Zorizko bektoreak II

Baterako, baldintzazko eta probabilitate-lege marginalak

Izan bedi (X, Z) zorizko aldagai bikotea, ondorengo banaketa duena:

- Z zorizko aldagaiak $\{0, 1\}$ balioak hartzen ditu, $p = P(Z = 1) = 1/3$ delarik.
- X zorizko aldagaiaren inguruan badakigu, Z jakinik, bere banaketa baldintzatuaren probabilitate-legea $5Z + 1$ parametrodun Poisson dela.

Ariketa: Zein probabilitate-banaketa ezagun jarraitzen du Z aldagaiak?

Bernoulli($\frac{1}{3}$) jarraitzen du.

Hasteko, definitu dezagun (X, Z)ren baterako probabilite-legea kalkulatuko duen funtzio bat. Horretarako, lehenengo Zren probabilitate legea definituko dugu:

```
getMarginalZ <- function (z) {
r <- dbinom(z, size=1, prob=1/3) #=Bernoulli(p=1/3)
return(r)
}
```

Bestalde, $X|Z$ aldagai baldintzatuaren banaketa Poisson banaketa bat da, beraz bere probabilitate legea honela definitu dezakegu:

```
getConditionalXGivenZ <- function (x, z) {
if (z!=0 & z!=1) {
stop("z-ren balioa 0 edo 1 izan behar du")
}
lambda <- 5*z + 1
return(dpois(x, lambda))
}
```

Ariketa: Bete aurreko kode blokean falta dena eta irudikatu $X|Z$ aldagaiaren probabilitate-legea.

Behin bi probabilitate-lege hauek definituta ditugula, teoriatik, badakigu:

