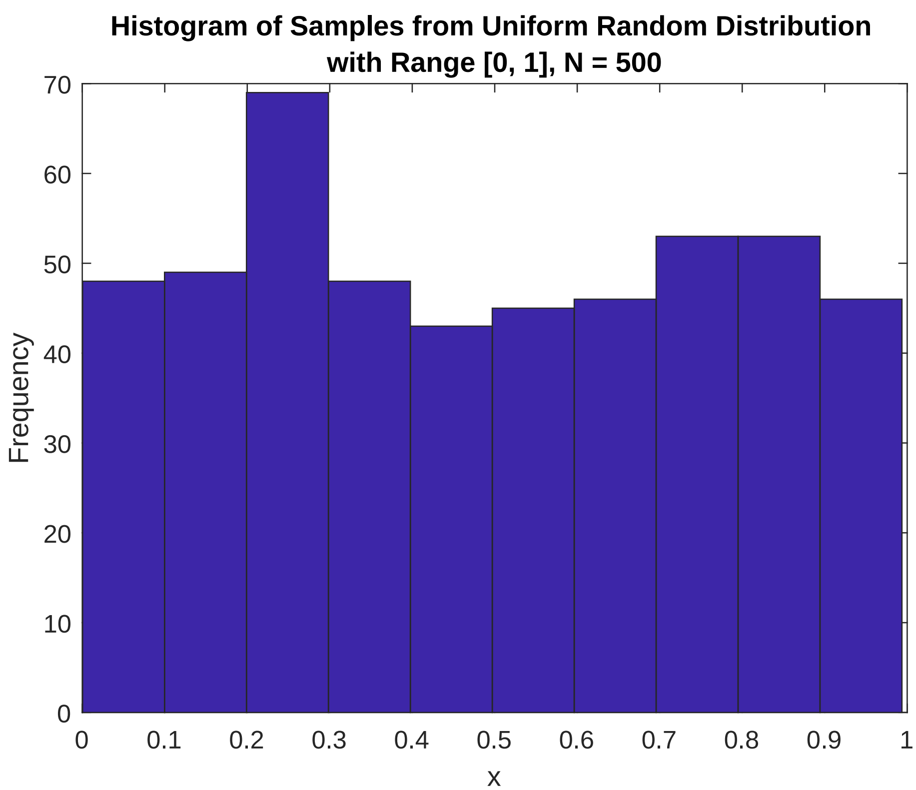

Figure 2: Histogram of Uniform Random Samples, $N = 500$

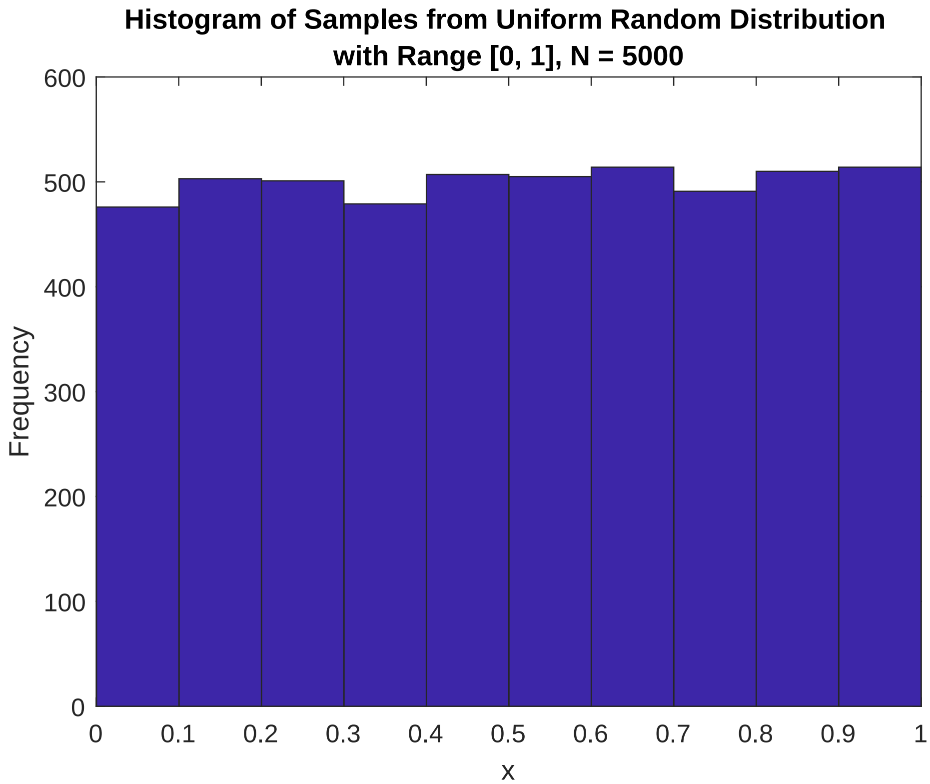

