

Exercise 1

In the game of blackjack, the player is initially dealt two cards from a deck of ordinary playing cards. Without going into all the game's details, it is enough to know the best possible hand for a player to receive on the initial deal is a combination of an ace of any suit and any face card or ten. What is the probability that a player will be dealt this combination?

The number of ways a single player can be dealt a blackjack (assuming there is only one player) is $\binom{4}{1}\binom{16}{1} = 64$. So the number of events in our event space (which we'll call A) is 64. The total number of ways to deal 2 cards to that player is $\binom{52}{2} = 1326$. So, the number of events in the whole sample space is 1326. the probability of a single player being dealt a blackjack is therefore

$$\frac{N_A}{N} = \frac{\binom{4}{1}}{\binom{16}{1}} = \frac{64}{1326} = 0.048$$

Exercise 2

Discrete and random variables X and Y can each take on integer values 1, 3, and 5. The joint probability table of X and Y is given below.

Table 1: Joint Probabilities

X	Y=1	Y=3	Y=5
1	1/18	1/18	1/18
3	1/18	1/18	1/6
5	1/18	1/6	1/3

Problem (a)

Are the random variables X and Y independent?

No, the table clearly shows that the probability of X taking on certain values changes depending on the value of Y . For instance $P(X = 3|Y = 3) \neq P(X = 3|Y = 5)$ If X and Y were independent then $P(X = x|Y = y)$ would be the same for all values of y .

Problem (b)

Find the unconditional (marginal) probability $P(Y = 5)$.

The marginal probability $P(Y = 5) = P(Y = 5|X = 1) + P(Y = 5|X = 3) + P(Y = 5|X = 5) = (1/18) + (1/6) + (1/3) = (5/9)$.

Problem (c)

What is the conditional probability $P(Y = 5|X = 3)$?

Generally $P(A|B) = \frac{P(A,B)}{P(B)}$. This means we need to calculate the marginal probability $P(X = 3) = (1/18) + (1/18) + (1/6) = (5/18)$. With this information we can find

$$P(Y = 5|X = 3) = \frac{P(Y = 5, X = 3)}{P(X = 3)} = \frac{1/6}{5/18} = \frac{3}{5}$$

Exercise 3

Determine the value of a in the function

$$f_X(x) = \begin{cases} ax(1-x) & x \in [0, 1] \\ 0 & \text{otherwise} \end{cases}$$

so that $f_X(x)$ is a valid probability density function.

For $f_X(x)$ to be a valid probability density function $\int_{-\infty}^{\infty} f_X(x)dx$ must be equal to one.

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

Exercise 4

The probability density function of an exponentially distributed random variable is defined as follows

$$f_X(x) = \begin{cases} ae^{-ax} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

where $a \geq 0$.

Problem (a)

Find the probability distribution function of an exponentially distributed random variable.

$$\begin{aligned}
 P(a \leq X \leq b) &= \int_a^b f_X(x) dx \\
 &= \int_a^b a e^{-ax} dx \\
 &= -e^{-ax} \Big|_a^b
 \end{aligned}$$

assuming $0 < a \leq b$.

Problem (b)

Find the mean of an exponentially distributed random variable.

$$\begin{aligned}
 E[X] &= \int_{-\infty}^{\infty} x f_X(x) dx \\
 &= a \int_0^{\infty} x e^{-ax} dx \\
 &= a \left[\left(-\frac{x}{a} - \frac{1}{a^2} \right) e^{-ax} \right]_0^{\infty} \\
 &= \left[\left(-x - \frac{1}{a} \right) e^{-ax} \right]_0^{\infty} \\
 &= \left(-\infty - \frac{1}{a} \right) e^{-a\infty} + \frac{1}{a} e^0
 \end{aligned}$$

Because $e^{-\infty} = 0$, $E[X] = \frac{1}{a}$.

Problem (c)

Find the second moment of an exponentially distributed random variable.

$$\begin{aligned}
 E[X^2] &= \int_{-\infty}^{\infty} x^2 f_X(x) dx \\
 &= \int_0^{\infty} x^2 a e^{-ax} dx \\
 &= \frac{1}{a^2} \int_0^{\infty} t^2 e^{-t} dt \\
 &= \frac{1}{a^2} \Gamma(2 + 1) = \frac{2!}{a^2} \\
 &= \frac{2}{a^2}
 \end{aligned}$$

Problem (d)

Find the variance of an exponentially distributed random variable.

$$\begin{aligned}\sigma_x^2 &= E[X^2] - (E[X])^2 \\ &= \frac{2}{a^2} - \left(\frac{1}{a}\right)^2 \\ &= \frac{1}{a^2}\end{aligned}$$

Problem (e)

What is the probability that an exponentially distributed random variable takes on a value within one standard deviation of its mean?