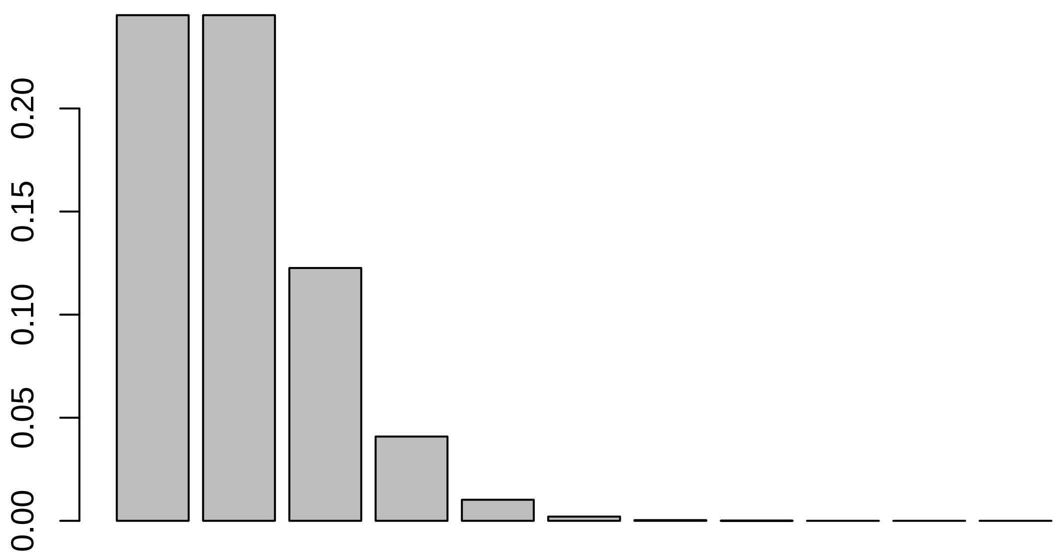
$$P(X, Z) = P(X|Z) \cdot P(Z)$$

Beraz,

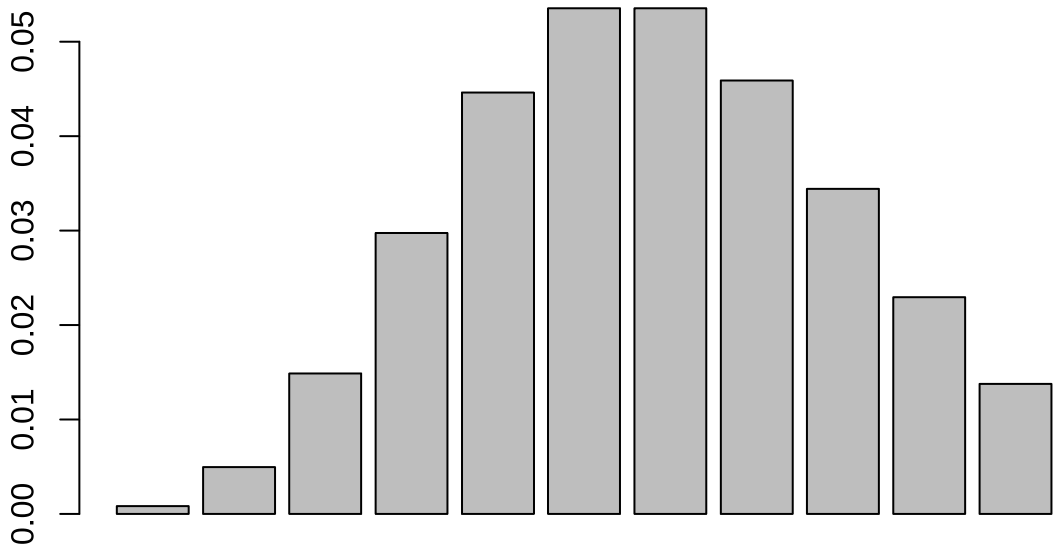
```r
getJointProbability <- function (x, z) {
return(getMarginalZ(z) * getConditionalXGivenZ(x, z))
}
```

Ariketa: Bete aurreko kode blokean falta dena eta irudikatu (X, Y) aldagaiaren probabilitate-legea hiru dimentsiotako grafiko bat erabiliz.

```r
#ardatzei balioak eman
x <- 0:10
z <- 0:1
prob <- outer(x, z, Vectorize(getJointProbability))

barplot(prob[, 1])
```




```r
barplot(prob[, 2])
```

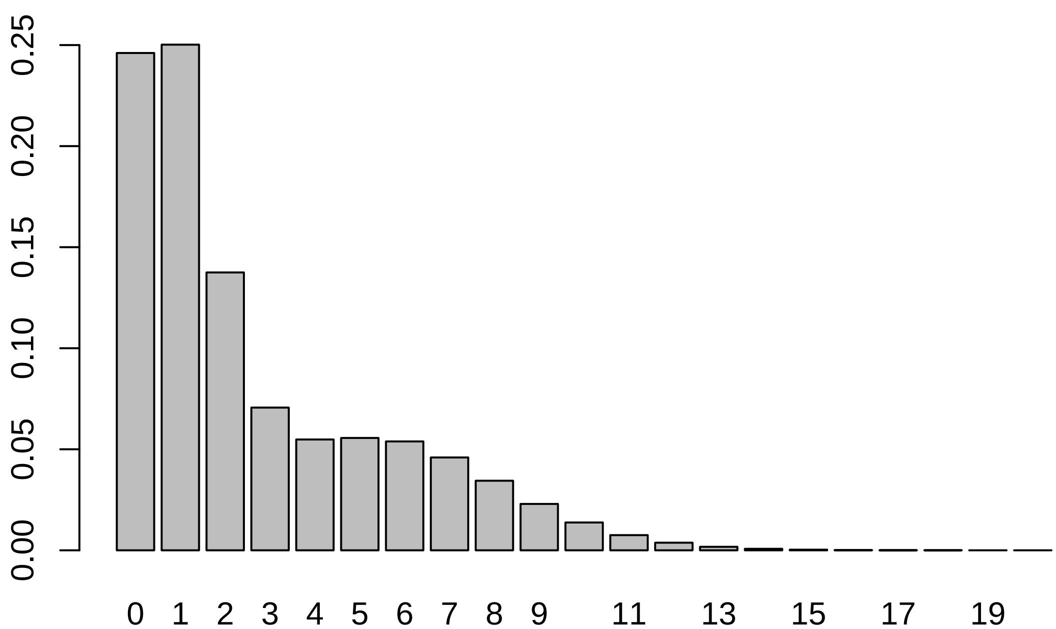


Ariketak

Aurreko adibidea jarraituz, erantzun honako galdera hauei (berrikusi klaseko gardenkiak, behar baduzu):

