

Investigation of the exponential distribution

Sérgio Cordeiro

Overview

The exponential distribution

Simulations

We ran 1000 independent simulations of exponentially distributed data then checked whether the result conform to the normal distribution. We choose 0.2 to lambda (a.k.a rate), the value that characterizes the exponential distribution, and 40 to sample size. For each sample we calculate the mean and the standard deviation; the expected value for both is 5 (the reciprocal of the rate). We could then compare the expected and the simulated values.

```
lambda <- 0.2
ssize <- 40
snum <- 1000
samples <- t(sapply(1:snum, function(i) { rexp(ssize, lambda) })))
avg <- sapply(1:snum, function(i) { mean(samples[i, ]) } )
sdev <- sapply(1:snum, function(i) { sd(samples[i, ]) } )
```

The histograms that follow show the distribution of the mean of the samples and also the distribution of three of them, randomly chosen. The plots show clearly that the distribution of the means is bell-shaped while the sample sets are not. The actual and theoretical values of the means are indicated too.

```

todisp <- round(runif(3, min = 1, max = snum))

# auxiliary functions to plot histograms
showhist <- function(data, comp1, comp2, comp3, tval, limits = NULL, tvar = 0, mode
= "m") {
  # data: vector with data to be displayed
  # comp: additional data to compare with
  # tval: theoretical value to display
  par(mfcol = c(2,2), lty = 1);
  plothist(data[[1]], data[[2]], data[[3]], tval, limits, tvar, mode)
  plothist(comp1[[1]], comp1[[2]], comp1[[3]], tval, limits, tvar, mode)
  plothist(comp2[[1]], comp2[[2]], comp2[[3]], tval, limits, tvar, mode, TRUE)
  plothist(comp3[[1]], comp3[[2]], comp3[[3]], tval, limits, tvar, mode)
}

plothist <- function(data, title, color, tval, limits, tvar = 0, mode = "m", leg = F
ALSE) {
  h <- hist(data, plot = FALSE)
  if(is.null(limits)) {
    limits = range(h$breaks)
  }
  hist(data, xlab = "", main = "", ylab = title, col = color, xlim = limits)
  avg <- mean(data)
  abline(v = tval, lty = 1, col = "black")
  abline(v = avg , lty = 1, col = "green")
  if(mode == "m") {
    # shows details for the means only
    txt <- "means"
    type <- 1
  }
  if(mode == "v") {
    # shows details for the variance
    type = 2
    fact <- c(-3, -2, -1, 1, 2, 3)
    for(i in fact) {
      abline(v = avg + i * sqrt(tvar), lty = 2, col = "black")
    }
    sdev = sd(data)
    for(i in fact) {
      abline(v = avg + i * sdev, lty = 2, col = "green")
    }
    txt <- paste("std. dev. = ", format(c(sdev, sqrt(tvar)), digits = 3))
  }
  if(mode == "o") {

