ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

Assignment 1

Exercise 1

Consider the following events in the toss of a single dice: A: Observe an odd number, B: Observe an even number, C: Observe 1 or 2.

(a) Are A and B independent events?

(b) Are A and C independent events?

Answer

(a) Two events A and B are called **independent** if and only if

i. $\boldsymbol{P(A \cap B) = P(A)P(B)}$

or

ii. $\boldsymbol{P(A|B) = P(A)\ and\ P(B|A) = P(B)}$

(The two statements above are equivalent).

In a single dice, the probability of A is:

$\boldsymbol{P(A) = \frac{3}{6} = 0.5,}$ because there are three odd numbers among the six number of the sample space (a fair dice is assumed).

By definition, the probability of event A, given event B is:

$$\boldsymbol{P(A|B) = \frac{P(A \cap B)}{P(B)}}$$

But $\boldsymbol{P(A \cap B) = 0}$, since $\boldsymbol{A \cap B} = \emptyset$ (a result cannot be odd and even at the same time).

$\Leftrightarrow \boldsymbol{P(A) \neq P(A|B)}$

Consequently, $\boldsymbol{A}$ and $\boldsymbol{B}$ are not independent events.

(b) In a single dice, the probability of A is:

$$P(A) = \frac{3}{6} = 0.5$$

By definition, the probability of event A, given event C is:

$$P(A|C) = \frac{P(A \cap C)}{P(C)}$$

But $P(A \cap C) = \frac{1}{6}$, since only result 1 satisfies both events A and C.

In addition, $P(C) = \frac{2}{6} = \frac{1}{3}$

$$\Leftrightarrow P(A|C) = \frac{P(A \cap C)}{P(C)} = \frac{\frac{1}{6}}{\frac{1}{3}} = \frac{3}{6} = \frac{1}{2} = P(A)$$

Consequently**,** A and C are independent events.

Exercise 2

Males and females are observed to react differently to a given set of circumstances. It has been observed that 65% of the females react positively to these circumstances, whereas only 45% of males react positively. A group of 50 people, 35 female and 15 male, was subjected to these circumstances, and the subjects were asked to describe their reactions on a written questionnaire. A response picked at random from the 50 people was negative. What is the probability that it was that of a male?

Answer

We get the following events and their respective symbols:

X1: Positive Reaction

X2: Negative Reaction

M: Male

F: Female

The information given by the exercise is:

$$P(X1|F) = 0.65 \leftrightarrow P(X2|F) = 0.35$$

$$P(X1|M) = 0.45 \leftrightarrow P(X2|M) = 0.55$$

(Since ***X1*** and ***X2*** are complementary events).

$$P(M) = \frac{15}{50} = 0.30$$

$$P(F) = 0.70$$

The probability that is asked to be calculated is $\boldsymbol{P(M|X2)}$

Using Bayes' Theorem, we get:

$$\boldsymbol{P(M|X2)} = \frac{\boldsymbol{P(M \cap X2)}}{\boldsymbol{P(X2)}} = \frac{\boldsymbol{P(X2|M)P(M)}}{\boldsymbol{P(X2)}}$$

The missing value to calculate the probability asked, is ***P(X2).***

$$\boldsymbol{P(X2) = P(X2|F)P(F) + P(X2|M)P(M) = 0.35 * 0.70 + 0.55 * 0.30 = 0.41}$$

Consequently, the probability that the response was picked by a male is:

$$\boldsymbol{P(M|X2)} = \frac{\boldsymbol{0.55 * 0.3}}{\boldsymbol{0.41}} \sim \boldsymbol{0.40}$$

(This result and the following results are calculated by rounding to the closest second decimal digit).

Exercise 3

An industry is planning to produce 2 new products, A and B. The probability that product A will be successful given that at the same time a competitor will produce an item similar to A, is 0.4, whereas the probability that the product A will be successful given that no other company will produce similar to A item is 0.7. The probability that a competitor will present a similar to A product is 0.3. The probability that the product B is going to be successful given that the product A is successful is 0.6. Finally, the probability that product B is going to be successful is 0.4.

(a) What is the probability that product A is going to be successful *and* a similar to A product is going to be presented by the competitor?