Figure 3: Histogram of Uniform Random Samples, $N = 5000$

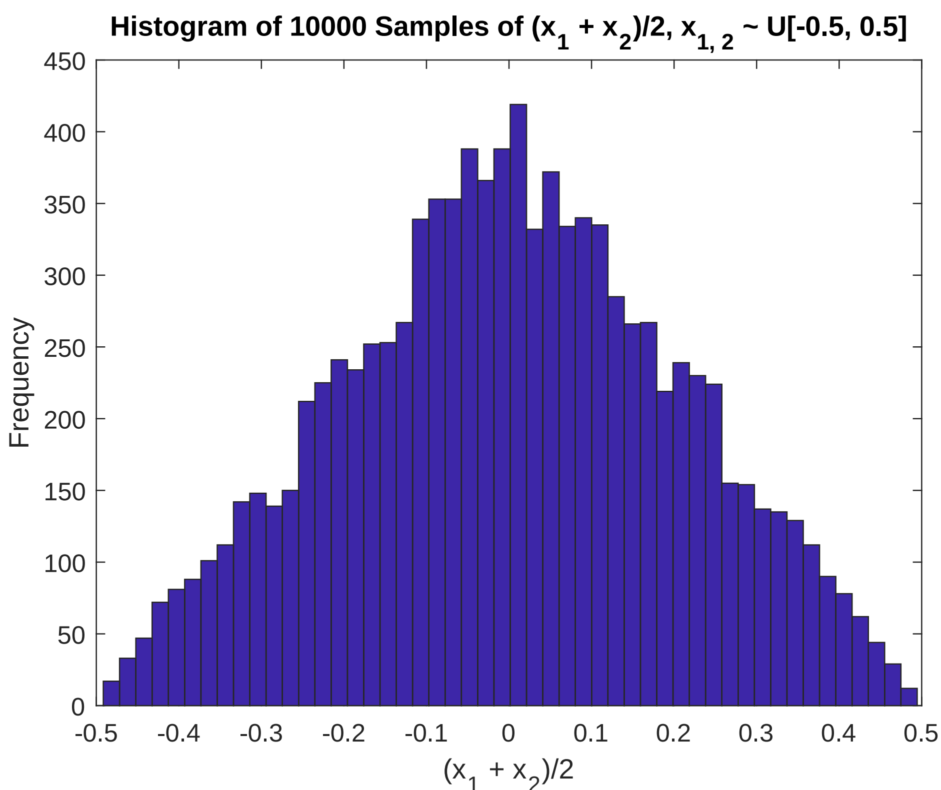

Figure 4: Histogram for $\frac{(x_1+x_2)}{2}$, $(x_1, x_2) \sim \mathcal{U}[-0.5, 0.5]$

