

Inferencia Estadística 2: Tarea 2

Sergio García Prado

sergio.garcia.prado@alumnos.uva.es

3 de diciembre de 2018

- **Enunciado:** Llevar a cabo un estudio de simulación para aproximar la potencia, en la alternativa $t(2)$ (*t de student con 2 grados de libertad*), de los tests de nivel $\alpha = 0,05$ de *Lilliefors*, *Cramer-Von-Mises*, *Anderson-Darling* y *Shapiro-Francia*, para contrastar la hipótesis nula de normalidad, basados en muestras de tamaños $n \in \{10, 30, 60, 100, 150, 300, 500\}$. Hacer una representación gráfica (potencia frente a tamaño muestral) que resuma los resultados obtenidos, asignando colores diferentes a cada test.

1. Introducción

Para el contraste de la hipótesis de normalidad se han desarrollado una gran cantidad de métodos y técnicas que tratan de responder a la pregunta sobre si los datos observados en una determinada muestra pueden asumirse normales, o por contra, siguen una distribución diferente.

$$H_0 : F \equiv \text{Normal}(\mu, \sigma^2)$$

$$H_1 : F \not\equiv \text{Normal}(\mu, \sigma^2)$$

Para ello, se puede definir un test de *bondad de ajuste* donde la hipótesis nula indica que la distribución F de los datos observados es Normal, frente a la alternativa de que no lo sea. Dicho contraste se indica en la Párrafo 1. Nótese que en este caso se ha llevado a cabo de manera compuesta (donde los parámetros de localización y escala son libres). Normalmente, cuando un contraste se define de esta manera, los parámetros μ y σ^2 son estimados a partir de los datos observados. Es en estos casos donde cobran importancia los test que se describen a continuación.

En este caso, vamos a comparar 4 tests de normalidad, los cuales describiremos a continuación: *Lilliefors*, *Cramer-Von-Mises*, *Anderson-Darling* y *Shapiro-Francia*. Posteriormente en la Sección 2 se describe la metodología llevada a cabo para la estimación de la potencia π de los distintos tests. Finalmente, en la Sección 3 se muestran los resultados obtenidos.

Lilliefors El test de bondad de ajuste de *Lilliefors* consiste en una variación del test de *Kolmogorov-Smirnov*, que estima los parámetros de localización y dispersión de la distribución normal a partir de los correspondientes valores muestrales \bar{x} y s^2 , permitiendo que la hipótesis nula sea compuesta.

$$\hat{D}_n = \sup_x | \hat{F}_n(x) - F_0(x) | \quad (1)$$

Por lo tanto, el estadístico test se define en la Ecuación 1, donde $\hat{F}_n(x)$ es la función de distribución empírica y $F_0(x)$ la de una distribución $\text{Normal}(\bar{x}, s^2)$. En este caso, se utilizan unos valores críticos menos restrictivos obtenidos por simulación montecarlo, lo cual hace que el test sea menos conservador que el original de *Kolmogorov-Smirnov*. El test rechaza la hipótesis nula para valores grandes del estadístico \hat{D}_n .

Cramer-Von-Mises El test de bondad de ajuste de *Cramer-Von-Mises* está basado en el estadístico descrito en la Ecuación 2. Rechaza la hipótesis nula para valores grandes del estadístico \hat{W}_n^2 y es más potente que el de *Lilliefors*.

$$\hat{W}_n^2 = n \int_{-\infty}^{\infty} \left(\hat{F}_n(x) - F_0(x) \right)^2 dF_0(x) \quad (2)$$

Anderson-Darling El test de bondad de ajuste de *Cramer-Von-Mises* está basado en el estadístico descrito en la Ecuación 3. Rechaza la hipótesis nula para valores grandes del estadístico \widehat{W}_n^2 y es muy similar en potencia al de *Cramer-Von-Mises*.

$$\widehat{W}_n^2 = n \int_{-\infty}^{\infty} \frac{(\widehat{F}_n(x) - F_0(x))^2}{F_0(x)(1 - F_0(x))} dF_0(x) \quad (3)$$

Shapiro-Francia Este test de bondad de ajuste se basa en una aproximación (más simple computacionalmente) del test de *Shapiro-Wilk*, basada en la correlación al cuadrado entre los valores ordenados de la muestra y los esperados de una distribución $Normal(\bar{x}, s^2)$. Dado que el estadístico consiste en una correlación, estará entre 0 y 1, rechazando la hipótesis nula para valores próximos a 0.

