STA5075: Practical 8

Jessica Stow (STWJES003@myuct.ac.za)

2025-02-07

Question 1:

Use a for loop to calculate cumulative probabilities from the counts produced by the hist() function below.

```
x <- rnorm(1000)
x.hist <- hist(x, plot = FALSE)
str(x.hist)
```

```
## List of 6
##  $ breaks  : num [1:13] -3 -2.5 -2 -1.5 -1 -0.5 0 0.5 1 1.5 ...
##  $ counts  : int [1:12] 4 21 55 87 129 188 216 136 97 44 ...
##  $ density : num [1:12] 0.008 0.042 0.11 0.174 0.258 0.376 0.432 0.272 0.194 0.088 ...
##  $ mids    : num [1:12] -2.75 -2.25 -1.75 -1.25 -0.75 -0.25 0.25 0.75 1.25 1.75 ...
##  $ xname   : chr "x"
##  $ equidist: logi TRUE
##  - attr(*, "class")= chr "histogram"
```

```
counts <- x.hist$counts/1000 # relative frequency counts

cum_prob <- numeric(length(counts)) # create a vector of same length

for (i in 2:length(counts)) {
  cum_prob[i] <- counts[i] + cum_prob[i-1]
}
```

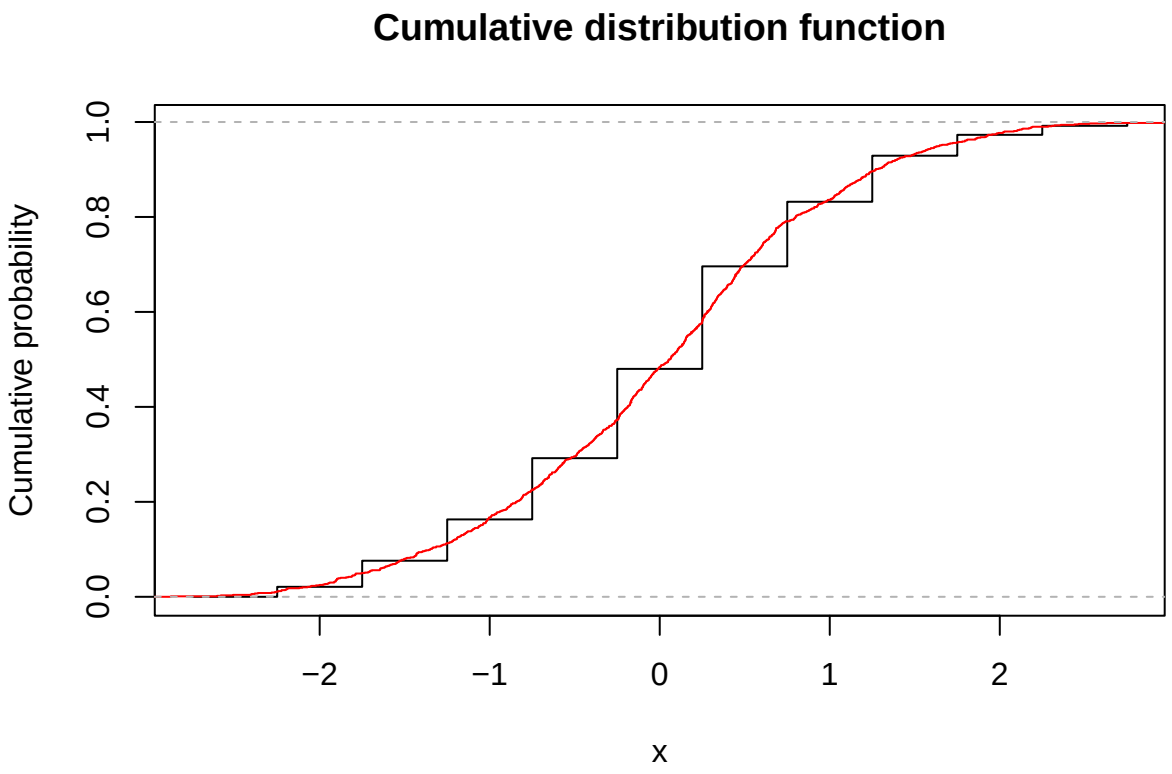
Question 2 & 3

Plot this empirical cumulative distribution function. Use a step function.

