CmpE 343- Assignment I

Karahan Sarıtaş - 2018400174

December 2, 2021

1 Answers

1. (a) Let's start by some definitions. (Source: CmpE 220 Discrete Math Handouts - Haluk Bingöl)

Definition 1. *The complement of A with respect to the universal set U : $\overline{A} \triangleq U \backslash A$. Formally $A^{\complement} = \{x \in U : x \notin A\}$.*

Definition 2. *The intersection of A and B defined as $A \cap B \triangleq \{x | x \in A \wedge x \in B\}$.*

Definition 3. *A and B are disjoint $\Leftrightarrow A \cap B = \emptyset$.*

Definition 4. *A is a proper subset of B, denoted by $A \subset B \Leftrightarrow \forall a(a \in A \rightarrow a \in B) \wedge \exists b(b \notin A \wedge b \in B)$.*

Definition 5. *The set of all possible outcomes for a statistical experiment is called the **sample space**.*

(Source: Probability & Statistics for Engineers & Scientists, 9th Edition)

By the definition of the sample space, we can infer that in the probability theory, universal set corresponds to the sample space. By the definition of the complement we know that $A^{\complement} = \{x \in \Omega : x \notin A\}$. Since $A \subset \Omega$, we can represent A as $\{x | x \in \Omega : x \in A)\}$.

First, let's prove that A and $A^{\complement}$ are disjoint events.

$$
\begin{aligned}
&A \cap A^{\complement} = \{x | x \in A \wedge x \in A^{\complement}\} \\
&A \cap A^{\complement} = \{x | (x \in \Omega : x \in A) \wedge (x \in \Omega : x \notin A)\} \\
&A \cap A^{\complement} = \{x | x \in \Omega : x \in A \wedge x \notin A\} \\
&A \cap A^{\complement} = \emptyset \quad \text{(by the contradiction law)}
\end{aligned}
$$

In conclusion, A and $A^{\complement}$ are disjoint events $\Leftrightarrow A \cap A^{\complement} = \emptyset$

Now, let's prove that $A \cup A^{\complement} = \Omega$.

$$\begin{aligned}
A \cup A^{\complement} &= \{x | x \in A \vee x \in A^{\complement}\} \\
A \cup A^{\complement} &= \{x | (x \in \Omega : x \in A) \vee (x \in \Omega : x \notin A)\} \\
A \cup A^{\complement} &= \{x | x \in \Omega : x \in A \vee x \notin A\} \\
A \cup A^{\complement} &= \{x | x \in \Omega\} \\
A \cup A^{\complement} &= \Omega
\end{aligned}$$

So, we know that $A \cup A^{\complement} = \Omega$. By the **countable additivity**:

$$\begin{aligned}
P(A) + P(A^{\complement}) &= P(A \cup A^{\complement}) \\
P(A) + P(A^{\complement}) &= P(\Omega) \\
P(A) + P(A^{\complement}) &= 1 \quad \text{by } \textbf{Normalization} \\
P(A) &= 1 - P(A^{\complement})
\end{aligned}$$

(b) Let's first show that if all E_i's are disjoint events, then all $E_i \cap A$'s are also disjoint events.

$$\text{For any } \; i, j \in \{1, 2, ..., n\} \; \text{ such that } \; i \neq j \quad \begin{aligned} E_i \cap A &= \{x | x \in A \wedge x \in E_i\} \\ E_j \cap A &= \{x | x \in A \wedge x \in E_j\} \end{aligned}$$

$$\begin{aligned}
(E_i \cap A) \cap (E_j \cap A) &= \{x | x \in (E_i \cap A) \wedge x \in (E_j \cap A)\} \\
&= \{x | (x \in A \wedge x \in E_i) \wedge (x \in A \wedge x \in E_j)\} \\
&= \{x | x \in A \wedge (x \in E_i \wedge x \in E_j)\} \\
E_i \cap E_j &= (x \in E_i \wedge x \in E_j) = \emptyset \quad \text{by the definition of disjoint events} \\
(E_i \cap A) \cap (E_j \cap A) &= \{x | x \in A \wedge \emptyset\} \\
(E_i \cap A) \cap (E_j \cap A) &= \emptyset \rightarrow \quad \text{All } E_i \cap A\text{'s are also disjoint events.}
\end{aligned}$$

Let's start by expanding the term:

$$\begin{aligned}
\sum_{i=1}^{n} P(A \cap E_i) &= P(\cup_{i=1}^{n} (A \cap E_i)) \quad \text{by } \textbf{countable additivity} \\
&= P((A \cap E_1) \cup (A \cap E_2) \cup ... \cup (A \cap E_n)) \\
&= P(A \cap (E_1 \cup E_2 \cup E_3 ... \cup E_n)) \quad \text{by distributivity} \\
&= P(A \cap (\cup_{i=1}^{n} (E_i))) \\
&= P(A \cap \Omega)* \\
&= P(A)
\end{aligned}$$