The standard deviation is $\sqrt{\frac{1}{a}} = \frac{1}{a}$, so we simply need to evaluate the probability distribution function from 0 to $\frac{2}{a}$:

$$\begin{aligned}P\left(0 < X \leq \frac{2}{a}\right) &= 1 - e^{-ax} \Big|_0^{2/a} \\ &= (1 - e^{-a \frac{2}{a}}) - (1 - e^0) \\ &= 1 - e^{-2}\end{aligned}$$

So the probability that X falls within one standard deviation of the mean is a constant, unrelated to the value of a .

Exercise 5

Generate $N = 50$ independent random numbers, each uniformly distributed between 0 and 1. Plot a histogram of the random numbers using 10 bins. What is the sample mean and standard deviation of the numbers that you generated? What would you expect to see for the mean and standard deviation? Repeat for $N = 500$ and $N = 5000$ random numbers. What changes do you see in the histogram as N increases?

Our uniformly distributed probability density function is defined as

$$f(x) = \begin{cases} c & x \in [0, 1] \\ 0 & \text{otherwise} \end{cases}$$

Because $\int_{-\infty}^{\infty} f(x) = 1$, $c = 1$. The expected value of this pdf is

$$\begin{aligned} E[X] &= \int_{-\infty}^{\infty} x f(x) dx \\ &= \int_0^1 x dx \\ &= \frac{1}{2} x^2 \Big|_0^1 \\ &= \frac{1}{2} \end{aligned}$$

and the expected value of X^2 is

$$\begin{aligned} E[X^2] &= \int_{-\infty}^{\infty} x^2 f(x) dx \\ &= \int_0^1 x^2 dx \\ &= \frac{1}{3} x^3 \Big|_0^1 \\ &= \frac{1}{3} \end{aligned}$$

So the variance can be calculated as

$$\sigma^2 = E[X^2] - (E[X])^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}$$

which means the standard deviation is $\sqrt{\sigma^2} = 0.2887$.

With a sample size $N = 50$ we get $\mu = 0.4899$ and $\sigma = 0.2786$, which is fairly close to the theoretical values. See figure 1, below, for a histogram of the experimental values. From the histogram we can see that the actual distribution of values is not terribly uniform.

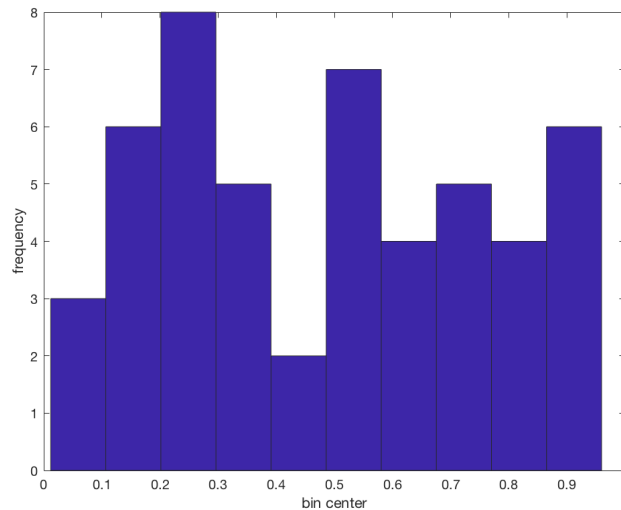


Figure 1: histogram for $N = 50$

With a sample size $N = 500$ we get $\mu = 0.4862$ and $\sigma = 0.2898$, which is fairly close to the theoretical values, though the μ value is actually further from the theoretical value than when $N = 50$. See figure 2, below, for a histogram of the experimental values. From the histogram we can see that the actual distribution of values is more uniform than in the previous case, though it still has distinct peaks that one would not expect with a truly uniform distribution.

