Statistical Inference

HW Author: Mohammad Arabzadeh, Nikoo Paknia
Instructor: Mohammadreza A. Dehaqani

Spring 2024

Homework 1

- If you have any questions about the homework, don't hesitate to drop an email to the HW Author.
- Feel free to use the class group to ask questions — our TA team will do their best to help out!
- Please consult the course page for important information on submission guidelines and delay policies to ensure your homework is turned in correctly and on time.
- This course aims to equip you with the skills to tackle all problems in this domain and encourages you to engage in independent research. Utilize your learnings to extend beyond the classroom teachings where necessary.

Problem 1: Visa

(a)

Note: In theory, the probability of winning per year is not independent of the previous years. However, since the probability of winning in some year given that the student has won in the previous year is 0, we can assume that the probabilities are independent, thus if you have assumed independence, you will get full credit.

The probability of winning the visa is winning one of the 3 lotteries.
V_X: winning the visa in year 202X.

$$1 - P(\bar{V_3}, \bar{V_2}, \bar{V_1}) =$$

$$1 - P(\bar{V_3}, \bar{V_2} | \bar{V_1}) \cdot P(\bar{V_1}) =$$

$$1 - P(\bar{V_3} | \bar{V_2}, \bar{V_1}) \cdot P(\bar{V_2} | \bar{V_1}) \cdot P(\bar{V_1})$$

$P(\bar{V_1}) = 1 - (P(\text{Normal pool}) + P(\text{graduate pool}|\overline{\text{Normal pool}}) \cdot P(\overline{\text{Normal pool}})) =$
$1 - (\frac{65000}{308613} + \frac{20000}{121000} \cdot (1 - \frac{65000}{308613})) = 0.6589$

$P(\bar{V_2}|\bar{V_1}) = 1 - (P(\text{Normal pool}) + P(\text{graduate pool}|\overline{\text{Normal pool}}) \cdot P(\overline{\text{Normal pool}})) =$

$1-(\frac{65000}{483927}+\frac{20000}{127600}\cdot(1-\frac{65000}{483927}))=0.72999$

$P(\bar{V_3}|\bar{V_2},\bar{V_1})=1-(P(\text{Normal pool})+P(\text{graduate pool}|\overline{\text{Normal pool}})\cdot P(\overline{\text{Normal pool}}))=$
$1-(\frac{65000}{780884}+\frac{20000}{314000}\cdot(1-\frac{65000}{780884}))=0.8583$

$$1-0.8583\cdot 0.72999\cdot 0.6589=0.5871$$

(b) For years 2021, 2022, and 2023:

$$P(V_1)=P(\text{Normal pool})+P(\text{graduate pool}|\overline{\text{Normal pool}})\cdot P(\overline{\text{Normal pool}})=$$

$$\frac{65000}{308613}+\frac{20000}{121000}\cdot(1-\frac{65000}{308613})=0.3411$$

$$P(V_2)=P(\text{Normal pool})+P(\text{graduate pool}|\overline{\text{Normal pool}})\cdot P(\overline{\text{Normal pool}})=$$

$$\frac{65000}{483927}+\frac{20000}{127600}\cdot(1-\frac{65000}{483927})=0.2700$$

$$P(V_3)=P(\text{Normal pool})+P(\text{graduate pool}|\overline{\text{Normal pool}})\cdot P(\overline{\text{Normal pool}})=$$

$$\frac{65000}{780884}+\frac{20000}{314000}\cdot(1-\frac{65000}{780884})=0.1417$$

The student's chances are not good, progressively getting worse each year. Hence they cannot rely much on H1-B visa to get to a green-card.

(c)
G_X: Getting green-card in year 202X

$$1-P(\bar{G_3},\bar{G_2},\bar{G_1})=$$

$$1-(P(\bar{G_3}|\bar{G_2},\bar{\bar{G_1}})\cdot P(G_2|\bar{G_1})\cdot P(\bar{G_1}))=$$

$P(\bar{G_3}|\bar{G_2},\bar{\bar{G_1}})\cdot P(\bar{G_2}|\bar{G_1})\cdot P(\bar{G_1})=(\frac{85}{100})^3=0.6141$

$$1-0.6141=0.3859$$

The student's chances of getting a marriage-based green-card are 38.59%.

Problem 2: Independence (Optional)

(a)

$$N = N_1 + N_2 \Rightarrow P(N = n) = \sum_{k=0}^{n} P(N_1 = k, N_2 = n - k) =$$

$$\sum_{k=0}^{n} P(N_1 = k)P(N_2 = n - k) =$$

$$\sum_{k=0}^{n} \frac{\lambda_1^k e^{-\lambda_1}}{k!} \cdot \frac{\lambda_2^{n-k} e^{-\lambda_2}}{(n-k)!} =$$

$$e^{-(\lambda_1+\lambda_2)} \sum_{k=0}^{n} \lambda_1^k \lambda_2^{n-k} \frac{1}{k!(n-k)!} =$$

$$e^{-(\lambda_1+\lambda_2)} \sum_{k=0}^{n} \lambda_1^k \lambda_2^{n-k} \frac{1}{n!} \binom{n}{k} =$$

$$e^{-(\lambda_1+\lambda_2)} \frac{1}{n!} \sum_{k=0}^{n} \binom{n}{k} \lambda_1^k \lambda_2^{n-k} =$$

$$e^{-(\lambda_1+\lambda_2)} \frac{1}{n!} (\lambda_1 + \lambda_2)^n = Poisson(\lambda_1 + \lambda_2)$$