1.1. ariketa Lortu Xren bazter banaketa eta irudikatu grafiko batean.

```
getMarginalX <- function(x){
  return(getJointProbability(x, 0) + getJointProbability(x, 1))
}
x <- 0:20
prob <- getMarginalX(x)

barplot(prob, names.arg = x)
```



1.2. ariketa Zein da Zren banaketa baldintzatua, Xa jakinda? Idatzi funtzio bat banaketa honen balio teorikoak itzuliko dituena.

$$P(Z|X) = \frac{P(Z) \cdot P(X|Z)}{P(X)} = \frac{P(X,Z)}{P(X)}$$

```
getConditionalZGivenX <- function(x, z){
  return(getJointProbability(x, z)/getMarginalX(x))
}
```

1.3. ariketa Kalkulatu $E(X|Z = z)$.

Badakigu $X \sim Poisson(5Z + 1)$ eta $E(Poisson(\lambda)) = \lambda$

```
getConditionalExpectedXgivenZ <- function(z){
  return(5*z+1)
}
```

1.4. ariketa Nola kalkulatuko zenuke $E(Z|X = x)$? Sortu funtzio bat $E(Z|X = x)$ balioa itzuliko duena x balio ezberdinetarako eta irudikatu barra diagrama bat erabiliz.

$$E(Z|X = x) = \sum_{z \in \Omega_z} z \cdot P(Z = z|X = x)$$
$$0 \cdot P(Z = 0|X = x) + 1 \cdot P(Z = 1|X = x) = P(Z = 1|X = x)$$

```
getConditionalExpectedZgivenX <- function(x){
  return(getConditionalZGivenX(x, 1))
}
```

Balio teoriko eta esperimentalen konparazioa

Tutorialaren bigarren atal honetan, (X, Z) aldagaiaren banaketatik lagin bat aterako dugu eta ondoren behaketa hauek probabilitate banaketa teorikoa jarraitzen dutela konprobatuko dugu.

Horretarako, simula ditzagun (X, Z) bikotearen $n = 5000$ behaketa. Horretarako, honako funtzio errekurtsiboa definituko dugu, behaketa bat ala gehiago sortzeko balioko duena.

```
sampleJointDistribution <- function (n.sim, p=1/3,
                                     lambda=function(z){5*z + 1}) {
  if (n.sim==1) {
    z <- runif(1) <= p
    x <- rpois(1, lambda(z))
    return (c(x, z))
  }else{
    samps <- sapply(1:n.sim,
                    FUN=function(i) {
                     return(sampleJointDistribution(1, p, lambda))
```

```
                    })
    df <- data.frame(X=samps[1, ], Z=samps[2, ])
    return(df)
  }
}
```

Ariketa: Azaldu aurreko kode blokeko 4. lerroa. Zer egiten du?

$r$ri balio bat esleitzen dio $poisson(lambda(z))$ banaketa jarraitzen duena.

Behin funtzioa definituta, lagina lortzeko erabili dezakegu:

```
n <- 5000

joint.sample <- sampleJointDistribution(n)
```

Simulatutako laginaren arabera kalkulatu ditzagun orain Zren maiztasun erlatiboak eta konparatu ditzagun probabilitate-lege teorikoarekin.

