

**UN ESTUDIO DE DIFERENTES PRUEBAS PARA EL PROBLEMA
GENERAL DE DOS MUESTRAS**

Por

CRISTIAN J. GÓMEZ HERAZO

Tesis sometida en cumplimiento parcial de los requerimientos para el grado de

MAESTRÍA EN CIENCIAS

en

MATEMÁTICAS ESTADÍSTICA

UNIVERSIDAD DE PUERTO RICO
RECINTO UNIVERSITARIO DE MAYAGÜEZ

2018

Aprobada por:

Edgardo Lorenzo González, Ph.D
Miembro, Comité Graduado

Fecha

Dámaris Santana Morant, Ph.D
Miembro, Comité Graduado

Fecha

Wolfgang A. Rolke, Ph.D
Presidente, Comité Graduado

Fecha

Sandra Coutin Rodicio, Ph.D
Representante de Estudios Graduados

Fecha

Víctor Ocasio González, Ph.D
Director del Departamento

Fecha

Resumen de Disertación Presentado a Escuela Graduada
de la Universidad de Puerto Rico como requisito parcial de los
Requerimientos para el grado de Maestría en Ciencias

UN ESTUDIO DE DIFERENTES PRUEBAS PARA EL PROBLEMA GENERAL DE DOS MUESTRAS

Por

CRISTIAN J. GÓMEZ HERAZO

2018

Consejero: Dr. Wolfgang A. Rolke
Departamento: Ciencias Matemáticas

En su forma más general, el problema de las dos muestras se refiere a si dos muestras independientes provienen de la misma distribución de probabilidad. Existe una gran variedad de métodos para probar esta hipótesis. Estudiaremos algunos de los más conocidos: Permutación, Wilcoxon, Rachas de Wald-Wolfowitz, Kolmogorov Smirnov, Anderson-Darling y Chi-Cuadrado. También desarrollamos una extensión del gráfico Cuantil-Cuantil agregando una banda de confianza para probar específicamente si las dos muestras provienen de una distribución uniforme. Estos métodos se basan en las estadísticas de orden, rango y función de distribución empírica. Compararemos las diferentes pruebas, estudiaremos su potencia para diferentes distribuciones, variando parámetros y utilizando diferentes tamaños de muestra. El poder de cada prueba se calculará mediante simulación, para lo cual creamos una aplicación web interactiva en línea (shinyapps) programada en R, disponible en: https://server-cristian.shinyapps.io/two_sample_problem/.

Los resultados muestran que no existe una prueba uniformemente más poderosa para el problema general de dos muestras. Esto se debe a que los poderes se ven afectados en gran medida por la definición de las distribuciones alternativas, sin embargo, podemos dar algunas pautas para usar las pruebas más apropiadas en una serie de casos estándares.

Abstract of Dissertation Presented to the Graduate School
of the University of Puerto Rico in Partial Fulfillment of the
Requirements for the Degree of Master of Sciences

**A STUDY OF DIFFERENT TESTS FOR THE GENERAL
TWO-SAMPLES PROBLEM**

By

CRISTIAN J. GÓMEZ HERAZO

2018

Chair: Dr. Wolfgang A. Rolke
Major Department: Mathematical Sciences

In its most general form the two-sample problem is concerned with the question whether two independent samples come from the same probability distribution. There are a wide variety of methods to test this hypothesis. We will study some of the most well known: the Permutation, Wilcoxon, Wald-Wolfowitz Runs, Kolmogorov Smirnov, Anderson-Darling and Chi-Square test. We also develop an extension to the Q-Q plot by adding a confidence band to test specifically whether the two samples come from a uniform distribution. These methods are based on the statistics of order, rank, and empirical distribution function. We will compare the different tests, studying their power for different distributions, varying parameters and using different sample sizes. The power of each test will be calculated by simulation, for which we created an interactive online web application (shinyapps) programmed in R, which is available at:

https://server-cristian.shinyapps.io/two_sample_problem/.

The results show that there is no uniformly best test for the general problem of two samples. This is because the powers are affected to a large extent by the definition of the alternate distributions, we can however give some guidelines to use the tests more appropriate in a number of standard cases.

Copyright © 2018

por

CRISTIAN J. GÓMEZ HERAZO

A la memoria de mi padre, Juvenal Gómez.

A mi madre, Martha Herazo, por su amor y apoyo incondicional.

AGRADECIMIENTOS

A Dios, en primer lugar, quien me sostiene.

Este trabajo no habría sido posible sin la influencia directa o indirecta de muchas personas a las que agradezco profundamente por estar presentes en las distintas etapas de su elaboración.

Le agradezco a mi asesor Dr. Wolfgang Rolke por manifestarme su interés en dirigir mi trabajo de grado, por su confianza, colaboración y apoyo.

A la Dra. Dámaris Santana Morant y al Dr. Edgardo Lorenzo, por su disponibilidad y colaboración.

Al Departamento de Ciencias Matemáticas del Recinto Universitario de Mayagüez por todo el apoyo brindado.

A mis amigos y compañeros que me acompañaron dentro del salón de clases, y los mejores momentos que viví durante esta etapa como estudiante de maestría, dentro y fuera del campus.

A mi familia y seres más queridos, a mi esposa Ana Cristina Moreno, por su amor y comprensión. A mi mamá, mis hermanos; ustedes me han dado su amor, apoyo y confianza.

A la memoria de mi padre. Juve, fuiste y seguirás siendo muy especial en mi vida.

Índice general

RESUMEN EN ESPAÑOL	II
ABSTRACT ENGLISH	IV
AGRADECIMIENTOS	VIII
Índice general	IX
Índice de tablas	XI
Índice de figuras	XIV
LISTA DE ABREVIATURAS	XVII
LISTA DE SIMBOLOS	XVIII
1. INTRODUCCIÓN: PROBLEMA DE DOS MUESTRAS	1
1.1. Ejemplo: Comparando distintas pruebas en poblaciones normales	2
2. MÉTODOS	4
2.1. Prueba de permutación	4
2.2. Prueba de Kolmogorov Smirnov	14
2.3. Prueba de Anderson-Darling	17
2.4. La prueba de rachas de Wald-Wolfowitz	18

2.5.	Prueba de suma de rango de Wilcoxon	23
2.6.	χ^2 para dos muestras	25
3.	GRÁFICO CUANTIL-CUANTIL PARA DOS MUESTRAS	28
	3.0.2. Construcción de los intervalos de confianza	30
	3.0.3. Comparación del poder	33
4.	SIMULACIONES	35
	4.0.1. Soporte infinito	38
	4.0.2. Soporte semi infinito	49
	4.0.3. Soporte finito (0,1)	55
	4.0.4. Elección del mejor número de intervalos (k) en la prueba de Chi-Cuadrado	64
5.	LA APLICACIÓN WEB INTERACTIVA	67
	5.1. Do you have your own data?	67
	5.2. Do you want to study the power?	71
6.	CONCLUSIONES Y TRABAJOS FUTUROS	75
	APENDICES	78
	A. PROPOSICIONES	79
	Bibliografía	82

Índice de tablas

2-1. Etiquetas originales del grupo	5
2-2. Etiquetas permutadas del grupo	6
2-3. Número de permutaciones (M) requeridas para la distribución nula exacta de las pruebas de permutación de dos muestras con tamaños $m = n$	9
3-1. Estimación del poder en función de <i>shape1</i>	34
4-1. Estimación del poder en función del tamaño de muestra. $N(\mu = 0, \sigma =$ 1) contra $t(df = 1)$	39
4-2. Estimación del poder en función del tamaño de muestra. $N(\mu = 0, \sigma =$ 1) contra $t(df = 2)$	40
4-3. Estimación del poder en función del tamaño de muestra. $N(\mu = 0, \sigma =$ 1) contra $t(df = 10)$	41
4-4. Estimación del poder en función de μ . $N(\mu = 0, \sigma = 1)$ contra $N(\mu \in$ $[-1, 1]), \sigma = 1)$	42
4-5. Estimación del poder en función de σ . $N(\mu = 0, \sigma = 1)$ contra $N(\mu =$ $0, \sigma \in (0, 2])$	43

4-6. Estimación del poder en función del tamaño de muestra. $N(\mu = 0, \sigma = 1, n \in [100, 1000])$ contra $t(df = 1, m = 1000)$	44
4-7. Estimación del poder en función del tamaño de muestra. $N(\mu = 0, \sigma = 1, n \in [100, 1000])$ contra $(df = 10, m = 1000)$	45
4-8. Estimación del poder en función μ . $N(\mu \in [-1, 1], \sigma = 1, n = 100)$ contra $N(\mu = 0, \sigma = 1, m = 1000)$	46
4-9. Estimación del poder en función de σ . $N(\mu = 0, \sigma \in (0, 2), n = 100)$ contra $N(\mu = 0, \sigma = 1, m = 1000)$	48
4-10. Estimación del poder en función del tamaño de muestra. $Exp(\lambda = 1)$ contra $F(df_1 = 2, df_2 = 5)$	49
4-11. Estimación del poder en función de <i>scale</i> . $Exp(\lambda = 1)$ contra $Weibull(shape = 1, scale \in [1, 2], shift = 0)$	50
4-12. Estimación del poder en función de λ . $Exp(\lambda = 1)$ contra $Exp(\lambda \in (0, 2])$	51
4-13. Estimación del poder en función de <i>shape</i> . $Exp(\lambda = 1)$ contra $Gamma(shape \in [0.25, 1.75])$	52
4-14. Estimación del poder en función de <i>shape</i> . $Exp(\lambda = 1, n = 100)$ contra $Gamma(shape \in [0.25, 1.75], m = 1000)$	53
4-15. Estimación del poder en función de <i>scale</i> . $Exp(\lambda = 1, n = 1000)$ contra $Weibull(shape = 1, scale \in [1, 2], shift = 0, m = 100)$	54

4-16. Estimación del poder en función de α . $U(a = 0, b = 1)$ contra $Beta(\alpha \in (0, 2], \beta = 1)$	55
4-17. Estimación del poder en función del tamaño de muestra. $U(a = 0, b = 1)$ contra $Lineal(Intercepto = 2)$	57
4-18. Estimación del poder en función del intercepto. $U(a = 0, b = 1)$ contra $Lineal(Intercepto \in [0, 2])$	58
4-19. Estimación del poder en función del tamaño de muestra. $Exponencial\ truncada(min = 0, max = 1, \lambda = 1)$ contra $Lineal(Intercepto = 2)$	59
4-20. Estimación del poder en función del tamaño de muestra. $Exponencial\ truncada(min = 0, max = 1, \lambda = 1)$ contra $Lineal(Intercepto = 1.5)$	60
4-21. Estimación del poder en función de α . $U(a = 0, b = 1, n = 1000)$ contra $Beta(\alpha \in (0, 2], \beta = 1, m = 100)$	61
4-22. Estimación del poder en función de $U(a = 0, b = 1, n = 1000)$ contra $Lineal(intercepto = 2, m \in [100, 1000])$	62
4-23. Estimación del poder en función de $U(a = 0, b = 1, n = 1000)$ contra $Lineal(intercepto \in [0, 2], m = 100)$	62
4-24. Elección del mejor k que resulta en una prueba más potente.	66
4-25. Comparación de distintas fórmulas para seleccionar el número de clases (se aproxima k al entero más cercano)	66