(b)

$$f_T(t) = \int_0^t f_{T_1}(u) f_{T_2}(t-u)\, du, \quad t \geq 0.$$

$$f_T(t) = \int_0^t \lambda_1 e^{-\lambda_1 u} \cdot \lambda_2 e^{-\lambda_2 (t-u)}\, du.$$

$$f_T(t) = \lambda_1 \lambda_2 e^{-\lambda_2 t} \int_0^t e^{-(\lambda_1-\lambda_2)u}\, du.$$

Evaluate the integral:

$$\int_0^t e^{-(\lambda_1-\lambda_2)u}\, du = \begin{cases} \frac{1-e^{-(\lambda_1-\lambda_2)t}}{\lambda_1-\lambda_2}, & \lambda_1 \neq \lambda_2, \\ t, & \lambda_1 = \lambda_2. \end{cases}$$

Substitute the integral into $f_T(t)$:

$$f_T(t) = \begin{cases} \frac{\lambda_1\lambda_2}{\lambda_1-\lambda_2} \left(e^{-\lambda_2 t} - e^{-\lambda_1 t}\right), & \lambda_1 \neq \lambda_2, \\ \lambda^2 t e^{-\lambda t}, & \lambda_1 = \lambda_2 = \lambda. \end{cases}$$

Problem 3: Mean, Median, and Mode

a

Note: There is a criteria missing in the question. here's the correct question:

> ”Show that in unimodel concave symmetric distributions, the mean, median, and mode are equal.”

If you have shown that the premise in the question is wrong, you will get the full credit of this section.

assuming that this criteria is present, we can prove this as follows:

The mean:

if the distribution is symmetric, there is a central point x_c such that the distribution is symmetric around this point. This means that for every x, there is a corresponding x' such that $x' = 2x_c - x$; with this perspective, there exists $\frac{n}{2}$ pairs (x_i, x_i') which are symmetric around x_c, hence:

$$\sum_{i=1}^{\frac{n}{2}} \overbrace{\frac{x_i + x_i'}{2}}^{x_c} = \frac{n}{2} x_c \Rightarrow x_c = \frac{1}{n} \sum_{i=1}^{\frac{n}{2}} x_i + x_i' = \frac{1}{n} \sum_{i=1}^{n} x_i = \mu$$

The median:

from the difinition of median: M is the point that:

$$F(M) = P(X \leq M) = \frac{1}{2}$$

and for any point greater than M, the probability of being greater than M is:

$$P(X > M) = 1 - P(X \leq M) = \frac{1}{2}$$

and since the distribution is symmetric, the probability of being less than M is also $\frac{1}{2}$, hence M is the point that divides the distribution into two equal parts, so M is the same as x_c and μ.

The mode: The mode is the point that has the highest density in the distribution. If the distribution is unimodal and concave (i.e. mode exists and is unique) then the central point x_c is the peak point in the distribution, because if x_c is not the peak point, then the mode either doesn't exist or is not unique. So mode is equal to x_c and in symmetric distributions, $x_c = \mu$.

---

b.

No, in general the mean, median, are not necessarily equal in a multimodal, symmetric distribution. Figure 2 shows an example of a bimodal distribution where the mean and median are equal but the modes are different. This happens when the multimodal distribution doesn't have a unique peak point, so the mean and median are equal but the modes are different.

Problem 4: A geometric point of view

(a)

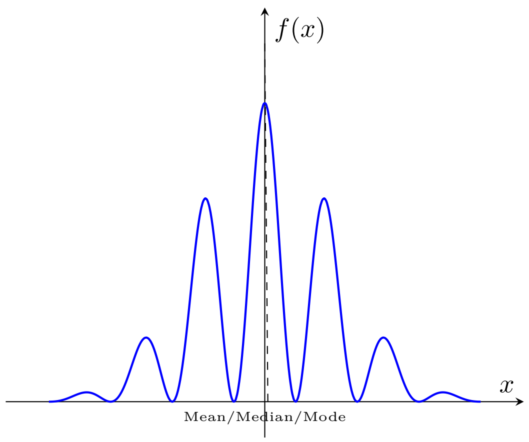

Figure 1: Example of mean, median, and mode appearing at the same point.

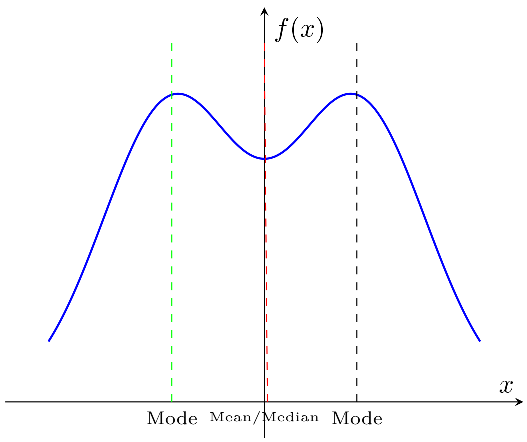

Figure 2: Mean and median at the same point but two modes at different points.