(b) What is the probability that product A is going to be successful?

(c) What is the probability that a similar product to A is going to be presented by the competitor, given that the product A is successful?

Answer

a) We get the following events and their respective symbols:

AS: Product A will be successful

AU: Product A will be unsuccessful

BS: Product B will be successful

BU: Product B will be unsuccessful

ACP: Competitor produces product similar to A

ACN: Competitor does not produce product similar to A

BCP: Competitor Produces product similar to B

BCN: Competitor Produces product similar to B

The information given by the exercise is:

$\boldsymbol{P(AS|ACP) = 0.4}$

$\boldsymbol{P(AS|ACN) = 0.7}$

$\boldsymbol{P(ACP) = 0.3}$

$\boldsymbol{P(BS|AS) = 0.6}$

$\boldsymbol{P(BS) = 0.4}$

The probability that is asked to be calculated is $\boldsymbol{P(AS \cap ACP)}$**.**

Using Bayes' Theorem, we get:

$$\boldsymbol{P(AS \cap ACP) = P(AS|ACP)P(ACP) = 0.4 * 0.3 = 0.12}$$

b) The probability that is asked to be calculated is $\boldsymbol{P(AS)}$.

$$\boldsymbol{P(AS) = P(AS|ACP)P(ACP) + (P(AS|ACN)P(ACN)}$$

$$\Leftrightarrow \boldsymbol{P(AS) = P(AS|ACP)P(ACP) + (P(AS|ACN) * \big(1 - P(ACP)\big) =}$$

$$\boldsymbol{0.4 * 0.3 + 0.7 * (1 - 0.3) = 0.61}$$

c) The probability that is asked to be calculated is $\boldsymbol{P(ACP|AS)}$.

Using Bayes' Theorem, we get:

$$\boldsymbol{P(ACP|AS) = \frac{P(AS|ACP)P(ACP)}{P(AS)} = \frac{0.4 * 0.3}{0.61} \sim 0.20}$$

Exercise 4

The daily weight of biological waste (in tones) that a biological purification facility uses is a continuous random variable X that has a probability density function given by

$$\boldsymbol{f(x) = \begin{cases} c(4 - 2x) & , 1 \leq x \leq 2 \\ 0 & , otherwise \end{cases}}$$

a) What is the value of c?

b) Find the cumulative distribution function of X.

c) What is the probability that the weight of waste is 1.5 tones at most?

d) Compute the mean value of X.

Answer

a) From the definition of the probability density function:

$$\int_{-\propto}^{\propto} \boldsymbol{f(x)dx = 1}$$

$$\Leftrightarrow \int_1^2 \boldsymbol{c(4-2x)dx = 1}$$

$$\Leftrightarrow \boldsymbol{c[4x - x^2]_1^2 = 1}$$

$$\Leftrightarrow \boldsymbol{c[4*2 - 2^2 - 4*1 - (-1^2)] = 1}$$

$$\Leftrightarrow \boldsymbol{c = 1}$$

Consequently, the probability density function is $\boldsymbol{f(x)} = \begin{cases} \boldsymbol{(4-2x)} & \boldsymbol{, 1 \leq x \leq 2} \\ \boldsymbol{0} & \boldsymbol{, otherwise} \end{cases}$

b) $\boldsymbol{F(x) = P(X \leq x) = \int_{-\propto}^{x}(4-2x)dx = \int_1^x(4-2x)dx = [4x - x^2]_1^x = 4x - x^2 - 3}$ **,** $\boldsymbol{1 \leq x \leq 2}$

$\boldsymbol{F(x) = 0\,, x \leq 1}$

$\boldsymbol{F(x) = 1\,, 2 \leq x}$

c) $\boldsymbol{F(1.5) = 4*1.5 - 1.5^2 - 3 = 0.75}$

d) $\boldsymbol{E[X] = \int_{-\infty}^{\infty} xf(x)dx = \int_1^2(4x - 2x^2)dx = [2x^2 - \frac{2}{3}x^3]_1^2 = \frac{4}{3} \sim 1.33}$

