

# Project FSML-2

jose lise

9/29/2019

## Exercise 1

(a) Let  $\mathbf{X} \sim \mathcal{N}(-1, 0.01)$ , 0.01 being the variance, Compute:

i.  $P(X \leq -0.98)$  is given by the function `pnorm()` the distribution function.

```
pnorm(q=-0.98, mean = -1, sd=sqrt(0.01))
```

```
## [1] 0.5792597
```

ii.  $P(X \leq -1.02)$

```
pnorm(q=-1.02, mean = -1, sd=sqrt(0.01))
```

```
## [1] 0.4207403
```

iii.  $P(X \geq 0.82) = 1 - P(X \leq 0.82)$

```
1. - pnorm(q=0.82, mean = -1, sd=sqrt(0.01))
```

```
## [1] 0
```

iv.  $P(X \in [-1.22; -0.96]) = P(X \leq -0.96) - P(X \leq -1.22)$

```
pnorm(q=-0.96, mean = -1, sd=sqrt(0.01)) - pnorm(q=-1.22, mean = -1, sd=sqrt(0.01))
```

```
## [1] 0.6415183
```

(b) Let  $\mathbf{X} \sim \mathcal{N}(0, 1)$ . Determine  $t$  such that:

i.  $P(X \leq t) = 0.9$

We don't need to specify mean parameter as default is mean=0. And also for the standard deviation, by default sd=1.  $t$  is given by the quantile function `qnorm(p=0.9) = 1.281552.`

```
qnorm(p=0.9)
```

```
## [1] 1.281552
```

ii.  $P(X \leq t) = 0.2$  then  $t = -0.8416212$

```
qnorm(p=0.2)
```

```
## [1] -0.8416212
```

iii.  $P(X \in [-t; t]) = 0.95$  As the normal distribution is symmetric, we have:

$$P(X \in [-t; t]) = 2 * P(X \in [0; t])$$

This gives  $2 * (P(X \leq t) - P(X \leq 0))$  and  $P(X \leq 0) = 0.5$  Therefore  $2 * P(X \leq t) - 1 = 0.95$  And  $P(X \leq t) = (1 + 0.95)/2$  then  $t = 1.959964$

```
qnorm(p=1.95/2)
```

```
## [1] 1.959964
```

## Exercise 2

(a) A density function  $f$  is defined by:

$$P(X = x) = \int_{-\infty}^x f(t)dt$$

(b) An unbiased estimator  $\hat{\theta}_n$  of  $\theta$  is such that:

$$E[\hat{\theta}_n] = \theta$$

(c) Calculation of expectation and variance of  $\overline{X}_n$

$$\overline{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

$$E[\overline{X}_n] = E\left[\frac{1}{n} \sum_{i=1}^n X_i\right]$$

As expectation is a linear function, expectation of a sum is the sum of expectation, then:

$$E[\overline{X}_n] = \sum_{i=1}^n \frac{1}{n} E[X_i]$$

$$E[\overline{X}_n] = \frac{1}{n} \sum_{i=1}^n E[X_i]$$

And as  $E[X_i] = \mu$  we have:  $E[\overline{X}_n] = \frac{1}{n} n\mu = \mu$

$$V[\overline{X}_n] = V\left[\frac{1}{n} \sum_{i=1}^n X_i\right] \text{ We know that:}$$

$$V[\overline{X}_n] = \frac{1}{n^2} V\left[\sum_{i=1}^n X_i\right]$$

And as the  $X_i$  are independent, we have:

$$V[\overline{X}_n] = \frac{1}{n^2} \sum_{i=1}^n V[X_i]$$

And as  $V[X_i] = \sigma^2$  we have:

$$V[\overline{X}_n] = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}$$

(d) Let  $X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2)$  sample.

$$\text{An unbiased estimator of } \sigma \text{ is } \hat{\sigma}_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \overline{X}_n)^2$$

As the  $X_i$  are Gaussian independent random variables, we know that  $\frac{n-1}{\sigma^2} \hat{\sigma}_n^2$  follow a Chi-square distribution with  $n-1$  degrees of freedom.

$$\text{Therefore } E\left[\frac{n-1}{\sigma^2} \hat{\sigma}_n^2\right] = n-1$$

$$\text{We deduce that: } E[\hat{\sigma}_n^2] = \sigma^2$$

Therefore  $\hat{\sigma}_n^2$  is an unbiased estimator of  $\sigma^2$ .

