CMPE 343: Introduction to Probability and Statistics for Computer Engineers

Fall 2021

Homework 1

Altay Acar – 2018400084

30.11.2021

1. a)

The events A and A^c are considered and A^c is the complement of the event A. The intersection of the sets A and B, denoted $A \cap B$, is the set of all objects that are members of both A and . When it comes to the terms of events, since it is not possible that an event can occur and do not occur at the same time, from the definition of intersection we know that:

$$A \cap A^c = \emptyset$$

This makes the events A and A^c disjoint. From the axiom of "Countable Additivity:"

$$P(A \cup A^c) = P(A) + P(A^c) \qquad (1)$$

Considering a sample space Ω that is for the event A, A will either occur, or not occur. Thus, the probability of the sample space will equal to the union of probability that the event A occurs and does not occur.

$$P(\Omega) = P(A \cup A^c) \quad (2)$$

Replacing (1) in (2):

$$P(\Omega) = P(A) + P(A^c) \qquad (3)$$

From the "Normalization" axiom the probability of entire sample space is:

$$P(\Omega) = 1 \qquad (4)$$

Replacing (4) in (3):

$$1 = P(A) + P(A^c) \quad | - P(A^c)\ from\ both\ sides$$

$$P(A) = 1 - P(A^c)$$

b)

Considering a sample space Ω containing event A, from the "Normalization" axiom the probability of entire sample space is:

$$P(\Omega) = 1$$

The intersection of the event A and the sample space Ω does therefore mean that event A and all the possible events that constitutes Ω occur at the same time. Thus, we can say that:

$$A \cap \Omega = A$$

When it comes to the terms of probability:

$$P(A \cap \Omega) = P(A) \qquad (1)$$

From the "Countable Additivity" axiom we know that for a countable collection of disjoint events $E_1, E_2, \dots\ E_n$ we have:

$$P\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{i=1}^{n} P(E_i) \qquad (2)$$

We also know that the union of these disjoint events constitutes the sample space Ω:

$$\bigcup_{i=1}^{n} E_i = \Omega$$

Also, when it comes to terms of probability:

$$P\left(\bigcup_{i=1}^{n} E_i\right) = P(\Omega) \qquad (3)$$

Substituting (3) in (2) we have:

$$P(\Omega) = \sum_{i=1}^{n} P(E_i) \qquad (4)$$

Substituting (4) in (1) we get:

$$P(A) \cap \sum_{i=1}^{n} P(E_i) = P(A)$$

Using the distributive property of intersection operation, we get:

$$P(A) = \sum_{i=1}^{n} P(A \cap E_i)$$

c)

Considering two events A and B. We can express the term $A \cup (B \cap A^c)$ using the distributive property of union operation as follows:

$$A \cup (B \cap A^c) = (A \cup B) \cap (A \cup A^c)$$

And, from the section (a) we know that:

$$A \cup A^c = 1$$

If we use the identity law, we get:

$$A \cup (B \cap A^c) = A \cup B$$

From the "Countable Additivity" axiom we get:

$$P(A) + P(B \cap A^c) = P(A \cup B) \qquad (1)$$

From section (a) we know that $A \cup A^c = 1$. Combining it with the identity law of intersection operation we can express the event B as follows:

$$B = B \cap (A \cup A^c)$$

Using the distributive property of intersection operation:

$$B = B \cap (A \cup A^c) = (B \cap A) \cup (B \cap A^c)$$

Applying the "Countable Additivity" axiom we get:

$$P(B) = P(B \cap A) + P(B \cap A^c) \qquad | - P(B \cap A)\ from\ both\ sides$$

$$P(B) - P(B \cap A) = P(B \cap A^c) \qquad (2)$$

Substituting (2) in (1):

$$P(A \cup B) = P(A) + P(B) - P(B \cap A) \qquad (3)$$

Multiplying both sides of equation (3) with -1:

$$-P(A \cup B) = -P(A) - P(B) + P(B \cap A) \qquad (4)$$

We know that the probability of the entire sample space is 1 from "Normalization" axiom. So, we can say that the probability of union of two events A and B of a sample space can be at most 1 and from the "Nonnegativity" axiom, we know that it is greater than 0:

$$0 \leq P(A \cup B) \leq 1$$

Multiplying each side with -1:

$$0 \geq -P(A \cup B) \geq -1 \qquad (5)$$

Substituting (4) in (5) and considering only the lower bound:

$$-P(A) - P(B) + P(B \cap A) \geq -1 \quad (6)$$

Adding $P(A)$ and $P(B)$ to both sides of (6) we get:

$$P(B \cap A) \geq P(A) + P(B) - 1$$

From the associativity property of intersection operation, we get:

$$P(A \cap B) \geq P(A) + P(B) - 1$$

2.

Two fair dice are thrown and the probability distribution of random variables X and Y are given as follows:

$$X = \begin{cases} 1, & \mathit{if\ the\ sum\ of\ two\ numbers}\ \leq 5 \\ 0, & \mathit{otherwise} \end{cases}$$

$$Y = \begin{cases} 1, & \mathit{if\ the\ product\ of\ two\ numbers\ is\ odd} \\ 0, & \mathit{otherwise} \end{cases}$$

Putting every possible result of this dice throw and the corresponding value of random variables X and Y into a table (Table 1):

Dice throw	X	Y
(1, 1)	1	1
(1, 2)	1	0
(1, 3)	1	1
(1, 4)	1	0
(1, 5)	0	1
(1, 6)	0	0
(2, 1)	1	0
(2, 2)	1	0
(2, 3)	1	0
(2, 4)	0	0
(2, 5)	0	0
(2, 6)	0	0
(3, 1)	1	1
(3, 2)	1	0
(3, 3)	0	1
(3, 4)	0	0
(3, 5)	0	1
(3, 6)	0	0
(4, 1)	1	0
(4, 2)	0	0
(4, 3)	0	0
(4, 4)	0	0
(4, 5)	0	0
(4, 6)	0	0
(5, 1)	0	1
(5, 2)	0	0
(5, 3)	0	1
(5, 4)	0	0
(5, 5)	0	1
(5, 6)	0	0
(6, 1)	0	0
(6, 2)	0	0
(6, 3)	0	0
(6, 4)	0	0

(6, 5)	0	0
(6, 6)	0	0

From this table we can construct the joint probability function (Table 2):

$f(x,y)$		Y	
		1	0
X	1	$\frac{3}{36}$	$\frac{7}{36}$
	0	$\frac{6}{36}$	$\frac{20}{36}$

The random variables X and Y are discrete random variables. Because the number of possible values for these variables are countable, namely 1 and 0 for both. On the contrary if they were continuous random variables, they would take on all the values in some interval of numbers.

Knowing both random variables X and Y are discrete, we can first calculate the mean of both as follows:

$$\mu_X = E(X) = \sum\nolimits_x x * f(x)$$

$$\mu_Y = E(Y) = \sum\nolimits_y y * f(y)$$

For X from the "Table 1" above we know that the probability of the random variable X to be 1 is $\frac{10}{36}$ and probability of it being 0 is $\frac{26}{36}$.

For Y from the "Table 1" above we know that the probability of the random variable Y to be 1 is $\frac{9}{36}$ and probability of it being 0 is $\frac{27}{36}$.

Putting these values in the formulas for calculating mean, we get:

$$\mu_X = E(X) = \sum\nolimits_x x * f(x) = 1 * \frac{10}{36} + 0 * \frac{26}{36} = \frac{10}{36}$$

$$\mu_Y = E(Y) = \sum\nolimits_y y * f(y) = 1 * \frac{9}{36} + 0 * \frac{27}{36} = \frac{9}{36}$$

Now, we can compute the covariance of X and Y with given covariance formula for discrete random variables:

$$E[(X - \mu_X)(Y - \mu_Y)] = \sum_x \sum_y (x - \mu_X) * (y - \mu_Y) * f(x, y)$$

Putting the mean values and possible values for X and Y with the corresponding value of probability function $f(x, y)$ from "Table 2" in the above equation for covariance, we get:

$$\sigma_{XY} = \left(1-\frac{10}{36}\right)*\left(1-\frac{9}{36}\right)*\frac{3}{36}+\left(1-\frac{10}{36}\right)*\left(0-\frac{9}{36}\right)*\frac{7}{36}+\left(0-\frac{10}{36}\right) *\left(1-\frac{9}{36}\right)*\frac{6}{36}+\left(0-\frac{10}{36}\right)*\left(0-\frac{9}{36}\right)*\frac{20}{36}$$
$$= \frac{26}{36}*\frac{27}{36}*\frac{3}{36}+\frac{26}{36}*\left(-\frac{9}{36}\right)*\frac{7}{36}+\left(-\frac{10}{36}\right)*\frac{27}{36}*\frac{6}{36}+\left(-\frac{10}{36}\right) *\left(-\frac{9}{36}\right)*\frac{20}{36}=\frac{13}{288}-\frac{91}{2592}-\frac{5}{144}+\frac{25}{648}=\frac{1}{72}$$

Thus, as a result the covariance of the given random variables X and Y is:

$$Cov(X,Y) = \sigma_{XY} = \frac{1}{72}$$

3.

