Statistical Method Homework 2

Kai Chien Chen(H24101222)

September 29, 2024

1 Question 1

1.1 Question

You can use theoretical method or simulation method to solve the following questions. The probability of having typhoid fever after drinking from a well water in a villager is 0.45. If ten villagers drank out of the well, what is the probability that:

(a) one villager will have typhoid fever;
(b) two villagers will have typhoid fever;
(c) at most two villagers will have typhoid fever;
(d) at least two villagers will have typhoid fever; and
(e) hence, calculate the number of villagers that expected to have typhoid fever.

1.2 Solution

(a) This is the binomial distribution example. Now we know that the pmf of binomial distribution is: $p(X = x) = \binom{n}{x} p^x(1-p)^{n-x}$ for x = 0,1,...,n

and we know $n = 10$, $p = 0.45$, and $x = 1$,

therefore $p(X = 1) = \binom{10}{1} 0.45^1 0.55^9 = \mathbf{0.0207}$

(b) This is the same concept! we use binomial distribution with $n = 10$, $p = 0.45$, and $x = 2$. The answer is:

$p(X = 2) = \binom{10}{2} 0.45^2 0.55^8 = \mathbf{0.0763}$

(c) It still binomial! But the question ask how to know **at most** to villagers. This mean that we need to calculate:

$p(X \leq 2) = p(X = 0) + p(X = 1) + p(X = 2)$ So the answer is $p(X \leq 2) =$ **0.0996**

(d) It is the same concept as (c). But it a little different because it require **at least** two villagers. Therefore, this mean we need to calculate $p(X \geq 2) = 1 - p(X < 2) =$ **0.9767**

(e) This question mean that we need to know the expected number of random variable X, and we know the expected number of the binomial distribution is $E(X) = \sum_0^n x \binom{n}{x} p^x (1-p)^{n-x} = np$ and we know $n = 10$ and $p = 0.45$. Therefore, the expected value is **4.5**.

2 Question 2

2.1 Question

Give the sample space (domain), probability density (mass) function, mean, and variance of the following distributions:

(a) Poisson distribution with rate $\lambda = 3$.

(b) Geometric distribution with probability $p = 1/3$.

(c) Normal distribution with $\mu = 0$ and $\sigma = 1$.

(d) Exponential distribution with rate $\lambda = 2$.

(e) Laplace distribution with location parameter $\mu = 0$ and **scale parameter** $b = 1/2$.

(f) Gamma distribution with shape parameter $\alpha = 10$ and **rate parameter** $\lambda = 1/2$.

(g) Chi-square distribution with degree of freedom 20.

2.2 Solution

(a) Poisson distribution

- Sample space

$$x = 0, 1, 2, 3, ...$$

- Probability mass function

$$p(X = x) = \tfrac{e^{-\lambda}\lambda^x}{x!} = \tfrac{e^{-3}3^x}{x!}$$

- Mean

$$E(X) = \lambda = 3$$

- Variance

$$Var(X) = \lambda = 3$$

(b) Geometric distribution

- Sample space

$$x = 1, 2, 3, ...$$

- Probability mass function

$$(1-p)^{x-1}p = \frac{2}{3}^{x-1}\frac{1}{3}$$

- Mean

$$E(X) = \frac{1}{p} = 3$$

- Variance

$$Var(X) = \frac{1-p}{p} = 2$$

(c) Normal distribution

- domain

$$x \in R$$

- Probability density function

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

- Mean

$$E(X) = \mu = 0$$

- Variance

$$Var(X) = \sigma^2 = 1$$

(d) Exponential distribution

- Domain

$$x \in [0, \infty)$$

- Probability density function

$$f(x) = \lambda e^{-\lambda x} = 2e^{-2x}$$

- Mean

$$E(X) = \frac{1}{\lambda} = \frac{1}{2}$$

- Variance

$$Var(X) = \frac{1}{\lambda^2} = \frac{1}{4}$$

(e) Laplace distribution

- Domain

$$x \in R$$

- Probability density function

$$f(x) = \frac{1}{2b} exp(-\frac{|x-\mu|}{b}) = exp(-2|x|)$$

- Mean

$$E(X) = \mu = 0$$

- Variance

$$Var(X) = 2b^2 = \frac{1}{2}$$

(f) Gamma distribution

- Domain

$$x \in (0, \infty)$$

- Probability density function

$$f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} = \frac{(1/2)^{10}}{\Gamma(10)} x^9 e^{-(1/2)x}$$