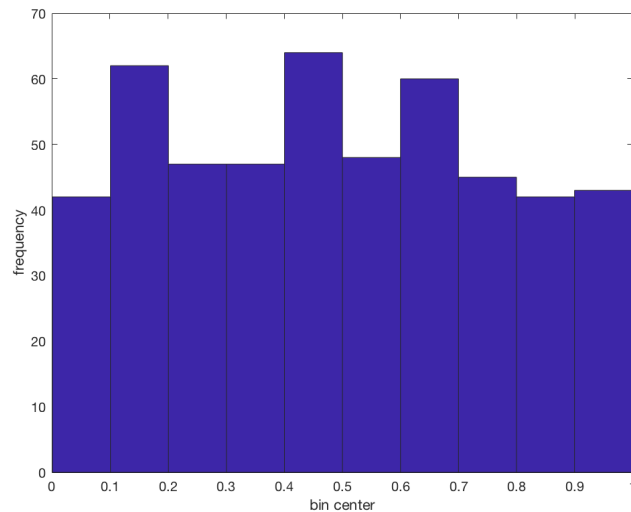


Figure 2: histogram for $N = 500$

With a sample size $N = 5000$ we get $\mu = 0.4999$ and $\sigma = 0.2892$, which is nearly the same as the theoretical values. See figure 3, below, for a histogram of the experimental values. From the histogram we can see that the actual distribution of values is very nearly uniform, with each of the 10 bins containing about 500 values.

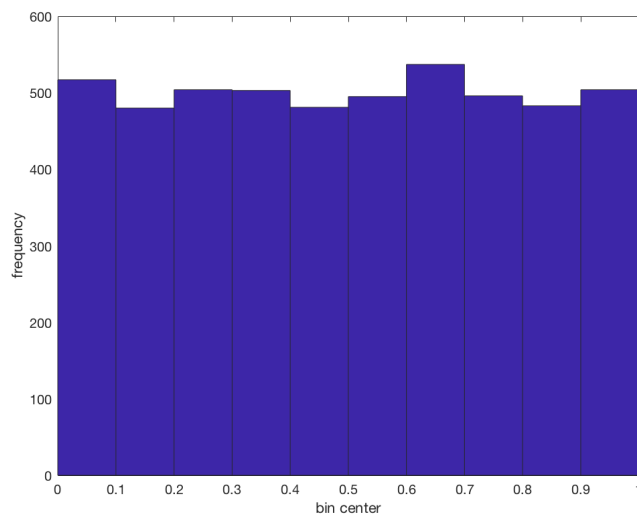


Figure 3: histogram for $N = 5000$

Exercise 6

Generate 10000 samples of $(x_1 + x_2)/2$, where each x_i is a random number uniformly distributed on $[-1/2, +1/2]$. Plot the samples in a 50-bin histogram. Repeat for $(x_1 + x_2 + x_3 + x_4)/4$. Describe the difference between the two histograms.

The plot below shows a histogram for 10000 samples of $(x_1 + x_2)/2$. Because each value in this distribution is the mean of two samples from a uniform probability distribution one would expect the distribution to be more heavily weighted toward the mean and to taper off linearly as x diverges from the mean. This is exactly what we see in figure 4, below.

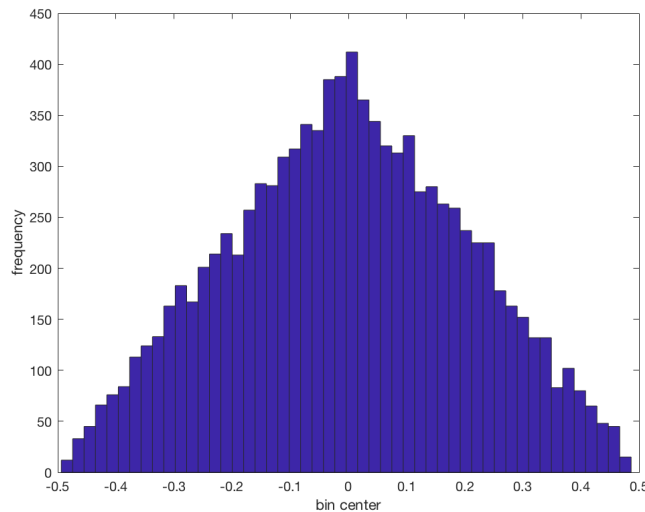


Figure 4: histogram for $(x_1 + x_2)/2$

The plot below shows a histogram for 10000 samples of $(x_1 + x_2 + x_3 + x_4)/4$. Because each value in this distribution is the mean of four samples from a uniform probability distribution one would expect the distribution to be more heavily weighted toward the mean as in the previous case. However, instead of tapering linearly we would expect the tapering to approximate a polynomial shape. Again, this is what we see in figure 5, below.