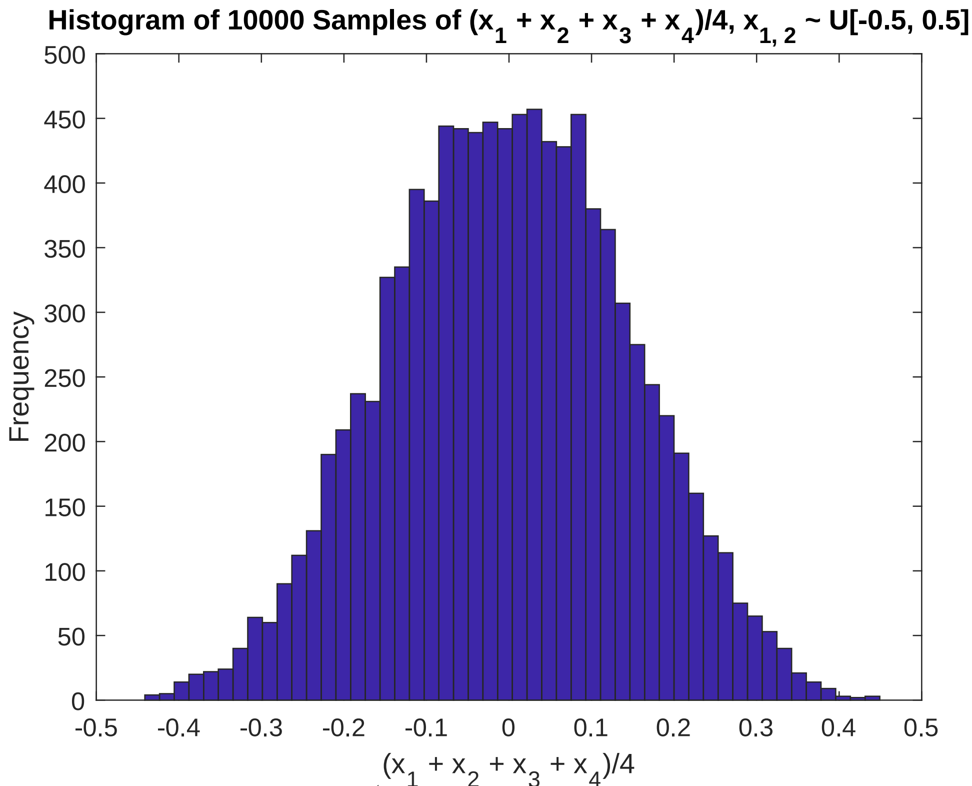

Figure 5: Histogram for $\frac{(x_1+x_2+x_3+x_4)}{4}$, $(x_1, x_2, x_3, x_4) \sim \mathcal{U}[-0.5, 0.5]$

6. The histograms for each averaging formula are shown in Figures 4 and 5. The histogram for $\frac{(x_1+x_2+x_3+x_4)}{4}$ has lower variance and is closer to normally distributed than that for $\frac{(x_1+x_2)}{2}$, which is expected per the Central Limit Theorem.

AQ1. (a) A valid PDF must satisfy

$$\int_{-\infty}^{\infty} f_X(x)dx = 1$$

Thus

$$\int_{-\infty}^{\infty} \frac{ab}{b^2+x^2}dx = \left[a\tan^{-1}\frac{x}{b}\right]_{-\infty}^{\infty} = a\pi = 1 \implies a = \frac{1}{\pi}$$

(b) The mean of a PDF is given by

$$\int_{-\infty}^{\infty} x f_X(x)dx = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{bx}{b^2+x^2}dx$$

This is a two-sided improper integral, and must be evaluated by taking the limits to infinity and negative infinity separately. Thus,

$$\begin{aligned}
\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{bx}{b^2+x^2}dx &= \lim_{d\to\infty}\frac{1}{\pi}\int_c^d \frac{bx}{b^2+x^2}dx + \lim_{d\to-\infty}\frac{1}{\pi}\int_d^c \frac{bx}{b^2+x^2}dx \\
&= \lim_{d\to\infty}\frac{1}{\pi}\left[\frac{1}{2}\ln(b^2+x^2)\right]_c^d + \lim_{d\to-\infty}\frac{1}{\pi}\left[\frac{1}{2}\ln(b^2+x^2)\right]_d^c \\
&= \lim_{d\to\infty}\frac{1}{\pi}\left[\frac{1}{2}\ln(b^2+x^2)\right]_c^d + \lim_{d\to-\infty}\frac{1}{\pi}\left[\frac{1}{2}\ln(b^2+x^2)\right]_d^c
\end{aligned}$$

where $c \in \mathbb{R}$ is some finite constant. Since

$$\lim_{x \to \infty} \ln(b^2 + x^2) \to \infty$$

and

$$\lim_{x \to \infty} \ln(b^2 + x^2) \to \infty$$

neither of the one-sided integrals converge, and therefore the mean of the Cauchy distribution is undefined.

AQ2. (a) First observe that the mean of each of the individual mixands is given by

$$\mu_i = \int_{-a_i}^{b_i} x\mathcal{U}[a_i, b_i]dx$$

and the variances of each mixand are given by

$$\begin{aligned}
\sigma_i^2 &= \int_{-a_i}^{b_i} (x - \mu_i)^2 \mathcal{U}[a_i, b_i]dx \\
&= \int_{-a_i}^{b_i} (x^2 - 2\mu_i x + \mu_i^2)\mathcal{U}[a_i, b_i]dx \\
&= \int_{-a_i}^{b_i} x^2\mathcal{U}[a_i, b_i]dx - 2\mu_i \int_{-a_i}^{b_i} x\mathcal{U}[a_i, b_i]dx + \mu_i^2 \int_{-a_i}^{b_i} \mathcal{U}[a_i, b_i]dx \\
&= \int_{-a_i}^{b_i} x^2\mathcal{U}[a_i, b_i]dx - 2\mu_i^2 + \mu_i^2 \\
&= \int_{-a_i}^{b_i} x^2\mathcal{U}[a_i, b_i]dx - \mu_i^2
\end{aligned}$$