* = Let's prove that $A \subset \Omega$ implies $A \cap \Omega = A$. Notice that this is actually if-and-only-if relationship but for the question we do not have to prove right

to left part.

$$
\begin{aligned}
x \in A \cap \Omega \to x \in A \wedge x \in \Omega \\
A \cap \Omega \subseteq A \\
x \in A \to x \in \Omega \quad (\text{by } A \subset \Omega) \\
x \in \Omega \wedge x \in A \to x \in A \cap \Omega \\
A \subseteq A \cap \Omega \\
\text{to conclude} \quad A \cap \Omega \subseteq A \wedge A \subseteq A \cap \Omega \longrightarrow A = A \cap \Omega
\end{aligned}
$$

(c) In the first question, we proved that $P(A) = 1 - P(A^\complement)$. By **nonnegativity** and **normalization**, we know that:

$$
\begin{aligned}
\forall A \subset \Omega \quad P(A^\complement) &\geq 0 \\
-P(A^\complement) &\leq 0 \\
1 - P(A^\complement) &\leq 1 \\
P(A) &\leq 1
\end{aligned}
$$

Now, let reorder the inequality given in the question.

$$
\begin{aligned}
P(A \cap B) &\geq P(A) + P(B) - 1 \\
1 &\geq P(A) + P(B) - P(A \cap B)
\end{aligned}
$$

Let's show that $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. To begin with, we have to represent $A \cup B$ as the union of two disjoint sets.

$$
\begin{aligned}
P(A \cup B) &= P(\{x | x \in A \vee x \in B\}) \\
A \cup B &= (A \cup B) \cap \Omega \qquad (A \cup B \subseteq \Omega) \\
&= (A \cup B) \cap (A \cup A^\complement) \\
&= A \cup (B \cap A^\complement) \\
A \cap (B \cap A^\complement) &= A \cap A^\complement \cap B = \emptyset \cap B = \emptyset \longrightarrow \; A \text{ and } B \cap A^\complement \text{ are disjoint events.} \\
P(A \cup (B \cap A^\complement)) &= \color{red}{P(A) + P(B \cap A^\complement)} \;\; \text{by } \textbf{countable additivity} \\
B &= B \cap (A \cup A^\complement) = (B \cap A) \cup (B \cap A^\complement) \\
(B \cap A) \cap (B \cap A^\complement) &= B \cap A \cap A^\complement = \emptyset \longrightarrow \; (B \cap A) \text{ and } (B \cap A^\complement) \text{ are disjoint events.} \\
P(B) &= P(B \cap A) + P(B \cap A^\complement) \\
P(B) - P(B \cap A) &= \color{red}{P(B \cap A^\complement)}
\end{aligned}
$$

to conclude $P(A \cup B) = P(A) + P(B \cap A^\complement) = P(A) + P(B) - P(A \cap B)$

We proved above that $\forall A \subset \Omega \quad P(A) \leq 1$.

$$
\begin{aligned}
A \cup B \subset \Omega \longrightarrow P(A \cup B) &\leq 1 \\
P(A) + P(B) - P(A \cap B) &\leq 1 \\
P(A \cap B) &\geq P(A) + P(B) - 1
\end{aligned}
$$

2. $X = 1$ if the sum of the numbers ≤ 5. Let's count the possible combinations: $(1,1)$,$(1,2)$,$(1,3)$,$(1,4)$,$(2,1)$,$(2,2)$,$(2,3)$,$(3,1)$,$(3,2)$,$(4,1)$. There are 10 different outcomes that satisfy the condition. In total there are $6 \cdot 6 = 36$ outcomes. $P(X = 1) = \frac{5}{18}$.

   $Y = 1$ if the product of the number is odd. To have a odd product, we have to multiply two odds. Therefore we have to find the number of pairs consisting only of odd numbers. On a dice, there are 3 odds (1, 3, 5) and 3 evens (2, 4, 6). $3 \cdot 3 = 9$ different combinations of odd numbers can be found. $P(Y = 1) = \frac{1}{4}$.