We know that the probability distribution of the Poisson random variable X is:

$$p(x;\ \lambda t) = \frac{e^{-\lambda t}(\lambda t)^x}{x!} \qquad for\ x = 0, 1, 2, 3, \ldots$$

The expected value of random variable X would give us the mean of the Poisson random variable:

$$E[X] = \mu_X = \sum_{x=0}^{\infty} x * p(x;\ \lambda t)$$

Putting the probability distribution in the mean formula, we get:

$$E[X] = \mu_X = \sum_{x=0}^{\infty} x * \frac{e^{-\lambda t}(\lambda t)^x}{x!}$$

We replace the $x!$ with $x * (x-1)!$ from the definition of factorial:

$$E[X] = \mu_X = \sum_{x=0}^{\infty} x * \frac{e^{-\lambda t}(\lambda t)^x}{x * (x-1)!}$$

The x's inside the sum cancel each other. So, we get:

$$E[X] = \mu_X = \sum_{x=0}^{\infty} \frac{e^{-\lambda t}(\lambda t)^x}{(x-1)!}$$

We multiply and divide the term inside the sum with λt:

$$E[X] = \mu_X = \sum_{x=0}^{\infty} \frac{e^{-\lambda t}(\lambda t)^x (\lambda t)(\lambda t)^{-1}}{(x-1)!}$$

Because λt does not depend on X, we can take that term outside of the summation. Also, we combine $(\lambda t)^x$ with $(\lambda t)^{-1}$:

$$E[X] = \mu_X = \lambda t \sum_{x=0}^{\infty} \frac{e^{-\lambda t}(\lambda t)^{x-1}}{(x-1)!} \qquad (1)$$

From the Poisson distribution function, we know that for $(x-1)$:

$$p((x-1);\ \lambda t) = \frac{e^{-\lambda t}(\lambda t)^{x-1}}{(x-1)!} \qquad (2)$$

Replacing (2) in (1), we get:

$$E[X] = \mu_X = \lambda t \sum_{x=0}^{\infty} p((x-1);\ \lambda t)$$

Since this term of $p((x-1);\ \lambda t)$ is a probability distribution, all the probabilities should sum up to 1, as the "Normalization" axiom states:

$$E[X] = \mu_X = \lambda t * 1 = \lambda t$$

So, the mean of the Poisson distribution is λt.

4. a)

To prove the given binomial theorem:

$$(1+x)^N = \sum_{m=0}^{N} \binom{N}{m} * x^m \qquad (1)$$

we will use mathematical induction method, starting from $N = 1$:

$$(1+x)^1 = \sum_{m=0}^{1} \binom{1}{m} * x^m = \binom{1}{0} * x^0 + \binom{1}{1} * x^1 = 1 + x$$

So, we can clearly see that for the case of $N = 1$, this equality holds. Assuming the given equality is true, we can search for the case of $N \longleftarrow N + 1$:

$$(1+x)^{N+1} = \sum_{m=0}^{N+1} \binom{N+1}{m} * x^m$$

We separate the exponential of the LHS:

$$(1+x)^N * (1+x) = \sum_{m=0}^{N+1} \binom{N+1}{m} * x^m \qquad (2)$$

We substitute (1) in (2):

$$(1+x) * \sum_{m=0}^{N} \binom{N}{m} * x^m = \sum_{m=0}^{N+1} \binom{N+1}{m} * x^m$$

We expand the summation of both sides:

$$(1+x) * \left[\binom{N}{0} * x^0 + \binom{N}{1} * x^1 + \binom{N}{2} * x^2 + \cdots \binom{N}{N} * x^N\right] = \left[\binom{N+1}{0} * x^0 + \binom{N+1}{1} * x^1 + \binom{N+1}{2} * x^2 + \cdots \binom{N+1}{N+1} * x^{N+1}\right]$$

Using the distributive property of * operation:

$$\left[\binom{N}{0} * x^0 + \binom{N}{1} * x^1 + \binom{N}{2} * x^2 + \cdots \binom{N}{N} * x^N + \binom{N}{0} * x^1 + \binom{N}{1} * x^2 + \binom{N}{2} * x^3 + \cdots \binom{N}{N} * x^{N+1}\right] = \left[\binom{N+1}{0} * x^0 + \binom{N+1}{1} * x^1 + \binom{N+1}{2} * x^2 + \cdots \binom{N+1}{N+1} * x^{N+1}\right]$$

Using the distributive property of * operation again, we can combine the terms with the same exponential of x variable for LHS:

$$\left[\binom{N}{0} * x^0 + \left(\binom{N}{1} + \binom{N}{0}\right) * x^1 + \left(\binom{N}{2} + \binom{N}{1}\right) * x^2 + \cdots \left(\binom{N}{N} + \binom{N}{N-1}\right) * x^N + \binom{N}{N} * x^{N+1}\right] = \left[\binom{N+1}{0} * x^0 + \binom{N+1}{1} * x^1 + \binom{N+1}{2} * x^2 + \cdots \binom{N+1}{N+1} * x^{N+1}\right]$$

Using the Pascal’s identity for combinations:

$$\binom{N}{m} + \binom{N}{m-1} = \binom{N+1}{m}$$

We replace the terms with sum of two combination operations on the LHS:

$$\left[\binom{N}{0} * x^0 + \binom{N+1}{1} * x^1 + \binom{N+1}{2} * x^2 + \cdots \binom{N+1}{N} * x^N + \binom{N}{N} * x^{N+1}\right] = \left[\binom{N+1}{0} * x^0 + \binom{N+1}{1} * x^1 + \binom{N+1}{2} * x^2 + \cdots \binom{N+1}{N+1} * x^{N+1}\right]$$

$\binom{N}{0} * x^0 = 1$ and $\binom{N+1}{0} * x^0 = 1$ making $\binom{N}{0} * x^0 = \binom{N+1}{0} * x^0$

$\binom{N}{N} * x^{N+1} = x^{N+1}$ and $\binom{N+1}{N+1} * x^{N+1} = x^{N+1}$ making $\binom{N}{N} * x^{N+1} = \binom{N+1}{N+1} * x^{N+1}$

Also, all the other elements of the sum for both sides are identical. Thus, we proved that this equation also holds for the term $N+1$.

Proving the equation for $(N+1)$th term and the first term concludes binomial theorem is true in the light of the induction method.

The binomial distribution is given:

$$\sum_{m=0}^{N} \binom{N}{m} * p^m * (1-p)^{(N-m)} = 1$$

We will separate the exponential of the term $(1-p)^{(N-m)}$ as follows:

$$\sum_{m=0}^{N} \binom{N}{m} * p^m * \frac{(1-p)^N}{(1-p)^m} = 1$$

Because the term $(1-p)^N$ is not dependent on m we can take it to the outside of the summation:

$$(1-p)^N \sum_{m=0}^{N} \binom{N}{m} * \frac{p^m}{(1-p)^m} = 1$$

Because the exponential of the terms p^m and $(1-p)^m$ are the same, we can combine them into a one exponential term:

$$(1-p)^N \sum_{m=0}^{N} \binom{N}{m} * \left(\frac{p}{1-p}\right)^m = 1 \qquad (3)$$

From the binomial theorem proved above, we know that:

$$(1+x)^N = \sum_{m=0}^{N} \binom{N}{m} * x^m$$

Replacing x with $\frac{p}{1-p}$ **we get:**

$$\left(1+\left(\frac{\boldsymbol{p}}{\boldsymbol{1-p}}\right)\right)^{N}=\sum_{m=0}^{N}\binom{N}{m}*\left(\frac{\boldsymbol{p}}{\boldsymbol{1-p}}\right)^{m} \qquad (4)$$

Substituting (4) in (3):

$$(\boldsymbol{1-p})^{\boldsymbol{N}}*\left(1+\left(\frac{\boldsymbol{p}}{\boldsymbol{1-p}}\right)\right)^{N}=1$$

For the term $\left(1+\left(\frac{\boldsymbol{p}}{\boldsymbol{1-p}}\right)\right)^{N}$ we combine the whole term under the same divisor, namely $\boldsymbol{1-p}$**:**

$$(\boldsymbol{1-p})^{\boldsymbol{N}}*\left(\frac{1-p}{1-p}+\frac{\boldsymbol{p}}{\boldsymbol{1-p}}\right)^{N}=1$$

$$(\boldsymbol{1-p})^{\boldsymbol{N}}*\left(\frac{1-p+p}{1-p}\right)^{N}=1$$

$$(\boldsymbol{1-p})^{\boldsymbol{N}}*\left(\frac{1}{1-p}\right)^{N}=1$$

On the LHS, both terms have the same exponential. Thus, we can combine them under the same exponential, which is N:

$$\left(\frac{1-p}{1-p}\right)^{N}=1$$

$1-p=1-p$, making the equation:

$$1^{N}=1$$

For every exponential of 1, the result is 1. So, the equation holds. Thus, we can say that the binomial distribution is normalized.

b)

For a uniform random variable $x \sim U(0,1)$ the probability density function is:

$$p(x)=\frac{1}{U-L} \qquad (1)$$

where U is the upper bound and L is the lower bound. In this case we know the values of mean and variance of the uniform random variable:

$$\mu = 0 \quad and \quad \sigma^2 = 1$$

We also know that for a uniform random variable, the equation for mean and variance depending on the upper and lower bound as follows:

$$\mu = \frac{U + L}{2}$$

$$\sigma^2 = \frac{(U - L)^2}{12}$$

whereas L is the lower bound and U is the upper bound. To find the values of lower and upper bounds, we replace mean and variance in above equations with their values and then do the calculations:

$$0 = \frac{U + L}{2} \qquad 1 = \frac{(U - L)^2}{12}$$

$$0 = U + L \qquad 12 = (U - L)^2$$

$$0 = U + L \qquad \sqrt{12} = U - L$$

$$0 = U + L \qquad 2\sqrt{3} = U - L \qquad (2)$$

Then we add two equations in (2):

$$2\sqrt{3} = 2U \qquad |\, divide\ both\ sides\ with\ 2$$

$$U = \sqrt{3}$$

From the equation $0 = U + L$ we find that $L = -\sqrt{3}$. Now we know the upper and lower bounds. We replace them in the equation (1) to find the probability distribution function of the random variable:

$$p(x) = \frac{1}{\sqrt{3} - (-\sqrt{3})} = \frac{1}{\sqrt{3} + \sqrt{3}} = \frac{1}{2\sqrt{3}} = \frac{\sqrt{3}}{6}$$