The expected value of x is then given by

$$\begin{aligned}
\mathbb{E}[x] &= \int_{-\infty}^{\infty} xp(x)dx = \int_{-\infty}^{\infty} x \sum_{i=1}^{M} w_i\mathcal{U}[a_i, b_i]dx \\
&= \sum_{i=1}^{M} w_i \int_{-\infty}^{\infty} x\mathcal{U}[a_i, b_i]dx = \sum_{i=1}^{M} w_i \int_{-a_i}^{b_i} x\mathcal{U}[a_i, b_i]dx \\
&= \sum_{i=1}^{M} w_i\mu_i
\end{aligned}$$

and the variance of x is given by

$$\begin{aligned}
\mathrm{var}[x] &= \int_{-\infty}^{\infty} (x-\mu)^2 p(x)dx \\
&= \int_{-\infty}^{\infty} (x^2 - 2\mu x + \mu^2) p(x)dx \\
&= \int_{-\infty}^{\infty} x^2 p(x)dx - 2\mu \int_{-\infty}^{\infty} x p(x)dx + \mu^2 \int_{-\infty}^{\infty} p(x)dx \\
&= \int_{-\infty}^{\infty} x^2 \sum_{i=1}^{M} w_i \mathcal{U}[a_i, b_i]dx - 2\mu^2 + \mu^2 \\
&= \sum_{i=1}^{M} w_i \int_{-a_i}^{b_i} x^2 U[a_i, b_i]dx - \mu^2 \\
&= \sum_{i=1}^{M} w_i \left(\int_{-a_i}^{b_i} x^2 U[a_i, b_i]dx - \mu_i^2 + \mu_i^2 \right) - \mu^2 \\
&= \sum_{i=1}^{M} w_i \left(\sigma_i^2 + \mu_i^2 \right) - \mu^2
\end{aligned}$$

The PDFs of each mixand are defined as follows

$$\mathcal{U}[a_i, b_i] = \begin{cases} \frac{1}{b_i - a_i}, & a_i \leq x \leq b_i \\ 0 & \text{otherwise} \end{cases}$$

(b) The PDF of the mixture model is shown in Figure 6.

(c) The grid-based and analytical (if applicable) mean, variance, differential entropy, and K-L divergence are shown in Table 2. Note that the expressions provided for these quantities in the problem statement use a commonplace abuse of notation, where x, the domain of the PDF $p(x)$, and X the random variable distributed about $p(x)$ are used interchangeably. This is an important distinction to make when interpreting the formulas for differential entropy and K-L divergence, in that the correct interpretation of the formulas is $\mathcal{H}[p(X)] = \mathbb{E}[-\ln p(X)]$ and $\mathrm{KL}[p(X)||q(X)] = \mathbb{E}[-\ln \frac{p(X)}{q(X)}]$, otherwise the formulas would imply taking expectations of deterministic functions of x, which is of little interest.

The analytical mean and variance can be computed by finding expressions for the mean and variance of each of the mixands in terms of a_i and b_i and substituting into the formulas derived in Part (a). The mean of each mixand is given by

$$\begin{aligned}
\mu_i &= \int_{-a_i}^{b_i} x \mathcal{U}[a_i, b_i]dx = \int_{-a_i}^{b_i} \frac{x}{b_i - a_i} dx = \frac{1}{b_i - a_i} \left[\frac{1}{2} x^2 \right]_{a_i}^{b_i} \\
&= \frac{1}{2} \left(\frac{b_i^2 - a_i^2}{b_i - a_i} \right) = \frac{b_i + a_i}{2}
\end{aligned}$$

and the variances by

$$\sigma_i^2 = \int_{-a_i}^{b_i} (x - \mu_i)^2 \mathcal{U}[a_i, b_i]dx = \int_{-a_i}^{b_i} \frac{(x-\mu_i)^2}{b_i - a_i} dx$$