As the chance of being in any part of the ellipse is the same $XY \sim U(0, \text{Area of ellipse})$
Area of ellipse $= \pi ab$

$$f_{XY}(x,y) = \frac{1}{\pi ab}$$

$$\frac{b}{a}\sqrt{a^2 - x^2}$$

$$\text{Y is between } -\frac{b}{a}\sqrt{a^2-x^2} \text{ and } \frac{b}{a}\sqrt{a^2-x^2}$$

$$\text{X is between } -\frac{a}{b}\sqrt{b^2-y^2} \text{ and } \frac{a}{b}\sqrt{b^2-y^2}$$

$$\begin{aligned} f_X(x) &= \int_{-\frac{b}{a}\sqrt{a^2-x^2}}^{\frac{b}{a}\sqrt{a^2-x^2}} f_{XY}(x,y)\,dy = \frac{1}{\pi ab}\int_{-\frac{b}{a}\sqrt{a^2-x^2}}^{\frac{b}{a}\sqrt{a^2-x^2}} 1\,dy \\ &= \frac{1}{\pi ab}[y]_{-\frac{b}{a}\sqrt{a^2-x^2}}^{\frac{b}{a}\sqrt{a^2-x^2}} = \frac{1}{\pi ab}\left(\frac{2b}{a}\sqrt{a^2-x^2}\right) \\ &= \frac{2b}{\pi ab^2}\sqrt{a^2-x^2} \end{aligned}$$

$$\Rightarrow f_Y(y) = \frac{2}{\pi a^2}\sqrt{a^2-x^2}$$

(b)

$$f_{XY}(x,y) = 1 \cdot \frac{1}{\pi ab}$$

(c)

$$A = \int_{-\frac{a}{2}}^{\frac{a}{2}} \frac{2}{a^2}\sqrt{a^2 - x^2}\,dx$$

$$\frac{2}{a}\int_{-\frac{a}{2}}^{\frac{a}{2}} \sqrt{1 - \frac{x^2}{a^2}}\,dx$$

$$x = a\sin u \Rightarrow \frac{x}{a} = \sin u \Rightarrow dx = a\cos u\,du$$

$$-\frac{a}{2} \leq x \leq \frac{a}{2} \Rightarrow -\frac{\pi}{6} \leq u \leq \frac{\pi}{6}$$

$$A = \frac{2}{a}\int_{-\frac{\pi}{6}}^{\frac{\pi}{6}} \sqrt{1 - \sin^2 u}\,a\cos u\,du$$

$$= \frac{2}{a}\int_{-\frac{\pi}{6}}^{\frac{\pi}{6}} a\cos^2 u\,du$$

$$= \frac{2}{a}\cdot a\int_{-\frac{\pi}{6}}^{\frac{\pi}{6}} \cos^2 u\,du$$

$$= 2\int_{-\frac{\pi}{6}}^{\frac{\pi}{6}} \cos^2 u\,du$$

$$= 2\left(\frac{\pi}{6} - \frac{1}{2}\sin\frac{\pi}{3}\right)$$

$$= 2\left(\frac{\pi}{6} - \frac{1}{2}\cdot\frac{\sqrt{3}}{2}\right)$$

$$= 2\left(\frac{\pi}{6} - \frac{\sqrt{3}}{4}\right)$$

$$= \frac{\pi}{3} - \frac{\sqrt{3}}{2}$$

$$\Rightarrow \rho\left(|X| < \frac{a}{2}\right) = \frac{0.8}{\pi}$$

Problem 5: Inequalities

(a)

1.

$$\mathbb{P}(X \geq k) \leq \frac{\mathbb{E}[X]}{k} = \frac{5}{10} = 0.5$$

2.

The Markov inequality provides an upper limit on $P(X \geq a)$ because it relies solely on the mean $\mathbb{E}[X]$ without considering the detailed distribution of X. It represents a worst-case estimate, making it cautious and potentially overestimating the probability. As a result, the actual value of $P(X \geq a)$ might be significantly smaller but will never exceed this bound.

3.

For example, the distribution of X might highly skewed, with a long tail on the right side. In such cases, the bound provided by the Markov inequality might be significantly higher than the actual probability of $X \geq a$. This discrepancy arises because the Markov inequality does not consider the detailed distribution of X and only relies on the mean.

(b) Suppose X is a random variable with $\mathbb{E}[X] = 20$ and $\mathrm{Var}(X) = 36$.
1. Use the Chebyshev inequality to find an upper bound for $P(|X - 20| \geq 12)$.

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

$$P(|X - 20| \geq k \cdot 6) \leq \frac{1}{k^2}$$

As $6k = 12 \Rightarrow k = 2$:

$$P(|X - 20| \geq 12) \leq \frac{1}{2^2} = \frac{1}{4}$$

2.

The Chebyshev inequality says that the probability of a random variable deviating from its mean by more than k standard deviations is at most $\frac{1}{k^2}$. Take for example, the normal distribution, where 95% of the data lies within 2 standard deviations of the mean. By choosing $k = 2$, we are ensuring that the probability of X deviating from its mean (being out of its reach) by more than 2 standard deviations is at most $\frac{1}{4}$.

3.

Since the Chebyshev inequality, just the Markov one, doesn't specify anything about the shape and skewness of the distribution of X, it provides a worst-case estimate. Therefore, the actual probability of $|X - 20| \geq 12$ might be significantly lower than the upper bound provided by the Chebyshev inequality. You can draw a distribution to demonstrate this (optional)

(c)
1.

$$\mathbb{E}[XY] \leq \sqrt{\mathbb{E}[X^2]}\sqrt{\mathbb{E}[Y^2]}$$

$$\mathbb{E}[XY] \leq \sqrt{8}\sqrt{12} = 4\sqrt{3}$$

2.

One can write:

$$Cov(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] \Rightarrow \mathbb{E}[XY] = Cov(X,Y) + \mathbb{E}[X]\mathbb{E}[Y] \leq \sqrt{\mathbb{E}[X^2]}\sqrt{\mathbb{E}[Y^2]}$$

Now since the inequality doesn't assume anything about the covariance of the two random variables, if it were to be assumed X and Y are independent, the covariance would be 0, and $\mathbb{E}[XY]$ would be equal to $\mathbb{E}[X]\mathbb{E}[Y]$. In this case, the Cauchy-Schwarz inequality would provide a bound that is not tight.