```

```

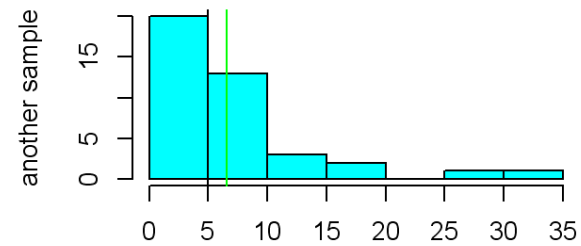
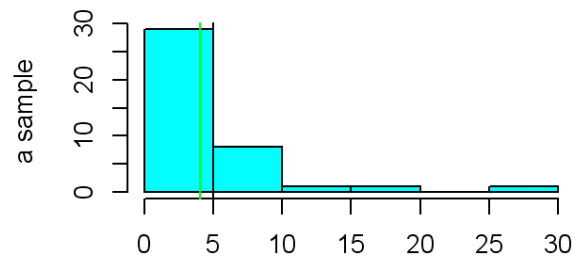
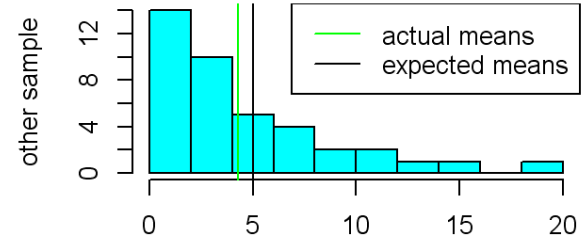
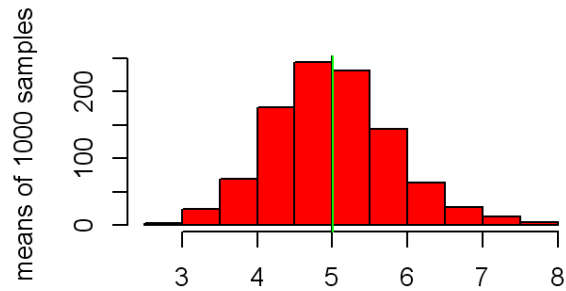
# lays a gaussian curve over the histogram
x <- seq(limits[1], limits[2], length = 50)
normal <- dnorm(x, tval, sqrt(tvar))
y <- normal * h$counts[1] / h$density[1]
lines(x, y, col = "blue", lwd = 2, yaxt = "n", lty = 1)
# adds median
abline(v = median(data), col = "green", lty = 2)
}

if(leg) {
  if((mode == "v") | (mode == "m")) {
    labels <- paste(c("actual", "expected"), txt)
    colors <- c("green", "black")
  }
  if(mode == "o") {
    labels <- c("standard normal", "median")
    colors <- c("blue", "green")
    type <- c(1, 2)
  }
  legend("topright", labels, col = colors, lty = type)
}

}

# defines values, title, color
df0 <- list(avg, "means of 1000 samples", "red")
df1 <- list(samples[todisp[1],], "a sample", "cyan")
df2 <- list(samples[todisp[2],], "other sample", "cyan")
df3 <- list(samples[todisp[3],], "another sample", "cyan")
# plots
showhist(df0, df1, df2, df3, 1/lambda)

```



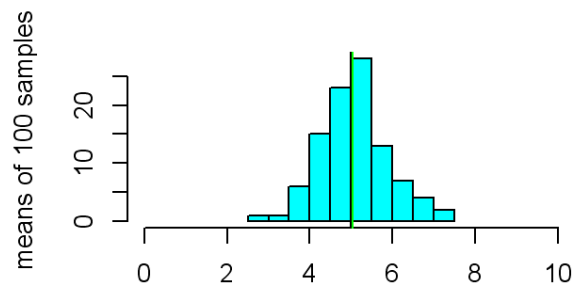
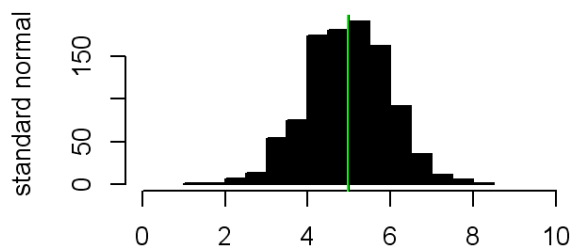
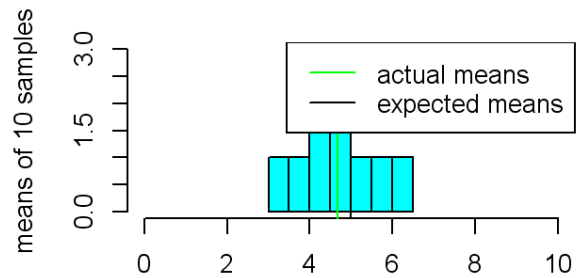
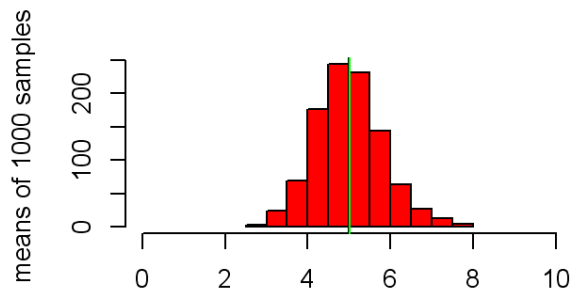
Sample Mean versus Theoretical Mean