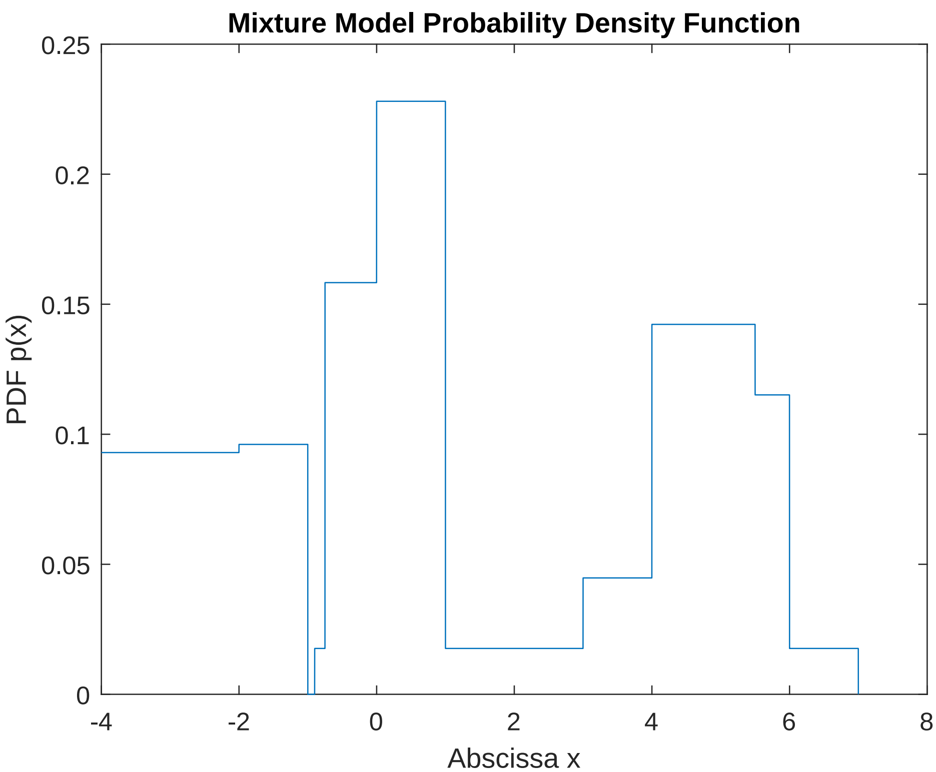

Figure 6: Mixture Model Probability Density Function

Making a change of variables to $x' \triangleq x - \mu_i \implies dx' = dx$

$$\begin{aligned}\sigma_i^2 &= \frac{1}{b_i - a_i}\int_{\frac{b_i-a_i}{2}}^{\frac{a_i-b_i}{2}} x'^2 dx' = \frac{1}{b_i - a_i}\left[\frac{1}{3}x'^3\right]_{\frac{b_i-a_i}{2}}^{\frac{a_i-b_i}{2}} \\ &= \frac{1}{b_i - a_i}\left(\frac{1}{3}\frac{(b_i-a_i)^3}{4}\right) = \frac{1}{12}(b_i - a_i)^2\end{aligned}$$

The full expression for the differential entropy of $p(x)$ is given by

$$\mathcal{H}[p(x)] = -\int_{-\infty}^{\infty} p(x)\ln p(x) dx$$

Limit of the integrand as $p(x) \to 0$ can be expressed as

$$\lim_{p(x)\to 0} p(x)\ln p(x) = \lim_{p(x)\to 0}\frac{\ln p(x)}{\frac{1}{p(x)}}$$

Using l'Hôpital's Rule,

$$\lim_{p(x)\to 0}\frac{\ln p(x)}{\frac{1}{p(x)}} = \lim_{p(x)\to 0}\frac{\frac{1}{p(x)}}{-\frac{1}{p(x)}} = \lim_{p(x)\to 0} -p(x) = 0$$

Table 2: Analytical and Grid-Based Approximations of Mixture Model Mean, Variance, Differential Entropy, and K-L Divergence

Quantity	Grid-Based	Analytical
Mean $\mathbb{E}[X]$	1.0513	1.0518
Variance var$[X]$	9.16319	9.16319
Differential Entropy $\mathcal{H}[p(X)]$	2.12655	
K-L Divergence KL$[p(X)\|\|q(X)]$	0.271343	

(d) The 1000-bin histograms of $p(x)$ for each of the sample set sizes is shown in Figures 7 to 9. The samples were generated by first partitioning the interval $[0, 1]$ into seven partitions with the partition boundaries defined by the provided weights. For each sample, a "selector" value was randomly generated from a uniform distribution over the interval $[0, 1]$, and the partition into which the selector value fell would determine which of the seven mixands would be sampled. A sample is then randomly generated from the PDF of the selected mixand.