Exercise 5

If X is an exponential random variable with parameter λ = 1 compute the probability density function of the random variable Y defined by Y = log X. [Hint: Work with the cumulative distribution function(cdf) and then compute the probability density function (pdf)]

Answer

The distribution function of Y is

$$\boldsymbol{F_Y(Y) = P(Y \le y) = P(logX \le y) = P(X \le e^y) = 1 - e^{-(e^y)}}$$

$$\Leftrightarrow \boldsymbol{f(y) = \frac{dF_Y(Y)}{dy} = e^{-(e^y)} * e^y = e^{-(e^y)+y}}$$

Exercise 6

1. A randomly chosen IQ test taker obtains a score that is a normal random variable with mean 100 and standard deviation 15. What is the probability that the score of such a person is (a) above 125 (b) between 90 and 110 (c) submit the R commands that was used to answer questions 1(a) and 1(b).

2. A multiple-choice examination has 15 questions, each with five possible answers, only one of which is correct. Suppose that one of the students who takes the examination answers each of the questions with an independent random guess. (a) What is the probability that he/she answers at least three questions correctly? (b) submit the R commands that was used to answer question 2(a).

Answer

1)

```
> Prob1 = 1 - pnorm (125,mean = 100, sd = 15)
> Prob1
```

[1] 0.04779035

```
> Prob2 = pnorm(110,mean = 100, sd = 15) - pnorm(90,mean = 100, sd = 15)
> Prob2
```

[1] 0.4950149

2)

```
> Prob3 = 1 - pbinom(2,size = 15, prob = 0.2)
> Prob3
```

[1] 0.6019768

Exercise 7

Use R in this exercise and submit your R code that was used to answer the questions.

From the list of distribution studied select one asymmetric discrete, call it X, and one asymmetric continuous, call it Y. For each of the two distributions generate two random samples of size 100 and 10000. Each time you generate a random sample use the command set.seed(.) using the same unique number of your choice (so that your results are reproducible). For all four generated data sets:

(a) Provide a graphical representation (visualization) of your data.

(b) Get an estimate of the: mean, sd, median, IQR. What are the theoretic values of these statistics when the true distribution is used?

(c) What is the proportion of the sample data in the region:

[min(mean, median), max(mean, median)]

What is the theoretic proportion of this region when the true distribution is used?

(d) What is the quantile of the sample data that are at or below the lowest 1%? What is the respective theoretic quantile when the true distribution is used?

(e) What is the quantile of the sample data that are at or above the upper 1%? What is the respective theoretic quantile when the true distribution is used?

(f) Comment on the discrepancies between theoretic and sample evaluated statistics in the previous questions (b)-(e).