Índice de figuras

2-1. Valores de sd con distintos B	10
2-2. La estimación de la densidad (tipo núcleo) del gen $NM152299$, $sample1$ son las medidas en adenomas no progresados y $sample2$ los carcinomas con su boxplot y qqplot respectivos.	13
2-3. Distancia máxima (en negro) entre dos funciones de densidad empíricas.	15
3-1. Gráfico Cuantil-Cuantil para dos muestras uniforme ($n=m=100$)	29
3-2. Gráfica Cuantil-Cuantil para dos muestras uniformes en $(0, 1)$ de tamaño 100, con las bandas de confianza	32
3-3. Gráfica Cuantil-Cuantil $U(0, 1)$ versus $beta(shape1 = 2, shape2 = 0)$ de tamaño 100, con las bandas de confianza	33
3-4. Pruebas más potentes, note que la prueba de ENV es superior.	34
4-1. Pruebas más potentes para $N(\mu = 0, \sigma = 1)$ contra $t(df = 1)$	39
4-2. Pruebas más potentes para $N(\mu = 0, \sigma = 1)$ contra $t(df = 2)$	40
4-3. Potencias para $N(\mu = 0, \sigma = 1)$ contra $t(df = 10)$	41
4-4. Potencias para $N(\mu = 0, \sigma = 1)$ contra $N(\mu \in [-1, 1], \sigma = 1)$	42

4-5. Potencias para $N(\mu = 0, \sigma = 1)$ contra $N(\mu = 0, \sigma \in (0, 2])$	43
4-6. Pruebas más potentes. $N(\mu = 0, \sigma = 1, n \in [100, 1000])$ contra $t(df = 1, m = 1000)$	44
4-7. Pruebas más potentes. $N(\mu = 0, \sigma = 1, n \in [100, 1000])$ contra $(df = 10, m = 1000)$	45
4-8. Pruebas más potentes. $N(\mu \in [-1, 1], \sigma = 1, n = 100)$ contra $N(\mu = 0, \sigma = 1, m = 1000)$	47
4-9. Pruebas más potentes. $N(\mu = 0, \sigma \in (0, 2), n = 100)$ contra $N(\mu = 0, \sigma = 1, m = 1000)$	48
4-10. Potencias para $Exp(\lambda = 1)$ contra $F(df_1 = 2, df_2 = 5)$	49
4-11. Potencias para $Exp(\lambda = 1)$ contra $Weibull(shape = 1, scale \in [1, 2], shift = 0)$	50
4-12. Potencias para $Exp(\lambda = 1)$ contra $Exp(\lambda \in (0, 2])$	51
4-13. Potencias para $Exp(\lambda = 1)$ contra $Gamma(shape \in [0.25, 1.75])$	52
4-14. Potencias para $Exp(\lambda = 1, n = 100)$ contra $Gamma(shape \in [0.25, 1.75], m = 1000)$	53
4-15. Potencias para $Exp(\lambda = 1, n = 1000)$ contra $Weibull(shape = 1, scale \in [1, 2], shift = 0, m = 100)$	54
4-16. Potencias para $U(a = 0, b = 1)$ contra $Beta(\alpha \in (0, 2], \beta = 1)$	56

4-17.Potencias para $U(a = 0, b = 1)$ contra $Lineal(Intercepto = 2)$	57
4-18.Potencias para $U(a = 0, b = 1)$ contra $Lineal(Intercepto \in [0, 2])$. . .	58
4-19.Potencias para $Exponencial\ truncada(min = 0, max = 1, \lambda = 1)$ contra $Lineal(Intercepto = 2)$	59
4-20.Potencias para $U(a = 0, b = 1, n = 1000)$ contra $Beta(\alpha \in (0, 2], \beta =$ $1, m = 100)$	61
4-21.Potencias para $U(a = 0, b = 1, n = 1000)$ contra $Lineal(intercepto \in$ $[0, 2], m = 100)$	63

LISTA DE ABREVIATURAS

FDE	Función de Densidad Empírica.
MC	Simulación Monte Carlo.
H_0	Hipótesis nula.
H_1	Hipótesis alterna.
<i>iid</i>	Independiente e idénticamente distribuidas.
PT	Prueba de Permutación.
KS	Prueba de Kolmogorov Smirnov.
WT	Prueba de Suma de Rango de Wilcoxon.
AD	Prueba de Anderson Darling.
WW	Prueba de Rachas de Wald Wolfowitz.
CH	Prueba de Chi-Cuadrado.

LISTA DE SIMBOLOS

μ	Media de una variable aleatoria.
σ	Desviación estándar de una variable aleatoria.
$F(x)$	Función de distribución acumulada.
$\hat{F}(x)$	Función de distribución empírica acumulada.
$N(\mu, \sigma)$	Función de densidad normal con media μ y desviación estándar σ .
$Exp(\lambda)$	Función de densidad exponencial con parámetro λ .
$U(a, b)$	Función de densidad uniforme definida en (a, b)
$X_{(n)}$	Estadística de orden n .
α	Nivel de significancia de la prueba estadística.

Capítulo 1

INTRODUCCIÓN: PROBLEMA DE DOS MUESTRAS

En muchas áreas de la investigación científica nos encontramos con la situación que hay dos conjuntos de datos, y deseamos saber si tienen la misma distribución de probabilidad. Por ejemplo, los datos podrían ser las mediciones tomadas en dos lugares o en dos momentos diferentes. Una versión moderna de este problema se presenta de la siguiente manera: un conjunto de datos es ‘real’, también conocido como mediciones reales de un experimento, y el otro conjunto de datos es Monte Carlo (MC), diseñado para imitar el verdadero. Debemos determinar si los datos generados MC son equivalentes probabilísticamente. Esta variación del problema a menudo tiene una característica que no se encuentra generalmente en los conjuntos de datos reales, es decir, que los conjuntos de datos MC a menudo son muy grandes, mucho más grandes que los datos reales correspondientes, lo que lleva a la necesidad de comparar dos conjuntos de datos de tamaños muy diferentes.

La situación anterior corresponde al llamado problema general de dos muestras (en la literatura estadística no hay un nombre genérico para este tipo de problema). Más específicamente:

Supongamos que tenemos dos muestras independientes de dos poblaciones. Sea X_1, X_2, \dots, X_n y Y_1, Y_2, \dots, Y_m , las n y m observaciones muestrales con función de distribución acumulada F_X y F_Y respectivamente. Nuestra hipótesis de interés es

que las dos muestras provienen de la misma distribución, esto es:

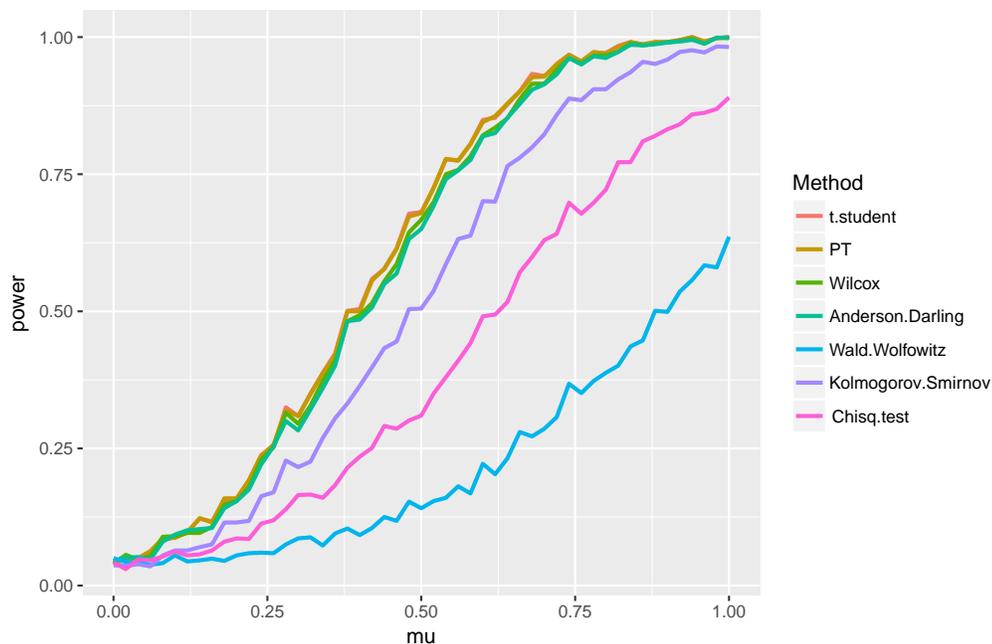
$$H_0 : F_X(x) = F_Y(x) \text{ para todo } x \in S \text{ vs } H_1 : F_X(x) \neq F_Y(x) \text{ para algún } x \in S.$$

Obsérvese que en la hipótesis nula, la distribución de F_X y F_Y no está especificada. En este estudio sólo consideramos $S \subseteq \mathbb{R}$ y datos con distribución teórica continuas.

Existe una amplia lista de métodos para probar estas hipótesis, tales como: la prueba de permutación, la prueba de Wilcoxon, Wald-Wolfowitz Runs, Kolmogorov Smirnov, Anderson-Darling, la prueba de Chi-Cuadrado. También se puede emplear métodos gráficos como el QQplot para dos muestras; estos métodos se basan en los estadísticos del orden, el rango y en la función de distribución empírica.

1.1. Ejemplo: Comparando distintas pruebas en poblaciones normales

A continuación se comparan las potencias de las pruebas para dos poblaciones normales de tamaño 50, se empieza con dos normales estándar, luego se varía el parámetro de la media (μ) en una de ellas hasta llegar a $\mu = 1$. Se realizan 1,000 simulaciones en cada una de ellas, se calculan los “p-values” y se promedia los casos en que se rechaza la hipótesis nula, usamos nivel de significancia $\alpha = 0.05$. Se obtiene el siguiente gráfico del poder para las distintas pruebas:



Note que a pesar de ser este ejemplo paramétrico (en el sentido que se conocen las distribuciones de las poblaciones), al usar la prueba de t-student se obtienen los mismos resultados en términos del poder, que usando el test de permutaciones, esto es una ventaja porque no necesitamos el supuesto de normalidad para este escenario (Bajo un contexto paramétrico, uno de los puestos fundamentales es normalidad en los datos, es decir, los datos provienen de una distribución normal). Las pruebas menos potentes en esta situación resultaron ser la prueba de Chi-Cuadrado y Wald Wolfowitz.

Capítulo 2

MÉTODOS

En este capítulo se estudian varias pruebas de hipótesis para determinar si dos distribuciones muestreadas (mutuamente excluyentes) proviene de la misma distribución.