It can be seen clearly that with an increasing number of samples, the shape of the histogram approaches that of the analytical PDF shown in Figure 6.

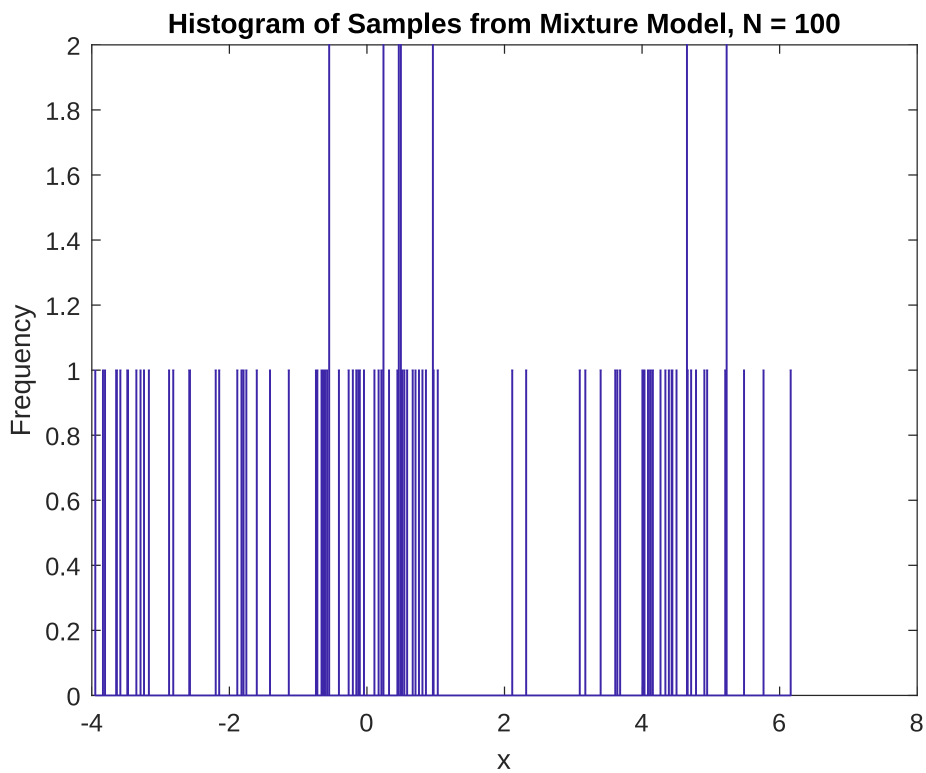

Figure 7: Mixture Model Histogram, $N = 100$

(e) The Monte-Carlo approximations of the mean, variance, differential entropy, and K-L divergence are provided in Table 3. As the number of samples increases,