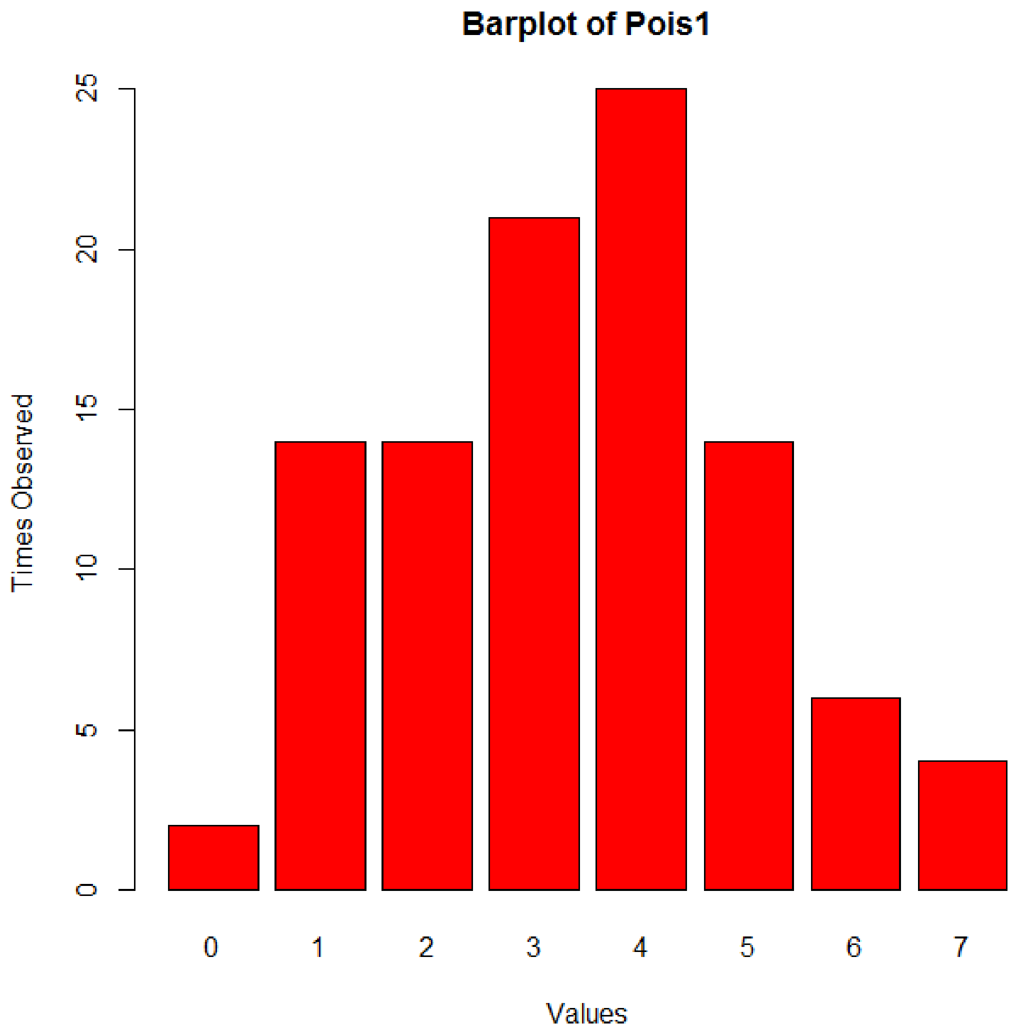
Answer

a)

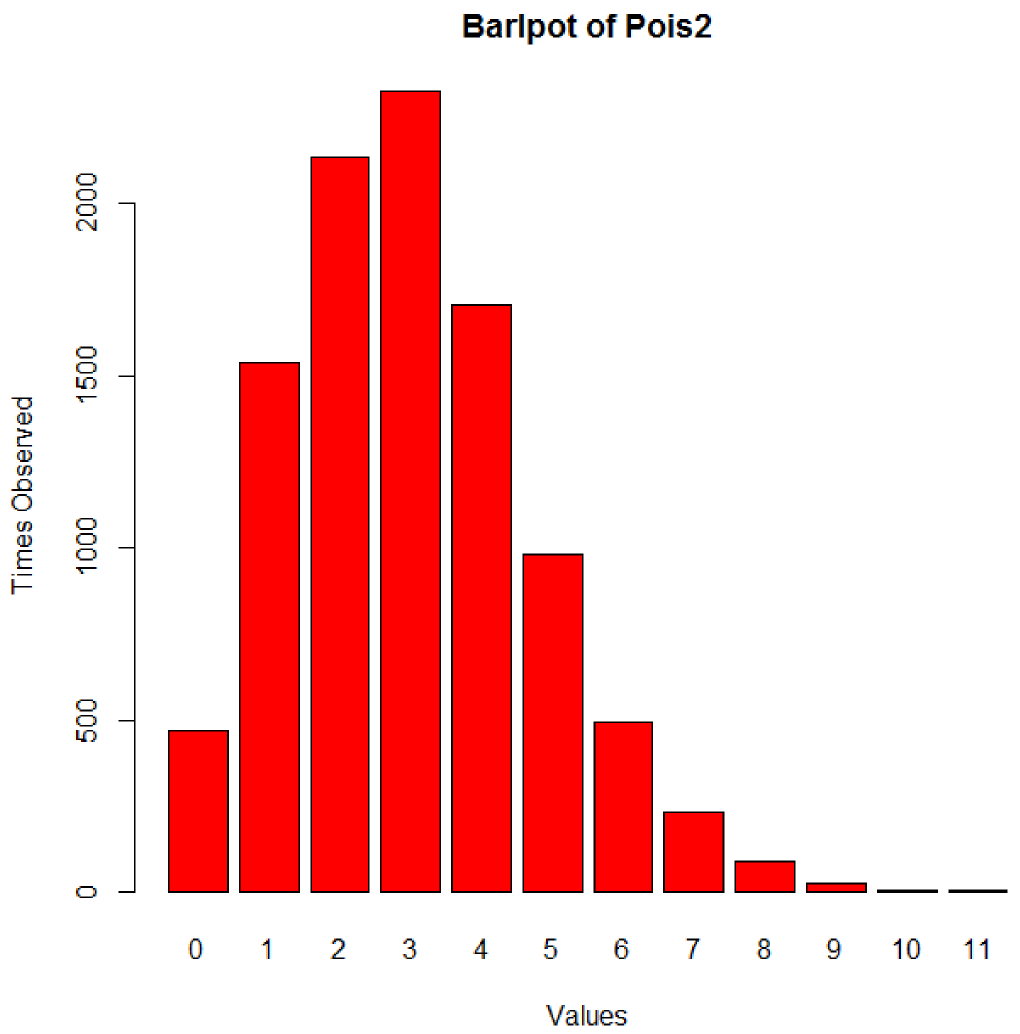
Let **Poisson** with $\boldsymbol{\lambda = 3}$ be the asymmetric discrete distribution and **Exponential** with $\boldsymbol{\lambda = 15}$ be the asymmetric Continuous one, while seed chosen is **424242**:

```
> set.seed(424242)
> pois1=rpois(100,3)
> set.seed(424242)
> pois2=rpois(10000,3)
> set.seed(424242)
> exp1=rexp(100, 15)
> set.seed(424242)
> exp2=rexp(10000, 15)
```