In order to further investigate how well the distribution of the means fits to the normal distribution, we plotted them side by side. We included two other examples to indicate the dependence of that approximation to the number of samples. The horizontal scale is the same in all plots to make the comparison easier. We can see the distribution approaches to the normal as the number of samples increase.

```

# chooses random sets of samples
ex.snum <- c(10,100)
todisp <- runif(ex.snum[1], min = 1, max = snum)
ex.samples1 <- samples[todisp,]
avg.1 <- sapply(1:ex.snum[1], function(i) { mean(ex.samples1[i,]) } )
todisp <- runif(ex.snum[2], min = 1, max = snum)
ex.samples2 <- samples[todisp,]
avg.2 <- sapply(1:ex.snum[2], function(i) { mean(ex.samples2[i,]) } )
# creates normal distribution
normal <- rnorm(1000, 1/lambda, 1)
# defines values, title, color
df0 <- list(avg, "means of 1000 samples", "red")
df1 <- list(normal, "standard normal", "black")
df2 <- list(avg.1, "means of 10 samples", "cyan")
df3 <- list(avg.2, "means of 100 samples", "cyan")
# plots
showhist(df0, df1, df2, df3, 1/lambda, limits = c(0,2/lambda))

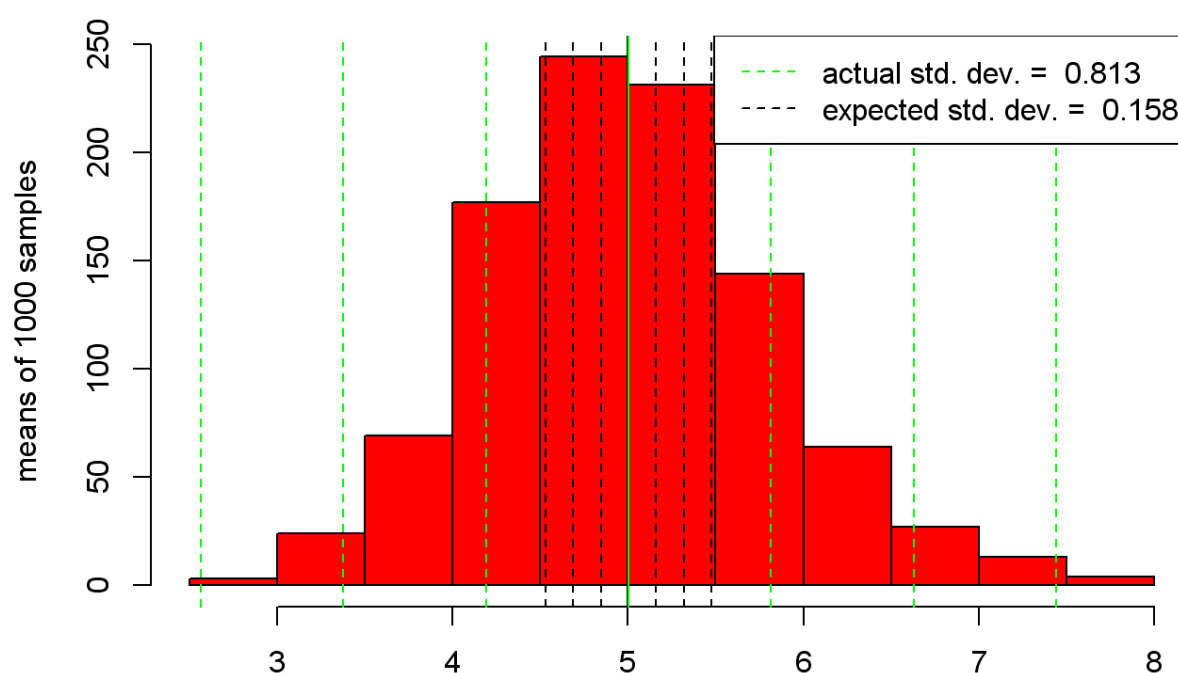
```



Sample Variance versus Theoretical Variance

The variance could be shown in the same plots as the means but this could make some details hard to see so a specific plot is provided. It shows where the actual and expected multiples of the standard deviation lie in the distribution. According to the Central Limit Theorem, the expected variance is the population variance divided by the number of samples. The population variance is that of the exponential distribution, i.e. the square of its standard deviation, given by $1/\lambda$. We can see that the number of samples high enough to result in so low a dispersion than that of a perfect normal.

```
par(mfcol = c(1,1));
variance <- (1/lambda)^2 / 1000
plothist(df0[[1]], df0[[2]], df0[[3]], 1/lambda, limits = NULL, tvar = variance , mode = "v", leg = TRUE)
```



Distribution

Finally, we can plot the normal distribution over the histogram to have another visual clue to how well both distributions fit. We added the median line that shows the distribution is slightly skewed.

```

plothist(df0[[1]], df0[[2]], df0[[3]], 1/lambda, limits = NULL, tvar = variance, mod
e = "o", leg = TRUE)

```