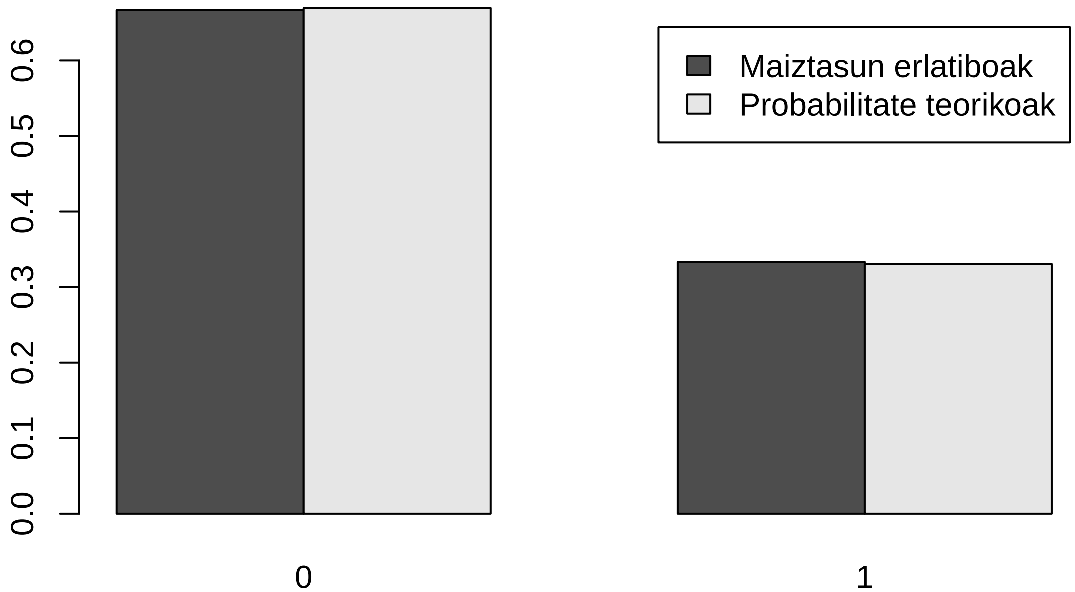
```
empirical.freq <- table(joint.sample$Z) / nrow(joint.sample)
```

Orain, laginean agertzen diren balioetarako, ikus dezagun zein den probabilitate marginal teorikoa:

```
z.values <- as.numeric(names(empirical.freq))
theorical.prob <- sapply(z.values, FUN=getMarginalZ)
```

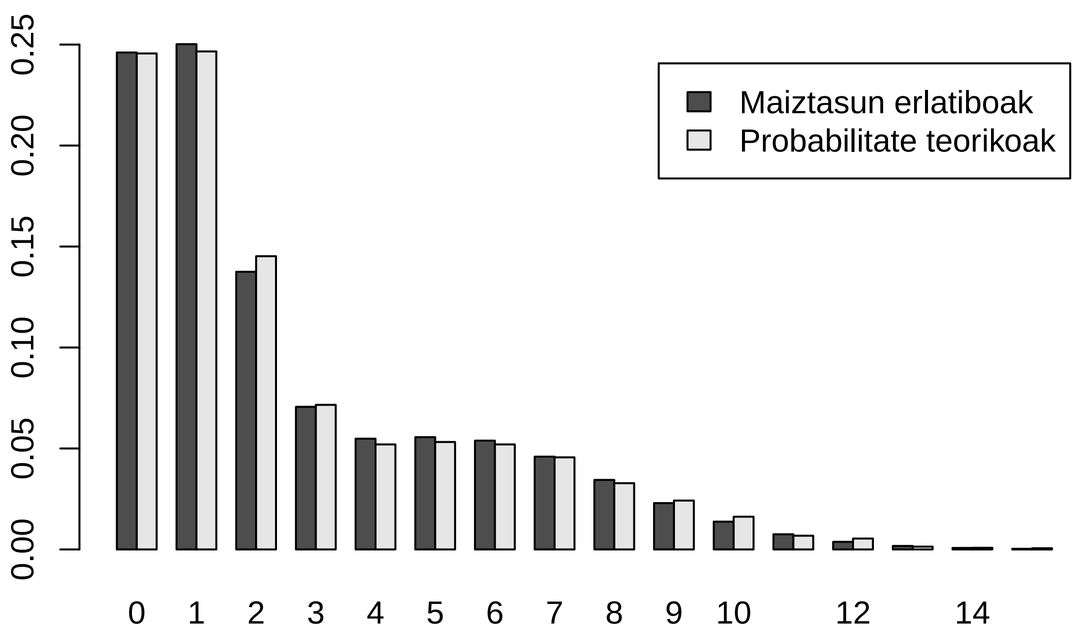
Ariketa: Maiztasun erlatiboak eta probabilitate teorikoak konparatu barra-diagrama bat erabiliz.

```
barplot(rbind(theorical.prob, empirical.freq), beside=TRUE,
        legend.text = c("Maiztasun erlatiboak", "Probabilitate teorikoak"),
        names.arg = names(empirical.freq))
```