Las pruebas estadísticas para el problema de dos muestras usualmente se clasifican en pruebas no paramétricas en el sentido que no se requiere suposiciones acerca de las distribuciones en los conjuntos de datos. En contraste, suponiendo que F y G son distribuciones normales que sólo se diferencian en su media, la prueba paramétrica t-Student es la mejor prueba conocida (en términos del poder). Si no se garantiza la normalidad de los datos, la prueba para medias es sensible al no cumplimiento de este supuesto y resulta en una prueba que pierde potencia. Bajo este escenario las pruebas no paramétricas, resultan idóneas.

Otra característica de las pruebas no paramétricas es que utilizan estadísticos cuya distribución no depende de la distribución muestreada y se conoce como pruebas de *distribución libre*

2.1. Prueba de permutación

Las pruebas de permutación fueron iniciadas por Fisher (1935) y posteriormente desarrolladas por Pitman (1937) . Estas pruebas no suponen propiedades de distribución de la población, por lo que son pruebas de *distribución libre*, de hecho, toda la información requerida para el análisis está contenida en los datos.

El argumento de esta prueba es como sigue:

Supongamos X_1, X_2, \dots, X_n y Y_1, Y_2, \dots, Y_m las observaciones muestrales, podemos etiquetar los datos de tal forma que se diferencie a qué conjunto de datos pertenece cada observación, por ejemplo: asignando 1 como etiqueta a los elementos del primer conjunto de datos y asignando 2 a todos la etiqueta de los datos en la segunda muestra.

X	Etiquetado original	Y	Etiquetado original
X_1	1	Y_1	2
X_2	1	Y_2	2
X_3	1	Y_3	2
\vdots	\vdots	\vdots	\vdots
X_n	1	Y_n	2

Tabla 2-1: Etiquetas originales del grupo

Este es el ‘agrupamiento o etiquetado original’. Al calcular un estadístico de prueba que mejor distinga entre la hipótesis alterna y la nula, la pregunta natural es si este valor (del estadístico) es demasiado grande o demasiado pequeño en comparación con lo que se espera bajo la hipótesis nula.

Supongamos que se eligió el estadístico de prueba t :

$$T = \left| \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{S_x^2}{n} + \frac{S_y^2}{m}}} \right|$$

Bajo la hipótesis alterna se espera grandes valores de t (porque \bar{X} y \bar{Y} difieren), por lo tanto se debe investigar si t es singularmente grande, es decir, debemos calcular la probabilidad:

$$p = P(T \geq t) \tag{2.1}$$

bajo H_0 .

Si la hipótesis nula es cierta, el etiquetado original no es informativo. Consideremos la nueva etiqueta de agrupación:

X	Etiquetado original	Y	Etiquetado original
X_1	1	Y_1	2
X_2	2	Y_2	1
X_3	1	Y_3	2
\vdots	\vdots	\vdots	\vdots
X_n	1	Y_n	2

Tabla 2-2: Etiquetas permutadas del grupo

Aquí, la segunda observación del primer grupo se intercambia con la segunda observación del segundo grupo. Si H_0 es cierta, entonces el valor de X_2 e Y_2 no depende en que grupo estén, y por lo tanto habría sido tan probable que se observara X_2 en el segundo grupo y Y_2 en el primero. Por tanto, el estadístico de prueba, calculada en el conjunto de datos permutados t^* , es tan probable como el estadístico de prueba calculada en los datos originales, t , al menos bajo H_0 . Este razonamiento se cumple para todas las $\binom{m+n}{m}$ permutaciones de las etiquetas sobre los dos grupos.

Denotamos el valor de la estadística de prueba calculada en la i -ésima permutación como $t^{(i)}$ ($i = 1, \dots, \binom{m+n}{m}$). Todos los valores de la estadística de prueba calculados para estos conjuntos de datos permutados son igualmente probables bajo H_0 . En este sentido, cada una de las $t^{(i)}$ en el conjunto $\{t^{(1)}, \dots, t^{(\binom{m+n}{m})}\}$ obtiene la misma probabilidad $\frac{1}{\binom{m+n}{m}}$ de ser la estadística calculada, lo que da como resultado la distribución de T . Note que en la Ecuación 2.1 definimos el “p-value” que ahora se puede calcular en base a la distribución nula exacta como:

$$p = P(T \geq t) = \frac{\text{número de } t^{(i)} \geq t}{\text{número total de } t^{(i)}} = \frac{1}{\binom{m+n}{m}} \text{número de } t^{(i)} \geq t.$$

El término permutación se refiere a la propiedad de que bajo la hipótesis nula la distribución conjunta de las observaciones de muestra no cambia cuando los subíndices (las ‘etiquetas’ que identifican las cantidades) de las observaciones se permutan. Esta propiedad se conoce como la *intercambiabilidad* de las observaciones.

Construcción de la prueba de permutación

Para la construcción de la prueba, empleamos la notación usada por (Hoeffding, 1952). En el problema de dos muestras $Z = (Z_1, \dots, Z_{(m+n)}) = (X_1, \dots, X_m, Y_1, \dots, Y_n)$ son las observaciones de la muestra agrupada, sea Ω denota el espacio muestral de Z , entonces el espacio muestral de las $m + n$ observaciones $\mathbf{z}^t = (z_1, \dots, z_{m+n})$ es igual a Ω^{m+n} . Sea $T(\mathbf{Z})$ cualquier estadístico para probar H_0 . Sea G un grupo finito de transformaciones de Ω^{m+n} en sí mismo, es decir, para cada $g \in G$ y para cada $\mathbf{z}^t = (z_1, \dots, z_{m+n}) \in \Omega^{m+n}$, $g\mathbf{z} = g(\mathbf{z}) \in \Omega^{m+n}$.

Comenzamos definiendo una función de prueba ϕ que mapea el espacio muestral Ω^{m+n} de la muestra en $[0, 1]$, que para una muestra dada, da la probabilidad con la cual la hipótesis nula debe rechazarse en el nivel de significancia α .

Denotemos por $T^{(1)}(\mathbf{z}) \leq T^{(2)}(\mathbf{z}) \leq \dots \leq T^{(M)}(\mathbf{z})$ los M valores ordenados de $T(g\mathbf{z})$, g variando en G y cardinal de G igual a M . Dado un número α , $0 < \alpha < 1$, se define k como

$$k = M - [M\alpha]$$

donde $[M\alpha]$ denota el entero más grande menor o igual a $M\alpha$. Sea $M^+(\mathbf{z})$ y $M^0(\mathbf{z})$ los números de valores $T^j(\mathbf{z})$, ($j = 1, \dots, M$) mayores que $T^k(\mathbf{z})$ e iguales a $T^k(\mathbf{z})$, respectivamente. Sea

$$a(\mathbf{z}) = \frac{M\alpha - M^+(\mathbf{z})}{M^0(\mathbf{z})}.$$

Notemos que $M^+(\mathbf{z}) \leq M - k \leq M\alpha$ y $M^+(\mathbf{z}) + M^0(\mathbf{z}) \geq M - k + 1 > M\alpha$, lo cual implica $0 \leq M\alpha - M^+(\mathbf{z}) < M^0(\mathbf{z})$, es decir

$$0 \leq a(\mathbf{z}) < 1.$$

Se define la función prueba ϕ

$$\phi(\mathbf{z}) = \begin{cases} 1 & \text{si } T(\mathbf{z}) > T^{(k)}(\mathbf{z}), \\ \alpha(z) & \text{si } T(\mathbf{z}) = T^{(k)}(\mathbf{z}), \\ 0 & \text{si } T(\mathbf{z}) < T^{(k)}(\mathbf{z}). \end{cases} \quad (2.2)$$

El siguiente teorema garantiza que en una prueba de permutación, la probabilidad de rechazar H_0 dado que H_0 es verdadera es igual a α .

Teorema 1. *Sea \mathbf{Z} la muestra agrupada de tamaño $m + n$, y sea ϕ la función de prueba (2.2) definida en términos de la estadística de prueba $T(\mathbf{Z})$. Supongamos que la hipótesis nula implica que la distribución de \mathbf{Z} es invariante bajo el grupo finito de transformaciones G bajo la hipótesis nula. Entonces,*

$$P_0(\text{rechazar } H_0) = E_0\{\phi(\mathbf{Z})\} = \alpha.$$

Demostración. Tenemos que para todo $\mathbf{z} \in \Omega^{m+n}$,

$$\sum_{g \in G} \phi(g\mathbf{z}) = M^+(\mathbf{z}) + a(\mathbf{z})M^0(\mathbf{z}) = M\alpha.$$

Dado que $M\alpha$ es constante para tamaños de muestra fijos m y n , tenemos que $M\alpha = E_0\{M\alpha\}$, de la anterior igualdad se tiene:

$$\begin{aligned} M\alpha &= E_0\{M\alpha\} = E_0\left\{\sum_{g \in G} \phi(g\mathbf{z})\right\} \\ &= \sum_{g \in G} E_0\{\phi(g\mathbf{z})\} \\ &= \sum_{g \in G} E_0\{\phi(\mathbf{z})\} \\ &= ME_0\{\phi(\mathbf{Z})\} \end{aligned} \quad (2.3)$$

$$\therefore E_0\{\phi(\mathbf{Z})\} = \alpha. \quad \square$$

Aproximación de la distribución nula exacta en la prueba de permutación

Para muestras grandes, el número de permutaciones requerida para calcular la distribución nula exacta, crece rápidamente. Para el caso $m = n$ tenemos:

$n = m$	$M = \binom{m+n}{m}$
5	252
10	184756
15	155117520
20	$1.378465288 \times 10^{11}$
30	$1.182645816 \times 10^{17}$
40	$1.075072087 \times 10^{23}$
50	$1.008913445 \times 10^{29}$

Tabla 2–3: Número de permutaciones (M) requeridas para la distribución nula exacta de las pruebas de permutación de dos muestras con tamaños $m = n$

Con las herramientas computacionales actuales no es posible calcular la distribución nula exacta para tamaños de muestras grandes, una alternativa para aproximar dicha distribución se denomina diversamente como: ‘remuestreo’, ‘aleatorización’, ‘aleatorización aproximada’, ‘permutaciones muestreadas’, ‘re-aleatorización’, o simplemente mediante simulaciones de Monte Carlo, en la que se examina un subconjunto relativamente pequeño de todas las permutaciones posibles. En lugar de enumerar todas las permutaciones M , se muestrean aleatoriamente un gran número de permutaciones de G , teniendo cada permutación la misma probabilidad de ser seleccionada. Digamos que realizamos B de tales permutaciones. Este procedimiento se repite B veces, y para cada repetición se calcula la estadística de prueba. Del mismo modo que para la construcción de la distribución nula exacta, dejemos que $T^{(1)} \leq T^{(2)} \leq \dots \leq T^{(B)}$ denote las estadísticas de prueba ordenadas. A partir de

ahora, los cálculos son como antes, pero M reemplazado por B , que suele ser mucho más pequeño que M .