To calculate the entropy of the uniform random variable, we will use the given formula:

$$H(x) = -\int_x p(x) * lnp(x) * dx$$

Replacing $p(x)$ with $\frac{\sqrt{3}}{6}$, according to the above calculation of probability density function of uniform random variable:

$$H(x) = -\int_x \frac{\sqrt{3}}{6} * ln\frac{\sqrt{3}}{6} * dx$$

Since $ln1$ is a constant variable, we can take it to the outside of the integral:

$$H(x) = -\frac{\sqrt{3}}{6} * ln\frac{\sqrt{3}}{6}\int_x 1dx$$

Since the lower bound is $-\sqrt{3}$, and the upper bound is $\sqrt{3}$, we can calculate the integral as follows:

$$H(x) = -\frac{\sqrt{3}}{6} * ln\frac{\sqrt{3}}{6}\int_{-\sqrt{3}}^{\sqrt{3}} 1dx = -\frac{\sqrt{3}}{6} * ln\frac{\sqrt{3}}{6} * \left(x|_{-\sqrt{3}}^{\sqrt{3}}\right)$$
$$= -\frac{\sqrt{3}}{6} * ln\frac{\sqrt{3}}{6} * \left(\sqrt{3} - (-\sqrt{3})\right) = -\frac{\sqrt{3}}{6} * ln\frac{\sqrt{3}}{6} * 2\sqrt{3} = -ln\frac{\sqrt{3}}{6}$$
$$= 1.242453325$$

So, we can say that the entropy of uniform random variable $x \sim U(0, 1)$ is 1.242453325.

For a Gaussian random variable $z \sim \mathcal{N}(0, 1)$ the probability density function is:

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} * e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

To calculate the entropy of the Gaussian random variable, we will use the given formula (slightly changing it by putting the -1 to the inside of the integral) by replacing $p(x)$ with the above equation:

$$H(x) = \int_x p(x) * -\ln\left(\frac{1}{\sqrt{2\pi\sigma^2}} * e^{-\frac{(x-\mu)^2}{2\sigma^2}}\right) * dx$$

Calculating the logarithm:

$$H(x) = \int_x p(x) * \left(\frac{1}{2}\ln(2\pi\sigma^2) + \frac{1}{2\sigma^2} * (x-\mu)^2\right) dx$$

We distribute $p(x)$ to the next term:

$$H(x) = \int_x p(x) * \left(\frac{1}{2}\ln(2\pi\sigma^2)\right) + p(x) * \left(\frac{1}{2\sigma^2} * (x-\mu)^2\right) dx$$

After that, we can distribute the integral over the summation:

$$H(x) = \int_x p(x) * \left(\frac{1}{2}\ln(2\pi\sigma^2)\right) dx + \int_x p(x) * \left(\frac{1}{2\sigma^2} * (x-\mu)^2\right) dx$$

Since the terms $\frac{1}{2}\ln(2\pi\sigma^2)$ and $\frac{1}{2\sigma^2}$ are not dependent on the variable x we can put them outside of the integrals:

$$H(x) = \frac{1}{2}\ln(2\pi\sigma^2) \int_x p(x)\, dx + \frac{1}{2\sigma^2} \int_x p(x) * (x-\mu)^2 dx$$

From the "Normalization" axiom, we know that the integral $\int_x p(x)\, dx = 1$. Also, from the definition of variance σ^2, we know that $\int_x p(x) * (x-\mu)^2 dx = \sigma^2$. Replacing them accordingly:

$$H(x) = \frac{1}{2}\ln(2\pi\sigma^2) * 1 + \frac{1}{2\sigma^2} * \sigma^2$$

$$H(x) = \frac{1}{2}\ln(2\pi\sigma^2) + \frac{1}{2} = \frac{1}{2} * (\ln(2\pi\sigma^2) + 1) \qquad (1)$$

Here the given parameters of distribution indicate mean and variance respectively. In this case the mean of the Gaussian distribution is 0, and the variance of the Gaussian distribution is 1.

$$\mu = 0$$

$$\sigma^2 = 1$$

Replacing the value of variance in the equation (1), we get:

$$H(x) = \frac{1}{2} * (\ln(2\pi 1) + 1) = \frac{1}{2} * (\ln(2\pi) + 1) = 1.418938533$$

So, we can say that the entropy of Gaussian random variable $z \sim \mathcal{N}(0,1)$ is 1.418938533.

Comparing the entropies of both uniform normal variable and Gaussian normal variable, we can clearly say that the Gaussian normal variable has a greater entropy than the uniform normal variable by:

$$H(x)_U = 1.242453325$$

$$H(x)_{\mathcal{N}} = 1.418938533$$

$$\Rightarrow H(x)_{\mathcal{N}} > H(x)_U$$

c) i.

The KL-divergence for a given distribution can be calculated as follows:

$$KL(p||q) = \int_{-\infty}^{\infty} p(x) * \ln\left(\frac{p(x)}{q(x)}\right) * dx$$

To calculate the KL-divergence between equal distributions we can change the distribution $q(x)$ according to $q(x) = p(x)$:

$$KL(p||q) = \int_{-\infty}^{\infty} p(x) * \ln\left(\frac{p(x)}{p(x)}\right) * dx$$

Since $\frac{p(x)}{p(x)} = 1$ the logarithmic term becomes $ln1$, which is equal to zero:

$$KL(p||q) = \int_{-\infty}^{\infty} p(x) * 0 * dx = 0\int_{-\infty}^{\infty} p(x) * dx$$

So, whatever the integral of the distribution $p(x)$ results because it will be multiplied with 0 after the integral is resolved, the KL-divergence between equal distributions will yield 0.

$$KL(p||p) = 0$$

ii.

KL-divergence is not symmetric, that is $KL(p||q) \neq KL(q||p)$. This can be shown by a simple example as follows:

We first define two discrete probability distributions $P(x)$ and $Q(x)$:

$$P(x) = \begin{cases} 1 & , with\ probability\ \frac{1}{2} \\ -1 & , with\ probability\ \frac{1}{2} \end{cases}$$

$$Q(x) = \begin{cases} 1 & , with\ probability\ \frac{1}{3} \\ -1 & , with\ probability\ \frac{2}{3} \end{cases}$$

For discrete probability distributions the KL-divergence can be calculated as follows:

$$D_{KL}(P||Q) = \sum_{x \in X} P(x) * \ln\left(\frac{P(X)}{Q(X)}\right)$$

For $D_{KL}(P||Q)$ the calculation will follow:

$$D_{KL}(P||Q) = \frac{1}{2} * \ln\left(\frac{\frac{1}{2}}{\frac{1}{3}}\right) + \frac{1}{2} * \ln\left(\frac{\frac{1}{2}}{\frac{2}{3}}\right) = \frac{1}{2} * \ln\left(\frac{3}{2}\right) + \frac{1}{2} * \ln\left(\frac{3}{4}\right) = 0.05889151783$$

For $D_{KL}(Q||P)$, which is symmetric to the $D_{KL}(P||Q)$ the calculation will follow:

$$D_{KL}(Q||P) = \frac{1}{3} * \ln\left(\frac{\frac{1}{3}}{\frac{1}{2}}\right) + \frac{2}{3} * \ln\left(\frac{\frac{2}{3}}{\frac{1}{2}}\right) = \frac{1}{3} * \ln\left(\frac{2}{3}\right) + \frac{2}{3} * \ln\left(\frac{4}{3}\right) = 0.05663301227$$

The results show us that $D_{KL}(P||Q)$ is not equal to the $D_{KL}(Q||P)$. Thus, the KL-divergence is not symmetric.

$$D_{KL}(P||Q) = 0.05889151783$$

$$D_{KL}(Q||P) = 0.05663301227$$

$$\Longrightarrow\ D_{KL}(P||Q) \neq D_{KL}(Q||P)$$

iii.