- Mean

$$E(X) = \frac{\alpha}{\lambda} = 20$$

- Variance

$$Var(X) = \frac{\alpha}{\lambda^2} = 40$$

(g) Chi-square distribution

- Domain

$$x \in [0, +\infty)$$

- Probability density function

$$f(x) = \frac{\frac{1}{2}^{n/2}}{\Gamma(n/2)} x^{\frac{n}{2}-1} e^{-\frac{1}{2}x} = \frac{\frac{1}{2}^{10}}{\Gamma(10)} x^9 e^{-\frac{1}{2}x}$$

- Mean

$$E(X) = n = 20$$

- Variance

$$Var(X) = 2n = 40$$

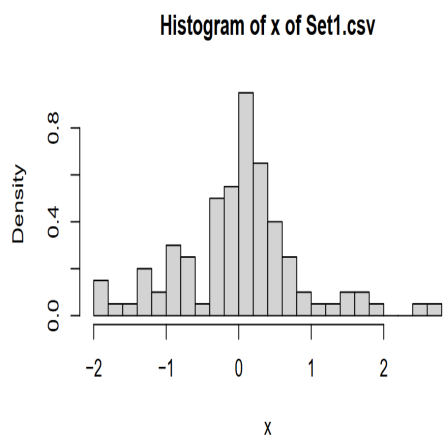
3 Question 3

3.1 Question

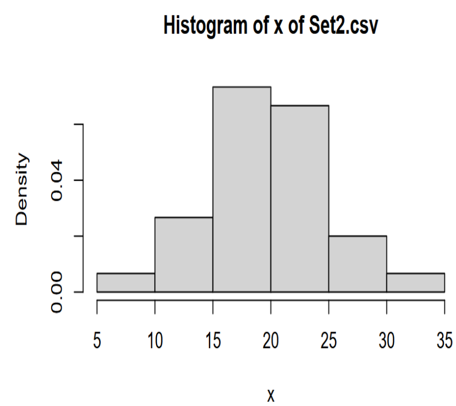
Given two datasets,
(a) Please provide the histograms of two datasets.
(b) For each dataset, add the probability density functions of the given distributions (c)-(g) in **Question 2** to the figures in **Question 3(a)**. Try to select suitable distributions to the data based on your observation.

3.2 Solution

(a) Figure 1 is a plot that we want to create.



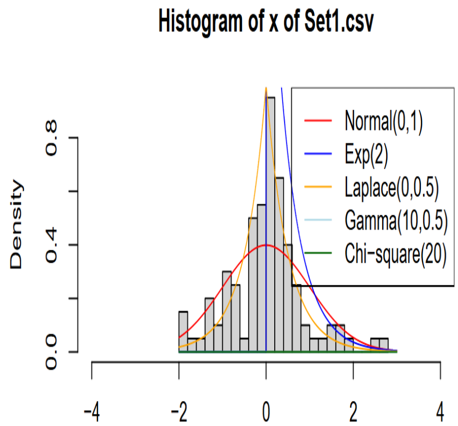
(a) Histogram of Set1



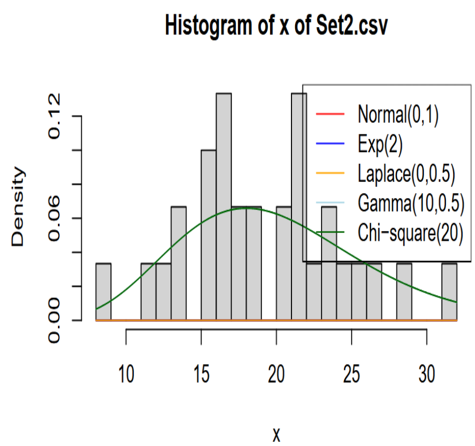
(b) Histogram of Set1

Figure 1: Two histograms which we want to observed

(b) Next, we want to add some pdf in two histograms.



(a) Histogram of Set1 and pdf



(b) Histogram of Set2 and pdf

Figure 2: Two histograms and pdf which we want to observed

Besides, we know $\chi^2(n) = \Gamma(\frac{n}{2}, \frac{1}{2})$, so $\chi^2(20) = \Gamma(10, 0.5)$. In addition, by the plot we can know that laplace distribution seems the most similar distribution compare to other distributions in Set1.csv.

However, Gamma distribution and Chi-square distribution seem to be suitable distribution to data, Set2.csv if you compare to other distribution.