### Exercise 3

(a) As the uniform distribution is a continuous distribution, the easiest test to use is the Kolmogorov test - function `ks.test` with parameter 'punif' for the cumulative distribution function of uniform distribution and boundaries :

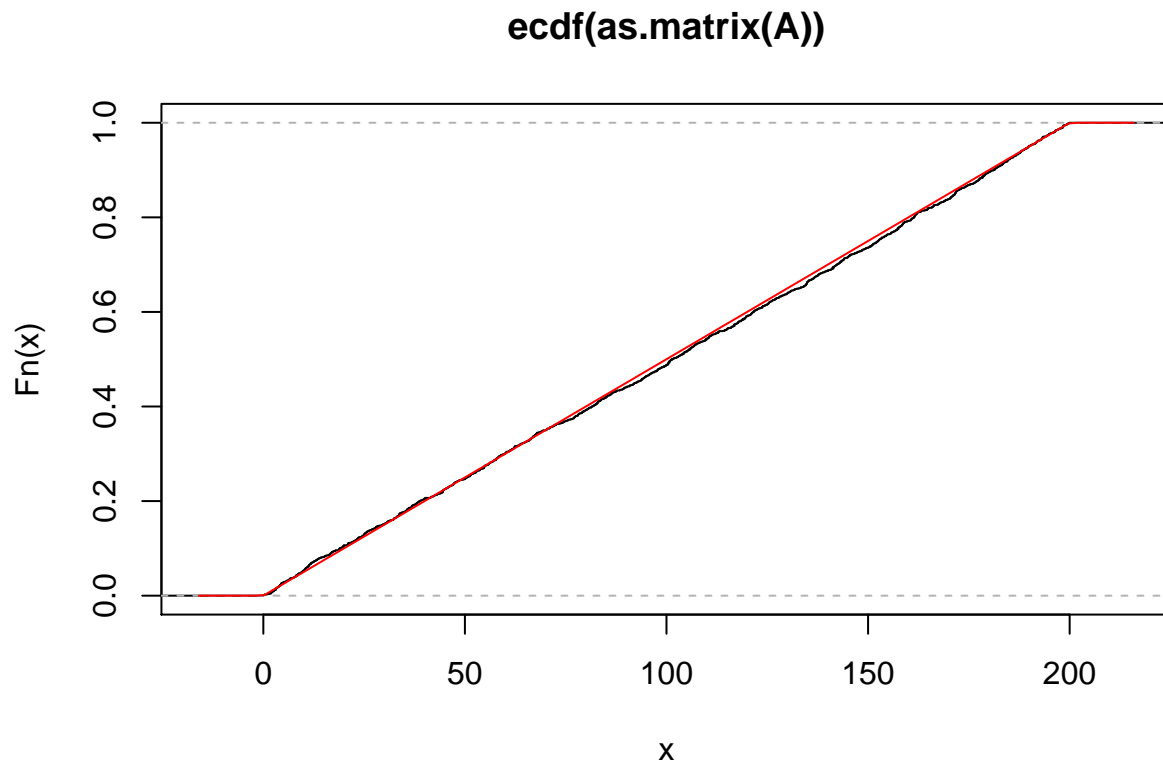
```
A=read.table('dataexam.txt')
t <- ks.test(A, 'punif',0,200)
t$p.value
```

```
## [1] 0.7310588
```

P-Value = 0.731  $\geq$  of the significance level (0.05), therefore we do not reject the null hypothesis. And The data in the file follows an uniform distribution between 0 and 200.