There may be other examples where the bound provided by the Cauchy-Schwarz inequality is not tight.

Problem 6: Sort, sort, sort

Note: Your implementation might be different from the one provided here, based on the randomness, hence the results might vary.

a Applying the said algorithms on a list-size of 1000 items, 100 times, the duration distribution is shown in Figure 3.

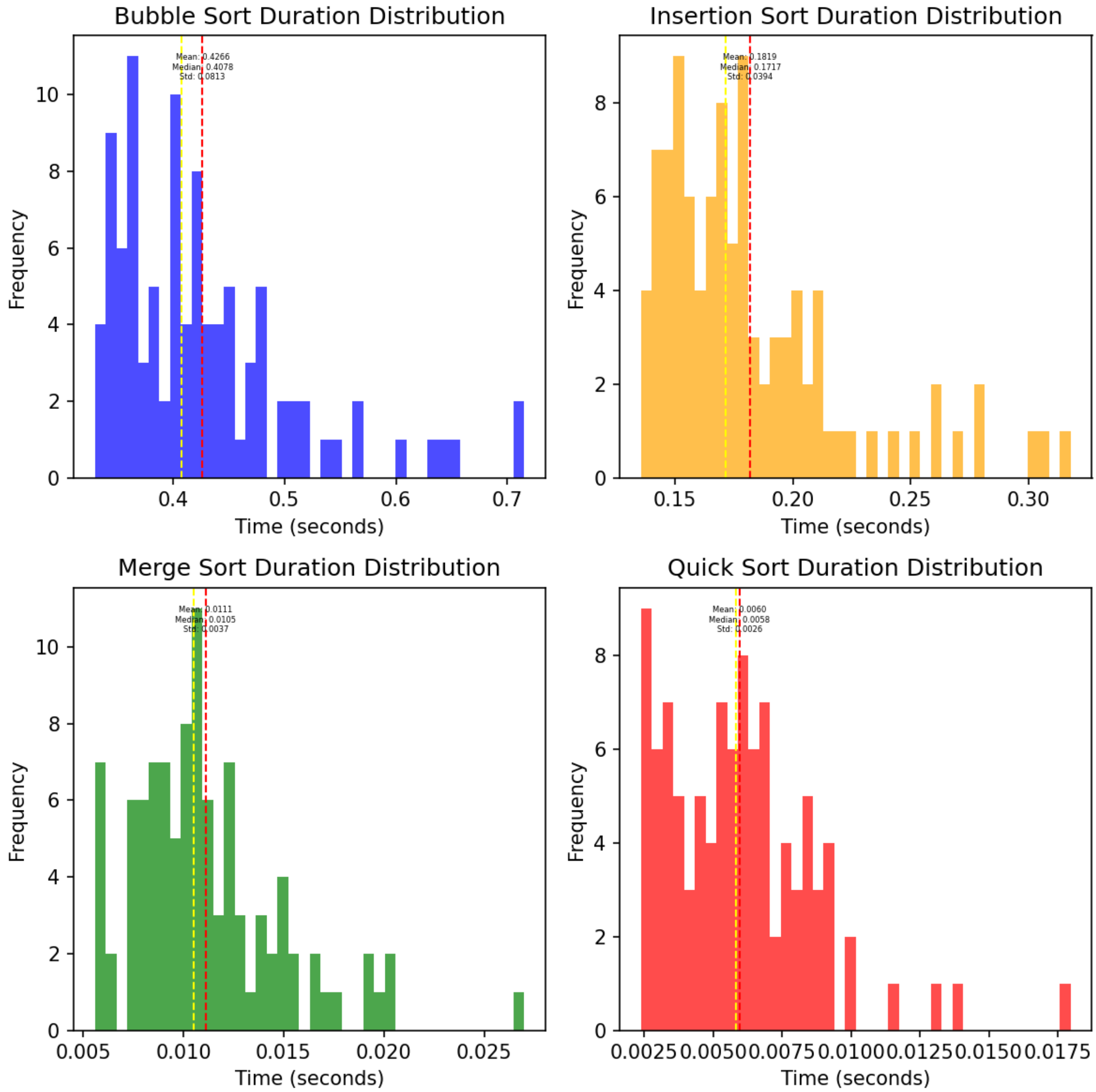

Figure 3: Duration Distribution of Sorting Algorithms

Sorting the algorithms by mean duration, the order is as follows:

1. Quick Sort: 0.006 seconds
2. Merge Sort: 0.0111 seconds
3. Insertion Sort: 0.1819 seconds
4. Bubble Sort: 0.4266 seconds

Also, by looking at the standard deviation, the quicksort algorithm is the most consistent, having the least standard deviation and the bubble sort algorithm is the most volatile, having the highest standard deviation.

b.

Calculating Skewness and Kurtosis for the duration of the sorting algorithms, Table 1 is obtained.

Algorithm	**Kurtosis**	**Skewness**
Bubble Sort	2.448	1.535
Insertion Sort	2.122	1.537
Merge Sort	2.427	1.250
Quick Sort	3.467	1.360

Table 1: Kurtosis and Skewness values for sorting algorithms.

Insertion sort has the highest skewness and Merge sort has the least skewness.

Kurtosis is a measure of the "tailedness" of the probability distribution of a real-valued random variable. It describes the shape of the distribution. A high kurtosis distribution has a sharper peak and longer, fatter tails, while a low kurtosis distribution has a more rounded peak and shorter, thinner tails.