   $P(X = 1, Y = 1)$ can be found by checking the combinations that satisfy both demands. Those combinations are: $(1,1)$, $(1,3)$, $(3,1)$.
   Therefore $P(X = 1, Y = 1) = \frac{3}{36}$.
   Let's find the combinations that do not satisfy both of the requirements: $(1,6)$, $(2,4)$, $(2,5)$, $(2,6)$, $(3,4)$, $(3,6)$, $(4,2)$, $(4,3)$, $(4,4)$, $(4,5)$, $(4,6)$, $(5,2)$, $(5,4)$, $(5,6)$, $(6,1)$,$(6,2)$,$(6,3)$,$(6,4)$,$(6,5)$,$(6,6)$. Thus, $P(X = 0, Y = 0) = \frac{20}{36}$.
   Combinations that satisfy first condition, but not the second one: $(1,2)$, $(1,4)$, $(2,1)$, $(2,2)$, $(2,3)$, $(3,2)$, $(4,1)$. Thus, $P(X = 1, Y = 0) = \frac{7}{36}$.
   Combinations that satisfy second condition, but not the first one:$(1,5)$, $(3,3)$, $(3,5)$, $(5,1)$, $(5,3)$, $(5,5)$. Thus, $P(X = 0, Y = 1) = \frac{6}{36}$.

   $\text{Cov}(X, Y) = E[(X - \mu_x)(Y - \mu_y)] = \sum_x \sum_y (x - \mu_x)(y - \mu_y) f(x, y)$.

   Let's first find the expected value of X and Y respectively.

$$\mu_x = (1) \cdot (\frac{5}{18}) + (0) \cdot (\frac{13}{18})$$
$$\mu_x = \frac{5}{18}$$
$$\mu_y = (1) \cdot (\frac{1}{4}) + (0) \cdot (\frac{3}{4})$$
$$\mu_y = \frac{1}{4}$$

   Now, let's do the calculations to find covariance.

$$\begin{aligned}\sum_y (x - \mu_x)(y - \mu_y) f(x, y) &= (x - \frac{5}{18})(0 - \frac{1}{4}).f(x, 0) + (x - \frac{5}{18})(1 - \frac{1}{4}).f(x, 1) \\ &= (x - \frac{5}{18})(-\frac{1}{4} \cdot f(x, 0) + \frac{3}{4} \cdot f(x, 1))\end{aligned}$$

$$\begin{aligned}\sum_x (x - \frac{5}{18})(-\frac{1}{4} \cdot f(x, 0) + \frac{3}{4} \cdot f(x, 1)) &= (0 - \frac{5}{18})(-\frac{1}{4} \cdot f(0, 0) + \frac{3}{4} \cdot f(0, 1)) + \\ &\quad (1 - \frac{5}{18})(-\frac{1}{4} \cdot f(1, 0) + \frac{3}{4} \cdot f(1, 1)) \\ &= -\frac{5}{18}(-\frac{1}{4} \cdot \frac{20}{36} + \frac{3}{4} \cdot \frac{6}{36}) + \frac{13}{18}(-\frac{1}{4} \cdot \frac{7}{36} + \frac{3}{4} \cdot \frac{3}{36}) \\ &= \frac{1}{72}\end{aligned}$$

3. Let's derive the mean of the Poisson distribution.

$$
\begin{aligned}
\mu &= \sum_{x=0}^{\infty} x \frac{(\lambda t)^x e^{-\lambda t}}{x!} = 0 + \sum_{x=1}^{\infty} x \frac{(\lambda t)^x e^{-\lambda t}}{x!} \\
&= \sum_{x=1}^{\infty} \frac{(\lambda t)^x e^{-\lambda t}}{(x-1)!} \\
&= \lambda t \sum_{x=1}^{\infty} \frac{(\lambda t)^{x-1} e^{-\lambda t}}{(x-1)!} \qquad \text{substitute x - 1 with y} \\
&= \lambda t \sum_{y=0}^{\infty} \frac{(\lambda t)^y e^{-\lambda t}}{(y)!} \\
&= \lambda t \cdot \sum_{y=0}^{\infty} p(y; \lambda t) \quad \text{(normalization)} \\
\mu &= \lambda t
\end{aligned}
$$

4. Let's prove $(a+b)^N = \sum_{m=0}^{N} \binom{N}{m} a^m b^{N-m}$. Then we can easily prove the given equality and normalization as well.

