1

Τετάρτη 10/03/21:

Test του Cramer - von Mises

Θα δείξουμε πως καταλήγουμε από την σχέση

$$\hat{W} = n \int \{ \hat{F}(x) - F(x,\hat{\theta}) \}^2 dF(x,\hat{\theta})$$ στην σχέση

$$\hat{W} = \sum_{i=1}^{n} \left\{ \frac{2i-1}{2n} - F(x_{(i)},\hat{\theta}) \right\}^2 + \frac{1}{12n} \quad (1)$$

Αρχικά θυμόμαστε ότι η εμπειρική αθροιστική συνάρτηση κατανομής δίνεται από την σχέση (2) $\hat{F}(x) = n^{-1} \sum_{i=1}^{n} I\{x_i \leq x\} = \begin{cases} 0, & x \leq X_{(1)} \\ \frac{i}{n}, & X_{(i)} \leq x \leq X_{(i+1)} \\ 1, & X_{(n)} \leq x \end{cases}$

όπου $X_{(1)}, \dots, X_{(n)}$ το διατεταγμένο δείγμα $X_1, \dots X_n$ (κατά αύξουσα σειρά)

Θέτουμε $F(x,\theta) \equiv F_0(x)$ για ευκολία, τότε

$$\int \{ \hat{F}(x) - F_0(x) \}^2 dF_0(x) = \int_{-\infty}^{X_{(1)}} \{ -F_0(x) \}^2 dF_0(x) + \sum_{i=1}^{n-1} \int_{X_{(i)}}^{X_{(i+1)}} \left\{ \frac{i}{n} - F_0(x) \right\}^2 dF_0 + \int_{X_{(n)}}^{+\infty} \{ 1 - F_0(x) \}^2 dF_0(x) \quad (3)$$

Έστω $U_i = F_0 X_{(i)}$, $i = 1, \dots n$ έχουμε

$$\int_{-\infty}^{X_{(1)}} \{ -F_0(x) \}^2 dF_0(x) = \int_{-\infty}^{X_{(1)}} F_0^2(x) dF_0(x) = \left. \frac{F_0^3(x)}{3} \right|_{-\infty}^{X_{(1)}} = \frac{F_0^3(X_{(1)})}{3} = \frac{U_1^3}{3} \quad (4)$$

Επίσης $$\int_{X_{(n)}}^{\infty} \{ 1 - F_0(x) \}^2 dF_0(x) = \left. \frac{\{F_0(x) - 1\}^3}{3} \right|_{X_{(n)}}^{+\infty} = - \frac{\{F_0(X_{(n)}) - 1\}^3}{3}$$

$$= - \frac{(U_n - 1)^3}{3} \quad (5)$$

$$\sum_{i=1}^{n-1}\int_{x(i)}^{x(i+1)}\left\{\frac{i}{n}-F_0(x)\right\}^2 dF_0(x) = \frac{1}{3}\sum_{i=1}^{n-1}\left[\left(u_{i+1}-\frac{i}{n}\right)^3-\left(u_i-\frac{i}{n}\right)^3\right]$$

$$= \frac{1}{3}\sum_{i=1}^{n-1}\left[u_{i+1}^3-u_i^3+\frac{3i^2}{n^2}(u_{i+1}-u_i)-\frac{3i}{n}(u_{i+1}^2-u_i^2)\right] \quad (6)$$

Παρατηρούμε ότι $\sum_{i=1}^{n-1} i^2(u_{i+1}-u_i) = \sum_{i=2}^{n}(i-1)^2 u_i - \sum_{i=1}^{n-1} i^2 u_i =$

$$= (n-1)^2 u_n - u_1 - \sum_{i=2}^{n-1}(2i-1)u_i = n^2 u_n - \sum_{i=1}^{n}(2i-1)u_i \quad (7)$$

Βάσει της (7), η (6) γίνεται

$$\frac{1}{3}\sum_{i=1}^{n-1}\left[u_{i+1}^3-u_i^3+\frac{3i^2}{n^2}(u_{i+1}-u_i)-\frac{3i}{n}(u_{i+1}^2-u_i^2)\right] =$$

$$= \frac{1}{3}u_n^3-\frac{1}{3}u_1^3+\left(u_n-\sum_{i=1}^{n}\frac{2i-1}{n^2}u_i\right)-\left(u_n^2-\sum_{i=1}^{n}\frac{u_i^2}{n}\right) \quad (8)$$

