University of Bremen

Institute for Statistics

Daniel Ochieng
MZH, Room 7240
E-Mail: dochieng@uni-bremen.de

Exercises for the lecture

Mathematics III

Summer Semester 2023

Solutions to Exercise 09

Tasks

33. **Chebyshev's inequality.** Let X denote the number of molecules that goes into one of the bottles, then $X \sim Bin(n, \theta)$ where

$$n = 0.26 * 10^{23},$$
$$\theta = 0.5,$$
$$E(X) = n\theta,$$
$$= 1.3 * 10^{22},$$

and

$$Var(X) = n\theta(1-\theta),$$
$$= 6.5 * 10^{21}.$$

If the random variable X has a mean μ and variance σ^2, then, for every $k \geq 1$,

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

We need to find,

$$P\Big(X \geq 0.13 * 10^{23}(1+10^{-8})\Big),$$

$$= P\Big((X-\mu) \geq 0.13 * 10^{23}(1+10^{-8}) - 1.3 * 10^{22}\Big),$$

$$= P\Big(|X-\mu| \geq 1.3 * 10^{14}\Big),$$

$$= P\Big(|X-\mu| \geq k * \sigma = 1.3 * 10^{14}\Big) \leq \frac{1}{k^2} = \frac{1}{1612.452} = 3.8461 * 10^{-7}.$$

34. **Statistical modelling**

(a) The measure space is given by $(X, \mathcal{F}, \mu)$ where

$$X = \{(x_1, \ldots, x_{49}) \in \{0, \ldots, 520\}^{49} : \sum_{j=1}^{49} X_j = 6 * 520\},$$

$$\mathcal{F} = 2^x \text{ is the power set of } X, \text{ and}$$

$$\mu \text{ is a measure on the measurable space } (X, \mathcal{F}).$$

(b) The measure space is given by $(X, \mathcal{F}, \mu)$ where

$$X = \{0,1\}^{n_1} \times \{0,1\}^{n_2} = \{0,1\}^{n_1+n_2},$$

$$\mathcal{F} = 2^x \text{ is the power set of } X.$$

$$\mu \text{ is a measure on the measurable space } (X, \mathcal{F}).$$

which can be described by two Bernoulli distributions: The first n_1 observations are stochastically independent and identically distributed (i.i.d) as Bernoulli(θ_1) and the other n_2 observations are also stochastically i.i.d as Bernoulli(θ_2).

35. **Programming task: Monte Carlo integration.**

Task 35

Daniel Ochieng

2023-06-29

```r
### Task 35

## (a)
f <- function(y) {y^2*exp(-y)}
res=integrate(f, lower=0, upper=Inf)
res
```

```
## 2 with absolute error < 7.1e-05
```