Ariketa: Egin ezazu gauza bera X aldagaiarentzako.

```
empirical.freq <- table(joint.sample$X) / nrow(joint.sample)
x.values <- as.numeric(names(empirical.freq))
theorical.prob <- sapply(x.values, FUN=getMarginalX)

barplot(rbind(theorical.prob, empirical.freq), beside=TRUE,
        legend.text = c("Maiztasun erlatiboak", "Probabilitate teorikoak"),
        names.arg = names(empirical.freq))
```

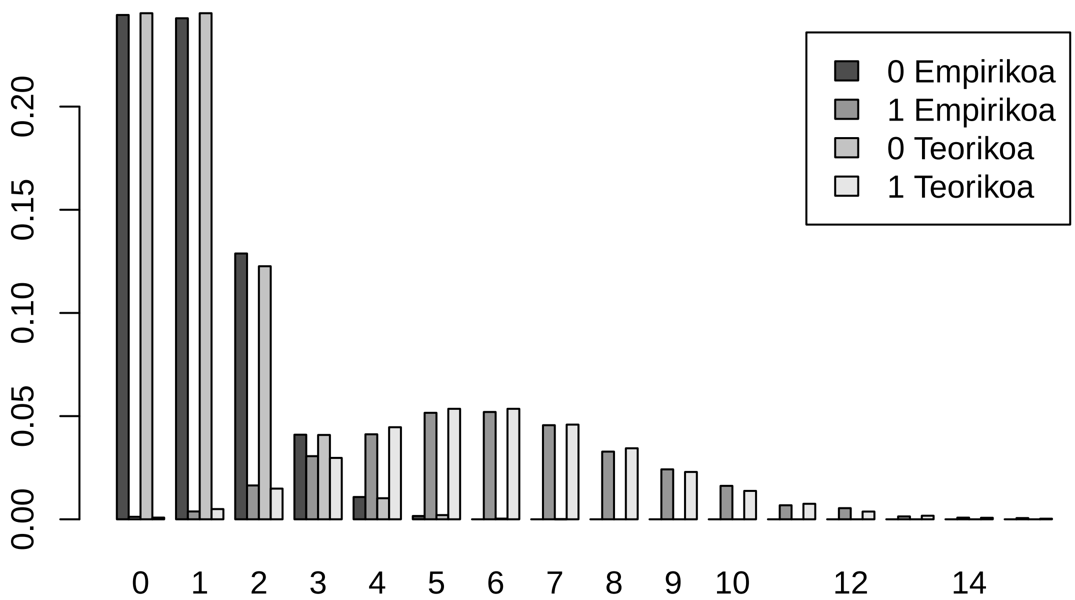


Ariketak

Aurreko adibideak jarraituz, erantzun itzazu honako galdera hauek:

2.1. ariketa Kalkulatu itzazu behatutako lagineko (x, z) konbinazio bakoitzeko bikotearen maiztasun erlatiboak. Konparatu lortutako emaitzak 1.1. atalean kalkulatu duzun baterako probabilitate balio teorikoekin. Zer ondorioztatzen duzu?

```
empirical.freq <- table(joint.sample) / nrow(joint.sample)
x.values <- as.numeric(rownames(empirical.freq))
z.values <- as.numeric(colnames(empirical.freq))
theorical.prob = outer(X=x.values, Y=z.values, FUN = Vectorize(getJointProbability))
barplot(rbind(t(empirical.freq), t(theorical.prob)), beside=TRUE,
        legend.text = c("0 Empirikoa", "1 Empirikoa", "0 Teorikoa", "1 Teorikoa"))
```