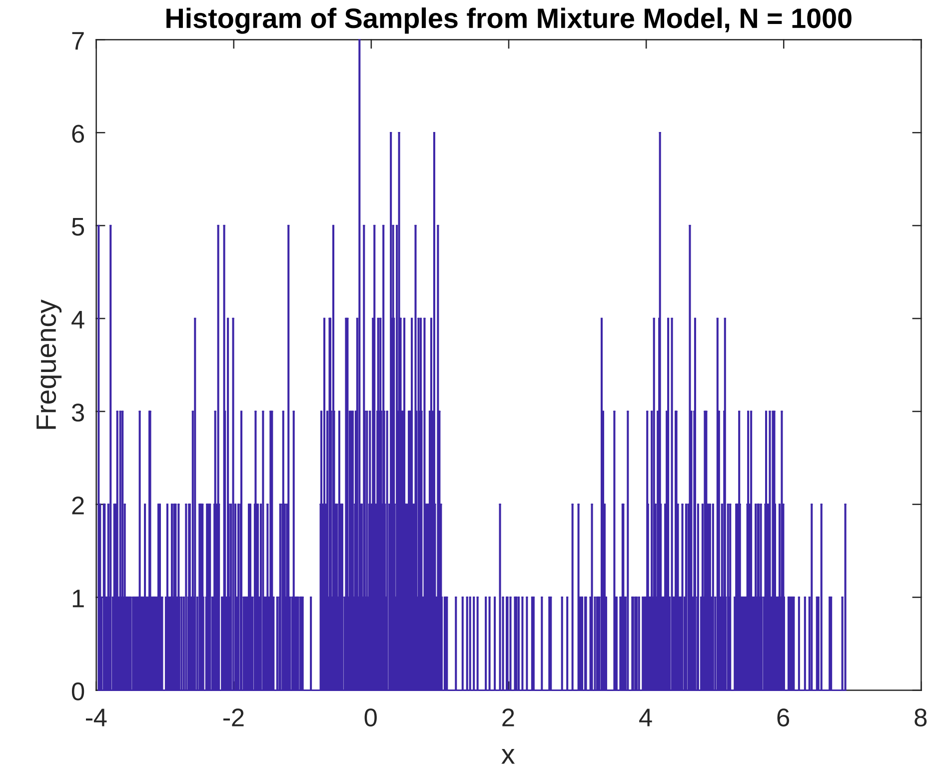

Figure 8: Mixture Model Histogram, $N = 1000$

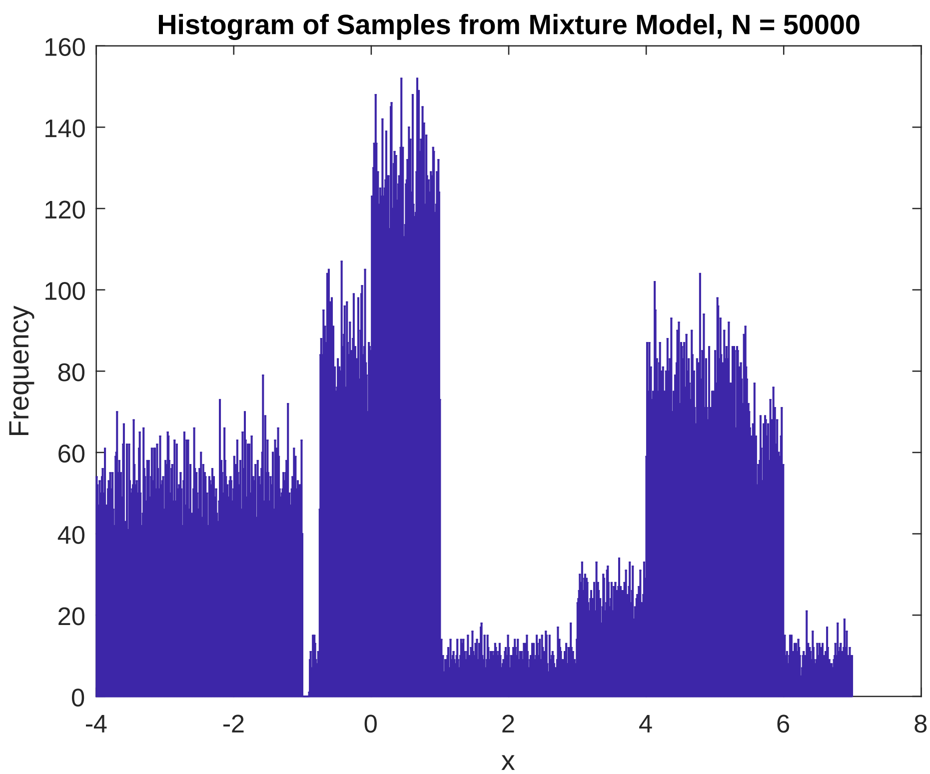

Figure 9: Mixture Model Histogram, $N = 50000$

the Monte-Carlo approximations of these quantities get closer to their grid-based approximations and analytical values.

Table 3: Monte-Carlo Approximations of Mixture Model Mean, Variance, Differential Entropy, and K-L Divergence

Quantity	$N = 100$	$N = 1000$	$N = 50000$
Mean $\mathbb{E}[X]$	0.842645	0.950543	1.05104
Variance $\mathrm{var}[X]$	8.15262	8.95561	9.16787
Differential Entropy $\mathcal{H}[p(X)]$	2.07886	2.14162	2.1293
K-L Divergence $\mathrm{KL}[p(X)\|\|q(X)]$	0.319031	0.256273	0.268592