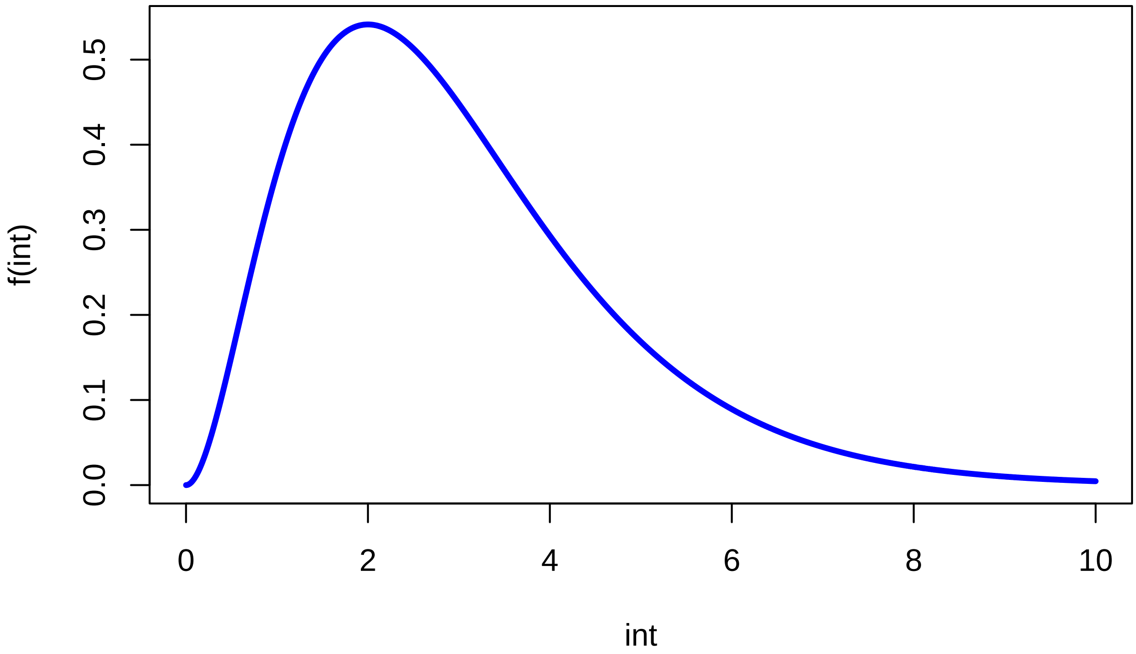
```r
str(res)
```

```
## List of 5
##  $ value       : num 2
##  $ abs.error   : num 7.08e-05
##  $ subdivisions: int 4
##  $ message     : chr "OK"
##  $ call        : language integrate(f = f, lower = 0, upper = Inf)
##  - attr(*, "class")= chr "integrate"
```

```r
res$value
```

```
## [1] 2
```

```r
int<-seq(0,10,0.001)
plot(int,f(int),type="l",col="blue",lwd="3")
```




```
### (b)
f<-function(y) {y}
set.seed(7)

# n=1000

n<-1000

# generate data from gamma (2,1)
x <- rgamma(n, shape=2, rate=1)

# compute the MC approximation
res2<-sum(sapply(x, f))/n
res2
```

```
## [1] 2.028509
```

```
# n=10000

n<-10000

# generate data from gamma (3,1)
x <- rgamma(n, shape=2, rate=1)

# compute the MC approximation
res2<-sum(sapply(x, f))/n
res2
```

```
## [1] 1.999642
```

```r
# n=100000
f1=function(){
n<-100000
# generate data from gamma (2,1)
x <- rgamma(n, shape=2, rate=1)
# compute the MC approximation
res2<-sum(sapply(x, f))/n
return(res2)
}
replicate(1,f1())
```

```
## [1] 1.995258
```

```r
## (c)

f2=function(){
# n=100000
n<-100000
# generate data from gamma (1,1)
x <- rgamma(n, shape=1, rate=1)
# compute the MC approximation
res3<-2*sum(sapply(x, f))/n
return(res3)
}
replicate(1,f2())
```

```
## [1] 1.993922
```

```r
## (d)

# case b

set.seed(7)
b=100
m1=replicate(b,f1())
mse1=mean((m1-2)^2)
mse1
```

```
## [1] 2.248775e-05
```

```r
# case c

set.seed(7)
b=100
m2=replicate(b,f2())
mse2=mean((m2-2)^2)
mse2
```

```
## [1] 3.426723e-05
```

```r
mse1<mse2
```

```
## [1] TRUE
```

```r
#method (b) gives most accurate result.
```

In task 35 c, let

$$u = y^2, \implies du = 2ydy,$$
$$dv = e^{-y}, \implies v = -e^{-y},$$
$$\therefore \int_0^\infty y^2 e^{-y} dy = -y^2 e^{-y}\Big|_0^\infty + 2\int_0^\infty ye^{-y}dy,$$
$$= 2\int_0^\infty ye^{-y}dy.$$

36. **Multiple select task**

(a) TRUE. If the underlying distribution function F is continuous, the empirical distribution function F_n will converge to F, and with probability 1, it will decrease from $Y_1, Y_2, \ldots, Y_n$ to any value in the range of $[0, 1]$ that can be represented as k/n, where k is an integer between 0 and n.

(b) TRUE. $\widehat{F}_X(x) = \frac{Y_X}{n}$ where Y_X = number of $X_i \leq x$, hence $Y_X \sim Bin(n, F_X(x))$.

(c) TRUE.The empirical distribution function F_n assigns a probability to each value based on the relative frequency of occurrence in the sample. For a given value y_i in the sample, the probability assigned by F_n is $1/n$.

Therefore, the expected value of $G(X)$ can be computed as follows:

$$E[G(X)] = \sum_{i=1}^{n}[G(y_i) * Pr(X = y_i)],$$
$$= \sum_{i=1}^{n}[G(y_i) * (1/n)],$$
$$= (1/n) * \sum_{i=1}^{n} G(y_i).$$

In the last step, we can pull out the common factor $(1/n)$ since it is the same for all terms in the summation. So, the expected value of $G(X)$ is given by $(1/n)$ times the sum of $G(y_i)$ for each value in the sample y. Therefore, the expression for the expected value of G(X) is:

$$E[G(X)] = (1/n) * \sum_{i=1}^{n} G(yi).$$

(d) TRUE. When the sample size n is an odd number and the observations $y_1, y_2, \ldots, y_n$ are all distinct, the median induced by the empirical distribution is uniquely determined. However, if the sample size is even, the median is not uniquely determined due to the presence of two observations at the center.