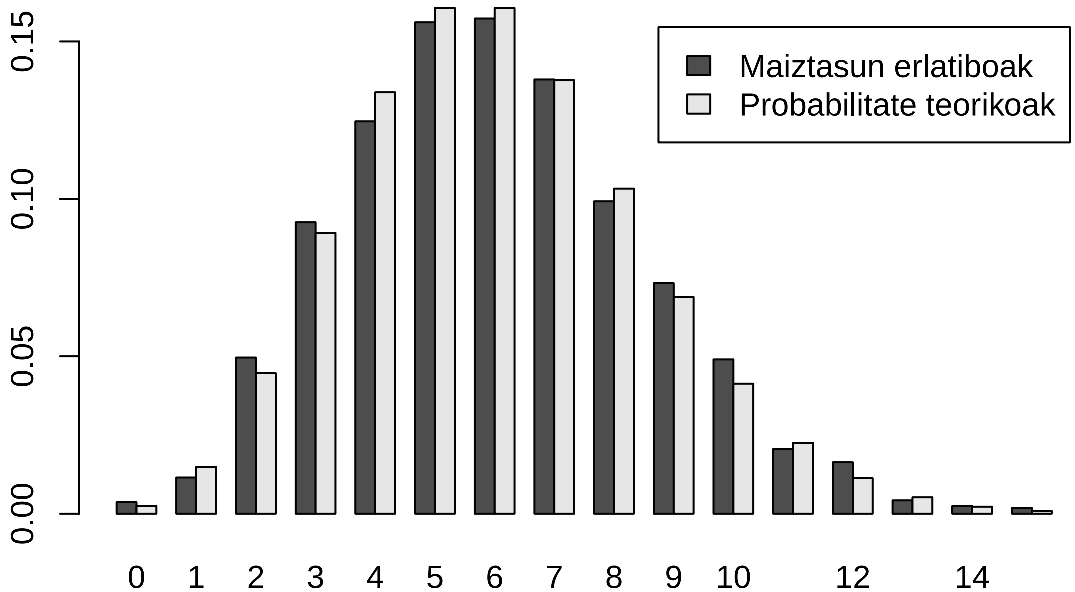


Laginaren maiztasun erlatiboak probabilitate teorikoa jarraitzen du.

2.2. ariketa Aurreko ataleko `data.frame`etik filtratu $z = 1$ betetzen duten kasuak eta, ondoren, X-en balio posible bakoitzerako irudikatu bere maiztasun erlatiboa. Zein banaketari dagokio? Gehitu iezaiozu grafikoari banaketa teorikoa.

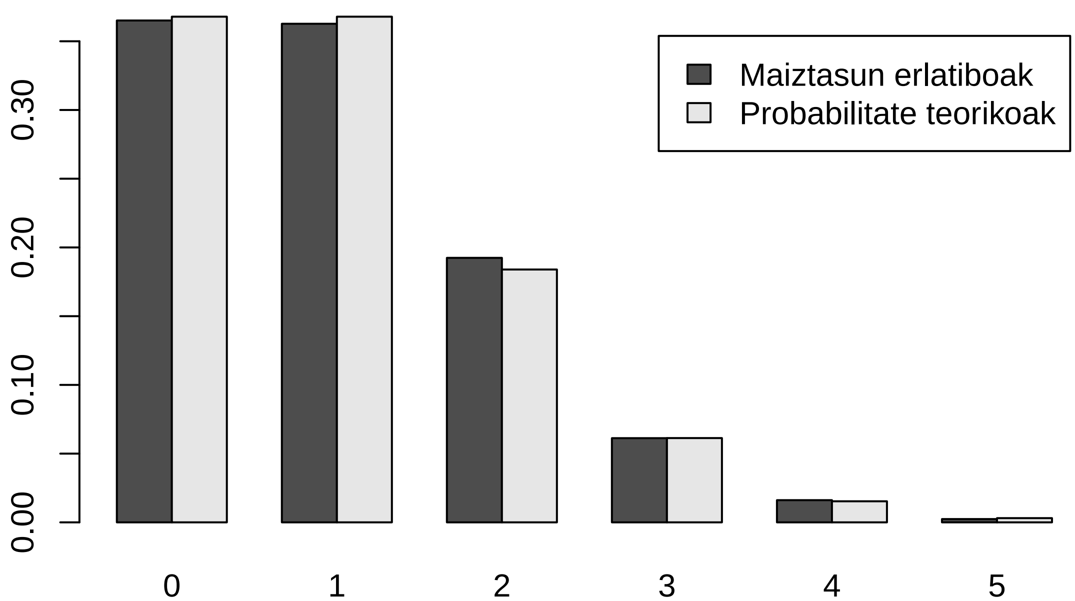
Banaketa baldintzatuari dagokio $P(X|Z = 1)$.

```
sample.z1 <- joint.sample[joint.sample$Z==1,]
empirical.freq <- table(sample.z1$X) / nrow(sample.z1)
x.values = as.numeric(names(empirical.freq))
theorical.prob <- sapply(x.values, FUN=function(i){return(getConditionalXGivenZ(i, 1))})

barplot(rbind(empirical.freq, theorical.prob), beside=TRUE,
        legend.text = c("Maiztasun erlatiboak", "Probabilitate teorikoak"))
```