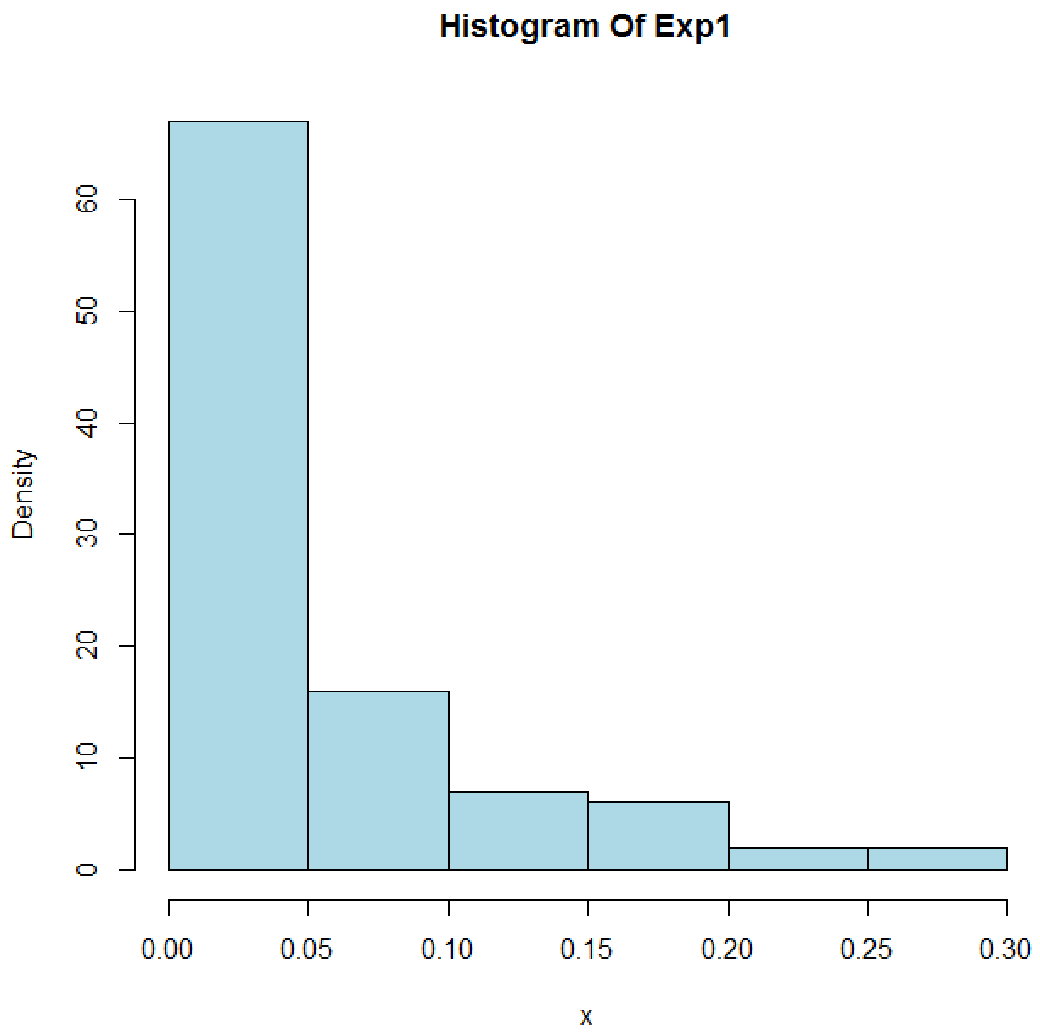
```
> barplot(table(pois1), axes = TRUE , col  = "red" , xlab = "Values" , ylab
= "Times Observed" , , main = "Barplot of Pois1")
```



