# CSCI 410: Modeling and Simulation

# Written Assignment 3 Solutions

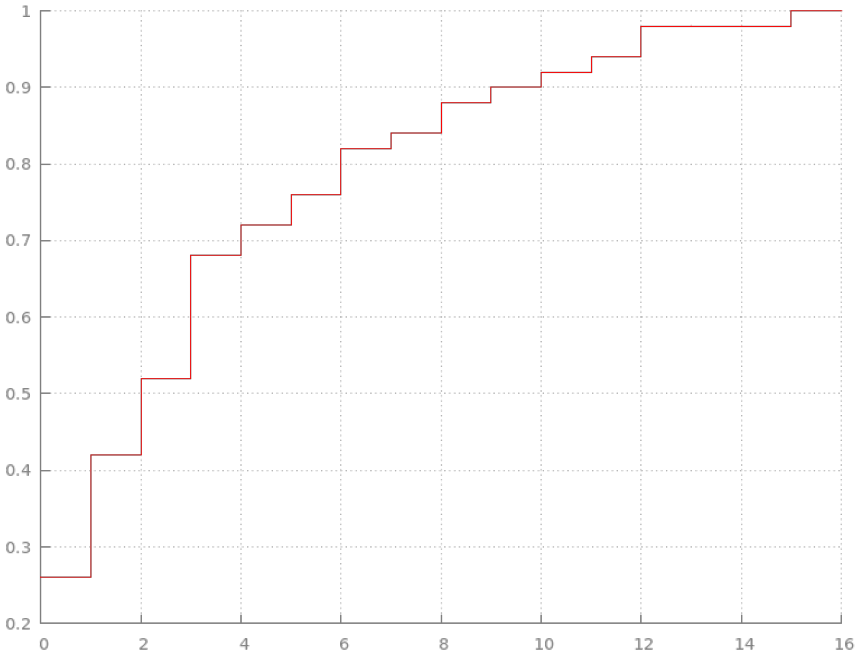
1. The following data are drawn from a discrete distribution.
   1, 11, 3, 1, 2, 4, 2, 1, 1, 4
   1, 7, 2, 10, 4, 2, 2, 4, 1, 13
   2, 4, 4, 9, 1, 12, 7, 3, 9, 4
   1, 2, 3, 1, 3, 8, 16, 2, 5, 6
   4, 1, 1, 3, 13, 1, 5, 1, 7, 6
   (1) Define an empirical distribution and plot it. **(10 pts)**
   (2) Plot its histogram and calculate the summary statistics as shown in Section 6.4. Use the histogram and the statistics to hypothesize a distribution for the sample data. **(10 pts)**

Answer:

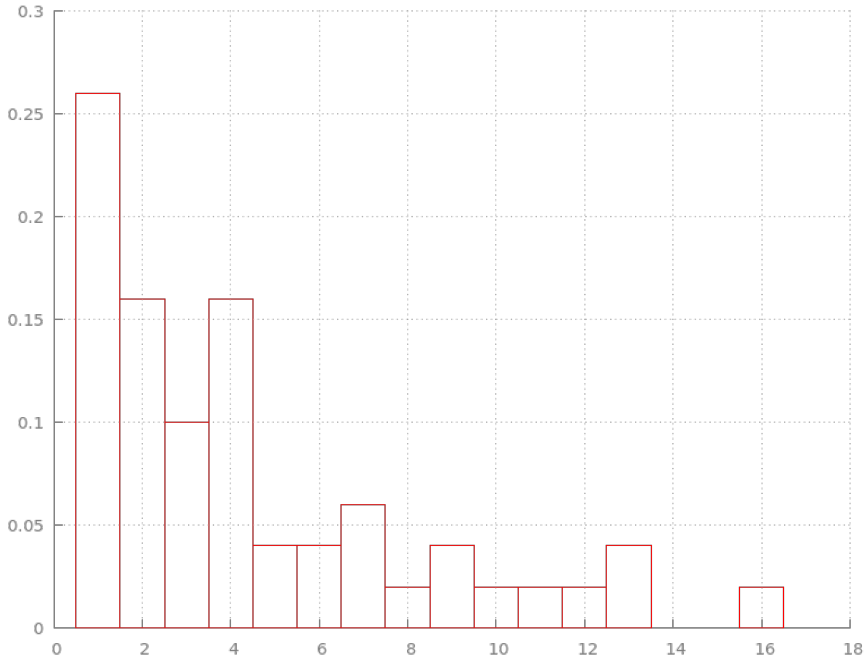
(1) Empirical distribution:

The minimum value of the sample value is 1, the maximum is 16. Therefore, the empirical distribution only specifies the probability of the values in [1, 16].