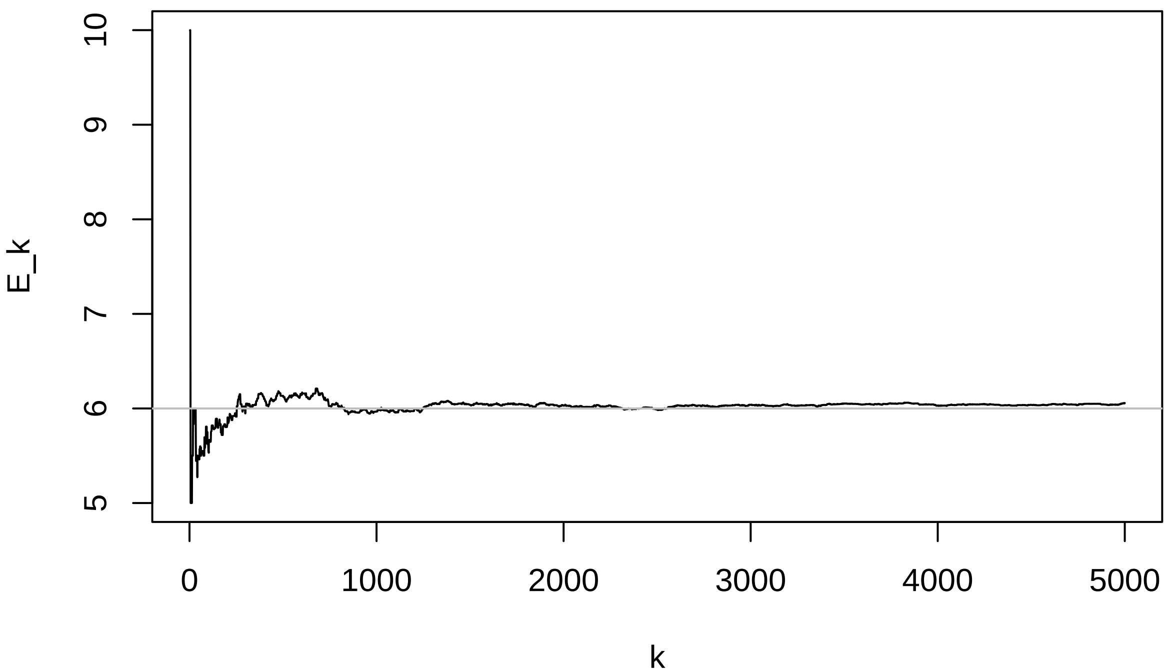
2.3. ariketa Errepika ezazu aurreko atala $z = 0$ den kasurako.

```
sample.z0 <- joint.sample[joint.sample$Z==0,]
empirical.freq <- table(sample.z0$X) / nrow(sample.z0)
x.values = as.numeric(names(empirical.freq))
theorical.prob <- sapply(x.values, FUN=function(i){return(getConditionalXGivenZ(i, 0))})

barplot(rbind(empirical.freq, theorical.prob), beside=TRUE,
        legend.text = c("Maiztasun erlatiboak", "Probabilitate teorikoak"))
```



2.4. ariketa Adieraz ezazu grafikoki $E_k = \frac{\sum_{i=1}^{k} x_i z_i}{\sum_i z_i}$, $k = 1, 2, \ldots, n$-rako, (x_i, z_i) sortu dugun `joint.sample` `data.frame`eko i. lerroko balioak izanik. Zer ikusten duzu? Arrazoitu ikusten den konbergentzia.