Para los cálculos del p – *valor* basados en B permutaciones muestreadas, la aproximación normal asintótica a la distribución binomial se puede aplicar para el cálculo de las desviaciones estándar en el valor p estimado, a saber, $\sqrt{\frac{p(1-p)}{B}}$.

La siguiente gráfica muestra los valores de la desviación estándar para distintos valores de p , con $B = 100, 1000, 10000, 100000$

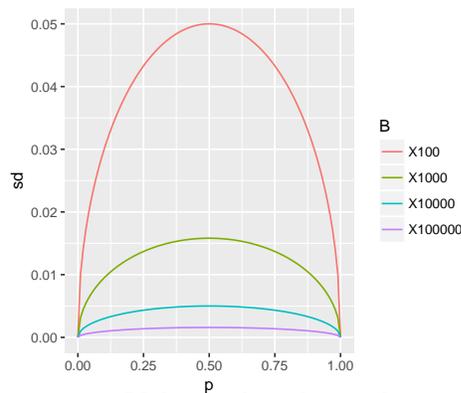


Figura 2–1: Valores de sd con distintos B

Como se desconoce el p – *valor* del experimento, el tamaño (B) de las permutaciones muestreadas se puede emplear considerando la situación más conservadora de $p = 0.5$, por tanto para con $B = 10000$, el intervalo de confianza más conservador al 95 % en p – *valor* tiene un ancho de ± 0.0096

Prueba de permutación en R Project

El siguiente código calcula el p – *valor* exacto en la prueba de permutación usando el estadístico de la prueba t para la media.

```
permutation_test <- function(data1,data2){
  library(gtools)
  T = abs((mean(data1)-mean(data2)) /
    sqrt( sd(data1)^2/length(data1) +sd(data2)^2 /length(data2)))
  mx = max(length(data1),length(data2))
  data = c(data1,data2)
```

```

C = t(combn(x = data,mx))
tvalue = rep(0,length(C[,1]))
for(i in 1:length(C[,1])){
  tvalue[i] = abs(( mean(C[i,])-mean(data[-which((data)%in%
C[i,]])]))/sqrt( sd(C[i,])^2/length(data1) + sd((data)
[-which((data)%in% C[i,])])^2/length(data2) ) )
}
pvalue = length(tvalue[tvalue>=T])/length(C[,1])
pvalue
}

```

Existe una función llamada `oneway_test` del paquete `coin` que permite aproximar los p valores en la prueba de permutación.

Ejemplo 2.1. *Expresión génica en pacientes con cáncer colorrectal*

Los datos que se usan, se recogieron en el VU-University Medical Center (VUmc), Amsterdam, Países Bajos. El objetivo del estudio fue determinar qué genes están involucrados en la progresión de adenoma a carcinoma en cáncer colorrectal. Un experimento de microarrays se realizó en ARN aislado de 68 muestras de tumor colorrectal congeladas a presión: 37 adenomas no progresados y 31 carcinomas. El microarreglo midió los niveles de expresión de 28,830 genes únicos.

Para este ejemplo sólo se ha seleccionado un gen que tiene como referencia de secuencia: NM152299. Para más detalles consulte (Thas, 2010).

```

adenomas=c(0.28500107,1.24538199,1.52518563,0.31864190,-0.08486848,
0.46999286,0.64884110,0.05890883,0.21892938,0.22627216,0.40573796,
0.41136084,1.01347959,0.57529560,0.22322844,0.41095242,1.23572024,
0.14057838,0.05036994,0.37269292,0.59387943,0.63231699,0.78446628,
0.44282119,0.13731494,0.19536726,0.53703484,1.20069718,0.45929798,

```

```

0.61929182,-0.22365381,0.53616811,0.34665425,0.18537357,0.41278816,
0.06479304,0.42013539)

carcinomas=c(0.864974556,0.016643175,-0.782365883,0.216804514,
-0.723734031,1.154091044,0.263753875,0.590277488,1.341931844,
-0.690665240,0.404671347,-0.041318447,0.440559521,-0.316895804,
-0.197012430,-0.060731077,-0.145421451,-0.090790097,0.001298894,
-0.121174484,-0.011491208,-0.226738583,-0.344762000,0.082418631,
0.695980246,-0.895347070,1.350557128,1.271523688,0.673752397,
-0.327500633,1.495886888)

d = as.data.frame( cbind(dat=c(adenomas,carcinomas),level=c(rep(
  1,length(adenomas)),rep(2,length(carcinomas)))))

library(coin)

## Loading required package: survival

pvaluePME = oneway_test(d$dat~as.factor(d$level),data=d,distribution
                        =approximate(B=10000))

pvaluePME

##

## Approximative Two-Sample Fisher-Pitman Permutation Test

##

## data:  d$dat by as.factor(d$level) (1, 2)

## Z = 2.0641, p-value = 0.0376

## alternative hypothesis: true mu is not equal to 0

```

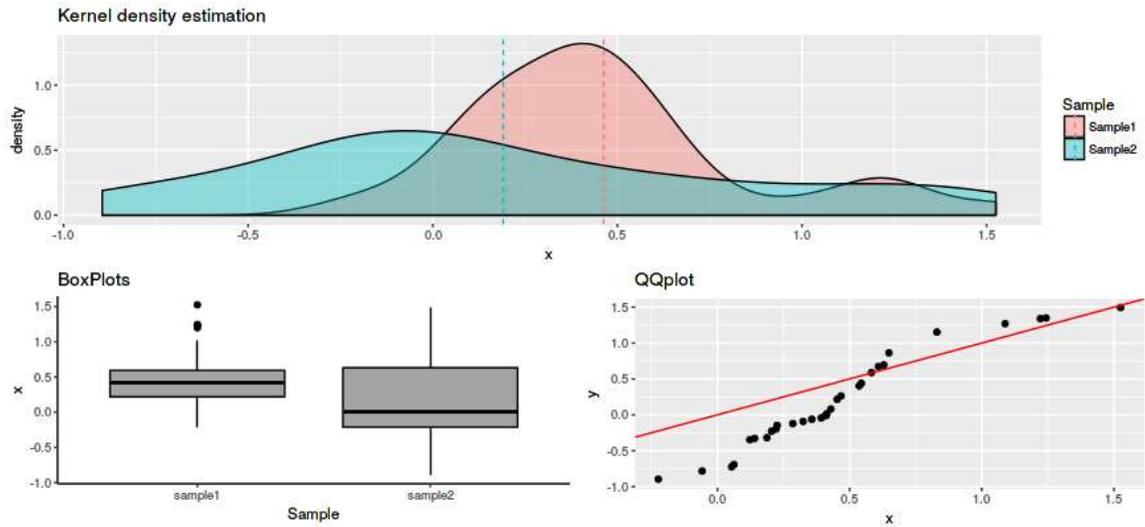


Figura 2–2: La estimación de la densidad (tipo núcleo) del gen *NM152299*, sample1 son las medidas en adenomas no progresados y sample2 los carcinomas con su boxplot y qqplot respectivos.

2.2. Prueba de Kolmogorov Smirnov

La prueba de Kolmogorov-Smirnov, denotada como *pruebas KS* utiliza la estimación muestral natural (función de distribución empírica FDE) de la función de distribución acumulada $F(x)$, la cual cuenta la proporción de observaciones menores o iguales a x . Es posible definir la FDE usando las estadísticas del orden

Definición 2.2.1. (*Estadística del orden*) Sea X una variable aleatoria continua con n observaciones.

La estadística de orden $X_{(i)}$ ($i = 1, \dots, n$) es igual al valor i -ésimo más pequeño en X .

Definición 2.2.2. (*Función de distribución empírica*) Sea X una variable aleatoria continua con n observaciones iid. La FDE se define como

$$\begin{cases} \hat{F}_n(x) = 0 & \text{si } x < X_{(1)} \\ \hat{F}_n(x) = \frac{i}{n} & \text{si } X_{(i)} \leq x < X_{(i+1)}, \quad i = 1, \dots, n-1 \\ \hat{F}_n(x) = 1 & \text{si } X_{(n)} \leq x. \end{cases}$$

El estadístico de Kolmogorov

Sea $\hat{F}_m(x)$ la FDE de la muestra aleatoria X_1, \dots, X_m y sea $\hat{G}_n(x)$ la FDE de la muestra aleatoria Y_1, \dots, Y_n . El estadístico de prueba, denotado por $D_{m,n}$ se define como la diferencia máxima absoluta entre las dos distribuciones empíricas

$$D_{m,n} = \max_x \left| \hat{F}_m(x) - \hat{G}_n(x) \right|$$

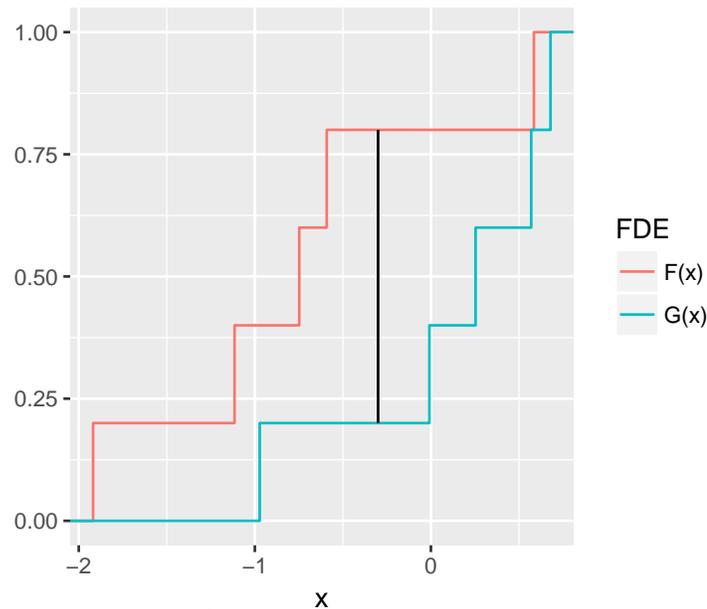


Figura 2–3: Distancia máxima (en negro) entre dos funciones de densidad empíricas.

Si las dos muestras provienen de la misma distribución, se puede esperar que las distribuciones acumuladas de ambas muestras sean bastante cercanas entre sí, si existe alguna discrepancia sólo deberían atribuirse a desviaciones aleatorias de la distribución de la población. Si las dos distribuciones acumulativas de muestra están ‘demasiado separadas’ en cualquier punto, esto sugiere que las muestras provienen de poblaciones diferentes. Por lo tanto, un valor de $D_{m,n}$ lo suficientemente grande entre las dos funciones de densidad empíricas es evidencia para rechazar H_0 .

Dado que sólo se consideran las magnitudes y no las direcciones, el estadístico es apropiado para hipótesis alternas del tipo $F \neq G$. El p-value es:

$$P(D_{m,n} \geq D_o | H_o)$$

donde D_o es el valor observado de la estadística de prueba $D_{m,n}$.

La distribución exacta del estadístico de prueba fue atribuida por (Gnedenko, 1954) y (Korolyuk, 1961). (Gibbons, 2003) describe un método geométrico que calcula el p-value de la prueba para tamaños de muestra pequeños.