Add lines(ecdf(x), col = "red") to compare.

```
# using the middle of the bins for plotting
plot(x.hist$mids, cum_prob,
     type = "s", # right side step function
     ylab = "Cumulative probability",
     xlab = "x",
     main = "Cumulative distribution function")

# Add lines(ecdf(x), col = "red") to compare.
lines(ecdf(x), col = "red")
```



Question 4

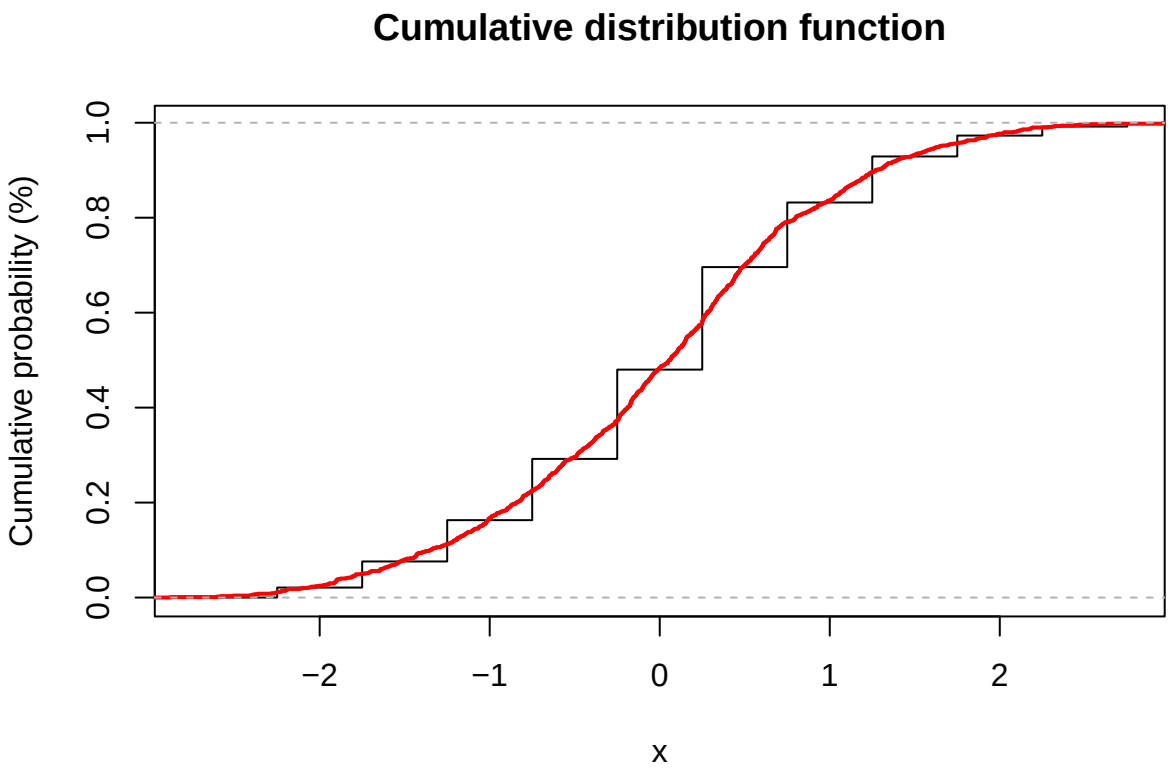
Make this into a function and apply the function to the vector x.

```r
# make function
cdf <- function(x) {
  x.hist <- hist(x, plot = FALSE)
  counts <- x.hist$counts/length(x) # relative frequency counts
  n <- length(counts)

  cum_prob <- numeric(n) # create a vector of length n

  for (i in 2:n) {
    cum_prob[i] <- counts[i] + cum_prob[i-1]
  }
  plot(x.hist$mids, cum_prob,
     type = "s", # right side step function
     xlab = "x",
     ylab = "Cumulative probability (%)",
     main = "Cumulative distribution function")
  lines(ecdf(x), col = "red", lwd = 2)
}

# apply the function to the vector x.
cdf(x)
```



Question 5

Test your function with other data.

```
uuu <- runif(50000)
cdf(uuu)
```