To confirm that the data in the file follow a uniform distribution, we can use the `ecdf` function:

```
plot(ecdf(as.matrix(A)))
curve(punif(x, 0, 200), add=TRUE, col='red' )
```



```
# Another alternative is to use the chi square test.
# But this test is designed for discrete distributions.
# Therefore before to use it, we will need to change the
# continuous distribution into a discrete distribution.
setwd('D:/OneDrive - Data ScienceTech Institute/DSTI/FoundationOfStatisticsAndML2/Test')
A=read.table('dataexam.txt')
#
# We make 100 classes for the uniform distribution in the data set
C = cut(as.matrix(A), 100)
```

```

# We create a table
C=table(C)

# As we know that the uniform function is between 0 and 200
# we create a vector between those values by step 2
x=seq(0,200,2)
# cp is the cumulative probability distribution
cp = punif(x,min=0, max=200)
# We compute the probability distribution:  $p(i) = cp(i+1) - cp(i)$ 
p = diff(cp)

# n is the number of observations
n=sum(C)
# tc is the theoretical counts
tc=n*p
oc = C
# we apply the definition of cn below
cn=sum((oc - tc)^2/tc )

c=qchisq(0.95,99)

cn

```

```
## [1] 113.1538
```

```
c
```

```
## [1] 123.2252
```

```

# cn = 113.1538 < c = 123.2252
# Therefore we do not reject the null hypothesis and accept
# that the distribution is an uniform distribution in [0,200]
#
#We can also confirm the result using the chisq.test() function:

```

```
chisq.test(oc,p=p)
```

```

##
## Chi-squared test for given probabilities
##
## data:  oc
## X-squared = 113.15, df = 99, p-value = 0.1566
# p-value is p-value = 0.1566 > 0.05
# therefore we do not reject the null hypothesis

```

## (b) Time between 2 events distribution

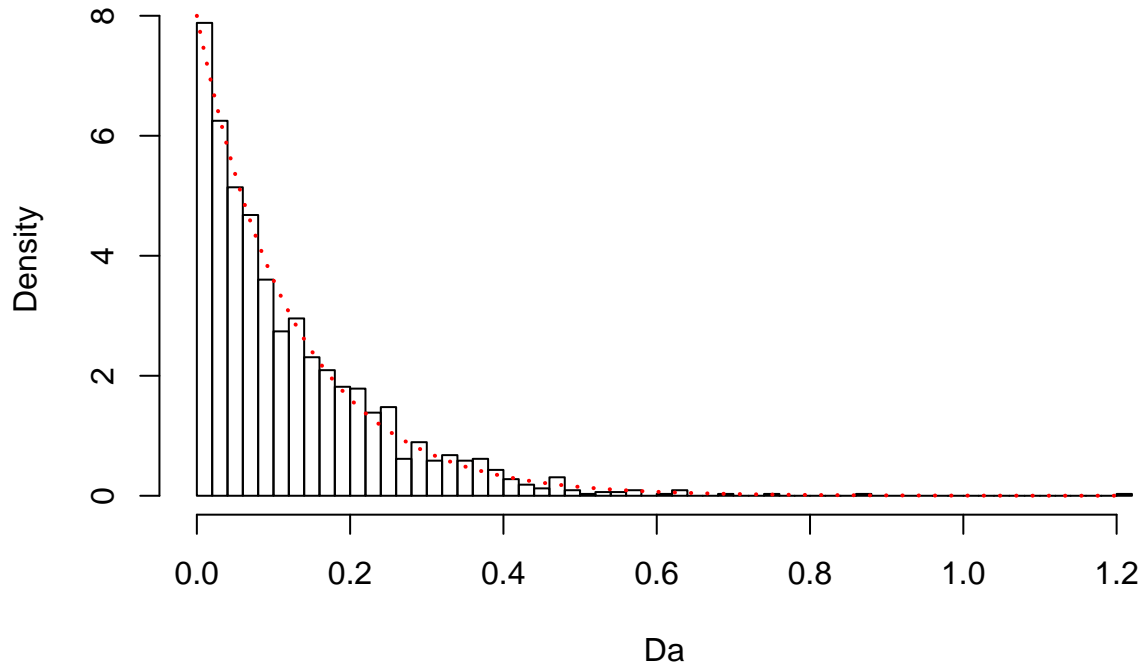
```

# A is the time dataframe read from the file and used in the previous
# question. We will compute below the difference between 2 events and
# try to guess the distribution
B = as.matrix(A)
D = as.vector(B)
Da = diff(D)
hist(Da,freq=FALSE,breaks=50, main="Time between 2 events")
# Exponential distribution is plotted in red with parameter 8