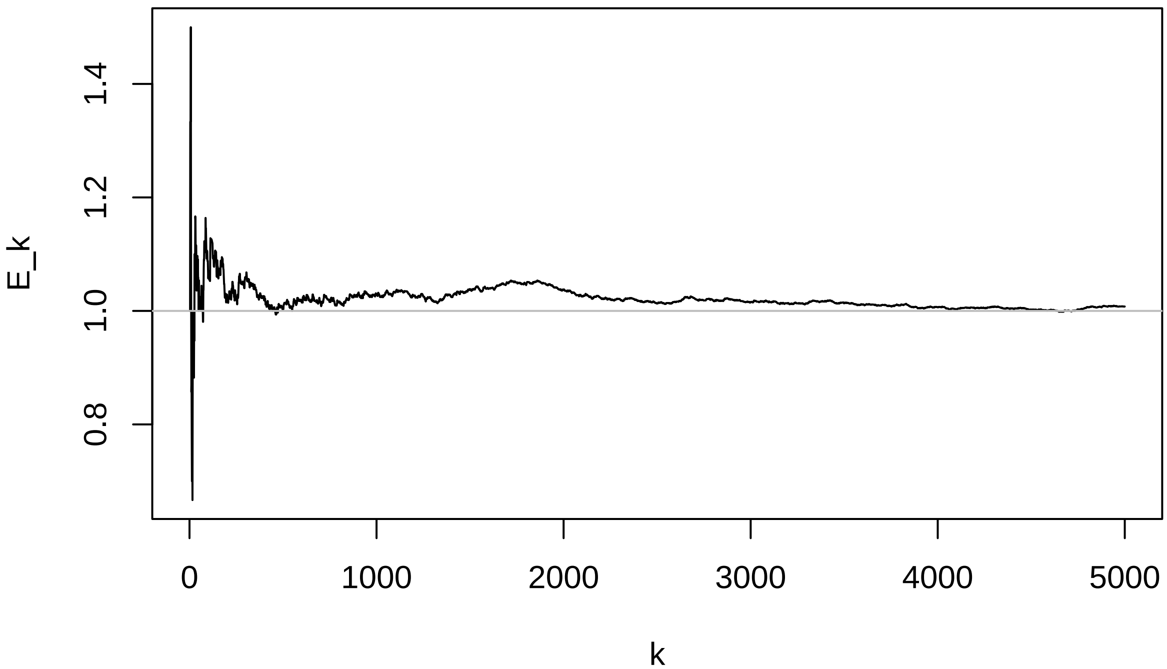
```
k <- 1:n
xz <- joint.sample$X[k]*joint.sample$Z[k]

E_k <- cumsum(joint.sample$X*joint.sample$Z)/cumsum(joint.sample$Z)

plot(k, E_k, type = "l")
abline(h=getConditionalExpectedXgivenZ(1), col = "gray")
```



Konbergentzia $E(X|Z=1)$ da, izan ere $x_k z_k = 0$ da $z_k = 0$ denean, ondorioz $Z = 1$ denean $x$ren bataz bestekoa daukagu.

2.5. ariketa Errepikatu aurreko atala $F_k = \frac{\sum_{i=1}^{k} x_i(1-z_i)}{\sum_i (1-z_i)}$ baliorako.

```
k <- 1:n
xz <- joint.sample$X[k]*(1 - joint.sample$Z[k])

E_k <- cumsum(joint.sample$X*(1 - joint.sample$Z))/cumsum(1 - joint.sample$Z)

plot(k, E_k, type = "l")
abline(h=getConditionalExpectedXgivenZ(0), col = "gray")
```


E_k
1.4
1.2
1.0
0.8
0
1000
2000
3000
4000
5000
k