To calculate the KL-divergence between $p(x)\sim\mathcal{N}(\mu_1, \sigma_1^2)$ and $q(x)\sim\mathcal{N}(\mu_2, \sigma_2^2)$, where $\mu_1 = 2, \mu_2 = 1.8, \sigma_1^2 = 1.5, \sigma_2^2 = 0.2$ we first start with the general formula for KL-divergence:

$$KL(p||q) = \int_{-\infty}^{\infty} p(x) * \ln\left(\frac{p(x)}{q(x)}\right) * dx$$

Then we separate the inside of the logarithmic term by using the quotient rule of logarithm:

$$KL(p||q) = \int_{-\infty}^{\infty} p(x) * \left[\ln\big(p(x)\big) - \ln\big(q(x)\big)\right] * dx$$
$$= \int_{-\infty}^{\infty} p(x) * \ln\big(p(x)\big) * dx - p(x) * \ln\big(q(x)\big) * dx$$
$$= \int_{-\infty}^{\infty} p(x) * \ln\big(p(x)\big) * dx - \int_{-\infty}^{\infty} p(x) * \ln\big(q(x)\big) * dx$$

We first resolve the first integral with replacing $p(x)$ with its probability distribution formula inside the logarithm:

$$p(x) = \frac{1}{\sqrt{2\pi\sigma_1^2}} * e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}$$

$$KL(p||q) = \int_{-\infty}^{\infty} p(x) * \ln\left(\frac{1}{\sqrt{2\pi\sigma_1^2}} * e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}\right) * dx - \int_{-\infty}^{\infty} p(x) * \ln\big(q(x)\big) * dx$$
$$= \int_{-\infty}^{\infty} p(x) * \ln\left((2\pi\sigma_1^2)^{-\frac{1}{2}} * e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}\right) * dx$$
$$- \int_{-\infty}^{\infty} p(x) * \ln\big(q(x)\big) * dx$$

We separate the terms inside the logarithm by product rule:

$$KL(p||q) = \int_{-\infty}^{\infty} p(x) * \ln\left((2\pi\sigma_1^2)^{-\frac{1}{2}} * e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}\right) * dx - \int_{-\infty}^{\infty} p(x) * \ln\big(q(x)\big) * dx$$
$$= \int_{-\infty}^{\infty} p(x) * [\ln\left((2\pi\sigma_1^2)^{-\frac{1}{2}}\right) + \ln\left(e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}\right)] * dx$$
$$- \int_{-\infty}^{\infty} p(x) * \ln\big(q(x)\big) * dx$$

We rearrange the first integral as follows:

$$KL(p||q) = \int_{-\infty}^{\infty} p(x) * \ln\left((2\pi\sigma_1^2)^{-\frac{1}{2}}\right) * dx + \int_{-\infty}^{\infty} p(x) * \ln\left(e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}\right) * dx$$
$$- \int_{-\infty}^{\infty} p(x) * \ln\big(q(x)\big) * dx$$

Using the properties of natural logarithm and natural exponential and the quotient rule of logarithm:

$$KL(p||q) = \int_{-\infty}^{\infty} p(x) * \left(-\frac{\ln(2\pi\sigma_1^2)}{2}\right) * dx + \int_{-\infty}^{\infty} p(x) * \left(-\frac{(x-\mu_1)^2}{2\sigma_1^2}\right) * dx$$
$$- \int_{-\infty}^{\infty} p(x) * \ln\big(q(x)\big) * dx$$

The term $\left(-\frac{\ln(2\pi\sigma_1^2)}{2}\right)$ is not dependent on x. So, we can put it outside of the integral and we know that $\int_{-\infty}^{\infty} p(x) * dx = 1$ from the "Normalization" axiom:

$$KL(p||q) = -\frac{\ln(2\pi\sigma_1^2)}{2} * \int_{-\infty}^{\infty} p(x) * dx + \int_{-\infty}^{\infty} p(x) * \left(-\frac{(x-\mu_1)^2}{2\sigma_1^2}\right) * dx$$
$$- \int_{-\infty}^{\infty} p(x) * \ln\big(q(x)\big) * dx$$
$$= -\frac{\ln(2\pi\sigma_1^2)}{2} * 1 + \int_{-\infty}^{\infty} p(x) * \left(-\frac{(x-\mu_1)^2}{2\sigma_1^2}\right) * dx$$
$$- \int_{-\infty}^{\infty} p(x) * \ln\big(q(x)\big) * dx$$
$$= -\frac{\ln(2\pi\sigma_1^2)}{2} + \int_{-\infty}^{\infty} p(x) * \left(-\frac{(x-\mu_1)^2}{2\sigma_1^2}\right) * dx$$
$$- \int_{-\infty}^{\infty} p(x) * \ln\big(q(x)\big) * dx$$

We know that $(x-\mu_1)^2 = x^2 - 2x\mu_1 + \mu_1^2$. So, we replace it in the above equation. Also, we distribute the $p(x)$ in the integral to the next term and put the -1 to outside of the integral:

$$KL(p||q) = -\frac{\ln(2\pi\sigma_1^2)}{2}$$
$$- \int_{-\infty}^{\infty} \left(\frac{p(x) * x^2 - p(x) * 2x\mu_1 + p(x) * \mu_1^2}{2\sigma_1^2}\right) * dx$$
$$- \int_{-\infty}^{\infty} p(x) * \ln\big(q(x)\big) * dx$$

If we put the non-x-depending term $\frac{1}{2\sigma_1^2}$ to the outside of the integral and distribute the integral over the addition and subtraction, we get:

$$KL(p||q) = -\frac{\ln(2\pi\sigma_1^2)}{2} - \frac{1}{2\sigma_1^2}$$
$$* \left(\int_{-\infty}^{\infty} p(x) * x^2 * dx - \int_{-\infty}^{\infty} p(x) * 2x\mu_1 * dx + \int_{-\infty}^{\infty} p(x) * \mu_1^2 * dx\right)$$
$$- \int_{-\infty}^{\infty} p(x) * \ln\big(q(x)\big) * dx$$

We can put the term μ_1^2 outside of the third integral and the term $2\mu_1$ outside of the second integral. Because they do not depend on x:

$$KL(p||q) = -\frac{\ln(2\pi\sigma_1^2)}{2} - \frac{1}{2\sigma_1^2}$$
$$* \left(\int_{-\infty}^{\infty} p(x) * x^2 * dx - 2\mu_1 \int_{-\infty}^{\infty} p(x) * x * dx + \mu_1^2 * \int_{-\infty}^{\infty} p(x) * dx\right)$$
$$- \int_{-\infty}^{\infty} p(x) * \ln\big(q(x)\big) * dx$$

We know that expectation of x is given as $E(x) = \int_{-\infty}^{\infty} x * p(x) * dx$. So, we can say that:

$$E(x^2) = \int_{-\infty}^{\infty} x^2 * p(x) * dx$$

$$E(x) = \int_{-\infty}^{\infty} x * p(x) * dx$$

We know that the variance $Var(x) = E(x^2) - (E(x))^2$. If we rearrange this equation, we get: $E(x^2) = Var(x) + (E(x))^2$. We also know that $E(x) = \mu$ and $Var(x) = \sigma^2$. Replacing these in the previous equation and renaming them as μ_1 and σ_1, we get:

$$\begin{aligned}
KL(p||q) &= -\frac{\ln(2\pi\sigma_1^2)}{2} - \frac{1}{2\sigma_1^2} * (E(x^2) - 2\mu_1 * E(x) + \mu_1^2 * 1) \\
&- \int_{-\infty}^{\infty} p(x) * \ln(q(x)) * dx = \\
&= -\frac{\ln(2\pi\sigma_1^2)}{2} - \frac{1}{2\sigma_1^2} * (\sigma_1^2 + \mu_1^2 - 2\mu_1 * \mu_1 + \mu_1^2) \\
&- \int_{-\infty}^{\infty} p(x) * \ln(q(x)) * dx = \\
&= -\frac{\ln(2\pi\sigma_1^2)}{2} - \frac{1}{2\sigma_1^2} * (\sigma_1^2 + 2\mu_1^2 - 2\mu_1^2) - \int_{-\infty}^{\infty} p(x) * \ln(q(x)) * dx \\
&= -\frac{\ln(2\pi\sigma_1^2)}{2} - \frac{1}{2\sigma_1^2} * \sigma_1^2 - \int_{-\infty}^{\infty} p(x) * \ln(q(x)) * dx = -\frac{\ln(2\pi\sigma_1^2)}{2} \\
&- \frac{1}{2} \\
&- \int_{-\infty}^{\infty} p(x) * \ln(q(x)) * dx \\
&= -\frac{1}{2} * (\ln(2\pi\sigma_1^2) + 1) - \int_{-\infty}^{\infty} p(x) * \ln(q(x)) * dx
\end{aligned}$$

The probability distribution of the normal random variable $q(x) \sim \mathcal{N}(\mu_2, \sigma_2^2)$ is as follows:

$$q(x) = \frac{1}{\sqrt{2\pi\sigma_2^2}} * e^{-\frac{(x-\mu_2)^2}{2\sigma_2^2}}$$

We can replace the $q(x)$ with its equation in the KL-divergence equation:

$$KL(p||q) = -\frac{1}{2} * (\ln(2\pi\sigma_1^2) + 1) - \int_{-\infty}^{\infty} p(x) * \ln\left(\frac{1}{\sqrt{2\pi\sigma_2^2}} * e^{-\frac{(x-\mu_2)^2}{2\sigma_2^2}}\right) * dx$$

Then, we separate the inside of the ln term according to the quotient rule of logarithm and rearrange the equation:

$$KL(p||q) = -\frac{1}{2} * (\ln(2\pi\sigma_1^2) + 1)$$
$$- \int_{-\infty}^{\infty} p(x) * \left[\ln\left(e^{-\frac{(x-\mu_2)^2}{2\sigma_2^2}}\right) - \ln\left(\frac{1}{\sqrt{2\pi\sigma_2^2}}\right)\right] * dx$$
$$= -\frac{1}{2} * (\ln(2\pi\sigma_1^2) + 1)$$
$$- \int_{-\infty}^{\infty} p(x) * \ln\left(e^{-\frac{(x-\mu_2)^2}{2\sigma_2^2}}\right) * dx - \int_{-\infty}^{\infty} p(x) * \ln\left(\frac{1}{\sqrt{2\pi\sigma_2^2}}\right) * dx$$

