24102023_Statistical_Learning

Mattia G.

2023-10-24

R Markdown

```
########################################
#    PROBABILITY DISTRIBUTIONS
# -- discrete Random Variables --
########################################

# sample from a URN of uppercase letters

urn <- LETTERS

sample(urn, 1)
```

```
## [1] "A"
```

```
sample(urn, 5)
```

```
## [1] "V" "M" "D" "N" "K"
```

```
urn <- LETTERS[1:5]
sample(urn, 5)
```

```
## [1] "A" "D" "E" "C" "B"
```

```
sample(urn, 5, replace=TRUE)
```

```
## [1] "C" "B" "B" "E" "E"
```

```
# Random seed: state of the random number generator in R
# set.seed(): function to specify  seeds

set.seed(123)
sample(urn, 1)
```

```
## [1] "C"
```

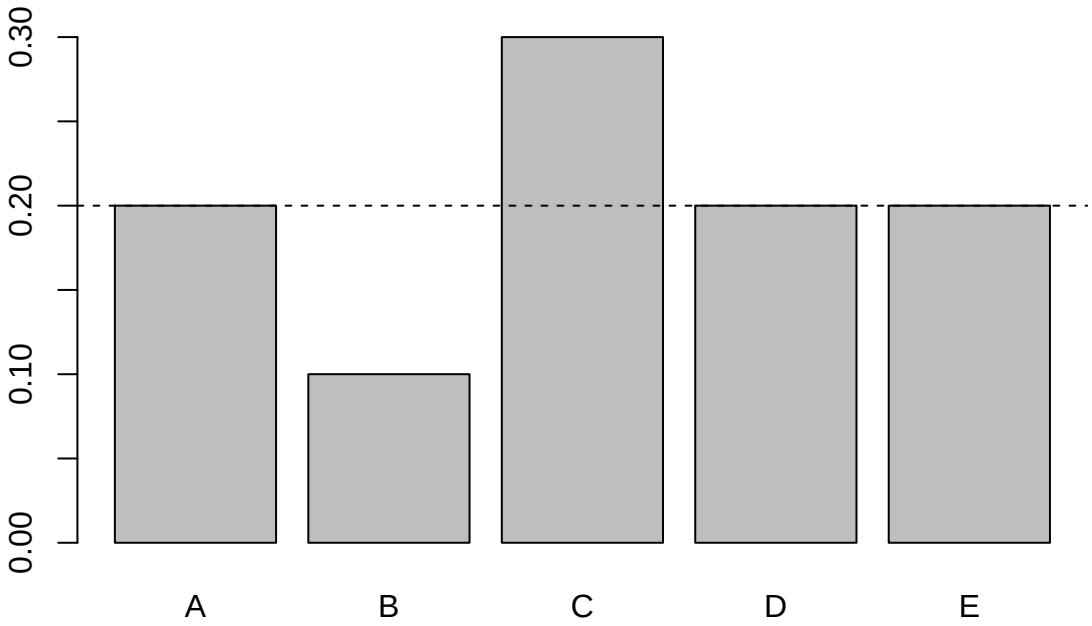
```
sample(urn, 5)
```

```
## [1] "C" "B" "D" "E" "A"
```

```
# compare empirical behavior with expected behavior

n <- 10 # increase up to 10^6
my.sample <- sample(urn, n, replace=TRUE)

barplot(table(my.sample)/n)
abline(h=0.20, lty=2)
```




```
# urn for Bernoulli random variable

urn <- c(rep(1, 7), rep(0,3))
urn
```

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

```
sample(urn, 1)
```

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

```
# (population) parameters

mu <- 0.7
sigma2 <- 0.7*0.3
sigma2
```

```
## [1] 0.21
```

```
sigma <- sqrt(sigma2)
sigma
```

```
## [1] 0.4582576
```

```
# sample

n <- 5
n <- 100
n <- 100000
out <- sample(urn, n, replace = TRUE)

# sample statistics

x.bar <- mean(out)
x.bar
```

```
## [1] 0.69823
```

```
mu-x.bar
```

```
## [1] 0.00177
```

```
s2 <- var(out)
s2
```

```
## [1] 0.210707
```

```
sigma2-s2
```

```
## [1] -0.0007069742
```

```
# binomial distribution

n <- 10
out <- sample(urn, n, replace=TRUE)
x <- sum(out)
x
```

```
## [1] 6
```

```
# rbinom() function

# binomial
rbinom(4, size=10, prob=0.7)
```

```
## [1] 6 8 6 5
```

```
# Bernoulli
rbinom(1, size=1, prob=0.7)
```

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