```

```
lines(seq(0, max(Da), 1/1000),dexp(seq(0, max(Da), 1/1000),8),col='red', lty=3,lwd=2)
```

## Time between 2 events



### (c) Parameter distribution and confidence interval

From the previous question we saw on the histogram chart that the distribution is quite similar to an exponential distribution with parameter equal to 8.

if  $X \sim \exp(\lambda)$  we know that:

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

And using the method of moments, we find that one estimator for  $\lambda$  is:

$$\hat{\lambda}_n = \frac{1}{\bar{X}_n}$$

We have seen in this case that asymptotically:  $P\left(\frac{\bar{X}_n}{1+\frac{T}{\sqrt{(n)}}} \leq \frac{1}{\lambda} \leq \frac{\bar{X}_n}{1-\frac{T}{\sqrt{(n)}}}\right) = 1 - \alpha$

If  $\alpha = 0.05$  then  $T = \text{qnorm}(0.975)$  then almost equal to 1.96

In our case  $n = 1624 \gg 30$  therefore we can assume that we are in the asymptotic case.

Therefore we have:

$$P(7.726847 \leq \lambda \leq 8.5168) = 95\%$$

## Exercise 4

### (a) Prove that:

$$\forall s, t > 0, P(X > t + s | X > t) = P(X > s)$$

$$\text{We have: } P(X > t + s | X > t) = \frac{P((X > t + s) \cap (X > t))}{P(X > t)}$$

But we have  $\{X > t + s\} \cap \{X > t\} = \{X > t + s\}$  therefore:

$$P(X > t + s | X > t) = \frac{P(X > t + s)}{P(X > t)}$$

As  $X$  follow an exponential distribution, the cumulative distribution is given by:

$$F(x) = P(X \leq x) = 1 - e^{-\lambda x}$$

As this is a continuous distribution, we can write:

$$P(X > x) = 1 - P(X \leq x) = 1 - F(x) = e^{-\lambda x}$$

Then we can write:

$$P(X > t + s | X > t) = \frac{e^{-\lambda(s+t)}}{e^{-\lambda t}} = e^{-\lambda s}$$

Therefore we have:

$$P(X > t + s | X > t) = P(X > s)$$

**(b)  $Y = E(X) + 1$  where  $E(x)$  is the biggest interger smaller or equal to  $x$ . Determine the distribution of  $Y$ .**

$E(X)$  only has values in  $\mathbb{N}$  then  $Y = E(X) + 1$  only has values in  $\mathbb{N}^*$ .

$Y = n, n \in \mathbb{N}^*$  means  $E(X) + 1 = n$  and  $E(X) = n - 1 \iff n - 1 \leq X < n$ .

Therefore  $P(Y = n) = P(n - 1 \leq X < n) = F(n) - F(n - 1)$ .

We have  $P(Y = n) = 1 - e^{-\lambda n} - (1 - e^{-\lambda(n-1)})$ .

And  $P(Y = n) = 1 - e^{-\lambda n} - 1 + e^{-\lambda(n-1)}$ .

$P(Y = n) = -e^{-\lambda n} + e^{-\lambda(n-1)} = (1 - e^{-\lambda})e^{-\lambda(n-1)}$ .

This is a geometric distribution with parameter  $p = 1 - e^{-\lambda}$ .

Indeed, we have:  $P(Y = n) = p(1 - p)^{n-1}$ .