Since the ln term in the first integral contains a natural (e) exponential, we can use the property of ln being the inverse function of natural exponential. Also, for the second integral the ln term is not dependent on x. So, we can treat it like a constant and put it outside of the integral. We rearrange the equation accordingly:

$$KL(p||q) = -\frac{1}{2} * (\ln(2\pi\sigma_1^2) + 1)$$
$$- \int_{-\infty}^{\infty} p(x) * \ln\left(e^{-\frac{(x-\mu_2)^2}{2\sigma_2^2}}\right) * dx - \int_{-\infty}^{\infty} p(x) * \ln\left(\frac{1}{\sqrt{2\pi\sigma_2^2}}\right) * dx$$
$$= -\frac{1}{2} * (\ln(2\pi\sigma_1^2) + 1) - \int_{-\infty}^{\infty} p(x) * \left(-\frac{(x-\mu_2)^2}{2\sigma_2^2}\right) * dx$$
$$- \ln\left(\frac{1}{\sqrt{2\pi\sigma_2^2}}\right) * \int_{-\infty}^{\infty} p(x) * dx = -\frac{1}{2} * (\ln(2\pi\sigma_1^2) + 1)$$
$$- \int_{-\infty}^{\infty} p(x) * \left(-\frac{(x-\mu_2)^2}{2\sigma_2^2}\right) * dx - \ln\left((2\pi\sigma_2^2)^{-\frac{1}{2}}\right) * \int_{-\infty}^{\infty} p(x) * dx$$

From the "Normalization" axiom, we know that $\int_{-\infty}^{\infty} p(x) * dx = 1$. Using this property, the power rule of the logarithm, and putting the -1 in the first integral to the outside of it we rearrange the equation as follows:

$$KL(p||q) = -\frac{1}{2} * (\ln(2\pi\sigma_1^2) + 1) - \int_{-\infty}^{\infty} p(x) * \left(-\frac{(x-\mu_2)^2}{2\sigma_2^2}\right) * dx$$
$$+ \ln\left((2\pi\sigma_2^2)^{-\frac{1}{2}}\right) * \int_{-\infty}^{\infty} p(x) * dx$$
$$= -\frac{1}{2} * (\ln(2\pi\sigma_1^2) + 1) + \int_{-\infty}^{\infty} p(x) * \left(\frac{(x-\mu_2)^2}{2\sigma_2^2}\right) * dx + \frac{1}{2}$$
$$* \ln(2\pi\sigma_2^2) * 1$$
$$= -\frac{1}{2} * (\ln(2\pi\sigma_1^2) + 1) + \int_{-\infty}^{\infty} p(x) * \left(\frac{(x-\mu_2)^2}{2\sigma_2^2}\right) * dx + \frac{1}{2}$$
$$* \ln(2\pi\sigma_2^2)$$

We know that $(x-\mu_2)^2 = x^2 - 2x\mu_2 + \mu_2^2$. So, we replace it in the above equation. Also, we distribute the $p(x)$ in the integral to the next term:

$$KL(p||q) = -\frac{1}{2} * (\ln(2\pi\sigma_1^2) + 1) + \int_{-\infty}^{\infty} p(x) * \left(\frac{(x-\mu_2)^2}{2\sigma_2^2}\right) * dx + \frac{1}{2} * \ln(2\pi\sigma_2^2)$$
$$= -\frac{1}{2} * (\ln(2\pi\sigma_1^2) + 1) + \int_{-\infty}^{\infty} p(x) * \left(\frac{x^2 - 2x\mu_2 + \mu_2^2}{2\sigma_2^2}\right) * dx + \frac{1}{2} * \ln(2\pi\sigma_2^2)$$
$$= -\frac{1}{2} * (\ln(2\pi\sigma_1^2) + 1) + \int_{-\infty}^{\infty} \frac{p(x) * x^2 * dx - p(x) * 2x\mu_2 * dx + p(x) * \mu_2^2 * dx}{2\sigma_2^2} + \frac{1}{2} * \ln(2\pi\sigma_2^2)$$

Since $2\sigma_2^2$ is not dependent on x we can treat it as a constant and put it outside of the integral. Also, we can distribute the integral over the summation and subtractions:

$$KL(p||q) = -\frac{1}{2} * (\ln(2\pi\sigma_1^2) + 1) + \frac{1}{2\sigma_2^2}\left(\int_{-\infty}^{\infty} p(x) * x^2 * dx - \int_{-\infty}^{\infty} p(x) * 2x\mu_2 * dx + \int_{-\infty}^{\infty} p(x) * \mu_2^2 * dx\right) + \frac{1}{2} * \ln(2\pi\sigma_2^2)$$

We know that expectation of x is given as $E(x) = \int_{-\infty}^{\infty} x * p(x) * dx$. So, we can say that:

$$E(x^2) = \int_{-\infty}^{\infty} x^2 * p(x) * dx$$

$$E(x) = \int_{-\infty}^{\infty} x * p(x) * dx$$

Also, we can put the term $2\mu_2$ to the outside of the second integral and term μ_2^2 to the outside of the third integral. Both because of the terms not depending on x. Putting all together, we rearrange the equation as given:

$$KL(p||q) = -\frac{1}{2} * (\ln(2\pi\sigma_1^2) + 1) + \frac{1}{2\sigma_2^2}(E(x^2) - 2\mu_2 * E(x) + \mu_2^2) + \frac{1}{2} * \ln(2\pi\sigma_2^2)$$

We know that the variance $Var(x) = E(x^2) - (E(x))^2$. If we rearrange this equation, we get: $E(x^2) = Var(x) + (E(x))^2$. We also know that $E(x) = \mu$ and $Var(x) = \sigma^2$. Replacing these in the previous equation and renaming them as μ_1 and σ_1, we get:

$$KL(p||q) = -\frac{1}{2} * (\ln(2\pi\sigma_1^2) + 1) + \frac{1}{2\sigma_2^2} * (\sigma_1^2 + \mu_1^2 - 2\mu_2\mu_1 + \mu_2^2) + \frac{1}{2} * \ln(2\pi\sigma_2^2)$$
$$= -\frac{1}{2} * (\ln(2\pi\sigma_1^2) + 1) + \frac{1}{2\sigma_2^2} * (\sigma_1^2 + (\mu_1 - \mu_2)^2) + \frac{1}{2}$$
$$* \ln(2\pi\sigma_2^2) = -\frac{1}{2} * (\ln(2\pi\sigma_1^2) + 1 - \ln(2\pi\sigma_2^2)) + \frac{1}{2\sigma_2^2}$$
$$* (\sigma_1^2 + \mu_1^2 - 2\mu_2\mu_1 + \mu_2^2)$$

Using the quotient rule of logarithm:

$$KL(p||q) = -\frac{1}{2} * (\ln(2\pi\sigma_1^2) - \ln(2\pi\sigma_2^2) + 1) + \frac{1}{2\sigma_2^2} * (\sigma_1^2 + \mu_1^2 - 2\mu_2\mu_1 + \mu_2^2)$$
$$= -\frac{1}{2} * \left(\ln\left(\frac{2\pi\sigma_1^2}{2\pi\sigma_2^2}\right) + 1\right) + \frac{1}{2\sigma_2^2} * (\sigma_1^2 + \mu_1^2 - 2\mu_2\mu_1 + \mu_2^2)$$

$$KL(p||q) = -\frac{1}{2} * \left(\ln\left(\frac{\sigma_1^2}{\sigma_2^2}\right) + 1\right) + \frac{1}{2\sigma_2^2} * (\sigma_1^2 + \mu_1^2 - 2\mu_2\mu_1 + \mu_2^2)$$
$$= -\frac{1}{2} * \left(2 * \ln\left(\frac{\sigma_1}{\sigma_2}\right) + 1\right) + \frac{1}{2\sigma_2^2} * (\sigma_1^2 + \mu_1^2 - 2\mu_2\mu_1 + \mu_2^2)$$
$$= -\ln\left(\frac{\sigma_1}{\sigma_2}\right) - \frac{1}{2} + \frac{1}{2\sigma_2^2} * (\sigma_1^2 + \mu_1^2 - 2\mu_2\mu_1 + \mu_2^2)$$
$$= \ln\left(\frac{\sigma_2}{\sigma_1}\right) - \frac{1}{2} + \frac{1}{2\sigma_2^2} * (\sigma_1^2 + \mu_1^2 - 2\mu_2\mu_1 + \mu_2^2)$$
$$= \ln\left(\frac{\sigma_2}{\sigma_1}\right) - \frac{1}{2} + \frac{1}{2\sigma_2^2} * (\sigma_1^2 + (\mu_1 - \mu_2)^2)$$

So, as a result the formula for KL-divergence depending on $\mu_1,\ \mu_2,\ \sigma_1,\ \sigma_2$ is as follows:

$$KL(p||q) = \ln\left(\frac{\sigma_2}{\sigma_1}\right) - \frac{1}{2} + \frac{1}{2\sigma_2^2} * (\sigma_1^2 + (\mu_1 - \mu_2)^2)$$

We know that $\mu_1 = 2, \mu_2 = 1.8, \sigma_1^2 = 1.5, \sigma_2^2 = 0.2$. So, we simply put all these values to the according places in the above equation and calculate:

$$KL(p||q) = \ln\left(\frac{\sqrt{0.2}}{\sqrt{1.5}}\right) - \frac{1}{2} + \frac{1}{2 * 0.2} * (1.5 + (2 - 1.8)^2) = 2.34254849$$

As a result, we can say that the KL-divergence between $p(x) \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $q(x) \sim \mathcal{N}(\mu_2, \sigma_2^2)$, where $\mu_1 = 2, \mu_2 = 1.8, \sigma_1^2 = 1.5, \sigma_2^2 = 0.2$ is 2.34254849.