Para la distribución nula asintótica, es decir, cuando m y n se aproximan al infinito, Smirnov, N. V. (1939), prueba que

$$\lim_{m,n \rightarrow \infty} P\left(\sqrt{\frac{mn}{m+n}} D_{m,n} \leq d\right) = 1 - 2 \sum_{i=1}^{\infty} (-1)^{i-1} e^{-2i^2 d^2}$$

Podemos aproximar la distribución nula de $D_{m,n}$ por simulación. Para esto, usamos la variable aleatoria uniforme estándar y la transformada integral de probabilidad.

Lema. Transformada Integral de Probabilidad

Sea X una variable aleatoria continua con función de distribución acumulada F y sea $U = F(X)$. Entonces

$$U \sim \text{Uniforme}[0, 1]$$

Prueba.

Sea $u \in [0, 1]$, como X es continua, existe $x_u \in \mathbb{R}$ tal que $F(x_u) = u$. Ahora, $F(u) = P(U < u) = P(F(X) < F(x_u)) = P(X < x_u) = F(x_u) = u$.

□

Para realizar la simulación de $D_{m,n}$:

1. Generar muestras aleatorias de tamaño n y m a partir de la distribución uniforme estándar $U[0, 1]$.
2. Encontrar la diferencia absoluta máxima entre $\hat{F}_m(u)$ y $\hat{G}_n(u)$.
3. Repetir esto B veces para obtener la distribución aproximada de $D_{m,n}$

R tiene incorporada una función llamada `ks.test` que realiza una prueba de *KS*.

Ejemplo 2.2. *Ilustramos el uso de la prueba Kolmogorov Smirnov en los datos de expresión genética del Ejemplo 2.1*

```

ks.test(carcinomas,adenomas)

##
## Two-sample Kolmogorov-Smirnov test
##
## data:  carcinomas and adenomas
## D = 0.49433, p-value = 0.0002497
## alternative hypothesis: two-sided

```

2.3. Prueba de Anderson-Darling

Darling propuso el estadístico

$$A_{mn}^2 = \frac{mn}{N} \int_{-\infty}^{\infty} \frac{(F_m(x) - G_n(x))^2}{H_N(x)(1 - H_N(x))} dH_N(x),$$

(donde $F_m(x)$, $G_n(x)$ y $H_N(x) = \{mF_m(x) + nG_n(x)\}/N$ con $N = m + n$, son la distribución muestral de la muestra X , la muestra Y y la muestra combinada respectivamente) como una modificación del estadístico empleado en el criterio de Cramer-von Misses para el caso de dos muestras (Darling, 1957). Este estadístico es estudiado en detalle por (Pittitt, 1976) quien da una fórmula computacional equivalente,

$$\frac{1}{mn} \sum_{i=1}^{N-1} \frac{(M_i N - ni)^2}{i(N-i)},$$

donde M_i es el número de observaciones en la primera muestra que no es mayor que el i -ésimo más pequeño en la muestra combinada.

La prueba resultante otorga más peso a las colas de la distribución que la prueba de KS.

La distribución límite del estadístico de Anderson-Darling para dos muestras ($n, m \rightarrow \infty, n/m \rightarrow \text{const.}$) coincide con la distribución límite del estadístico de Anderson-Darling para una muestra (Makarov y Simonova, 2017).

En R, el paquete K-sample que tiene función llamada `ad.test` realiza la prueba.

Ejemplo 2.3. *Ilustramos el uso de la prueba Anderson-Darling en los datos de expresión genética del Ejemplo 2.1*

```
library(kSamples)

## Loading required package: SuppDists

ad.test(carcinomas,adenomas,method = "exact")

## Anderson-Darling k-sample test.
## Number of samples: 2
## Sample sizes: 31, 37
## Number of ties: 0
## Mean of Anderson-Darling Criterion: 1
## Standard deviation of Anderson-Darling Criterion: 0.7409
## T.AD = ( Anderson-Darling Criterion - mean)/sigma
## Null Hypothesis: All samples come from a common population.
## Based on Nsim = 10000 simulations
##
##          AD  T.AD  asympt. P-value  sim. P-value
## version 1: 5.61 6.222          0.001261      0.0016
## version 2: 5.65 6.283          0.001203      0.0016
```

2.4. La prueba de rachas de Wald-Wolfowitz

La prueba de rachas de Wald-Wolfowitz (Wald y Wolfowitz, 1940) es una de las primeras pruebas en desarrollarse para el problema de dos muestras, su estadístico

se basa en el número de rachas en la sucesión dicotomizada de la muestra combinada ordenada.

Definición 2.4.1. (Muestra combinada) X_1, \dots, X_N , $N = m+n$, es la muestra combinada de las observaciones de las dos muestras, donde las primeras m representan la muestra de las X , es decir, $X_i = X_i, i = 1, \dots, m$ y las segundas n representan la muestra de las Y , es decir, $X_i = Y_i, i = m+1, \dots, N$.

Definición 2.4.2. (Racha) Una racha es una secuencia de eventos de un tipo tal que antes y después de esta secuencia ocurre un evento de otro tipo. $. Y Y Y X X$ y $1 1 1 0 0 0 0 1 1$ son sucesiones con 2 y 3 rachas respectivamente.

Sea $X_1, \dots, X_N, N = m+n$, la muestra combinada de las observaciones de las dos muestras.

Se representan las observaciones de la muestra 1, en la sucesión ordenada de la muestra combinada por unos, y las de la muestra 2 por ceros, de la manera siguiente:

$$T_i = \begin{cases} 1 & \text{si } X_{(j)} \text{ viene de una observación de la muestra 1} \\ 0 & \text{si } X_{(j)} \text{ viene de una observación de la muestra 2} \end{cases}, \quad (2.4)$$

donde $X_{(j)}$ representa la estadística del orden para la muestra combinada.

La sucesión de unos y ceros así construida se llama muestra *dicotomizada* y contiene una racha por cada sucesión de observaciones seguidas de la misma muestra, en la muestra combinada ordenada. Por ejemplo, con $m = 4$ y $n = 5$, una posible configuración para T , sería $T = 1 0 0 1 1 0 1 0 0$, lo que indica que en la muestra agrupada el elemento más pequeño es el elemento más pequeño de la muestra 1, el segundo más pequeño es el elemento más pequeño de la muestra 2, y así sucesivamente y por último el más grande es el mayor elemento de la muestra 2.

Bajo la hipótesis nula, esperamos que las rachas de unos y ceros estarán más o menos bien mezcladas en la muestra dicotomizada y habrá varias rachas en la sucesión dicotomizada, ya que las $m+n = N$ variables aleatorias constituyen una

sola muestra aleatoria de tamaño N de la población común. Wald y Wolfowitz utilizaron el número total de rachas R en la sucesión dicotomizada como estadística de prueba para la hipótesis, rechazando H_0 cuando R toma valores pequeños. Es decir que se rechaza H_0 en favor de la alternativa cuando $R \leq r_\alpha$ donde r_α obtiene de la distribución exacta o asintótica del número total de rachas, luego el p -value para la prueba de Wald- Wolfowitz viene dado por $P(R \leq R_0|H_0)$ donde R_0 es el valor observado de la estadística de prueba de rachas R . Se puede mostrar que bajo la hipótesis nula, la distribución del número total de rachas R formadas por m unos y n ceros, con m y n fijos en la muestra dicotomizada es:

$$P(R = r) = \begin{cases} \frac{{}_2 \binom{m-1}{\frac{r}{2}-1} \binom{n-1}{\frac{r}{2}-1}}{\binom{N}{m}} & \text{si } r \text{ es par} \\ \frac{\binom{m-1}{\frac{r-1}{2}} \binom{n-1}{\frac{r-3}{2}} + \binom{m-1}{\frac{r-3}{2}} \binom{n-1}{\frac{r-1}{2}}}{\binom{N}{m}} & \text{si } r \text{ es impar} \end{cases} \quad (2.5)$$

para $r = 2, 3, \dots, N$.

Además,

$$E(R) = 1 + \frac{2mn}{N} \quad (2.6)$$

$$Var(R) = \frac{2mn(2mn - m - n)}{N^2(N - 1)} \quad (2.7)$$

Para encontrar la distribución asintótica, suponemos que el tamaño total de la muestra N tiende a infinito de tal forma que $m/N \rightarrow \lambda$ y $n/N \rightarrow 1 - \lambda$, λ fijo,

$0 < \lambda < 1$. Para muestras grandes, la media y la varianza de R de 2.5 y 2.7 son

$$\lim_{N \rightarrow \infty} E(R/N) = 2\lambda(1 - \lambda) \quad \lim_{N \rightarrow \infty} \text{Var}(R/\sqrt{N}) = 4\lambda^2(1 - \lambda)^2.$$

Definamos la variable aleatoria

$$Z = \frac{R - 2N\lambda(1 - \lambda)}{2\sqrt{N}\lambda(1 - \lambda)},$$

sustituyendo R en términos de Z en 2.5, obtenemos la distribución de probabilidad estandarizada de R , o $f_Z(z)$. Si los factoriales en la expresión resultante son evaluados por la fórmula de Stirling, tenemos $\lim_{N \rightarrow \infty} f_Z(z) = -\ln\sqrt{2\pi} - 1/2z^2$, es decir,

$$Z \sim N(0, 1).$$

En R el paquete `randtests` tiene la función llamada `pruns` la cual realiza la prueba exacta. El siguiente código proporciona el pvalue para muestras grandes usando la distribución asintótica.

```

WWT<-function(d1,d2){
  if(length(d1)<=12 | length(d2)<=12){
    vector=sort(c(d1,d2))
    outVector <- rep(x = 0, times = length(vector))
    outVector[vector %in% d1] <- 1
    outVector[vector %in% d2] <- 0
    R=1
    for(i in 1:(length(outVector)-1)){
      if((outVector[i])!= (outVector[i+1])){R=R+1}
    }
    pvalueWW = pruns(R,length(d1),length(d2))
    return(pvalueWW)
  }
}

```

```

else {
  n1=length(d1)
  n2=length(d2)
  n=n1+n2
  vector=sort(c(d1,d2))
  outVector <- rep(x = 0, times = length(vector))
  outVector[vector %in% d1] <- 1
  outVector[vector %in% d2] <- 0
  cont=1
  for(i in 1:(length(outVector)-1)){
    if((outVector[i])!= (outVector[i+1])){cont=cont+1}
  }
  R = cont
  mu = ( (2*n1*n2) / (n1 + n2) ) + 1;
  sigmasq = ( (2*n1*n2) * (2*n1*n2-(n1+n2)) ) /
    ( ((n1+n2)^2) * (n1+n2-1) );
  sigma=sqrt(sigmasq)
  Z = (R - mu) / sigma
  pvalueWW=2*(1-pnorm(abs(Z)))
  return(pvalueWW)
}
}

```

Ejemplo 2.4. *Ilustramos el uso de la prueba de rachas de Wald-Wolfowitz en los datos de expresión genética del Ejemplo 2.1.*

```
WWT(carcinomas,adenomas)
```

```
## [1] 0.03142769
```

2.5. Prueba de suma de rango de Wilcoxon

Ésta es la única prueba tratada aquí, bajo la hipótesis alternativa de *localización*; es decir, poblaciones son de la misma forma pero con una medida diferente de tendencia central : $H_a : F_Y(x) = F_X(x - \theta) \forall x$ y $\theta \neq 0$.