2. Simulación

En este trabajo se ha comparado la potencia de los tests no paramétricos para la hipótesis de normalidad, frente a la alternativa, siendo esta una distribución *t de student* con 2 grados de libertad. Esto se va a llevar a cabo para distintos tamaños muestrales $n \in \{10, 30, 60, 100, 150, 300, 500\}$. En general, la potencia de un test se define como la capacidad para rechazar la hipótesis nula, cuando esta realmente es falsa. Mas matemáticamente esto es $\pi = P(\text{rechazar } H_0 \mid H_1 \text{ sea verdadero})$.

A pesar de que para algunos contrastes de hipótesis se puede calcular analíticamente una expresión que permite expresar la potencia de un test fijando el tamaño muestral n y el nivel de significancia α , en general esto no es posible. Sin embargo, esto puede ser calculado mediante métodos de simulación conocemos la distribución de los datos en la hipótesis alternativa (y por tanto podemos generar observaciones procedentes de esta).

Para la simulación se puede recurrir a distintas técnicas, entre las que se encuentra el método de la función de distribución inversa. Este método consiste en calcular la función inversa de la función de distribución de la distribución que se pretende simular. Esto es $F^{-1}(x)$ y a partir de esta función, transformar valores simulados de una distribución $U \sim Uniforme(0, 1)$ (que pueden obtenerse de manera sencilla con un ordenador). De esta manera se consiguen observaciones pertenecientes a la distribución F . Esto se debe a la propiedad que indica que $F^{-1}(U) \equiv F$, lo cual es cierto por el método jacobiano de transformación de variables.

Entonces, una vez se tiene una muestra de variables independientes de la distribución en la alternativa, es trivial calcular la potencia del test. Esto puede llevarse a cabo promediando el número de veces que la hipótesis nula H_0 es rechazada, que converge en probabilidad a la potencia del test.

En el Código Fuente 1 se indican las sentencias necesarias para realizar dicha simulación utilizando el lenguaje R.

3. Resultados

La simulación se ha llevado a cabo mediante la generación de 1000 para cada aproximación. De esta manera se han generado los resultados que se muestran en la Tabla 1, que se han representado de manera gráfica en la Figura 1.

En cuanto a la interpretación de los resultados, se puede apreciar que conforme el tamaño de muestra n aumenta, la potencia del test π tiene hacia 1. En este caso, cuando la muestra supera las $n = 150$ observaciones, la potencia del test de normalidad es prácticamente 1. Nótese que este valor depende de la distribución en la alternativa, que cuanto más similar sea a la distribución normal, más número de observaciones será necesario utilizar para que la potencia del test sea razonable.

```

rm(list = ls())

library(nortest)

library(magrittr)
library(dplyr, warn.conflicts = FALSE)
library(tidyr, warn.conflicts = FALSE)
library(reshape2, warn.conflicts = FALSE)
library(ggplot2)

AlternativeHypothesisGenerator <- function(n) {
  rt(n, df = 2) # Generador de valores procedentes de una distribución t_2.
}

PowerTest <- function(test, n, alpha = 0.05, nsim = 1000,
                      generator = AlternativeHypothesisGenerator){
  mean(replicate(nsim, test(generator(n))$p.value < alpha))
}

LillieforsPower <- function(n, ...) {
  PowerTest(lillie.test, n, ...)
}

CramerVonMisesPower <- function(n, ...) {
  suppressWarnings(PowerTest(cvm.test, n, ...))
}

AndersonDarlingPower <- function(n, ...) {
  PowerTest(ad.test, n, ...)
}

ShapiroFranciaPower <- function(n, ...) {
  PowerTest(sf.test, n, ...)
}

results <-
  data.frame(n = c(10, 30, 60, 100, 150, 300, 500)) %>%
  rowwise() %>%
  mutate(Lilliefors      = LillieforsPower(n),
         CramerVonMises  = CramerVonMisesPower(n),
         AndersonDarling = AndersonDarlingPower(n),
         ShapiroFrancia  = ShapiroFranciaPower(n))

results.plot <- results %>%
  gather(method, value, - n) %>%
  ggplot(aes(x = n, y = value, colour = method)) +
  geom_line() +
  theme(legend.position = "top") +
  xlab("Tamaño Muestral") +
  ylab("Potencia") +
  labs(colour = "Método")

write.csv(results, "normality-testing.csv", row.names = FALSE, quote = FALSE)
ggsave("normality-testing.png", results.plot, width = 7, height = 4, dpi = 120)