5. a)

To find the probability density function of $Z = X + Y$, where X and Y are independent standard Gaussian random variables we can use the convolution operation with X and Y:

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(\tau) * f_Y(z-\tau) d\tau$$

A standard Gaussian random variable has the following probability distribution:

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} * e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

And when the distribution of a Gaussian random variable is called standard, then it means that its mean is 0 and variance is 1. Putting these values to the above equation, we get:

$$p(x) = \frac{1}{\sqrt{2\pi 1}} * e^{-\frac{(x-0)^2}{2*1}} = \frac{1}{\sqrt{2\pi}} * e^{-\frac{x^2}{2}}$$

Replacing this probability distribution function in the convolution equation with parameters z and $(z-\tau)$:

$$f_Z(z) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} * e^{-\frac{\tau^2}{2}} * \frac{1}{\sqrt{2\pi}} * e^{-\frac{(z-\tau)^2}{2}} * d\tau = \int_{-\infty}^{\infty} \frac{1}{2\pi} * e^{-\frac{\tau^2}{2}} * e^{-\frac{(z-\tau)^2}{2}} * d\tau$$

Since the term $\frac{1}{2\pi}$ is not dependent on τ, we can put it outside of the integral:

$$f_Z(z) = \frac{1}{2\pi} * \int_{-\infty}^{\infty} e^{-\frac{\tau^2}{2}} * e^{-\frac{(z-\tau)^2}{2}} * d\tau$$

We can combine the natural exponential terms inside the integral under one exponential:

$$\begin{aligned} f_Z(z) &= \frac{1}{2\pi} * \int_{-\infty}^{\infty} e^{-\frac{\tau^2}{2}} * e^{-\frac{(z-\tau)^2}{2}} * d\tau = \frac{1}{2\pi} * \int_{-\infty}^{\infty} e^{\left(-\frac{\tau^2}{2}-\frac{(z-\tau)^2}{2}\right)} * d\tau \\ &= \frac{1}{2\pi} * \int_{-\infty}^{\infty} e^{\left(-\frac{\tau^2}{2}-\frac{(z-\tau)^2}{2}\right)} * d\tau = \frac{1}{2\pi} * \int_{-\infty}^{\infty} e^{\left(-\tau^2-\frac{z^2}{2}+z\tau\right)} * d\tau \\ &= \frac{1}{2\pi} * \int_{-\infty}^{\infty} e^{-\tau^2} * e^{-\frac{z^2}{2}} * e^{z\tau} * d\tau \end{aligned}$$

Since the term $e^{-\frac{z^2}{2}}$ is also independent of τ we can put it outside of the integral:

$$f_Z(z) = \frac{1}{2\pi} * e^{-\frac{z^2}{2}} * \int_{-\infty}^{\infty} e^{-\tau^2} * e^{z\tau} * d\tau = \frac{1}{2\pi} * e^{-\frac{z^2}{2}} * \int_{-\infty}^{\infty} e^{-\tau^2+z\tau} * d\tau$$

We know that $-\tau^2 + z\tau = \frac{z^2}{4} - (\tau - \frac{z}{2})^2$, since $-\left(\tau - \frac{z}{2}\right)^2 = -\tau^2 + z\tau - \frac{z^2}{4}$. So, we change the exponential of e accordingly:

$$f_Z(z) = \frac{1}{2\pi} * e^{-\frac{z^2}{2}} * \int_{-\infty}^{\infty} e^{\frac{z^2}{4}-\left(\tau-\frac{z}{2}\right)^2} * d\tau = \frac{1}{2\pi} * e^{-\frac{z^2}{2}} * \int_{-\infty}^{\infty} e^{\frac{z^2}{4}} * e^{-\left(\tau-\frac{z}{2}\right)^2} * d\tau$$

Since the term $e^{\frac{z^2}{4}}$ is also not dependent on τ we can put it outside of the integral:

$$f_Z(z) = \frac{1}{2\pi} * e^{-\frac{z^2}{2}} * e^{\frac{z^2}{4}} * \int_{-\infty}^{\infty} e^{-\left(\tau-\frac{z}{2}\right)^2} * d\tau = \frac{1}{2\pi} * e^{-\frac{z^2}{4}} * \int_{-\infty}^{\infty} e^{-\left(\tau-\frac{z}{2}\right)^2} * d\tau$$

If we substitute u in the exponential of e, with $u = \tau - \frac{z}{2}$, we will have:

$$f_Z(z) = \frac{1}{2\pi} * e^{-\frac{z^2}{4}} * \int_{-\infty}^{\infty} e^{-u^2} * du$$

We know that $\int_{-\infty}^{\infty} e^{-x^2}\, dx = \sqrt{\pi}$. So, we get:

$$f_Z(z) = \frac{1}{2\pi} * e^{-\frac{z^2}{4}} * \sqrt{\pi} = \frac{1}{2\sqrt{\pi}} * e^{-\frac{z^2}{4}}$$

Now we can reflect to our result. Probability distribution of a Gaussian random variable is:

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} * e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad (1)$$

And the probability distribution of the convolution of two standard Gaussian random variables is:

$$f_Z(z) = \frac{1}{2\sqrt{\pi}} * e^{-\frac{z^2}{4}} \qquad (2)$$

So, we can clearly see that the mean of the convolution of two standard Gaussian random variables is $\mu_Z = 0$ and the variance of the convolution of two standard Gaussian random variables is $\sigma_Z^2 = 2$. Putting these values in the equation (1) will yield us the equation (2).

Thus, it can be concluded that Z belongs to Gaussian distribution.

b)

We define three random variables $X, Y, and\ Z$. X is a standard Gaussian variable, Y is a discrete random variable, while $Z = XY$. The probability distribution for $X\ and\ Y$ would be:

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} * e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$Y = \begin{cases} 1 & , with\ probability\ \frac{1}{2} \\ -1 & , with\ probability\ \frac{1}{2} \end{cases}$$

We can show the probability function of Z as a cumulative probability function from $-\infty$ to z:

$$P(Z < z)$$

Knowing that $Z = XY$ this function can be expressed as follows:

$$P(XY < z)$$

We can express the cumulative probability function as the summation of each value that the discrete random variable Y can have:

$$P(XY < z) = \sum_i P(XY < z \mid Y = y_i) * P(Y = y_i)$$

As Y can be either 1 or -1 with the equal probability, we can rearrange the sum as follows:

$$P(XY < z) = P(XY < z \mid Y = 1) * P(Y = 1) + P(XY < z \mid Y = -1) * P(Y = -1)$$

As we know the Y values, we can substitute them as follows:

$$\begin{aligned} P(XY < z) &= P(X * 1 < z) * P(Y = 1) + P(X * (-1) < z) * P(Y = -1) \\ &= P(X < z) * P(Y = 1) + P(-X < z) * P(Y = -1) \end{aligned}$$

Since $P(Y = 1) = \frac{1}{2}$ and $P(Y = -1) = \frac{1}{2}$:

$$P(XY < z) = P(X < z) * \frac{1}{2} + P(-X < z) * \frac{1}{2}$$

$$\Rightarrow \qquad P(XY < z) = \frac{1}{2} * \big(P(X < z) + P(-X < z)\big)$$

For the term $P(-X < z)$ we can write $P(X > -z)$ as multiplying both sides in the inequality with -1.

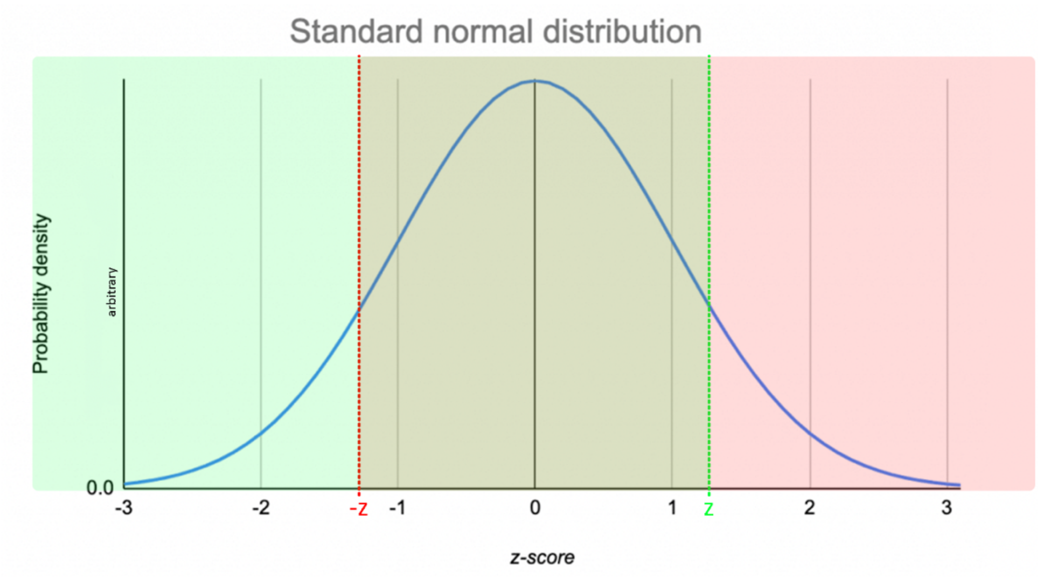
$$P(XY < z) = \frac{1}{2} * \big(P(X < z) + P(X > -z)\big) \quad (1)$$

Since X is a standard Gaussian random variable, we can express the cumulative distribution as the upper and lower bounds of the probability distribution function of X:

$$P(X < z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} * e^{-\frac{x^2}{2}} * dx$$

$$P(X > -z) = \int_{-z}^{\infty} \frac{1}{\sqrt{2\pi}} * e^{-\frac{x^2}{2}} * dx$$