Sea X_1, X_2, \dots, X_N la muestra combinada tal como se definió en 2.4.1. El rango o puesto de la i -ésima observación X_i en la muestra de tamaño N , denotado por Q_i , es el número de observaciones que son menores o iguales a X_i , es decir

$$Q_i = \sum_{j=1}^N I(X_j \leq X_i).$$

Note que

$$X_i = X_{(Q_i)}.$$

(Wilcoxon, 1954) se basa en los rangos de las $m + n = N$ observaciones de la muestra combinada. Cada observación tiene un rango: el más pequeño tiene rango 1, el segundo más pequeño rango 2, y así sucesivamente. La estadística de la prueba de suma de rangos de Wilcoxon es la suma de los rangos para las observaciones de una de las muestras. Si se considera la suma de rangos de la primera muestra, entonces el estadístico es

$$W = \sum_{i=1}^m Q_i$$

Si usamos las variables aleatorias T_i definidas en 2.4,

$$W_N = \sum_{i=1}^N iT_i.$$

Este estadístico hace parte de las llamadas estadísticas de rango lineal simple, de la forma

$$\mathfrak{T}_N(T) = \sum_{i=1}^N a_i T_i,$$

donde a_i son constantes llamadas pesos o puntajes. (Thas, 2010) prueba las siguientes propiedades

$$E(\mathfrak{T}_N) = m \sum_{i=1}^N \frac{a_i}{N}. \quad (2.8)$$

$$Var(\mathfrak{T}_N) = \frac{mn}{N^2(n-1)} \left[N \sum_{i=1}^N a_i^2 - \left(\sum_{i=1}^N a_i \right)^2 \right]. \quad (2.9)$$

$$\frac{\mathfrak{T}_n - \mu_n}{\sigma_n} \xrightarrow{d} N(0, 1). \quad (2.10)$$

Luego, para el estadístico de Wilcoxon,

$$E(W_N) = \frac{m(N+1)}{2}. \quad (2.11)$$

$$Var(W_N) = \frac{mn(N+1)}{12} \quad (2.12)$$

$$\frac{W_n - \mu_n}{\sigma_n} \xrightarrow{d} N(0, 1). \quad (2.13)$$

Para ejecutar la prueba en R usamos la función `wilcox.test` de la librería `stats`.

Ejemplo 2.5. *Ilustramos el uso de la prueba suma de rango de Wilcoxon en los datos de expresión genética del ejemplo 2.1*

```
a<- wilcox.test(carcinomas,adenomas)
n1 <- length(carcinomas)
  a$statistic <- a$statistic + n1*(n1+1)/2
  names(a$statistic) <- "T.W"
a
##
## Wilcoxon rank sum test
```

```
##
## data:  carcinomas and adenomas
## T.W = 871, p-value = 0.01409
## alternative hypothesis: true location shift is not equal to 0
```

2.6. χ^2 para dos muestras

Es bien conocido que si los datos consisten en frecuencias de categorías discretas, la prueba χ^2 puede utilizarse para determinar la importancia de las diferencias entre dos grupos independientes, esta prueba recibe el nombre de prueba χ^2 de *homocedasticidad* o *igualdad de proporciones*. Para poder realizar la prueba, discretizamos los datos de ambas muestras (los dos grupos) de tal forma que las categorías sean las *clases* o los *bins*, en cada clase contamos cuántos datos de cada muestra hay. Al final podemos calcular el estadístico cuya distribución es aproximadamente χ_{k-1}^2 , donde k es en número de clases.

Surge un problema que aún no tiene respuesta: ¿ cómo elegir el número de clases k y dónde ubicar los límites de las celdas (equiprobable, equiespaciado o alguna otra configuración)? En la literatura estadística existen varias opciones (reglas) para determinar el número de clases k por ejemplo: Square-root choice, Sturges' formula, Rice Rule, Doane's formula, Scott's normal reference rule, Freedman-Diaconis' choice, Ceiling function, entre otros. Nosotros consideramos celdas equiespaciadas y una variante de la regla de Rice, $k = 2n^{1/5}$. Una justificación de por qué tal k se encuentra en el Capítulo 4.

Ejemplo 2.6. *Ilustramos el uso de la prueba χ^2 en los datos de expresión genética del ejemplo 2.1.*

```

library(Hmisc)

chi.sqr.two.sample<- function(data1,data2){
k <- min(round(2*length(data1)^(1/5)),round(2*length(data2)^(1/5)))

R1 = max(data1)-min(data1)
R2 = max(data2)-min(data2)

if(R1>R2){

xbins =c( min(data1),rep(0,(k-1)) ,max(data1))

for(i in 1:(k-1)){xbins[i+1] = R1*(i/k) + min(data1)}

G=table(cut2(data1,cuts = xbins))

Gn=names(G)

Mcount = matrix(0,k,2)

for(i in 1:k){ Mcount[i,]= as.numeric(strsplit(substring(Gn[i],
                2, nchar(Gn[i]) - 1), split = ",")[[1]])}

county=rep(0,k)

for(i in 1:k){county[i]= length(data2[data2>=Mcount[i,1]&
                data2<Mcount[i,2]])}

}

else {

xbins =c( min(data2),rep(0,(k-1)) ,max(data2))

for(i in 1:(k-1)){xbins[i+1] = R2*(i/k) + min(data2)}

G=table(cut2(data2,cuts = xbins))

Gn=names(G)

Mcount = matrix(0,k,2)

for(i in 1:k){ Mcount[i,]= as.numeric(strsplit(substring(Gn[i],
                2, nchar(Gn[i]) - 1), split = ",")[[1]])}

county=rep(0,k)

for(i in 1:k){county[i]= length(data1[data1>=Mcount[i,1]&

```

```
data1<Mcount[i,2]])}  
  
}  
  
x=as.numeric( G)  
y=county  
return((chisq.test(rbind(x,y),simulate.p.value = TRUE))$p.value)  
}  
  
chi.sqrt.two.sample(carcinomas,adenomas)  
## [1] 0.00149925
```

Capítulo 3

GRÁFICO CUANTIL-CUANTIL PARA DOS MUESTRAS

El gráfico Cuantil-Cuantil , o también conocido como QQplot, es un método gráfico simple para comparar dos conjuntos de cuantiles correspondiente a las muestras.

Definición 3.0.1. *Sea X una variable aleatoria. El cuantil de orden p (denotado por $q_X(p)$) de X , donde $p \in (0, 1)$, se define como:*

$$P(X \leq q_X(p)) = p \tag{3.1}$$

Sea X_1, \dots, X_n una muestra de X , entonces el cuantil muestral de orden p es dado por:

$$\hat{q}_X(p) = X_{([np]+1)} \tag{3.2}$$

donde $X_{(\cdot)}$ denota el estadístico del orden definido en 2.2.1 y $[np]$ denota la parte entera de np .

Así, el j -ésimo valor más pequeño de la muestra $X_{(j)}$ es una estimación $\hat{q}_X(p)$ del cuantil de orden p donde $p = (j - 1)/n$, $(1 < j \leq n)$.

Consideremos nuestra segunda muestra Y_1, \dots, Y_m ; este también proporciona una estimación $\hat{q}_Y(p)$ de este mismo cuantil:

$$\hat{q}_Y(p) = Y_{([m(j-1)/n]+1)} \tag{3.3}$$

Si las dos muestras corresponden a la misma distribución de probabilidad, entonces $\hat{q}_X(p)$ y $\hat{q}_Y(p)$ deberían estar cerca.

Note que si $n = m$ en la ecuación 3.3, entonces $\hat{q}_Y(p) = Y_{([np]+1)}$. Por lo tanto, si consideramos el caso donde los tamaños de muestra para ambas muestras son los mismos, el gráfico Cuantil-Cuantil para dos muestras es solo una gráfico de $X_{(i)}$ contra $Y_{(i)}, i = 1, \dots, n$, (Stanford y Vardeman, 1994). Si las dos distribuciones son iguales, la gráfica será lineal con pendiente 1 e intercepto en 0.

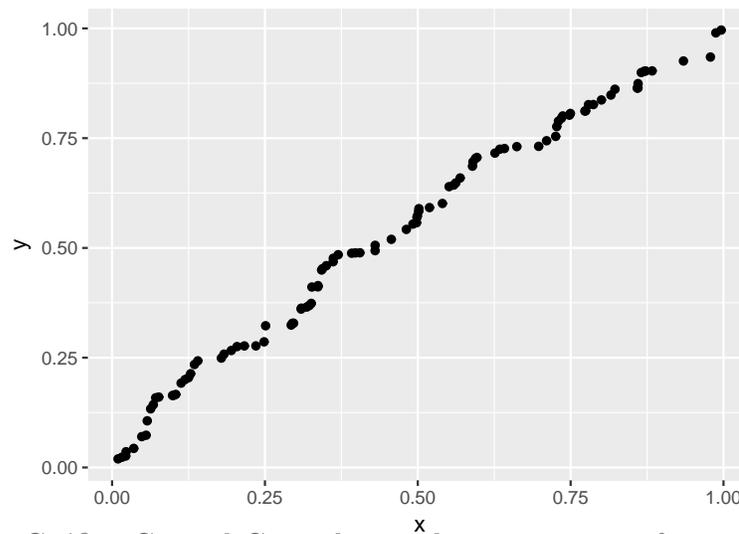


Figura 3-1: Gráfico Cuantil-Cuantil para dos muestras uniforme ($n=m=100$)

Este método gráfico es subjetivo. Discutiremos un método para eliminar esa subjetividad usando intervalos de confianza bajo la hipótesis nula H_0 : ambas distribuciones son uniformes en $(0, 1)$.

3.0.2. Construcción de los intervalos de confianza

Autores como (Noiman et al., 2013) y (Rolke) han propuesto métodos basados en la distribución de los estadísticos del orden para añadir bandas de confianza en los gráficos cuantil-cuantil de normalidad. Nosotros usamos algunas de esas ideas y las aplicamos en la construcción de bandas de confianza al graficar los cuantiles de dos distribuciones uniformes estándar.

Sea X_1, \dots, X_n una muestra *i.i.d.*. Si $X \sim U(0, 1)$, por el teorema A.0.2 se tiene que

$$X_{(i)} \sim \text{Beta}(i, n - i + 1) \quad (3.4)$$

Para obtener una probabilidad $1 - \alpha$ de no rechazar H_0 dado que es cierta, necesitamos encontrar los límites \mathbf{l} y \mathbf{h} tales que:

$$P(l_i \leq X_{(i)} \leq h_i; \quad i = 1, \dots, n) = 1 - \alpha \quad (3.5)$$

Usamos la Ecuación 3.4 para estimar los límites:

$$\begin{aligned} l_i(\alpha, n) &= \beta^{-1}(\alpha/2; i, n - i + 1) \\ h_i(\alpha, n) &= \beta^{-1}(1 - \alpha/2; i, n - i + 1) \end{aligned} \quad (3.6)$$

Usamos la simulación Monte Carlo para estimar límites apropiados para alcanzar la probabilidad $1 - \alpha$ deseada.