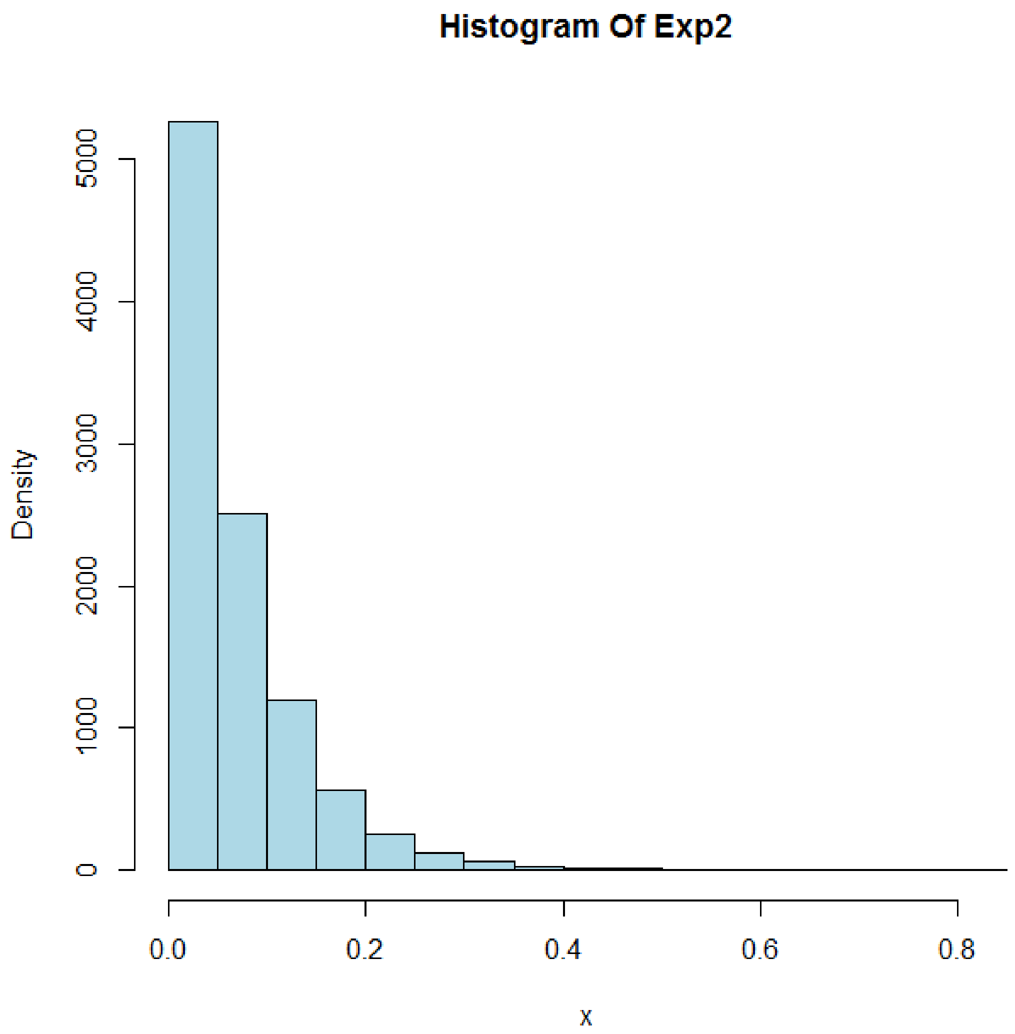

```
> barplot(table(poiss2), axes = TRUE , col  = "red" , xlab = "Values" , ylab = "Times Observed"
, main = "Barlpot of Pois2" )
```




```
> hist (exp1 , axes = TRUE, main = "Histogram Of Exp1", col = "lightblue" , xlab = "x" , ylab =
"Density")
```