Looking at Table 1, Bubblesort and Insertion sort have higher kurtosis values, which means they have longer tails, hence more outliers; what this means is that these algorithms are more volatile and have a higher probability of having outliers in their duration distribution. On the other hand, Quicksort and Mergesort have lower kurtosis values, which means they have shorter tails, hence less outliers; what this means is that these algorithms are more consistent and have a lower probability of having outliers in their duration distribution.

c.

Boxplots of the durations of the sorting algorithms are shown in Figure 4.

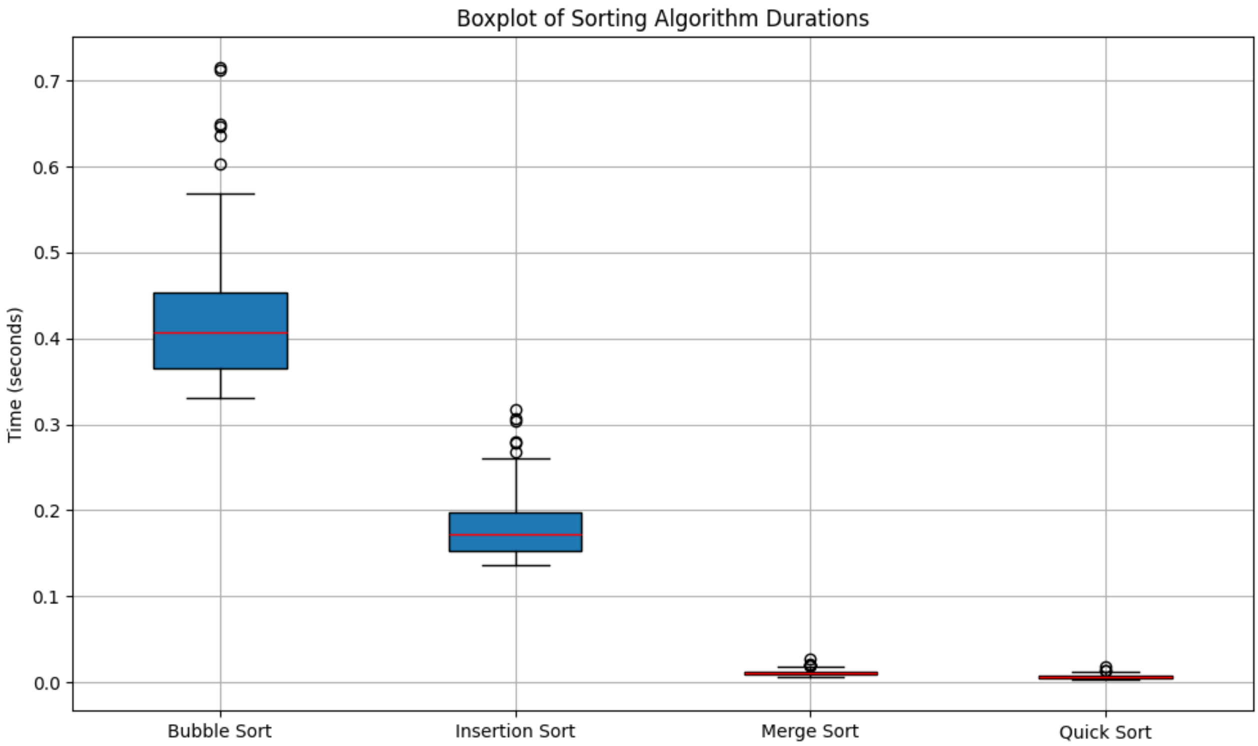

Figure 4: Boxplot of Sorting Algorithms

By looking at Figure 4, Insertion, Bubble and Merge sort have more outliers compared to Quicksort. The number of outliers for each algorithm is shown in Table 2.

Sorting Algorithm	**Number of Outliers**
Insertion Sort	6
Bubble Sort	6
Merge Sort	6
Quick Sort	3

Table 2: Number of Outliers for Different Sorting Algorithms

It is expected that if the sample size (list length) is to be increased, the difference between the number of outliers for each algorithm would be more visible, as bubble sort and insertion sort would have more distinct number of outliers than merge sort .

Problem 7: Cookie Monster

Note: Your implementation may vary based on the parameters loosely defined in this question, so the results may not be exactly the same as the ones shown in the solution.

a.

Applying the 3 noise types and 6 noise levels to the image, we get the following results:

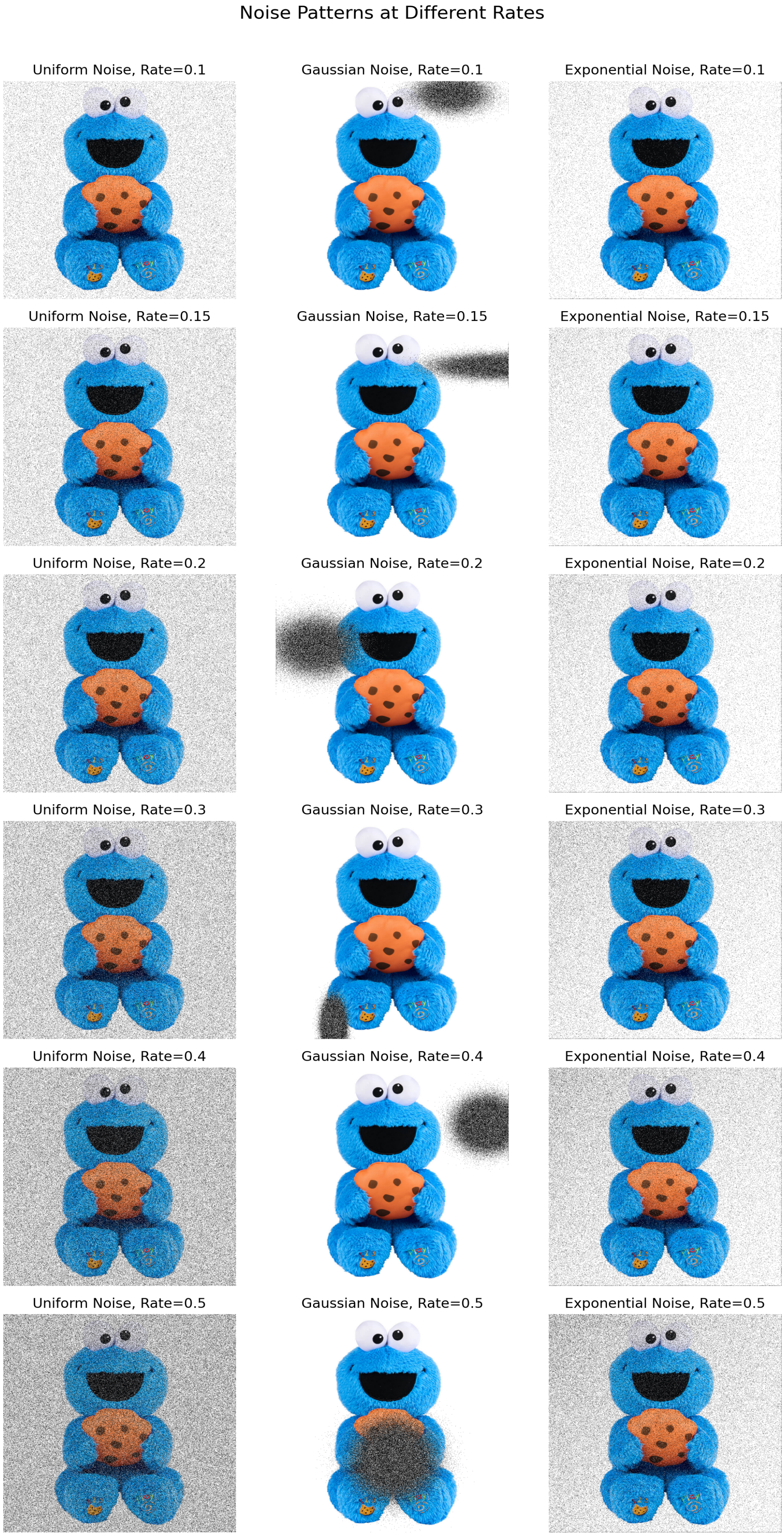

Figure 5: Cookie Monster

---

b.

What was done in section a is now going to be done a 100 times for each noise level and type. counting the corrupted pixels in each level and type are shown in Table 3.

Corruption Level	**Uniform**	**Gaussian**	**Exponential**
0.2	81	32	0
0.3	100	38	0
0.4	100	48	22
0.5	100	41	100

Table 3: Final corrupted counts for different noise types at various corruption levels.

Since the uniform noise makes the image uniformly corrupted, if the noise ratio was to be above 20 percent, then the face would be corrupted with ratios greater than 20, which is what can be seen in the table; all of the images in uniform noise are corrupted with a ratio of 100 percent. The Gaussian noise, on the other hand, is more random and does not necessarily corrupt the image uniformly. Depending on the mean and variance of the gaussian, the corruption rate can vary. The exponential noise, with a fixed $\lambda = 1$ parameter, is more dense near the left-top corner and less dense near the right-bottom corner, and because the face is almost center, the noise rate near the face is almost sparse, which is why the corruption rate is small for the first 3 levels and 100 percent for the last level. In this particular example (which the answer may be different with respect to your setup), the uniform noise is the most destructive noise type, and the exponential noise is the least destructive one.

---

c.

Applying the median filter to the corrupted images, we get the results in Figure 6:

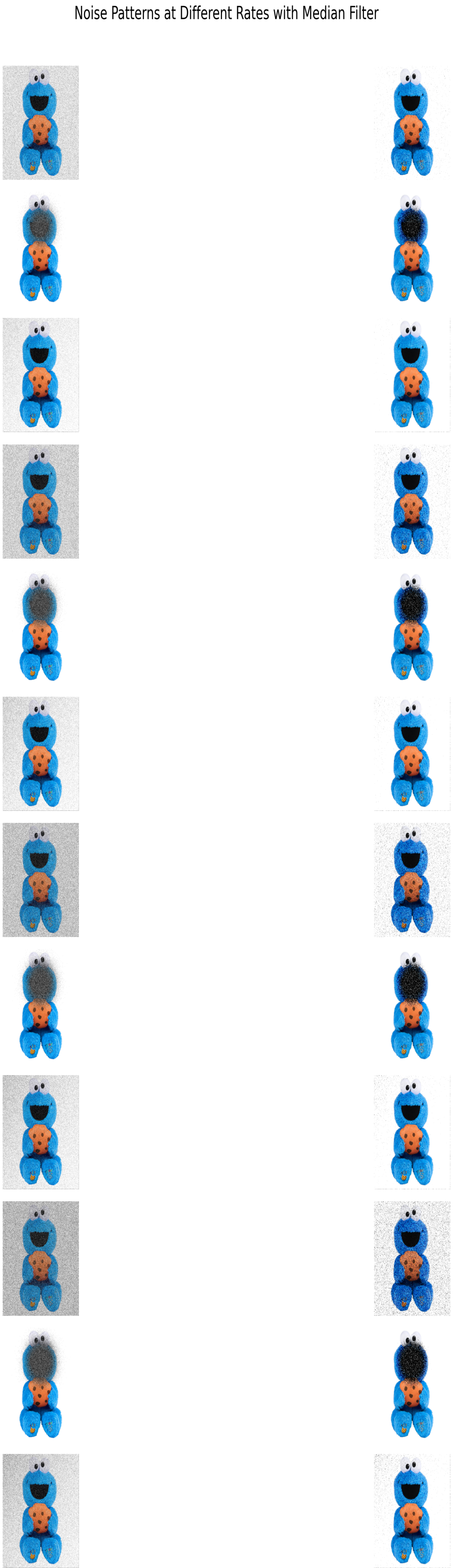

Figure 6: Median Filtered Images (Left: original image, right: median filtered image)