   (a) Base case:

$$
\begin{aligned}
(a+b)^1 &\stackrel{?}{=} \sum_{m=0}^{1} \binom{1}{m} a^m b^{1-m} \\
a+b &\stackrel{?}{=} \binom{1}{0} a^0 b^1 + \binom{1}{1} a^1 b^0 \\
a+b &\stackrel{?}{=} 1 \cdot a^0 b^1 + 1 \cdot a^1 b^0 \\
a+b &= a+b \quad \text{Base case is proven.}
\end{aligned}
$$

   Hypothesis: $(a+b)^k = \sum_{m=0}^{k} \binom{k}{m} a^m b^{k-m}$

Let's prove $(a+b)^{k+1} = \sum_{m=0}^{k+1} \binom{k}{m} a^m b^{k+1-m}$

$$(a+b)^{k+1} = (a+b)(a+b)^k$$

$$= (a+b) \cdot \sum_{m=0}^{k} \binom{k}{m} a^m b^{k-m}$$

$$= a \cdot \sum_{m=0}^{k} \binom{k}{m} a^m b^{k-m} + b \cdot \sum_{m=0}^{k} \binom{k}{m} a^m b^{k-m}$$

$$= \sum_{m=0}^{k} \binom{k}{m} a^{m+1} b^{k-m} + \sum_{m=0}^{k} \binom{k}{m} a^m b^{k-m+1}$$

$$= [\binom{k}{0} a^1 b^k + \binom{k}{1} a^2 b^{k-1} + \ldots + \binom{k}{k-1} a^k b^1 + \binom{k}{k} a^{k+1} b^0] +$$

$$[\binom{k}{0} a^0 b^{k+1} + \binom{k}{1} a^1 b^k + \ldots + \binom{k}{k-1} a^{k-1} b^2 + \binom{k}{k} a^k b^1]$$

Pair the same-coloured terms with each other and apply Pascal's Identity.

$$= \binom{k}{0} a^0 b^{k+1} + \binom{k+1}{1} a^1 b^k + \binom{k+1}{2} a^2 b^{k-1} + \ldots + \binom{k+1}{k} a^k b^1 + \binom{k}{k} a^{k+1} b^0$$

Observe that $\binom{k}{0} a^0 b^{k+1} = a^0 b^{k+1} = \binom{k+1}{0} a^0 b^{k+1}$

and $\binom{k}{k} a^{k+1} b^0 = a^{k+1} b^0 = \binom{k+1}{k+1} a^{k+1} b^0$

Therefore in overall:

$$(a+b)^{k+1} = \sum_{m=0}^{k+1} \binom{k+1}{m} a^m b^{k+1-m}$$

By showing the correctness of the hypothesis, we concluded that the equality given above is correct.

Now, replace a with x and b with 1.

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

$$(x+1^N = \sum_{m=0}^{N} \binom{N}{m} x^m \quad \text{Therefore, it's proven.}$$

We can also prove the normalization. Replace a with p and b with $1-p$.

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

$$1 = \sum_{m=0}^{N} \binom{N}{m} p^m (1-p)^{N-m} \quad \text{Therefore, it's proven.}$$

(b) Let's start with the probability density function of the continuous uniform random variable X on the interval $[A, B]$:

$$f(x; A, B) = \begin{cases} \frac{1}{B-A}, & A \leq x \leq B \\ 0, & \text{elsewhere} \end{cases}$$

Now, let's calculate the entropy of the uniform random variable $x \sim U(0, 1)$.

$$H(x) = -\int_0^1 \frac{1}{1-0} ln(\frac{1}{1-0})\, dx$$
$$H(x) = -\int_0^1 1 \cdot 0\, dx$$
$$H(x) = 0$$

Probability density function for Gaussian distribution:

$$X \sim \mathcal{N}(\mu,\, \sigma^2)$$
$$\mathcal{N}(\mu,\, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}}(\frac{x-\mu}{\sigma})^2 \qquad -\infty \leq x \leq \infty$$
$$\mathcal{N}(0,\, 1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \qquad -\infty \leq x \leq \infty$$

Let's calculate the entropy of the Gaussian random variable, without using the numerical values:

$$H(x) = -\int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}}(\frac{x-\mu}{\sigma})^2 ln(\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}}(\frac{x-\mu}{\sigma})^2)\, dx$$
$$H(x) = -\int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}}(\frac{x-\mu}{\sigma})^2 [ln(\frac{1}{\sigma\sqrt{2\pi}}) - \frac{1}{2}(\frac{x-\mu}{\sigma})^2]\, dx$$
$$H(x) = -ln(\frac{1}{\sigma\sqrt{2\pi}}) \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}}(\frac{x-\mu}{\sigma})^2 + \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}}(\frac{x-\mu}{\sigma})^2 \frac{1}{2\sigma^2}(x-\mu)^2$$
$$H(x) = -ln(\frac{1}{\sigma\sqrt{2\pi}}) \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}}(\frac{x-\mu}{\sigma})^2 + \frac{1}{2\sigma^2} \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}}(\frac{x-\mu}{\sigma})^2 (x-\mu)^2$$
$$H(x) = -ln(\frac{1}{\sigma\sqrt{2\pi}}) \int_{-\infty}^{\infty} p(x)\, dx + \frac{1}{2\sigma^2} \int_{-\infty}^{\infty} p(x)(x-\mu^2)\, dx$$