El método empleado para estimar la probabilidad en 3.5, se basa en la ley de los grandes números. Definimos una variable aleatoria de Bernoulli T que tiene un valor de 1 si cada observación de una muestra aleatoria uniforme, se encuentra dentro de la banda de confianza. Un valor de 0 de otro modo. Necesitamos que la tasa de éxito de T sea $p = 1 - \alpha$.

Definimos T_1, \dots, T_i , una sucesión de variables aleatorias de T *i.i.d.*. La media de cada T_i es su tasa de éxito, $1 - \alpha$. La ley fuerte de los grandes números establece

que

$$\lim_{i \rightarrow \infty} \frac{T_1 + \dots + T_i}{i} = 1 - \alpha \quad (3.7)$$

con probabilidad 1. (Révész, 1968)

Definimos la función:

$$\psi(p) = P(l_i(p, n) \leq X_{(i)} \leq h_i(p, n); \quad i = 1, \dots, n) \quad (3.8)$$

con límites de la forma 3.6, queremos encontrar un valor p tal que $\psi(p) = 1 - \alpha$ para un α fijo.

La función 3.8 es estrictamente decreciente en p ; si $p \rightarrow 0$ los límites $l_i(\alpha, n) \rightarrow -\infty$ y $h_i(\alpha, n) \rightarrow +\infty$ entonces $\psi(p) = 1$, cuando p aumenta cada intervalo se vuelve más pequeño, por lo tanto el valor de $\psi(p)$ se reduce. Esto nos asegura que existe una solución única para p tal que $\psi(p) = 1 - \alpha$. Usaremos un algoritmo de bisección para encontrar la solución.

Comenzamos con $pl = 0$ y $ph = 1$, para $m = (pl + ph)/2$ estimamos $\psi(m)$. Si es mayor a $1 - \alpha$ establecemos $pl = m$ ó $ph = m$ si es menor. Luego repetir hasta alcanzar $ph - pl < e$, para un e fijo. La estimación de $\psi(m)$ se realiza generando i muestras uniforme de tamaño n y luego usando 3.7, que es la proporción del número de muestras en el cual todas las observaciones ordenadas $U_{(i)}$ cayeron entre l_i y h_i .

Como tenemos dos muestras X e Y independientes, debemos estimar los intervalos de confianza para Y procediendo de forma análoga.

A continuación un gráfico Cuantil-Cuantil para dos muestras $U(0, 1)$ de tamaño 100, y un gráfico Cuantil-Cuantil para una distribución $U(0, 1)$ versus una $beta(shape1 = 2, shape2 = 0)$ de tamaño 100, con las bandas de confianza añadidas:

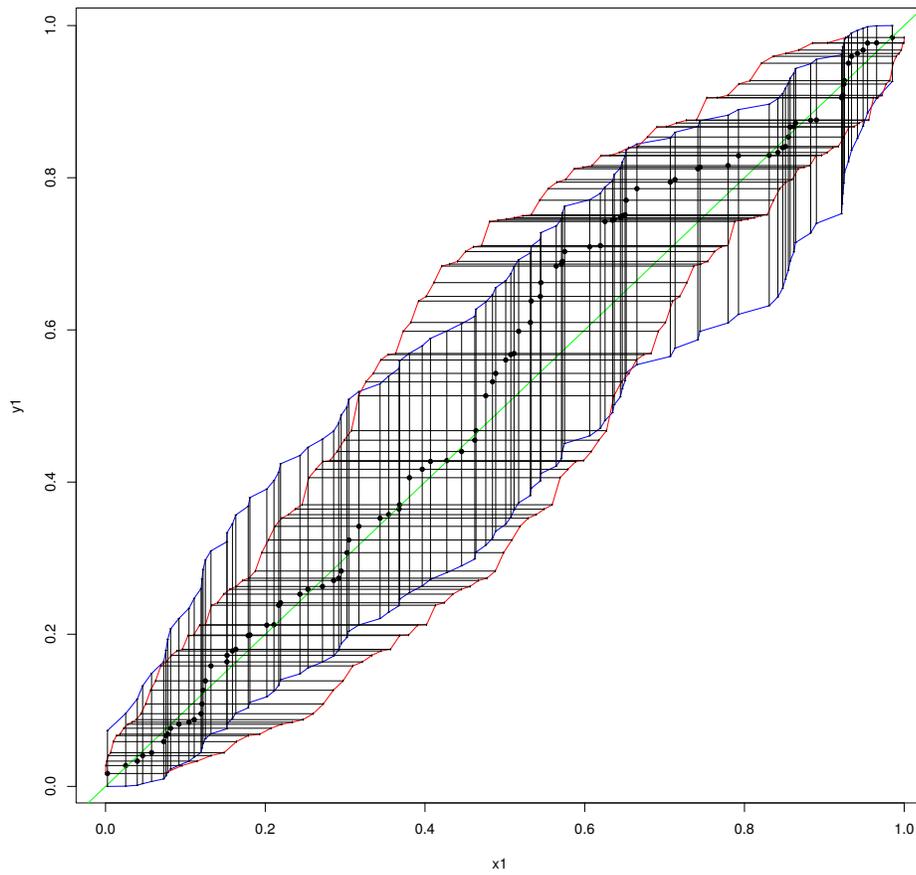


Figura 3–2: Gráfica Cuantil-Cuantil para dos muestras uniformes en $(0, 1)$ de tamaño 100, con las bandas de confianza

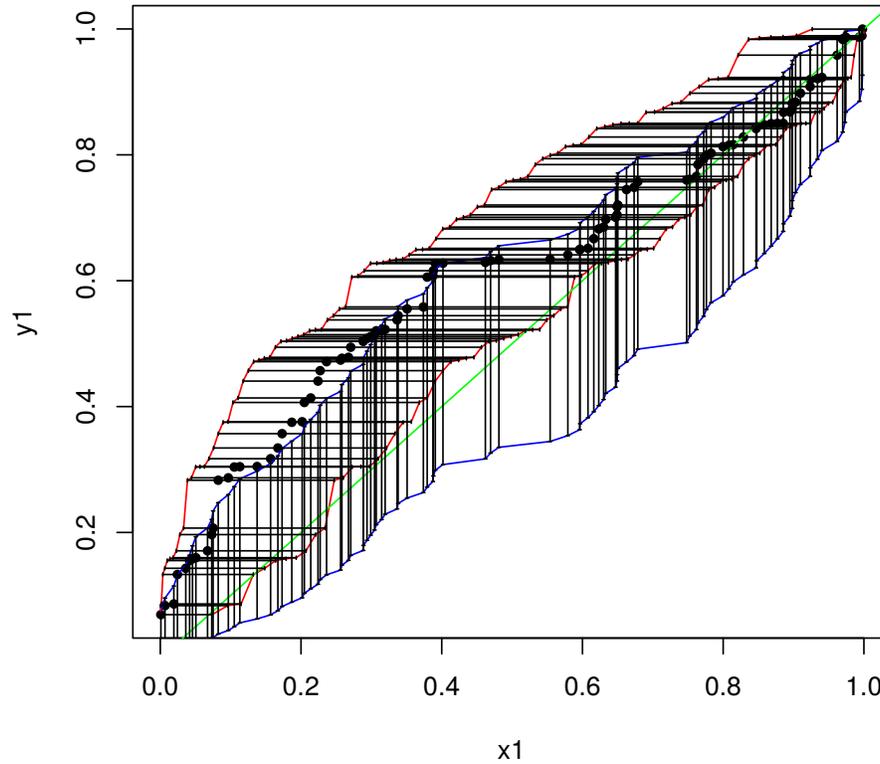


Figura 3–3: Gráfica Cuantil-Cuantil $U(0,1)$ versus $beta(shape1 = 2, shape2 = 0)$ de tamaño 100, con las bandas de confianza

3.0.3. Comparación del poder

Para comparar las pruebas (en términos del poder) descritas en el capítulo anterior, con el método descrito aquí, utilizamos dos muestras aleatorias de una distribuciones Beta y Uniforme, ambas de tamaño 100.

$$D_1 = Uniforme(min = 0, max = 1)$$

$$D_2 = Beta(shape1 \in (0, 2), shape2 = 1)$$

$shape1$ variando entre 0 y 2. Se usó un nivel de significancia $\alpha = 0.05$. Para contrastar D_1 y D_2 , se emplearon las pruebas de Permutación (PT), Kolmogorov Smirnov (KS), Suma de Rango de Wilcoxon (WT), Anderson Darling (AD), Prueba de Rachas de Wald Wolfowitz (WW), Chi-Cuadrado (CH) y el método descrito en este

capítulo (ENV). Para estimar los poderes se repitieron 1000 pruebas por cada punto en que varía *shape1* en cada una de las pruebas. Los tamaños de las muestras se fijaron en 100.

<i>shape1</i>	PT	AD	WT	KS	CH	WW	ENV
0.10	1.00	1.00	1.00	1.00	0.97	1.00	1.00
0.30	1.00	1.00	1.00	1.00	0.96	0.88	1.00
0.50	0.99	0.99	0.99	0.98	0.85	0.38	1.00
0.70	0.63	0.65	0.65	0.52	0.32	0.09	0.82
0.90	0.08	0.09	0.08	0.06	0.10	0.05	0.09
1.00	0.04	0.04	0.04	0.02	0.02	0.07	0.05
1.10	0.05	0.03	0.04	0.01	0.05	0.05	0.07
1.20	0.18	0.17	0.18	0.10	0.06	0.02	0.21
1.30	0.39	0.37	0.38	0.27	0.27	0.05	0.41
1.40	0.57	0.54	0.56	0.37	0.35	0.14	0.63
1.60	0.89	0.87	0.87	0.76	0.65	0.11	0.95
1.80	0.97	0.93	0.94	0.86	0.81	0.16	0.99
2.00	1.00	0.99	0.99	0.97	0.96	0.35	1.00

Tabla 3-1: Estimación del poder en función de *shape1*.