P(X = 1) = 13/50 = 0.26
P(X = 2) = 8/50 = 4/25 = 0.16
P(X = 3) = 5/50 = 1/10 = 0.1
P(X = 4) = 8/50 = 4/25 = 0.16
P(X = 5) = 2/50 = 1/25 = 0.04
P(X = 6) = 2/50 = 1/25 = 0.04
P(X = 7) = 3/50 = 0.06
P(X = 8) = 1/50 = 0.02
P(X = 9) = 2/50 = 1/25 = 0.04
P(X = 10) = 1/50 = 0.02
P(X = 11) = 1/50 = 0.02
P(X = 12) = 1/50 = 0.02
P(X = 13) = 2/50 = 1/25 = 0.04
P(X = 14) = 0
P(X = 15) = 0
P(X = 16) = 1/50 = 0.02




(2) Histogram:




Summary statistics:

| | |
|---|---|
| Minimum | 1 |
| Maximum | 16 |

| | |
|---|---|
| Mean | 4.4 |
| Median | 3 |
| Variance | 14.36734694 |
| Lexis ratio | 3.265306122 |
| Skewness | 1.328012392 |

Since the mean is larger than the median, the distribution is skewed to the right. The histogram also shows that. In addition, because the lexis ratio is larger than 1, the distribution is probably a negatively binomial distribution or geometric distribution (which is a special case of the negative binomial).

2. The following data are drawn from a continuous distribution.

2.0, 6.7, 5.7, 1.8, 2.0, 5.6, 3.3, 1.8, 1.9, 1.6,
5.1, 4.8, 4.0, 6.3, 6.6, 8.0, 2.4, 1.0, 0.6, 1.5,
7.0, 7.0, 1.4, 5.1, 1.6, 5.2, 7.0, 5.6, 5.3, 4.7,
7.0, 4.1, 6.7, 9.7, 1.5, 3.7, 6.4, 1.7, 0.4, 2.1,
5.7, 9.4, 8.9, 2.1, 4.1, 1.1, 3.8, 3.0, 2.5, 2.7

Group the number into 6 equal-width intervals.

(1) Use the group data to define an empirical distribution and plot it. **(10 pts)**
(2) Use the empirical distribution to calculate the probability of x ≤ 5.0. **(5 pts)**
(3) Plot its histogram and calculate the summary statistics as shown in Section 6.4. Use the histogram and the statistics to hypothesize a distribution for the sample data. **(10 pts)**

Answer:

(1) The width of the data range is 9.7 – 0.4 = 9.3. Each interval is of width 9.3 / 6 = 1.55.

The intervals are [0.4, 1.95), [1.95, 3,5), [3.5, 5.05), [5.05, 6.6), [6.6, 8.15), [8.15, 9.7].

The empirical distribution is: (Refer to the attached Excel sheet for values of $G(a_j)$'s.)

$F(x) = 0 \quad x < 0.4$

$F(x) = 0 + 0.168(x - 0.4) = 0.168x - 0.0671 \quad x \in [0.4, 1.95)$

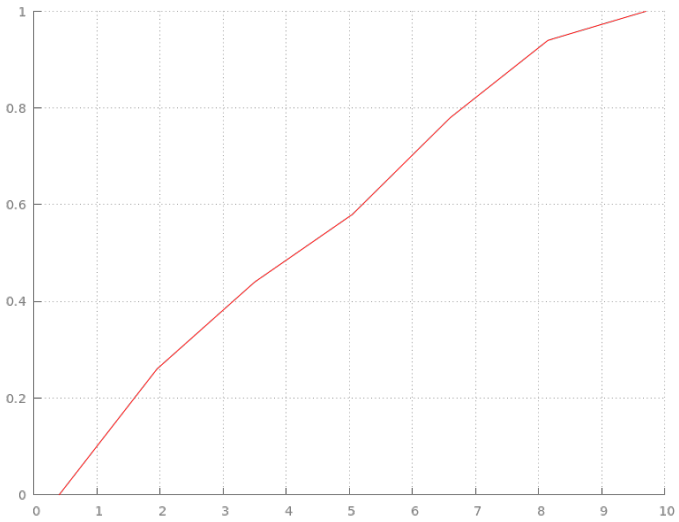
$F(x) = 0.26 + 0.116(x - 1.95) = 0.116x + 0.0335 \quad x \in [1.95,3.5)$