The images filtered from the uniform and exponential distributions are recovered well, but the image filtered from the Gaussian noise is not recovered well. The reason for the bad recovery can be explained by the low variance of the Gaussian noise. If there are a lot of noisy pixels in adjacency to each other, the median filter will not be able to recover the image well, as the median of the noisy pixels are noisy themselves. This is also visible in the exponential noise in higher noise rates where the top left corner of the picture is still a bit noisy. So the image recovery using the median filter depends on the density of the noise (variance).

Problem 8: Conditional Independence

(a)

$P(D_1 = 6) = \frac{1}{6}, P(D_2 = 5) = \frac{1}{6}$
$P(D_1 = 6, D_2 = 5) = \frac{1}{36}$
$P(D_1 = 6, D_2 = 5) = P(D_1 = 6) \times P(D_2 = 5) \Rightarrow D_1 = 6 \perp D_2 = 5$

(b)

$D_1 + D_2 > 10 = \{(5,6), (6,5), (6,6)\}$

$$\begin{cases} P(D_1 = 6, D_2 = 5 \mid D_1 + D_2 > 10) = \dfrac{1}{3}, \\ P(D_1 = 6 \mid D_1 + D_2 > 10) = \dfrac{2}{3}, \\ P(D_2 = 5 \mid D_1 + D_2 > 10) = \dfrac{1}{3}. \end{cases}$$

$$P(D_1 = 6, D_2 = 5|D_1 + D_2 > 10) \neq P(D_1 = 6|D_1 + D_2 > 10) \times P(D_2 = 5|D_1 + D_2 > 10)$$

$$\Rightarrow D_1 = 6|D_1 + D_2 > 10 \not\perp D_2 = 5|D_1 + D_2 > 10$$

(c)

$$A \subseteq B \subseteq C$$

If $A \perp C|B$:

$$\begin{cases} P(A \mid B) = \dfrac{P(A)}{P(B)}, \\ P(C \mid B) = 1, \\ P(A \cap C \mid B) = \dfrac{P(A)}{P(B)}. \end{cases}$$

$$P(A|B)P(C|B) = P(A \cap C|B) \Rightarrow A \perp C|B \checkmark$$

If $B \perp C|A$:

$$\begin{cases} P(B \mid A) = 1, \\ P(C \mid A) = 1, \\ P(B \cap C \mid A) = 1. \end{cases}$$

$$P(B \mid A)P(C \mid A) = P(B \cap C \mid A) \Rightarrow B \perp C|A\checkmark$$

(d)

If we consider $A \subseteq B \subseteq C$ as a Bayesian network head-to-tail structure, when B is observed, A and C become independent.

Problem 9: Limit Distributions

(a)

$X \sim Poisson(\lambda), P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$

$$\sum_{k=0}^{\infty} P(x = k) = \sum_{k=0}^{\infty} \frac{\lambda^k e^{-\lambda}}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda} = 1$$

(b) $Y \sim Binomial(n, p), P(Y = k) = \binom{n}{k} p^k (1-p)^{n-k}$

$$\lim_{n\to\infty, np\to\lambda} \frac{n!}{(n-k)!k!} p^k (1-p)^{n-k} = \frac{1}{k!} \lim_{n\to\infty, np\to\lambda} \frac{n^n}{n^{n-k}} p^k (1-p)^{n-k} =$$

$$\frac{1}{k!} \lim_{n\to\infty, np\to\lambda} (np)^k (1-p)^{n-k}$$

$$= \frac{\lambda^k}{k!} \lim_{n\to\infty, np\to\lambda} (1-p)^{n-k} = \frac{\lambda^k \cdot e^{\lambda}}{k!}$$

Problem 10: Memorylessness

$$P(X > k) = 1 - \sum_{x=0}^{k} (1-p)^x p = (1-p)^k$$

$$P(X > n + k - 1 \mid X > n - 1) =$$

$$\frac{P(X > n + k - 1 \cap X > n - 1)}{P(X > n - 1)} =$$

$$\frac{P(X > n + k - 1)}{P(X > n - 1)} =$$

$$\frac{(1-p)^{n+k-1}}{(1-p)^{n-1}} =$$
$$(1-p)^k$$

Problem 11: Transformation of Random Variables

(a)

$Y = g(X)$

$$\begin{aligned} F_Y(y) = P(Y \leq y) = P(g(X) \leq y) \\ = P(X \leq g^{-1}(y)) \\ = F_X(g^{-1}(y)) \end{aligned}$$

$$f_Y(y) = \frac{d}{dy} F_Y(y)$$
$$= \frac{d}{dy} F_X(g^{-1}(y))$$
$$= \frac{dF_X(g^{-1}(y))}{dg^{-1}(y)} \times \frac{dg^{-1}(y)}{dy}$$
$$= f_X(g^{-1}(y)) \times \frac{dx}{dy} =$$
$$\frac{f_X(g^{-1}(y))}{g'(g^{-1}(y))}$$