AQ 1

Consider the following probability density function:

$$f_X(x) = \frac{ab}{b^2 + x^2}, \quad b > 0$$

Problem (a)

Determine the value of a in the pdf so that $f_X(x)$ is a valid probability density function. Note: the correct value of a makes $f_X(x)$ a Cauchy pdf.

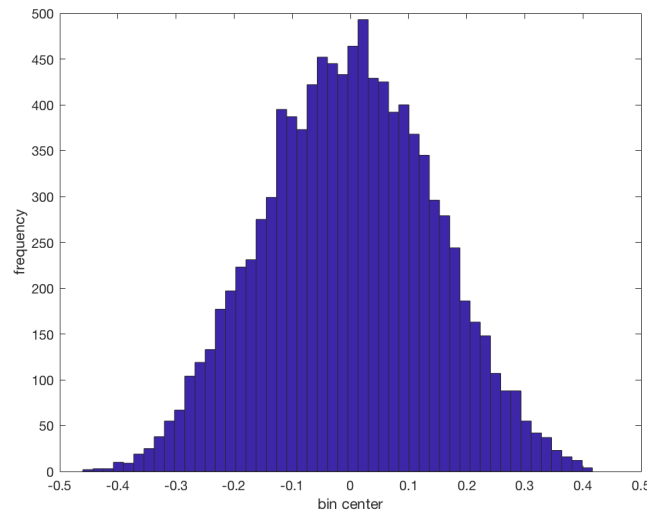


Figure 5: histogram for $(x_1 + x_2 + x_3 + x_4)/4$

Knowing $\int_{-\infty}^{\infty} f_X(x)dx = 1$ we can find a via the following:

$$\begin{aligned}
 1 &= \int_{-\infty}^{\infty} \frac{ab}{b^2 + x^2} dx \\
 &= ab \int_{-\infty}^{\infty} \frac{1}{b^2 + x^2} dx \\
 &= ab \left[\frac{1}{b} \tan^{-1} \frac{x}{b} \right]_{-\infty}^{\infty} \\
 &= a \tan^{-1} \frac{x}{b} \Big|_{-\infty}^{\infty} \\
 &= \frac{a\pi}{2} + \frac{a\pi}{2} \\
 &= a\pi \\
 a &= \frac{1}{\pi}
 \end{aligned}$$

Problem (b)

Find the mean of a Cauchy random variable.

If we try to find the mean of the Cauchy random variable as $E[X] = \int_{-\infty}^{\infty} x f_X(x) dx$ we find:

$$\begin{aligned}
 \int_{-\infty}^{\infty} x f_X(x) dx &= \frac{b}{\pi} \int_{-\infty}^{\infty} \frac{x}{b^2 + x^2} dx \\
 &= \frac{b}{2\pi} \ln |b^2 + x^2| \Big|_{-\infty}^{\infty}
 \end{aligned}$$

The above indefinite integral can be evaluated by breaking it up into two separate integrals:

$$\frac{b}{2\pi} \ln |b^2 + x^2| \Big|_{-\infty}^{\infty} = \frac{b}{2\pi} \ln |b^2 + x^2| \Big|_0^{\infty} + \frac{b}{2\pi} \ln |b^2 + x^2| \Big|_{-\infty}^0$$

However, both of the terms in this equation are infinite and have opposite signs so the mean of the Cauchy distribution is undefined.

AQ 2

Consider a continuous random variable x whose probability density function $p(x)$ is given by a mixture model consisting of M weighted uniform pdfs,

$$p(x) = \sum_{i=1}^M w_i U[a_i, b_i],$$

where $w_i \geq 0$, $\sum_{i=1}^M w_i = 1$, and $a_i < b_i$ for all mixture components ('mixands') $i = 1, \dots, M$.