This random variable X is a standard Gaussian random variable, meaning its probability distribution curve is symmetric about a vertical axis through the mean, which is $\mu = 0$. Since z and $-z$ has the same distance to the vertical symmetry axis at 0, we can say that the shapes of the curve from $-\infty$ to z and from $-z$ to ∞ are the same. Meaning the area under these upper and lower bounds are exactly same. If we try to show it via graphic, it would be like this:



The green colored area shows the from $-\infty$ to z part of the graph, while the red colored area shows from $-z$ to ∞ part. As it can be seen, because the distribution graph is symmetric to y-axis about $\mu = 0$, the areas under the curve in these borders are the same:

$$P(X < z) = P(X > -z) \qquad (2)$$

Substituting the equation (2) in (1), we get:

$$P(XY < z) = \frac{1}{2} * \big(P(X < z) + P(X < z)\big) = \frac{1}{2} * \big(2 * P(X < z)\big) = \frac{2}{2} * P(X < z)$$
$$= 1 * P(X < z) = P(X < z)$$

This calculation yields the result that the random variable $Z = XY$ is independent of Y and has the probability distribution as the standard Gaussian distribution.

c)

The random variable Y has the probability distribution function as follows:

$$Y = \begin{cases} k & , if\ X \geq k \\ 0 & , otherwise \end{cases}$$

We will have the condition when $X \geq k$:

$$\frac{X}{k} \geq 1 \implies X \geq k$$

When $X \geq k$, we know that $Y = k$. So, we can say that:

$$X \geq Y \qquad for\ the\ case\ of\ X \geq k$$

For the case of $X < k$, we know that X is non-negative and k is positive. So, $\frac{X}{k}$ can be at least 0:

$$\frac{X}{k} \geq 0 \implies X \geq 0$$

When $X < k$, we know that $Y = 0$. So, we can say that:

$$X \geq Y \qquad for\ the\ case\ of\ X < k$$

So, in either case we will have the same inequality:

$$X \geq Y$$

If a random variable's value is greater than or equal to another random variable, then we can say that the expected value of that random variable is also greater than the expected value of the other random variable. Applying this property of expected values, we can write the above equation as follows:

$$E[X] \geq E[Y] \qquad (1)$$

The expected value of the random variable Y can be calculated as the below summation:

$$E[Y] = \sum_{y} y * p(y) = 0 * p(0) + k * p(k) = k * p(k)$$

Probability of $p(k)$ can be expressed as the cumulative distribution function $P(X \geq k)$. Because as long as $X \geq k$, Y will be k:

$$E[Y] = k * P(X \geq k) \qquad (2)$$

Substituting (2) in the inequality (1), we get:

$$E[X] \geq k * P(X \geq k)$$

If we divide both sides with k, we get:

$$\frac{E[X]}{k} \geq \frac{k}{k} * P(X \geq k)$$

$$\frac{E[X]}{k} \geq 1 * P(X \geq k)$$

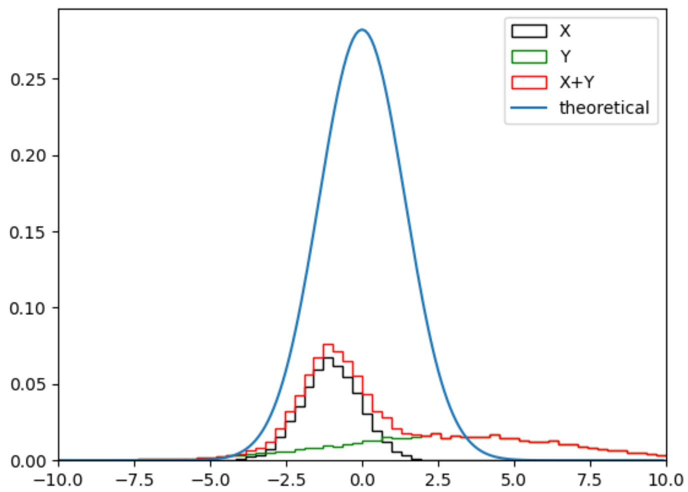
$$\frac{E[X]}{k} \geq P(X \geq k)$$

We can rearrange the inequality with smaller than or equal to sign:

$$P(X \geq k) \leq \frac{E[X]}{k}$$

6. a)

We have sampled two random variables X with $\mu_X = -1$ and $\sigma_X^2 = 1$ and Y with $\mu_Y = 3$ and $\sigma_Y^2 = 4$ over 100000 pairs. Then, we plot the resulting histograms in the graph shown below. Black histogram is for the random variable X and green histogram is for the random variable Y.



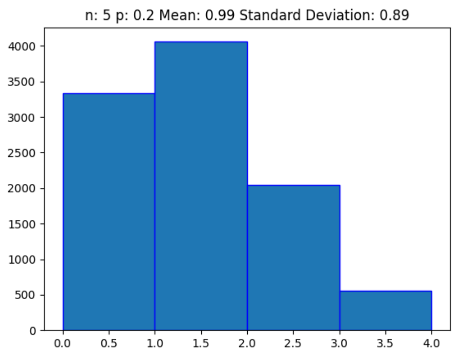
The red histogram shows the sum of these two random variables $X + Y$ in form of another histogram. The blue curve is the theoretical result we have found in the question 5. Section a) and it says the theoretical probability distribution of Z=$X + Y$ is as follows:

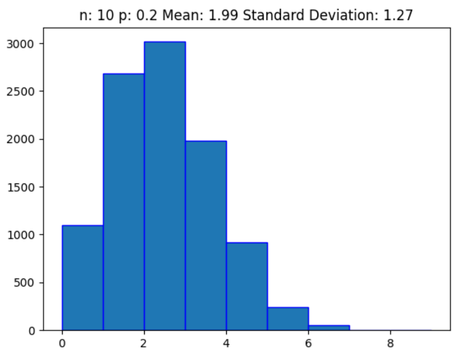
$$f_Z(z) = \frac{1}{2\sqrt{\pi}} * e^{-\frac{z^2}{4}}$$

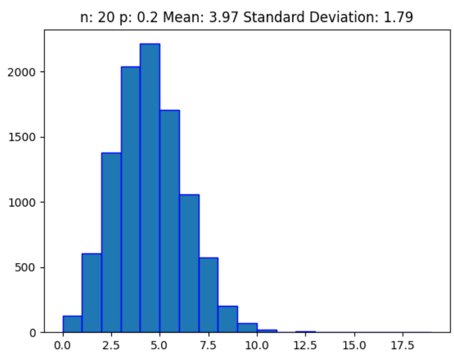
The result is very different than the sampled histogram result. The histogram of $X + Y$ is basically the sum of each histogram column of X and Y and does not show normal distribution properties: Such as it does not have any centered/symmetric behavior around an x-value, namely its mean, nor it does not sum up to 1. Because it is the sum of two histograms of two random variables, which are summing up to 1, making the concatenation histogram sum up to 2. But the graph of theoretical distribution shows a classic normal variable behavior: centered around 0, which is the mean of it and is symmetric around 0, plus the values of y will add up to 1.

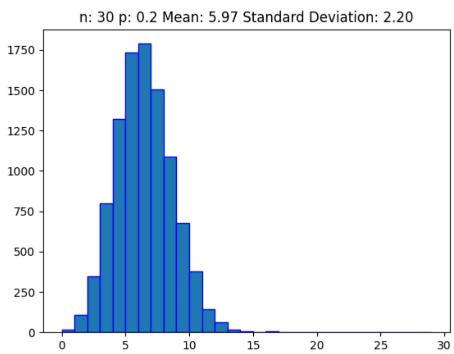
b)

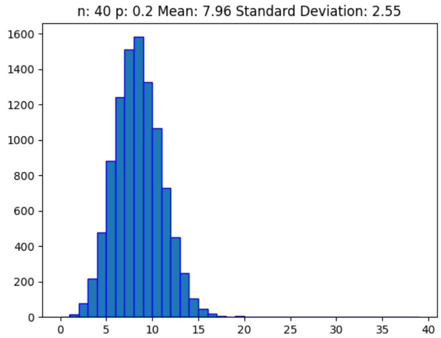
Considering the binomially distributed random variable $X\ B(n, p)$, we know that we can use the normal distribution as an approximation to the binomial distribution when n is large and/or p is close to 0.5. Below, 18 different results of a binomially distributed random variables' histograms and the corresponding n and p values are shown:

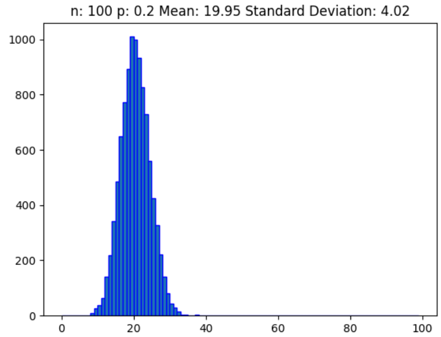




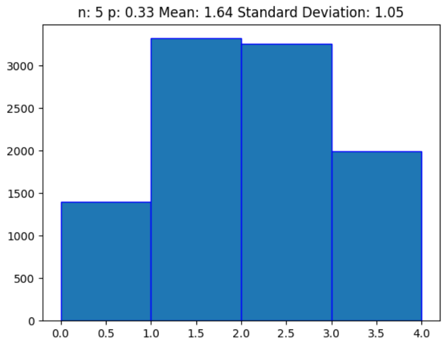


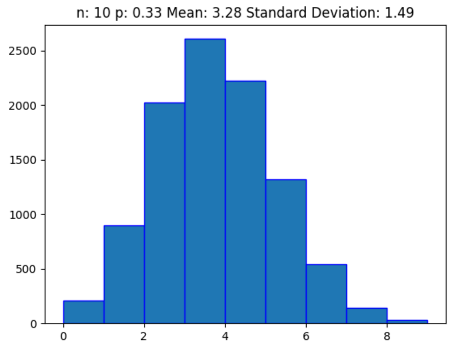


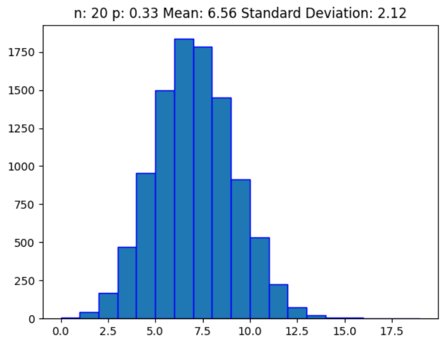


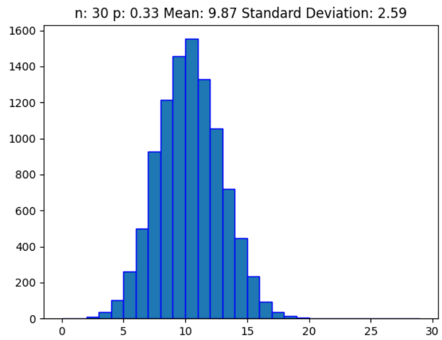


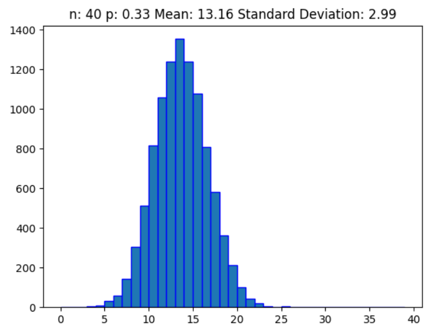
As we can see, p is constant throughout the whole sequence of six graphs above, and as n increases, the shape of the graph looks like a normally distributed variable's graph more and more. Although the mean values of the graphs are always close to the mean of the normal approximation's calculation $X \approx N(np, np(1-p))$, the standard deviation values are not close to the binomial distribution.

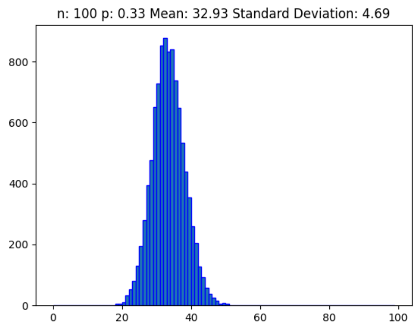




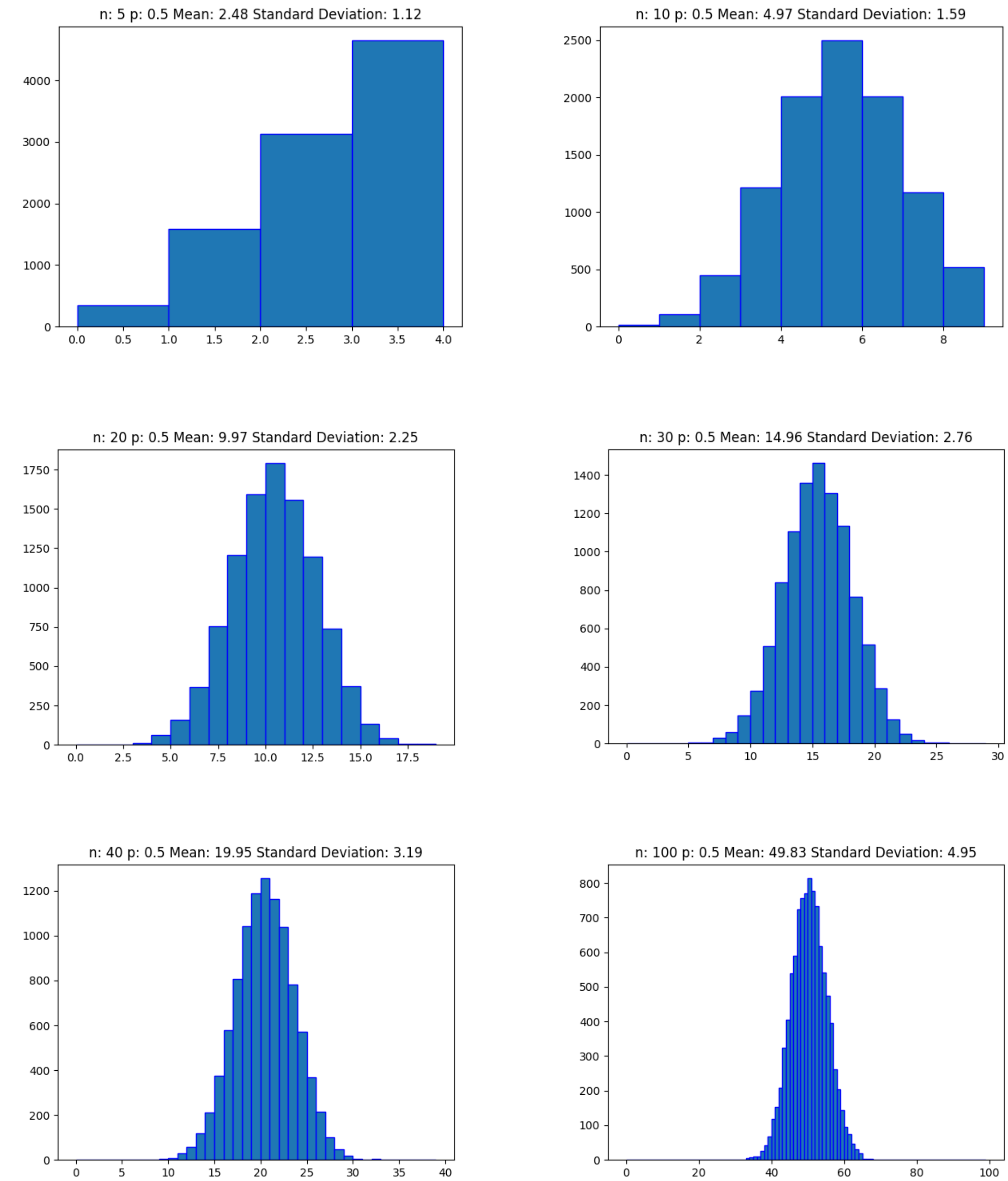








As we can see, p is constant throughout the whole sequence of six graphs above, and as n increases, the shape of the graph looks like a normally distributed variable's graph more and more just like when p was smaller but constant. Although the mean values of the graphs are always close to the mean of the normal approximation's calculation $X \approx N(np, np(1-p))$, the standard deviation values are still not that close. But when p is 0.33 is a better approximation to the normal distribution than p is 0.2.



As we can see, p is constant throughout the whole sequence of six graphs above, and as n increases, the shape of the graph looks like a normally distributed variable's graph more and more just like the previous two cases. In this case besides the mean values of the graphs, the standard derivation is also very similar to the normal approximation's calculation $X \approx N\big(np, np(1-p)\big)$. We can say that when p is 0.5, the calculation can be approximated by the normal distribution best.

c)

From the question 4. Section c) we have derived a closed formula for KL-divergence, dependent on $\mu_1,\ \mu_2,\ \sigma_1,\ \sigma_2$ and it is:

$$KL(p||q) = \ln\left(\frac{\sigma_2}{\sigma_1}\right) - \frac{1}{2} + \frac{1}{2\sigma_2^2} * (\sigma_1^2 + (\mu_1 - \mu_2)^2)$$

If we simply put the values of $\mu_1 = 0, \mu_2 = 0, \sigma_1^2 = 1, \sigma_2^2 = 4$, which are known from the distributions of normal random variables given in the function $p(x) = \mathcal{N}(0, 1)$ and $p(x) = \mathcal{N}(0, 4)$ in the above formula, we would get the KL-divergence value of given distributions of two normal random variables analytically:

$$KL(p||q) = \ln\left(\frac{\sqrt{1}}{\sqrt{4}}\right) - \frac{1}{2} + \frac{1}{2*2} * (1 + (0-0)^2) = \ln\left(\frac{1}{2}\right) - \frac{1}{2} * \frac{1}{4} * (1 + 0^2)$$
$$= \ln\left(\frac{1}{2}\right) - \frac{1}{8} = 0.3181471806$$

On the other hand, if we estimate the value of the KL-divergence by sampling 1000 samples from a Gaussian random variable with mean of 0 and variance of 1 and sum the result of $\ln\left(\frac{p(x)}{q(x)}\right)$, where x is the each sample value, for every sample, and divide the result by 1000 to calculate the average approximation, we would get the following results:

The program code has been run several times to calculate best approximation:

0.32194569

0.33746146

0.31380791

0.29679152

0.34402254

0.31812228

0.32544649

0.31392256

0.31758475

0.29892018

After 10 runs, the approximate result of the calculation yields 0.318802538 as the result of the KL-divergence calculation using the program code for estimation.

We can clearly see that the results are very similar. After the analytical calculation, we are provided the result of 0.3181471806, while after the computation we are provided the result of 0.318802538.

If we name the analytical result as *approximate value* and the computation result *exact value*, we will get the error rate of:

$$\frac{|approximate\ value - exact\ value|}{exact\ value} * 100 = \frac{|0.3181471806 - 0.318802538|}{0.318802538} * 100 = 0.2055684387 \cong \%0.2$$

So, we can say that we can find the KL-divergence value of given distributions of two normal random variables analytically using the above formula with the error rate of 0.2 percent compared to the exact value of the KL-divergence.