```

Código Fuente 1: Potencia de los tests de normalidad utilizando como alternativa una distribución *t de student* con 2 grados de libertad.

En cuanto a la comparación entre los distintos métodos utilizados, parece que el test de *Lilliefors* es el menos potente de los 4 comparados, mientras que el de *Shapiro-Francia* es el que más potencia presenta. Mientras que los métodos de *Cramer-Von-Mises* y *Anderson-Darling* se encuentran en la zona intermedia, proporcionando una potencia muy similar.

n	<i>Lilliefors</i>	<i>Cramer-Von-Mises</i>	<i>Anderson-Darling</i>	<i>Shapiro-Francia</i>
10	0.275	0.286	0.309	0.333
30	0.581	0.663	0.678	0.725
60	0.818	0.903	0.896	0.925
100	0.965	0.976	0.985	0.992
150	0.999	0.998	0.998	0.999
300	1	1	1	1
500	1	1	1	1

Tabla 1: Resultados de la simulación.

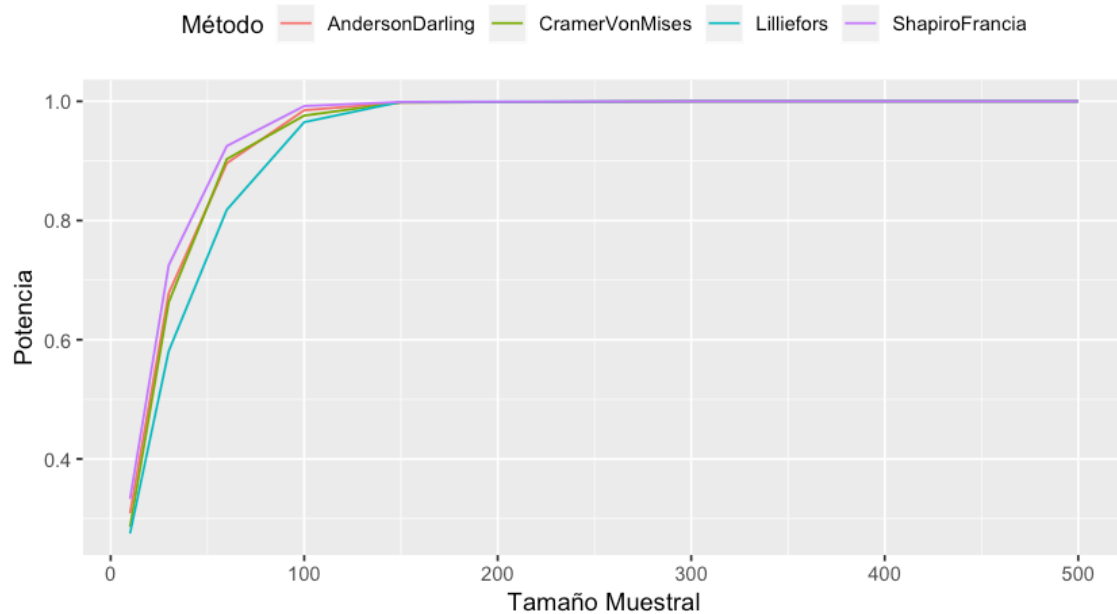


Figura 1: Resultados de la simulación.