$\int_{-\infty}^{\infty} p(x)\, dx = 1$ normalization $\int_{-\infty}^{\infty} p(x)(x-\mu^2)\, dx = \sigma^2$ variance

$$H(x) = -ln(\frac{1}{\sigma\sqrt{2\pi}}) \cdot 1 + \frac{1}{2\sigma^2} \cdot \sigma^2$$
$$H(x) = ln(\sigma\sqrt{2\pi}) + \frac{1}{2} = \frac{1}{2}(ln(\sigma^2 2\pi) + 1) \qquad \mu = 0, \quad \sigma^2 = 1$$
$$H(x) = \frac{1}{2}(ln(2\pi) + 1)$$

Entropy can be associated with the level of disorder in the system. Entropy of uniform random variable is zero, which means there is only one possible distribution, no disorder, no motion in other words. On the other hand, entropy of Gaussian random variable is very high compared to the entropy of uniform random variable. It means that there is relatively a high level of disorder in Gaussian distribution.

(c) i. Let's show that the KL-divergence between equal distributions is zero.

$$\text{Equal distributions} \quad \longrightarrow \quad p(x) = q(x)$$
$$\int_{-\infty}^{\infty} p(x)ln(\frac{p(x)}{q(x)})\,dx = \int_{-\infty}^{\infty} p(x)ln(1)\,dx = 0$$

ii. I've answered the third part of the question beforehand. Now we know the formula for the KL-divergence between two Gaussian distribution. Let's find $KL(p||q)$ and $KL(q||p)$ for the given values in the third part.

$$KL(p||q) = 2.35 \quad \text{(already found in the third part)}$$
$$KL(q||p) = \frac{1}{2}ln(\frac{\sigma_1^2}{\sigma_2^2}) - \frac{1}{2} + \frac{[\sigma_2^2 + (\mu_2 - \mu_1)^2]}{2\sigma_1^2}$$
$$KL(q||p) = \frac{1}{2}ln(\frac{1.5}{0.2}) - \frac{1}{2} + \frac{[0.2 + (1.8 - 2)^2]}{3.0}$$
$$KL(q||p) = \frac{1}{2}ln(\frac{\sigma_1^2}{\sigma_2^2}) - \frac{1}{2} + \frac{[\sigma_2^2 + (\mu_2 - \mu_1)^2]}{2\sigma_1^2}$$
$$KL(q||p) = 1 - \frac{1}{2} + 0.08$$
$$KL(q||p) = 0.58$$
$$KL(p||q) \neq KL(q||p)$$

iii. $KL(p||q)$ can be written like this:

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

We've already derived a closed form for the first part:

$$\int_{-\infty}^{\infty} p(x)ln(p(x))\,dx = -\frac{1}{2}(ln(\sigma_1^2 2\pi) + 1)$$

Let's find the closed form of the second one.

$$\int_{-\infty}^{\infty} p(x)ln(q(x))\,dx = -\int_{-\infty}^{\infty} \frac{1}{\sigma_1\sqrt{2\pi}}e^{-\frac{1}{2}}(\frac{x-\mu_1}{\sigma_1})^2[ln(\frac{1}{\sigma_2\sqrt{2\pi}}) - \frac{1}{2}(\frac{x-\mu_2}{\sigma_2})^2]\,dx$$
$$= ln(\frac{1}{\sigma_2\sqrt{2\pi}})\int_{-\infty}^{\infty} p(x)dx - \frac{1}{2\sigma_2^2}\int_{-\infty}^{\infty} p(x)(x-\mu_2^2)dx$$
$$= ln(\frac{1}{\sigma_2\sqrt{2\pi}}) - \frac{1}{2\sigma_2^2}[\int_{-\infty}^{\infty} p(x)x^2 + \int_{-\infty}^{\infty} p(x)\mu_2^2 - 2\mu_2\int_{-\infty}^{\infty} p(x)x]$$
$$= ln(\frac{1}{\sigma_2\sqrt{2\pi}}) - \frac{1}{2\sigma_2^2}[E(x^2) + \mu_2^2 - 2\mu_2\mu_1]$$

Now, let's leave the equation like this and find $E(x^2)$ for $p(x)$. Recall that:

$$\sigma_1^2 = E(x^2) - \mu_1^2$$
$$E(x^2) = \sigma_1^2 + \mu_1^2$$

Now, let's put it into the equation.