> hist (exp2 , axes = TRUE, main = "Histogram Of Exp2", col = "lightblue" , xlab = "x" , ylab = "Density")



b)

```
> c(mean(pois1), sd(pois1), median(pois1), IQR(pois1))
[1] 3.390000 1.656911 3.000000 2.000000
```

```
> c(mean(pois2), sd(pois2), median(pois2), IQR(pois2))
[1] 3.009700 1.724095 3.000000 2.000000
```

```
> c(mean(exp1), sd(exp1), median(exp1), IQR(exp1))
[1] 0.05566695 0.05966316 0.03568999 0.05884569
```

```
> c(mean(exp2), sd(exp2), median(exp2), IQR(exp2))
[1] 0.06633146 0.06639121 0.04590885 0.07401518
```

Poisson Distribution:

$$\boldsymbol{Mean[X] = \lambda = 3}$$

$$\boldsymbol{Var[X] = \sqrt{3} = 1.7320}$$

$$\boldsymbol{Median[X]: \lambda - ln2 \leq Median[X] < \lambda + \frac{1}{3} \Leftrightarrow 2.3068 \leq median < 3.3333}$$

$$\boldsymbol{IQR[X] = 2}$$

Exponential Distribution:

$$\boldsymbol{E[X] = \frac{1}{\lambda} = \frac{1}{15} = 0.0666}$$

$$\boldsymbol{Var[X] = \frac{1}{\lambda^2} = 0.044}$$

$$\boldsymbol{Median[X] = \frac{ln2}{\lambda} = 0.0462}$$

$$\boldsymbol{IQR[X] = \frac{ln3}{\lambda} = 0.0732}$$

Summary Tables:

POISSON	Sample1 (100)	Sample2(10000)	Expected
E[X]	3,390	3,009	3,000
SD[X]	1,656	1,724	1,732
Median[X]	3,000	3,000	3,000
IQR[X]	2,000	2,000	2,000

EXPONENTIAL	Sample1 (100)	Sample2(10000)	Expected
E[X]	0,0556	0,0663	0,0666
SD[X]	0,0596	0,0663	0,0666
Median[X]	0,0356	0,0459	0,0462
IQR[X]	0,0588	0,0740	0,0732

c)

```
> sum (min(mean(pois1),median(pois1)) <= pois1 & pois1 <=
max(mean(pois1),median(pois1)))/100
[1] 0.21
> sum (min(mean(pois2),median(pois2)) <= pois2 & pois2 <=
max(mean(pois2),median(pois2)))/10000
```

```
[1] 0.2327
> sum (min(mean(exp1),median(exp1)) < exp1 & exp1 < max(mean(exp1),median(exp1)))/100
[1] 0.21
> sum (min(mean(exp2),median(exp2)) < exp2 & exp2 < max(mean(exp2),median(exp2)))/10000
[1] 0.128
```

Poisson:

According to Poisson distribution there are $(\boldsymbol{\lambda} - \boldsymbol{ln2}) * \boldsymbol{SampleSize}$ elements between $\mathbf{min}(\boldsymbol{median}, \boldsymbol{mean}) = \boldsymbol{median}$ and $\mathbf{max}(\boldsymbol{median}, \boldsymbol{mean}) = \boldsymbol{mean} = \boldsymbol{\lambda}$ (for closed interval).