Αντικαθιστώντας τις (4), (5), (8) στην (3) παίρνουμε

$$\int\left\{F(x)-F_0(x)\right\}^2 dF_0(x) = \frac{u_1^3}{3}+\frac{1}{3}u_n^3-\frac{1}{3}u_1^3+\left(u_n-\sum_{i=1}^{n}\frac{2i-1}{n^2}u_i\right)-$$

$$-\left(u_n^2-\sum_{i=1}^{n}\frac{u_i^2}{n}\right)-\frac{(u_n-1)^3}{3} =$$

$$= \frac{1}{3}+\frac{1}{n}\sum_{i=1}^{n}\left(u_i^2-\frac{2i-1}{n}\right) = \frac{1}{3}+\frac{1}{n}\sum_{i=1}^{n}\left(u_i-\frac{2i-1}{2n}\right)^2-\frac{1}{n}\sum_{i=1}^{n}\left(\frac{2i-1}{2n}\right)^2$$

$$= \frac{1}{12n^2}+\frac{1}{n}\sum_{i=1}^{n}\left(u_i-\frac{2i-1}{2n}\right)^2$$

γιατί $\sum_{i=1}^{n}(2i-1)^2 = \frac{1}{3}n(4n^2-1)$.

Άσκηση: Θέτουμε $t = F(x,\theta)$ η οποία συνεπάγεται ότι $F^{-1}(t,\theta) = x$

οπότε $\hat{F}(x) = n^{-1} \sum_{i=1}^{n} I(X_i \le x) \overset{t=F(x,\theta)}{=} n^{-1} \sum_{i=1}^{n} I(X_i \le F^{-1}(t,\theta)) =$

$= n^{-1} \sum_{i=1}^{n} I(F(X_i,\theta) \le t, \theta) \equiv G_n(t)$ (10)

$t = F(x,\theta) \Rightarrow dt = dF(x,\theta)$ επομένως

$n \int (\hat{F}(x) - F(x,\theta))^2 dF(x,\theta) = n \int (G_n(t) - t)^2 dt.$ (11)

Για $Y_1 = F(X_1,\theta) \ldots Y_n = F(X_n,\theta)$ τότε

$G_n(t) = n^{-1} \sum_{k=1}^{n} I(Y_k \le t) = \frac{k}{n} I(Y_{(k)} \le t)$ για $G_n(t) \in \left[\frac{k}{n}, \frac{k+1}{n}\right]$ (12)

$n \int (G_n(t) - t^2)^2 dt = \frac{1}{n} \sum_{n=0}^{n-1} \int_{k/n}^{(k+1)/n} \left\{ k I(Y_{(k)} \le t) - t \right\}^2 dt.$ (13)

Μπορούμε να φτάσουμε από την (13) στην (9)?

Κατανομή του τεστ Cramer-von Mises.

Δ Ένας τρόπος για να εκτιμήσουμε την κατανομή του τεστ είναι να υπολογίσουμε π.χ. 1000 τιμές του τεστ και να χρησιμοποιήσουμε το ιστόγραμμα για να εκτιμήσουμε την σ.π.π.

Έστω π.χ. 1000 δεδομένα από τυπική κατανομή. Θα ελέγξουμε την ορθότητα της προσαρμογής στην τυπική κανονική α.σ.κ.

```
library(goftest); require(graphics)
n<-1000
TestVals<-vector(length=n, mode="numeric")
set.seed(4)
for (i in 1:n)
{
```

```
x <- norm(100)
TestVals[i] <- cvm.test(x, "pnorm") $ statistic
}
hist(TestVals, freq=F)
lines(density(TestVals), col=2)
```

Το ότι οι παράμετροι της κατανομής είναι σταθ. αριθμοί σημαίνει ότι είναι τα ίδια νούμερα και δεν αλλάζουν με το κάθε δείγμα που παράγαμε

Σε αντίθεση με το τεστ Kolmogorov-Smirnov για το τεστ Cramer-von Mises υπάρχει διόρθωση όταν οι παράμετροι εκτιμώνται από τα δεδομένα.

στην R.

```
n<-1000 ; set.seed(4)
TestValsie <- vector(length=n, mode="numeric")
for (i in 1:n)
{
  x <- rnorm(100)
  TestValsie[i] <- cvm.test(x, "pnorm", mean=mean(x), sd=sd(x),
                            estimated = TRUE) $ statistic.
}
hist(TestValsie, freq=F)
lines(density(TestValsie), col=2)
```

Test Anderson-Darling :

$$A = n \int \frac{\{F(x) - F(x,\theta)\}^2}{F(x)(1-F(x))} \, dF(x).$$

Χρησιμοποιώντας την $\hat{F}(x)$ στη θέση της $F(x)$ προκύπτει η εκτιμήτρια $\hat{A}$ της A.