$$\int_{-\infty}^{\infty} p(x)ln(q(x))\,dx = ln(\frac{1}{\sigma_2\sqrt{2\pi}}) - \frac{1}{2\sigma_2^2}[\sigma_1^2 + \mu_1^2 + \mu_2^2 - 2\mu_2\mu_1]$$
$$= -\frac{1}{2}ln(\sigma_2^2 2\pi) - \frac{1}{2\sigma_2^2}[\sigma_1^2 + (\mu_1 - \mu_2)^2]$$

Now, find $KL(p||q)$.

$$KL(p||q) = \int_{-\infty}^{\infty} p(x)ln(p(x))\,dx - \int_{-\infty}^{\infty} p(x)ln(q(x))\,dx$$
$$= -\frac{1}{2}(ln(\sigma_1^2 2\pi) + 1) + \frac{1}{2}ln(\sigma_2^2 2\pi) + \frac{1}{2\sigma_2^2}[\sigma_1^2 + (\mu_1 - \mu_2)^2]$$
$$= \frac{1}{2}ln(\frac{\sigma_2^2}{\sigma_1^2}) - \frac{1}{2} + \frac{[\sigma_1^2 + (\mu_1 - \mu_2)^2]}{2\sigma_2^2}$$
$$= ln(\frac{\sigma_2}{\sigma_1}) - \frac{1}{2} + \frac{[\sigma_1^2 + (\mu_1 - \mu_2)^2]}{2\sigma_2^2}$$

Now, let's calculate the result with the given values.

$$KL(p||q) = \frac{1}{2}ln(\frac{\sigma_2^2}{\sigma_1^2}) - \frac{1}{2} + \frac{[\sigma_1^2 + (\mu_1 - \mu_2)^2]}{2\sigma_2^2}$$
$$KL(p||q) = \frac{1}{2}ln(\frac{0.2}{1.5}) - \frac{1}{2} + \frac{[1.5 + (2 - 1.8)^2]}{0.4}$$
$$KL(p||q) = -1 - \frac{1}{2} + 3.85$$
$$KL(p||q) = 2.35$$

5. (a) Let's start with the definition of standard Gaussian random variable:

Definition 6. *The distribution of a normal random variable with mean 0 and variance 1 is called a standard normal distribution.*

(Source: Probability & Statistics for Engineers & Scientists, 9th Edition)

$$\mathcal{N}(0,\,1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \qquad -\infty \leq x \leq \infty$$

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(t) \cdot f_Y(z-t) dt$$
$$f_Z(z) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{(z-t)^2}{2}} dt$$
$$f_Z(z) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\cdot(t^2+(z-t)^2)} dt$$
$$f_Z(z) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\cdot(z^2-2tz+2t^2)} dt$$
$$f_Z(z) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-\frac{z^2}{2}} \cdot e^{tz} \cdot e^{-t^2} dt$$
$$f_Z(z) = \frac{1}{2\pi} e^{-\frac{z^2}{2}} \cdot \int_{-\infty}^{\infty} e^{tz} \cdot e^{-t^2} dt$$
$$f_Z(z) = \frac{1}{2\pi} e^{-\frac{z^2}{2}} \cdot \int_{-\infty}^{\infty} e^{-t^2} \cdot e^{tz} \cdot e^{\frac{-z^2}{4}} \cdot e^{\frac{z^2}{4}} dt$$
$$f_Z(z) = \frac{1}{2\pi} e^{-\frac{z^2}{2}} \cdot \int_{-\infty}^{\infty} e^{-(t-\frac{a}{2})^2} \cdot e^{\frac{z^2}{4}} dt$$
$$f_Z(z) = \frac{1}{2\pi} e^{-\frac{z^2}{2}} \cdot e^{\frac{z^2}{4}} \cdot \int_{-\infty}^{\infty} e^{-(t-\frac{a}{2})^2} \cdot dt$$
$$f_Z(z) = \frac{1}{2\pi} e^{-\frac{z^2}{4}} \cdot \int_{-\infty}^{\infty} e^{-(t-\frac{a}{2})^2} \cdot dt$$
$$\text{substitute} \quad t - \frac{a}{2} \quad \text{with} \quad n \to dt = dn$$
$$f_Z(z) = \frac{1}{2\pi} e^{-\frac{z^2}{4}} \cdot \int_{-\infty}^{\infty} e^{-n^2} \cdot dn$$
$$f_Z(z) = \frac{1}{2\sqrt{\pi}} e^{-\frac{z^2}{4}}$$