$F(x) = 0.44 + 0.09(x - 3.5) = 0.09x + 0.1239 \quad x \in [3.5,5.05)$

$F(x) = 0.58 + 0.129(x - 5.05) = 0.129x - 0.0716 \quad x \in [5.05,6.6)$

$F(x) = 0.78 + 0.103(x - 6.6) = 0.103x + 0.0987 \quad x \in [6.6,8.15)$

$F(x) = 0.94 + 0.039(x - 8.15) = 0.039x + 0.6245 \quad x \in [8.15,9.7)$

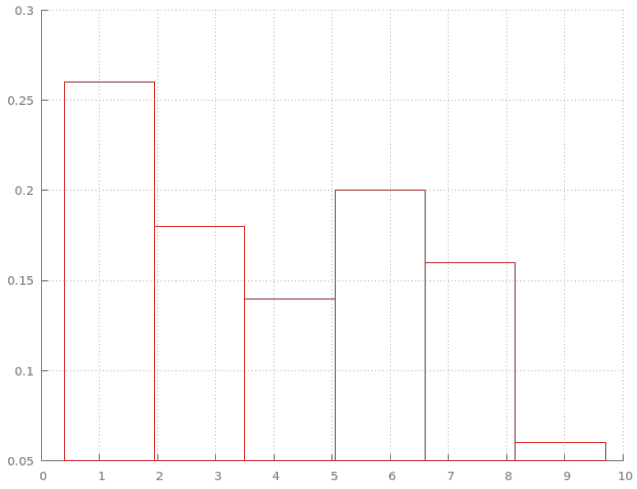



(2) Since 3.5 < 5 < 5.05, we use the following formula in the empirical distribution to calculate the probability.

$F(x) = 0.44 + 0.09(x - 3.5) = 0.09x + 0.1239 \quad x \in [3.5,5.05)$

$P(X \leq 5.0) = F(5) = 0.44 + 0.09(5 - 3.5) = 0.575$

(3) Histogram:




Summary statistics:

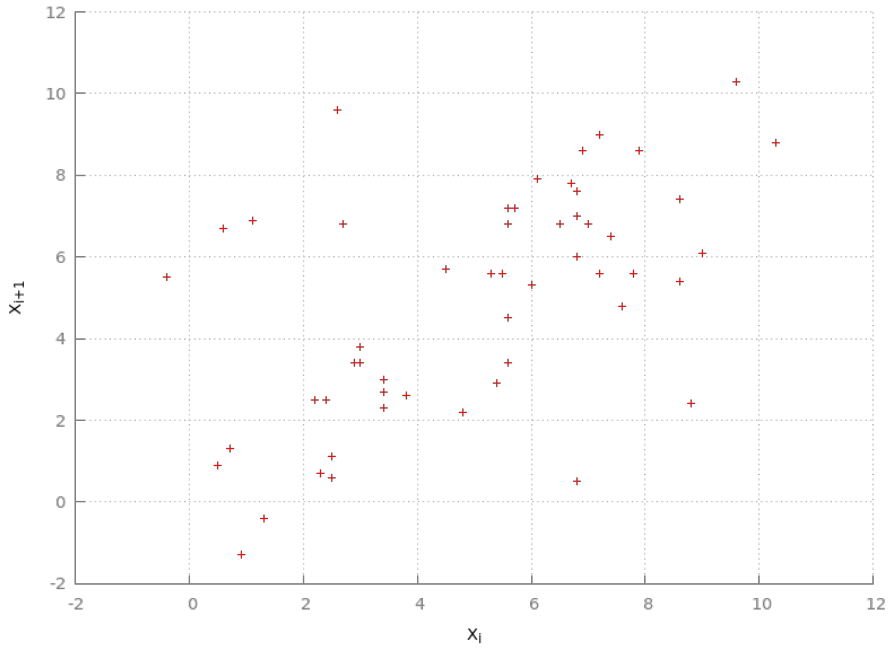
| | |
|---|---|
| Minimum | 0.4 |
| Maximum | 9.7 |
| Mean | 4.184 |
| Median | 4.05 |
| Variance | 6.15075918 |
| Coefficient of variation | 0.59275153 |
| Skewness | 0.38562824 |

Since the median is a little less than the mean, the distribution is skewed to the right a little bit. The coefficient of variation is less than 1. The distribution could be a Weibull or Gamma distribution with the shape parameter α > 1.

3. **(10 pts)** Use a scatter diagram to verify the independence among the following samples.

   3.0, 3.4, 2.7, 6.8, 7.6, 4.8, 2.2, 2.5, 0.6,
   6.7, 7.8, 5.6, 7.2, 9.0, 6.1, 7.9, 8.6, 7.4,
   6.5, 6.8, 6.0, 5.3, 5.6, 3.4, 3.0, 3.8, 2.6,
   9.6, 10.3, 8.8, 2.4, 2.5, 1.1, 6.9, 8.6, 5.4,
   2.9, 3.4, 2.3, 0.7, 1.3, -0.4, 5.5, 5.6, 4.5,
   5.7, 7.2, 5.6, 6.8, 7.0, 6.8, 0.5, 0.9, -1.3

Answer:




In the diagram, many points are closed to a line of positive slope. That means in the samples, a small value tends to be followed by a small value, and large value followed by large value. Therefore, independence exists among the samples.