(b)

$$E_Y[Y] = \int y f_x(y) dy =$$
$$\int y \frac{f_X(g^{-1}(y))}{g'(g^{-1}(y))} dy =$$
$$\int y \frac{f_X(x)}{g'(x)} dx =$$
$$\int g(x) \frac{f_X(x)}{\frac{dy}{dx}} dy =$$
$$\int g(x) f_X(x) dx = E_X[g(X)] =$$
$$E_X[Y]$$

Problem 12: What are AutoEncoders doing here?

$$E_Z[Z] = E_X[E_{Z \vee X}[Z]], Z \mid X \sim exp(En(x))$$

$$E_{Z \vee X}[Z] = \frac{1}{En(x)}$$

$$E_Z[Z] = E_X[\frac{1}{En(x)}] =$$

$$\int \frac{1}{En(x)} f_X(x)dx$$

Problem 13: Plants vs. Enemies

Case 1: The enemy is behind the line $x = c$:

$$\text{Farming area} = \frac{(1-c)^2}{2}, \quad P(\text{enemy not in the farming area}) = P(X < c).$$

As $X \sim U[0,1]$, we have:

$$P(\text{enemy not in the farming area}) = c.$$

—

Case 2: The enemy is behind the line $y = d$:

$$\text{Farming area} = \frac{(1-d)^2}{2}, \quad P(\text{enemy not in the farming area}) = P(Y < d).$$

Since $Y \mid X \sim U[0, 1-X]$, the marginal density of Y is:

$$f_Y(y) = \int_0^{1-y} f_{X,Y}(x,y)\, dx = \int_0^{1-y} f_{X,Y}(y,x) f_X(x)\, dx.$$

Substituting $f_X(x) = \frac{1}{1-x}$, we get:

$$f_Y(y) = \int_0^{1-y} \frac{1}{1-x}\, dx = -\ln(y).$$

Thus:

$$P(\text{enemy not in the farming area}) = P(Y < d) = \int_0^d -\ln(y)\, dy.$$

Solving the integral:

$$P(Y < d) = -d\ln(d) + d.$$

—

Conclusion:

Since the farming area and the probability of not facing the enemy in the farm are positively correlated in both cases, it is not possible to choose a line in either case as an optimal bound.

In such cases where there is no min-max game, the agent tries to maximize the expected successfully farmed area and compare it between the x-axis and y-axis.

$$f_X(x) = 1, f_{Y|X}(y \mid x) = \frac{1}{1-x} \Rightarrow f_Y(y) = -\ln(y)$$

$$E[\text{farming successfully along x-axis}] = P(enemy \leq C) \times \frac{(1-C)^2}{2} =$$

$$\int_0^1 \frac{C(1-C)^2}{2} dc =$$

$$\frac{1}{2} \int_0^1 (C^3 + c - 2C^2) dc =$$

$$\frac{1}{2}(\frac{1}{4}C^3 - \frac{2}{3}C^3 + \frac{1}{2}C^2)\Big|_0^1 =$$

$$\frac{1}{2} \cdot (\frac{1}{4} - \frac{2}{3} + \frac{1}{2}) = \frac{1}{24}$$

$$E[\text{farming successfully along y-axis}] = P(enemy \leq d) \times \frac{(1-d)^2}{2} =$$

$$\int_0^1 (\int_0^d -\ln(y) dy) \frac{(1-d)^2}{2} dd =$$

$$\frac{1}{12} + \frac{1}{4} - \frac{2}{9} + \frac{1}{12} = \frac{7}{36}$$

Problem 14: Random Genes (Optional)

The Beta distribution models the k-th order statistic from a uniform distribution, representing the k-th largest value in sorted samples.

Key Details

1. Shape Parameters:
   - $\alpha = k$: Represents the rank of the value (how many values are smaller).
   - $\beta = n - k + 1$: Represents the remaining samples (how many values are larger).
2. Uniform Distribution Basis:
   - The uniform distribution ensures equal probability for any value within $[0, 1]$.
   - The order statistic redistributes this uniformity, introducing skewness in the Beta distribution based on the rank k.
3. Relationship Between k, n, and Skewness:
   - When k is close to n:
     The k-th largest value is close to 1, and the Beta distribution is skewed towards 1.
   - When k is close to 1:
     The k-th largest value is close to 0, and the Beta distribution is skewed towards 0.

- When $k \approx \frac{n}{2}$:
  The Beta distribution is more symmetric.

The result of the above code execution is plotted in Figure 7

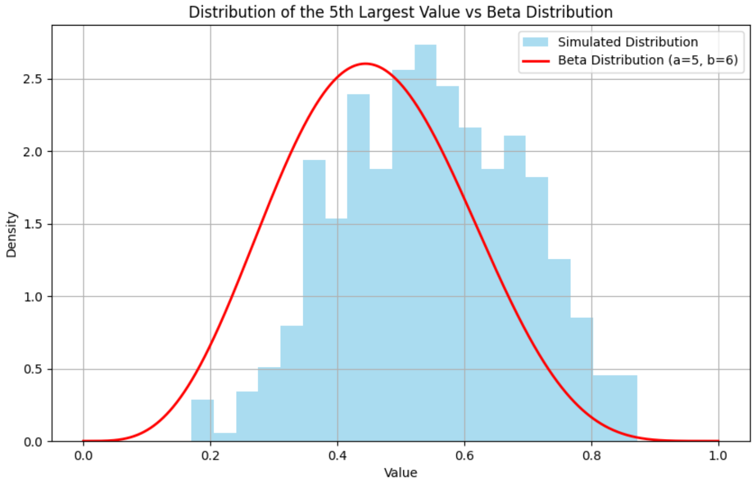

Figure 7: Distribution of the 5th Largest Value vs Beta Distribution