Since denominators are multiplied by $\sqrt{2}$, it looks like this function belongs to Gaussian distribution with $\mu = 0$ and $\sigma^2 = 2$. Let's find μ_Z and σ_Z^2 to prove our hypothesis.

$$
\begin{aligned}
\mu_Z &= \int_{-\infty}^{\infty} x \cdot f_Z(x)\, dx \\
\mu_Z &= \frac{1}{2\sqrt{\pi}} \int_{-\infty}^{\infty} x \cdot e^{-\frac{x^2}{4}}\, dx \\
e^{-\frac{x^2}{4}} &= u, \quad -\frac{x}{2} \cdot e^{-\frac{x^2}{4}} dx = du \\
\mu_Z &= \frac{1}{2\sqrt{\pi}} \int_{0}^{0} -2\, du \\
\mu_Z &= 0
\end{aligned}
$$

$$
\begin{aligned}
\sigma_Z^2 &= \int_{-\infty}^{\infty} (x-\mu_Z)^2 \cdot f_Z(x)\, dx \\
\sigma_Z^2 &= \frac{1}{2\sqrt{\pi}} \int_{-\infty}^{\infty} x^2 \cdot e^{-\frac{x^2}{4}}\, dx \\
\text{substitute} \quad \frac{x}{2} &= n, \quad \frac{1}{2} dx = dn \\
\sigma_Z^2 &= \frac{1}{2\sqrt{\pi}} \int_{-\infty}^{\infty} 8n^2 \cdot e^{-n^2}\, dn \\
\sigma_Z^2 &= \frac{4}{\sqrt{\pi}} \int_{-\infty}^{\infty} n^2 \cdot e^{-n^2}\, dn \\
\text{Integration by parts:} \quad u &= n, \quad dv = -2n \cdot e^{-n^2}\, dn \\
\sigma_Z^2 &= \frac{4}{\sqrt{\pi}} \cdot -\frac{1}{2} (n \cdot e^{-n^2} |_{-\infty}^{\infty} - \int_{-\infty}^{\infty} e^{-n^2} dn) \\
\sigma_Z^2 &= \frac{4}{\sqrt{\pi}} \cdot -\frac{1}{2} (0 - \sqrt{\pi}) \\
\sigma_Z^2 &= 2
\end{aligned}
$$

Well, indeed we successfully showed that this function belongs to the Gaussian distribution with $\mu = 0$ and $\sigma^2 = 2$.

(b) Let's show pdf of X and pmf of Y.

$$
\begin{aligned}
\mathcal{N}(0,\, 1) &= \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \qquad -\infty \le x \le \infty \\
f(y) &= \frac{1}{2}, \quad \text{for} \quad y = -1, 1
\end{aligned}
$$

Now, let's prove that $Z = XY$ is independent of Y. Observe that outcomes for Y is equally likely. When $Y = 1$, probability of $XY = z$ is the same as the probability of $X = z$, therefore it has the probability distribution $\frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$. When $Y = -1$, probability of $XY = z$ is the same as the probability of

$X = -z$, therefore it has the probability distribution $\frac{1}{\sqrt{2\pi}}e^{-\frac{(-z)^2}{2}}$. By Bayes' rule,

$$P(XY = z) = P(XY = z|Y = 1) \cdot P(Y = 1) + P(XY = z|Y = -1) \cdot P(Y = -1)$$

$$\begin{aligned} P(XY = z) &= \frac{1}{2} \cdot \frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}} + \frac{1}{2} \cdot \frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}} \\ &= \frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}} \end{aligned}$$

Now we know $P(XY = z)$. Let's prove the independence of XY and Y by showing that $P(XY = z \cap Y = y) = P(XY = z) \cdot P(Y = y)$ for each value of random variable Y.

$$\begin{aligned} P(XY = z \cap Y = 1) &\stackrel{?}{=} P(XY = z) \cdot P(Y = 1) && \text{for } Y = 1 \\ P(XY = z|Y = 1) \cdot P(Y = 1) &= P(XY = z) \cdot P(Y = 1) && \text{Bayes' rule} \\ \frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}} \cdot \frac{1}{2} &= P(XY = z) \cdot P(Y = 1) \\ \frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}} \cdot \frac{1}{2} &= \frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}} \cdot \frac{1}{2} && \text{(equality proved)} \end{aligned}$$

$$\begin{aligned} P(XY = z \cap Y = -1) &\stackrel{?}{=} P(XY = z) \cdot P(Y = -1) && \text{for } Y = -1 \\ P(XY = z|Y = -1) \cdot P(Y = -1) &= P(XY = z) \cdot P(Y = -1) && \text{Bayes' rule} \\ \frac{1}{\sqrt{2\pi}}e^{-\frac{(-z)^2}{2}} \cdot \frac{1}{2} &= P(XY = z) \cdot P(Y = 1) \\ \frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}} \cdot \frac{1}{2} &= \frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}} \cdot \frac{1}{2} && \text{(equality proved)} \end{aligned}$$

Therefore $Z = XY$ is independent of Y.