Problem (a)

Find general closed-form expressions for $E[x]$ and $\text{var}[x]$ in terms of the means u_i and variances σ_i^2 of the individual mixands $U[a_i, b_i]$.

$$\begin{aligned} E[x] &= \int_{-\infty}^{\infty} x \sum_{i=1}^M w_i U[a_i, b_i] dx \\ &= \sum_{i=1}^M w_i \frac{1}{b_i - a_i} \int_{a_i}^{b_i} x dx \\ &= \sum_{i=1}^M \frac{w_i}{b_i - a_i} \left[\frac{1}{2} x^2 \right]_{a_i}^{b_i} \\ &= \sum_{i=1}^M \frac{w_i}{b_i - a_i} \frac{b_i^2 - a_i^2}{2} \\ &= \sum_{i=1}^M \frac{w_i}{2} (a_i + b_i) \end{aligned}$$

$$\begin{aligned} \text{var}(x) &= \int_{-\infty}^{\infty} x^2 \sum_{i=1}^M w_i \frac{1}{b_i - a_i} dx - \mu^2 \\ &= \sum_{i=1}^M \frac{w_i}{b_i - a_i} \left[\frac{1}{3} x^3 \right]_{a_i}^{b_i} - \mu^2 \\ &= \sum_{i=1}^M \frac{w_i}{b_i - a_i} \frac{b_i^3 - a_i^3}{3} - \mu^2 \end{aligned}$$

Problem (b)

Suppose the (w_i, a_i, b_i) parameters for $M = 7$ mixands are specified as follows, in order for components $i = 1, \dots, 7$:

$(0.1859, -4, -2), (0.0961, -2, -1), (0.1055, -0.75, 0), (0.2104, 0, 1), (0.0678, 3, 5.5), (0.1950, 4, 6), (0.1393, -0.9, 7)$

Plot the resulting values for $p(x)$ with these mixture parameters over the range $x \in [-6, 8]$, using an abscissa step size of $\delta x = 0.001$ to evaluate at discrete grid points.

See figure 8 below for a plot of the resulting pdf for the given parameters.

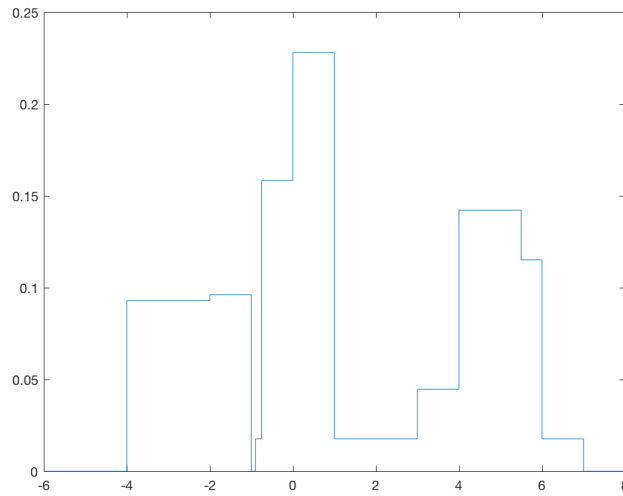


Figure 6: pdf for mixture model

Problem (c)

Use (b)'s results to compute the following via numerical grid-based approximations:

- $E[x]$ - how does this compare to the analytical result?
- $\text{var}(x)$ - how does this compare to the analytical result?
- $H[p(x)] = E[-\log p(x)]$
- $KL[p(x)||q(x)] = E[\log \frac{p(x)}{q(x)}]$

The expected value of $p(x)$ found by numerically integrating $xp(x)$ over $[-6, 8]$ is 1.0523. This agrees well with the analytical mean of 1.0518.

The variance of $p(x)$ found by numerically integrating $x^2p(x)$ over $[-6, 8]$ is 10.2705. This also agrees well with the analytical variance of 10.2695.

Because $\log(0) = -\infty$, I dropped any point where $p(x) = 0$ from the integration. The numerical integration of $-p(x) \log(p(x))$ resulted in a differential entropy of 2.1265.

I found the Kullback-Leibler divergence to be 0.2712 by numerically integrating $p(x) \log(p(x)/q(x))$, again dropping any values of p or q that would result in an infinite value for the log.

Problem (d)

Generate random sample sets of size $N = 100$, $N = 1000$, and $N = 50000$ from $p(x)$ and plot the results on histograms with 1000 bins to compare to (b)'s results.

When only 100 samples are drawn from the mixture distribution, a histogram of those samples, shown in figure 7, does not closely resemble the pdf shown in figure 8.