Παρατηρούμε ότι

$$\frac{1}{n}\hat{A} = \int \frac{\{\hat{F}(x) - F(x,\theta)\}^2}{F(x,\theta)(1-F(x,\theta))} dF(x,\theta) =$$

$$= \int_{-\infty}^{x(1)} \frac{\{0 - F(x,\theta)\}^2}{F(x,\theta)(1-F(x,\theta))} dF(x,\theta) + \int_{x(1)}^{x(2)} \frac{\{\hat{F}(x) - F(x,\theta)\}^2}{F(x,\theta)(1-F(x,\theta))} dF(x,\theta) + \dots +$$

$$+ \int_{x(n)}^{+\infty} \frac{\{1 - F(x,\theta)\}^2}{F(x,\theta)(1-F(x,\theta))} dF(x,\theta) \qquad (14)$$

Θέτω $F(x,\theta) = F_0(x)$ για συντομία τότε

$$\int_{-\infty}^{x(1)} \frac{F_0^2(x)}{F_0(x)(1-F_0(x))} dF_0(x) = \int_{-\infty}^{x(1)} \frac{F_0(x)}{1-F_0(x)} dF_0(x) = \int_0^{F_0(x(1))} \frac{u}{1-u} du =$$

$$= \Big[-u - \log(1-u)\Big]_0^{F_0(x(1))} \qquad (15)$$

$$\int_{x(n)}^{\infty} \frac{\{1 - F_0(x)\}^2}{F_0(x)(1-F_0(x))} dF_0(x) = \int_{x(n)}^{\infty} \frac{1-F_0(x)}{F_0(x)} dF_0(x) = \int_{F_0(x(n))}^{1} \frac{1-u}{u} du = \Big[\log u - u\Big]_{F_0(x(n))}^{1} \qquad (16)$$

Για $1 < i < n$.

$$\int_{x(i)}^{x(i+1)} \frac{\{\hat{F}(x) - F_0(x)\}^2}{F_0(x)(1-F_0(x))} dF_0(x) = \int_{F_0(x(i))}^{F_0(x(i+1))} \frac{(\frac{i}{n} - u)^2}{u(1-u)} du =$$

$$= \left[\frac{i^2}{n^2}\log u - \frac{(i-1)^2}{n^2}\log(1-u) - u\right]_{F_0(x(i))}^{F_0(x(i+1))} \qquad (17)$$

Αντικαθιστώντας τις (15), (16), (17) στην (14) έχουμε

$$\frac{1}{n}\hat{A} = \Big[-u - \log(1-u)\Big]_0^{F_0(X_{(1)})} + \sum_{i=2}^{n-1}\Big[\frac{i^2}{n^2}\log u - \frac{(i-1)^2}{n^2}\log(1-u) - u\Big]_{F_0(X_{(i)})}^{F_0(X_{(i+1)})}$$

$$+ \Big[\log u - u\Big]_{F_0(X_{(n)})}^{1}$$

$$= -F_0(X_{(1)}) - \log(1 - F_0(X_{(1)})) + \sum_{i=2}^{n-1}\Big[\frac{i^2}{n^2}\log u - \frac{(i-1)^2}{n^2}\log(1-u) - u\Big]_{F_0(X_{(i)})}^{F_0(X_{(i+1)})}$$

$$- 1 - \log F_0(X_{(n)}) + F_0(X_{(n)})$$

$$= -\log(1 - F_0(X_{(1)})) + \sum_{i=2}^{n-1}\Big[\frac{i^2}{n^2}\log u - \frac{(i-1)^2}{n^2}\log(1-u) - u\Big]_{F_0(X_{(i)})}^{F_0(X_{(i+1)})} - 1 - \log F_0(X_{(n)})$$

Μπορούμε από την παραπάνω εξίσωση να φτάσουμε στην

$$\hat{A} = \sum_{i=1}^{n}\frac{2i-1}{n}\Big[\log F(X_{(i)}, \theta) + \log\{1 - F(X_{(n+1-i)}; \theta)\}\Big]?$$

Κατανομή του test Anderson-Darling:

Όμοια με το Cramer von Mises εκτιμάμε την σ.π.π με ιστόγραμμα.

Ο κώδικας στη R είναι σε γενικές γραμμές ο ίδιος απλά αντί για cvm.test έχουμε την εντολή ad.test.