For samples 1 and 2 the expected value is $(\mathbf{3} - \boldsymbol{ln2}) \cong \mathbf{0.23}$.

Exponential:

The expected probability for both samples comes from the cumulative distribution function of exponential distribution:

$$\boldsymbol{F}(\mathbf{0.0666}) - \boldsymbol{F}(\mathbf{0.0462}) = \left(\mathbf{1} - \boldsymbol{e}^{-\mathbf{0.0666}*\mathbf{15}}\right) - \left(\mathbf{1} - \boldsymbol{e}^{-\mathbf{0.0462}*\mathbf{15}}\right) = \mathbf{0.1318261}$$

Summary Tables:

POISSON	Sample1 (100)	Sample2(10000)	Expected
Value	0.21	0.2327	0.23

EXPONENTIAL	Sample1 (100)	Sample2(10000)	Expected
Value	0.21	0.128	0.13

d)

```
> quantile(pois1, probs = 0.01)
1%
 0
> quantile(pois2, probs = 0.01)
1%
```

0

```
> quantile(exp1, probs = 0.01)
```

1%

0.000425281

```
> quantile(exp2, probs = 0.01)
```

1%

0.0007106257

Poisson:

The requested number will be the result of solving equation with $\boldsymbol{X_0}$ as variable:

$\boldsymbol{1-\sum_{\kappa=X_0}^{\kappa=\infty}\frac{\lambda^{X_0}e^{-\lambda}}{\kappa!}=0.01}$ (which means the total probability that the value will be less than $\boldsymbol{X_0}$ is **0.01**).

This equation is hard to solve with analysis.

We observe that for value $\boldsymbol{k=0}$:

$\frac{\lambda^k e^{-\lambda}}{\kappa!}=e^{-3}=\boldsymbol{0.04978}$ of the sample will have value 0. Since it is far bigger than **0.01** then the required value will be **0** (since 0.049 quantile will be 0).

Exponential:

We are looking for $\boldsymbol{X_1}$that satisfies (same value for both samples):

$$\boldsymbol{P(x<X_1)=0.01}$$

$$\Leftrightarrow \boldsymbol{X_1=-\frac{\ln(1-0.01)}{15}=0.00067}$$

Summary Tables:

POISSON	Sample1 (100)	Sample2(10000)	Expected
Value	0	0	0

EXPONENTIAL	Sample1 (100)	Sample2(10000)	Expected
Value	0.000425281	0.0007106257	0.00067

e)

```
> quantile(pois1, probs = 0.99)
```

99%

7

```
> quantile(pois2, probs = 0.99)
99%
 8
> quantile(exp1, probs = 0.99)
      99%
0.2550339
> quantile(exp2, probs = 0.99)
      99%
0.3047935
```

Poisson:

The requested number will be the result of solving equation with X_2 as variable:

$\mathbf{1} - \sum_{\kappa=X_2}^{\kappa=\infty} \frac{\lambda^{X_2} e^{-\lambda}}{\kappa!} = \mathbf{0.99}$ (which means the total probability that the value will be less than X_2 is **0.99**).

This equation is hard to solve with analysis.

Using computational methods we find that the expected value is **8.1587.**

Exponential:

We are looking for X_3that satisfies (same value for both samples):

$$P(x < X_3) = 0.99$$

$$\Leftrightarrow X_3 = -\frac{\ln(1-0.99)}{15} = 0.3070$$

Summary Tables:

POISSON	Sample1 (100)	Sample2(10000)	Expected
Value	7	8	8.1587

EXPONENTIAL	Sample1 (100)	Sample2(10000)	Expected
Value	0.2550359	0.3047935	0.3070

f) No matter how big our sample is, there will always be an error. In theory, the bigger the sample is, the better the results there will be, provided that all other parameters are the same between our samples.

Of course in statistics there is evidence that smaller samples can do better than large samples, under certain occasions.

If we try to observe how good the results are, we will find out that the large samples give results closer to the expected ones. This is something that confirms the statement above.

Despite this, there will always be some distance from the real value in practice. It will only be distinguished when the sample size approaches infinite or if we try to estimate a population of finite size, when the sample size equals the population size.