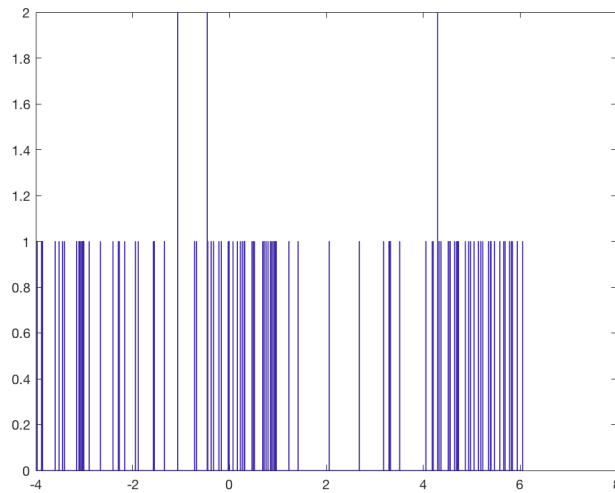


Figure 7: histogram of samples from mixture model with $N = 100$

When 1000 samples are drawn from the mixture distribution, a histogram of those samples, shown in figure ??, begins to resemble the pdf shown in figure 8, but the resemblance is still rough.

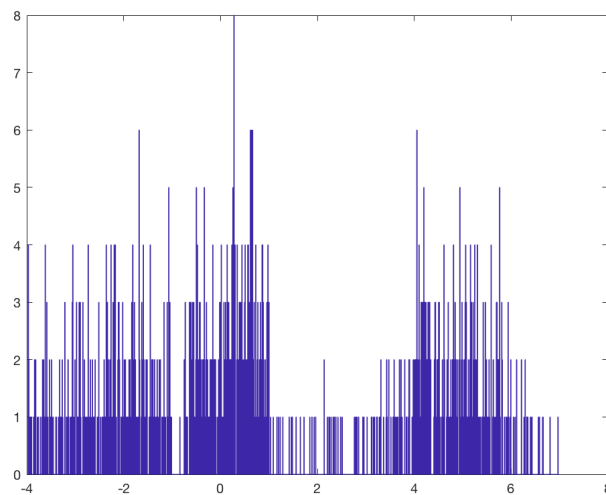


Figure 8: histogram of samples from mixture model with $N = 1000$

When 50000 samples are drawn from the mixture distribution, a histogram of those samples, shown in figure 9, now very closely matches the pdf shown in figure 8.

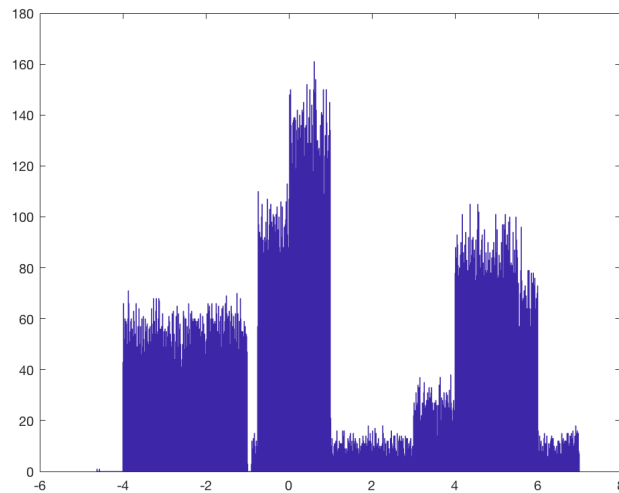


Figure 9: histogram of samples from mixture model with $N = 50000$

Problem (e)

Use each of the sample sets from (d) to produce direct Monte Carlo approximations of the quantities evaluated in (c). How do the results for each sample set size compare to each other and to the results from part (c)?

For a sample size of 100, the results can be seen below in table 2.

Table 2: Approximations with $n = 100$

$E[x]$	1.2142
$\text{var}(x)$	10.0645
$H[p(x)]$	4.446
$KL[p(x) q(x)]$	0.0

For a sample size of 1000, the results can be seen below in table 3.

Table 3: Approximations with $n = 1000$

$E[x]$	1.0743
$\text{var}(x)$	9.0373
$H[p(x)]$	3.483
$KL[p(x) q(x)]$	

For a sample size of 50000, the results can be seen below in table 4.

Table 4: Approximations with $n = 50000$

$E[x]$	1.0614
$\text{var}(x)$	9.1879
$H[p(x)]$	6.8013
$KL[p(x) q(x)]$	

In all cases the monte-carlo estimates of μ and σ^2 approximate the true values fairly well. Interestingly, the estimated value for μ converges to the true value as sample size increases but the estimated value for σ^2 diverges from the true value as sample size increases. My estimates for the differential entropy are of the same order of magnitude as the analytical result, but they are not very close. My method for estimating the differential entropy was likely incorrect. I was unable to find a way to estimate the Kullback Leibler divergence from sampled data.