Cramer-von Mises two sample test

Stefan Kapunac 157/2015

- Dva obeležja - $\mathbf{X}$ i $\mathbf{Y}$
- $x_1, ..., x_n, y_1, ..., y_m$ - objedinjeni uzorak
- $F_n(x)$ - empirijska funkcija za $\mathbf{X}$
- $G_m(x)$ - empirijska funkcija za $\mathbf{Y}$
- $H_{n+m}(x)$ - empirijska funkcija za objedinjen uzorak $\mathbf{X}$ i $\mathbf{Y}$
- T - test statistika

$$T = \frac{nm}{n+m} \int_{-\infty}^{\infty} ([F_n(x) - G_m(x)]^2 dH_{n+m}(x)) \tag{1}$$

$$T = \frac{nm}{(n+m)^2} (\sum_{i=1}^{n} [F_n(x_i) - G_m(x_i)]^2 + \sum_{i=1}^{m} [F_n(y_i) - G_m(y_i)]^2) \tag{2}$$

- r - rang $\mathbf{X}$ u objedinjenom sortiranom uzorku
- s - rang $\mathbf{Y}$ u objedinjenom sortiranom uzorku

$$F_n(x_i) - G_m(x_i) = \frac{i}{n} - \frac{r_i - i}{m}$$
$$F_n(y_i) - G_m(y_i) = \frac{s_i - i}{n} - \frac{i}{m}$$

$$T = \frac{nm}{(n+m)^2} (\sum_{i=1}^{n} [\frac{r_i}{m} - i(\frac{1}{n} + \frac{1}{m})]^2 + \sum_{i=1}^{m} [\frac{s_i}{n} - i(\frac{1}{n} + \frac{1}{m})]^2) \tag{3}$$

$$T = \frac{1}{(n+m)^2} (\frac{n}{m} \sum_{i=1}^{n} [r_i - i\frac{n+m}{n}]^2 + \frac{m}{n} \sum_{i=1}^{m} [s_i - i\frac{n+m}{m}]^2) \tag{4}$$

- za $m = n$:

$$T = \frac{1}{4n^2} (\sum_{i=1}^{n} [r_i - 2i]^2 + \sum_{i=1}^{n} [s_i - 2i]^2) \tag{5}$$

- *Literatura*: link

```r
cvm <- function(x, y)
{
  n = length(x)
  m = length(y)
  spojeno = c(x, y)
  # rang x
  r = rank(spojeno)[1:n]
  # rang y
  s = rank(spojeno)[(n+1):(n+m)]
  # m != n:
  a = 0
  a = sum( (r - (n + m) / n * (1:n))^2 )

  b = sum( (s - (n + m) / n * (1:m) )^2 )

  t1 = 1 / (n + m)^2 * (n/m * a + m/n * b)

  # m = n:
  # tmp = sum( (r - 2 * (1:i))^2 + (s - 2 * (1:i))^2 )
  #
  # t2 = 1 / (4 * n^2) * tmp


  return(t1)
  # return(c(t1, t2))
}
```

```r
x = runif(n = 30, min = 0, max = 1)
y = rnorm(n = 30, mean = 0.5, sd = 1 / sqrt(12))
cvm(x, y)
```

```
## [1] 10.25722
```