A MATLAB Script

```
%% ASEN5044 Assignment 4 Solutions
% Y. Shen
% 27 September 2018

%% Problem 5

N = [50 500 5000];              % Number of samples

for i = 1:length(N)
    x = rand(N(i), 1);
    figure;
    hist(x);
    xlabel('x');
    ylabel('Frequency');
    title(sprintf('Histogram of Samples from Uniform Random Distribution \nwith Range [0, 1

    fprintf('Mean of samples with N = %g: %g\n', N(i), mean(x));
    fprintf('Standard deviation of samples with N = %g: %g\n', N(i), std(x));
end

%% Problem 6

x = rand(2, 10000) - 0.5;

figure;
hist(sum(x)/2, 50);
xlabel('(x_1 + x_2)/2');
ylabel('Frequency');
title('Histogram of 10000 Samples of (x_1 + x_2)/2, x_{1, 2} ~ U[-0.5, 0.5]');


x = rand(4, 10000) - 0.5;

figure;
hist(sum(x)/4, 50);
xlabel('(x_1 + x_2 + x_3 + x_4)/4');
ylabel('Frequency');
title('Histogram of 10000 Samples of (x_1 + x_2 + x_3 + x_4)/4, x_{1, 2} ~ U[-0.5, 0.5]');


%% Advanced Question 2

% Part B

% Define generic uniform PDF
U = @(x, a, b) (x < b).*(x >= a)./(b - a);

% Define weights and ranges
w = [0.1859; 0.0961; 0.1055; 0.2104; 0.0678; 0.1950; 0.1393];
a = [-4; -2; -0.75; 0; 3; 4; -0.9];
```

```
b = [-2; -1; 0; 1; 5.5; 6; 7];

% Define mixture model
p = @(x) w'*U(x, a, b);

% Define grid
dx = 0.001;
x = -4:dx:7;
y = zeros(1, length(x));
for i = 1:length(x)
    y(i) = p(x(i));
end

% Plot
figure;
plot(x, y);
xlabel('Abscissa x');
ylabel('PDF p(x)');
title('Mixture Model Probability Density Function');

% Part C

% Compute mixture model mean, variance, differential entropy, and KL
% divergence using left endpoint rule
mu_grid = x*y'*dx;
var_grid = (x - mu_grid).^2*y'*dx;
H = 0;
for i = 1:length(x)
    if y(i) > 0
        H = H - y(i)*log(y(i))*dx;
    end
end
KL = 0;
for i = 1:length(x)
    if y(i) > 0
        KL = KL + y(i)*log(y(i)*11)*dx;
    end
end

% Compute analytical mean and variance
mui = (b + a)/2;
mu = w'*mui;
vari = 1/12*(b - a).^2;
va = w'*(vari + mui.^2) - mu^2;

fprintf('Grid-based mean: %g, analytical mean: %g\n', mu_grid, mu);
fprintf('Grid-based variance: %g, analytical variance: %g\n', var_grid, va);
fprintf('Grid-based differential entropy: %g\n', H);
fprintf('Grid-based K-L divergence: %g\n', KL);

% Part D

% Define sample sizes, define x as cell array
N = [100; 1000; 50000];
```

```
x = cell(3, 1);
part = [0; cumsum(w)];                    % Partition range of rand(), i.e. [0, 1], a
for i = 1:length(N)
    % Preallocate x
    x{i} = zeros(N(i), 1);

    % Compute mixand selectors and sample from mixands accordingly
    sel = rand(N(i), 1);                  % Mixand selectors
    for k = 1:N(i)
        comp = (sel(k) < part);           % Compare selector to partitions
        ind = find(comp(2:end) - comp(1:end - 1));
        x{i}(k) = a(ind) + (b(ind) - a(ind))*rand();
    end

    figure;
    hist(x{i}, 1000);
    xlabel('x');
    ylabel('Frequency');
    title(sprintf('Histogram of Samples from Mixture Model, N = %g', N(i)));
end

% Part E

for i = 1:length(N)
    mu_mc = mean(x{i});
    var_mc = var(x{i});

    y = zeros(N(i), 1);
    for k = 1:N(i)
        y(k) = p(x{i}(k));
    end
    H_mc = mean(-log(y));
    KL_mc = mean(log(y*11));

    fprintf('Mean of Monte Carlo approximation for N = %g: %g\n', N(i), mu_mc);
    fprintf('Variance of Monte Carlo approximation for N = %g: %g\n', N(i), var_mc);
    fprintf('Differential entropy of Monte Carlo approximation for N = %g: %g\n', N(i), H_m
    fprintf('K-L divergence of Monte Carlo approximation for N = %g: %g\n', N(i), KL_mc);
end
```