(c) Let's start by calculating $E[Y]$.

$$\begin{aligned} E[Y] &= \sum_y y.p(y) \\ E[Y] &= (0) \cdot P(X \leq k) + (k) \cdot P(X \geq k) \\ E[Y] &= k \cdot P(X \geq k) \end{aligned}$$

If X is smaller than k, Y becomes 0, and X may be 0 or a positive number. So for $X < k$, $0 = Y \leq X$ always holds.
If X is greater than or equal to k, Y becomes k, and X may be k or a larger number. So for $X \geq k$, $k = Y \leq X$ always holds.

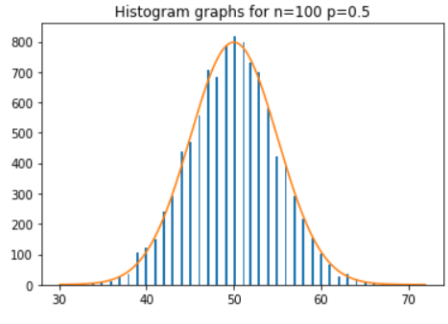
Notice that in each case, X always takes values greater than or equal to Y. Therefore we can conclude that expected value for X is greater than or equal to that of Y, $X \geq Y \rightarrow E[X] \geq E[Y]$. We can do this transformation because X and Y take non-negative values, therefore expectation function is a nondecreasing function. We can also show that $E[X] \geq E[Y]$ more formally by the inequality below. Reasoning is exactly the same for the discrete case:

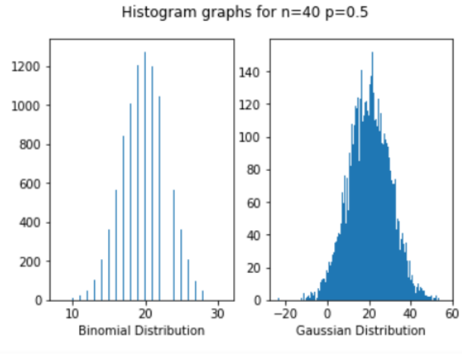
$$E[X] = \int_0^\infty xf(x)\,dx = \int_0^k xf(x)\,dx + \int_k^\infty xf(x)\,dx$$
$$E[X] = \int_0^k xf(x)\,dx + \int_k^\infty xf(x)\,dx \geq \int_0^k (0)f(x)\,dx + \int_k^\infty (k)f(x)\,dx = E[Y]$$

Therefore;

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

6. Python codes associated with this section can be found attached to this document.
   (a) X has a bell-shaped distribution. Frequency of the outcome is very high around X = -1, which is the mean of the distribution. It's variance is less than the variance of random variable Y, which can be observed from the histograms. Width of the histogram of X is narrower than that of Y. Peak point of the histogram of random variable X is greater than the peak point of Y. So, shape is perfectly consistent with what we've learned theoretically.
   (b) As stated in the question, as n increases and p is around 0.5, binomial distribution resembles the normal distribution. For p values not closer to 0.5, number of misaligned parts between the normal curve and binomial distribution increases. I've implemented two different functions to compare normal distribution and binomial distribution. In one function, normal curve and binomial distribution histogram are superimposed for comparison. Other function compares the distributions with two different histogram side by side. When p = 0.5 and n $\geq$ 40, resemblance becomes more evident.





(c) Let's calculate the $KL(p||q)$:

$$\begin{aligned}
KL(p||q) &= \frac{1}{2}ln(\frac{\sigma_2^2}{\sigma_1^2}) - \frac{1}{2} + \frac{[\sigma_1^2 + (\mu_1 - \mu_2)^2]}{2\sigma_2^2} \\
KL(p||q) &= \frac{1}{2}ln(4) - \frac{1}{2} + \frac{[1 + (0)^2]}{2 \cdot 4} \\
KL(p||q) &= \frac{1}{2}ln(4) - \frac{1}{2} + \frac{[1 + (0)^2]}{2 \cdot 4} \\
&= 0.693 - 0.5 + 0.125 \\
&= 0.318
\end{aligned}$$

My experimental result is approximately 0.320 (of course it varies from one run to another), and my analytical result is 0.318. Therefore result is consistent with the estimate. Furthermore, if I increase the sample size, it gets more and more closer to the analytical result.