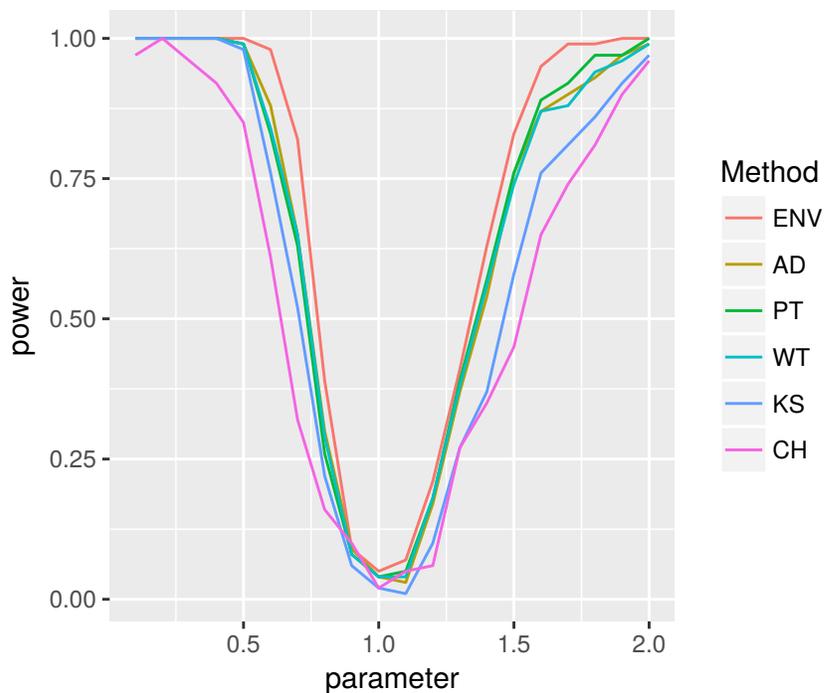


Figura 3-4: Pruebas más potentes, note que la prueba de ENV es superior.

Capítulo 4

SIMULACIONES

El poder de una prueba de hipótesis es la probabilidad de rechazar la hipótesis nula correctamente cuando la hipótesis alternativa es verdadera. Una prueba con un alto poder (alta tasa de rechazo) se considera un buen método de prueba. El poder deseado de una prueba es 1 ó 100 %; una prueba perfecta siempre rechazará la hipótesis nula cuando la hipótesis nula no es verdadera. Para el caso del problema de dos muestras, las pruebas deben rechazar que ambas muestras provienen de la misma distribución de probabilidad cuando en realidad las muestras son de distintas distribuciones. A medida que la potencia se aproxima a 1, la prueba puede considerarse una buena prueba.

Para comparar el rendimiento en términos del poder de las pruebas descritas en el Capítulo 2, el poder se estima utilizando la tasa de rechazo. El mismo procedimiento de prueba se repetirá k veces en cada prueba para probar k conjuntos de muestras pseudoaleatorias de las distribuciones en la hipótesis alternativa especificada (los dos conjuntos de datos se generan de dos distribuciones utilizando las funciones básicas de R como: `rnorm`, `runif`, `rexp`, entre otras). La tasa de rechazo entre estas k repeticiones, para cada prueba, será el poder estimado de las pruebas. El valor de k se establece en 1000, un valor de k superior a 1000, requiere de un gran esfuerzo computacional en términos del tiempo requerido para correr las simulaciones.

Se compara el poder de las pruebas cuando las distribuciones alternas son iguales, poco distintas y muy distintas, cuando tenemos tanto igual número de datos

en ambas muestras ($m = n$), como tamaños de muestras muy dispares en ambos conjuntos (por ejemplo $m = 10n$). Esto nos da una herramienta para observar el efecto del cambio en parámetros específicos y el tamaño sobre el rendimiento de las diferentes pruebas.

Se sigue el siguiente esquema en las simulaciones: Se especifican las distribuciones alternativas de acuerdo a la siguiente clasificación: soporte infinito, soporte semi infinito y soporte finito en $[0, 1]$. Se consideran tanto muestras de igual tamaño como muestras de tamaños dispares. En todos los casos se usa un nivel de significancia de $\alpha = 0.05$. En todas las gráficas, las curvas del poder están suavizadas mediante un ajuste a un modelo generalizado aditivo, para una mejor visualización. En las gráficas del poder, los nombres de las pruebas están etiquetadas de la siguiente manera:

1. PT : Prueba de permutación
2. KS : Prueba de Kolmogorov Smirnov
3. AD : Prueba de Anderson-Darling
4. WW : Prueba de rachas de Wald-Wolfowitz
5. WT : Prueba de suma de rango de Wilcoxon
6. CH : Prueba χ^2 para dos muestras

En cada gráfica, las etiquetas aparecen del lado izquierdo y en orden descendente de acuerdo al poder de las pruebas; la prueba más poderosa aparecerá etiquetada de primero, la segunda prueba más poderosa, se etiqueta de segunda y así sucesivamente. Se utiliza una [aplicación](#) interactiva para el estudio, la cual es descrita en detalle en el siguiente capítulo.

En las simulaciones, se usarán las siguientes distribuciones (hipótesis alternas):

1 Soporte infinito:

- 1.1 $X \sim N(\mu = 0, \sigma = 1, n = m), \quad Y \sim t(df = 1, m \in [20, 200])$
- 1.2 $X \sim N(\mu = 0, \sigma = 1, n = m), \quad Y \sim t(df = 2, m \in [20, 400])$
- 1.3 $X \sim N(\mu = 0, \sigma = 1, n = m), \quad Y \sim t(df = 10, m \in [20, 400])$
- 1.4 $X \sim N(\mu = 0, \sigma = 1, n = 50), \quad Y \sim N(\mu \in [-1, 1], \sigma = 1, m = 50)$
- 1.5 $X \sim N(\mu = 0, \sigma = 1, n = 50), \quad Y \sim N(\mu = 0, \sigma \in (0, 2], m = 50)$
- 1.6 $X \sim N(\mu = 0, \sigma = 1, n \in [100, 1000]), \quad Y \sim t(df = 1, m = 1000)$
- 1.7 $X \sim N(\mu = 0, \sigma = 1, n \in [100, 1000]), \quad Y \sim t(df = 10, m = 1000)$
- 1.8 $X \sim N(\mu \in [-1, 1], \sigma = 1, n = 100), \quad Y \sim N(\mu = 0, \sigma = 1, m = 1000)$
- 1.9 $X \sim N(\mu = 0, \sigma \in (0, 2], n = 100), \quad Y \sim N(\mu = 0, \sigma = 1, m = 1000)$

2 Soporte semi infinito:

- 2.1 $X \sim Exp(\lambda = 1, n = m), \quad Y \sim F(df_1 = 2, df_2 = 5, m \in [60, 200])$
- 2.2 $X \sim Exp(\lambda = 1, n = 50), \quad Y \sim Weibull(shape = 1, scale \in [1, 2], m = 50)$
- 2.3 $X \sim Exp(\lambda = 1, n = 50), \quad Y \sim Exp(\lambda \in (0, 2], m = 50)$
- 2.4 $X \sim Exp(\lambda = 1, n = 50), \quad Y \sim Gamma(shape \in [0.25, 1.75], rate = 1, m = 50)$
- 2.5 $X \sim Exp(\lambda = 1, n = 100), \quad Y \sim Gamma(shape \in [0.25, 1.75], rate = 1, m = 1000)$
- 2.6 $X \sim Exp(\lambda = 1, n = 1000), \quad Y \sim Weibull(shape = 1, scale \in [1.2], shift = 0, m = 100)$

3 Soporte en $[0, 1]$:

- 3.1 $X \sim U(a = 0, b = 1, n = 50), \quad Y \sim Beta(\alpha \in (0, 2], \beta = 1, m = 50)$
- 3.2 $X \sim U(a = 0, b = 1, n = m), \quad Y \sim Lineal(Intercepto = 2, m \in [20, 200])$
- 3.3 $X \sim U(a = 0, b = 1, n = 50), \quad Y \sim Lineal(Intercepto \in [0, 2], m = 50)$
- 3.4 $X \sim Exp\ truncada(min = 0, max = 1, \lambda = 1, n = m), \quad Y \sim Lineal(Intercepto \in [0, 2], m \in [20, 200])$

3.5 $X \sim \text{Exp truncada}(\min = 0, \max = 1, \lambda = 1, n = m), \quad Y \sim \text{Lineal}(\text{Intercepto} = 1.5, m = \{1000, 5000, 10000\})$

3.6 $X \sim U(a = 0, b = 1, n = 1000), \quad Y \sim \text{Beta}(\alpha \in (0, 2], \beta = 1, m = 100)$

3.7 $X \sim U(a = 0, b = 1, n = 1000), \quad Y \sim \text{Lineal}(\text{Intercepto} = 2, \beta = 1, m \in [100, 1000])$

3.8 $X \sim U(a = 0, b = 1, n = 1000), \quad Y \sim \text{Lineal}(\text{Intercepto} \in [0, 2], \beta = 1, m = 100)$

4.0.1. Soporte infinito

Muestras de igual tamaño

- $N(\mu = 0, \sigma = 1)$ contra $t(df = 1)$

La Tabla 4-1 muestra la potencia aproximada para diferentes tamaños de muestra. La prueba de mayor rendimiento al contrastar muestras de poblaciones $N(\mu = 0, \sigma = 1)$ y $t(df = 1)$ es la prueba de Anderson-Darling, que alcanza un poder de más del 90 % con un tamaño de muestra de 110. Las demás pruebas son bastante inferiores en términos del poder; este resultado se tiene porque al tener muestras grandes de una distribución $t(df = 1)$, esperamos ver observaciones demasiado grandes en comparación con una distribución normal estándar. Anderson-Darling detecta las diferencias en los extremos de las distribuciones. Las pruebas de Permutación y de Wilcoxon tienen un poder casi nulo porque están diseñadas para detectar diferencias en media y ambas poblaciones tienen media cero. La Figura 4-1 muestra las cuatro pruebas con el mejor rendimiento, es decir, la prueba de Anderson-Darling, la prueba Chi Cuadrado, la prueba de Wald-Wolfowitz y la prueba de Kolmogorov Smirnov.

n	AD	CH	WW	KS	WT	PT
20	0.11	0.12	0.06	0.05	0.05	0.05
30	0.18	0.18	0.17	0.07	0.07	0.06
40	0.26	0.23	0.17	0.07	0.04	0.04
50	0.40	0.32	0.27	0.14	0.06	0.05
60	0.47	0.38	0.26	0.17	0.04	0.04
70	0.61	0.47	0.33	0.16	0.06	0.06
80	0.69	0.48	0.33	0.21	0.06	0.05
90	0.81	0.50	0.40	0.27	0.06	0.04
100	0.87	0.47	0.48	0.29	0.04	0.04
110	0.90	0.53	0.42	0.36	0.05	0.06
120	0.94	0.51	0.51	0.41	0.04	0.05
130	0.98	0.54	0.60	0.59	0.06	0.05
140	0.99	0.52	0.56	0.61	0.04	0.04
150	0.99	0.56	0.62	0.63	0.05	0.04
160	0.99	0.63	0.65	0.67	0.05	0.04
170	0.99	0.61	0.62	0.71	0.05	0.05
180	1.00	0.60	0.65	0.79	0.06	0.04
190	1.00	0.61	0.72	0.83	0.05	0.05
200	1.00	0.61	0.73	0.84	0.06	0.04

Tabla 4-1: Estimación del poder en función del tamaño de muestra. $N(\mu = 0, \sigma = 1)$ contra $t(df = 1)$.

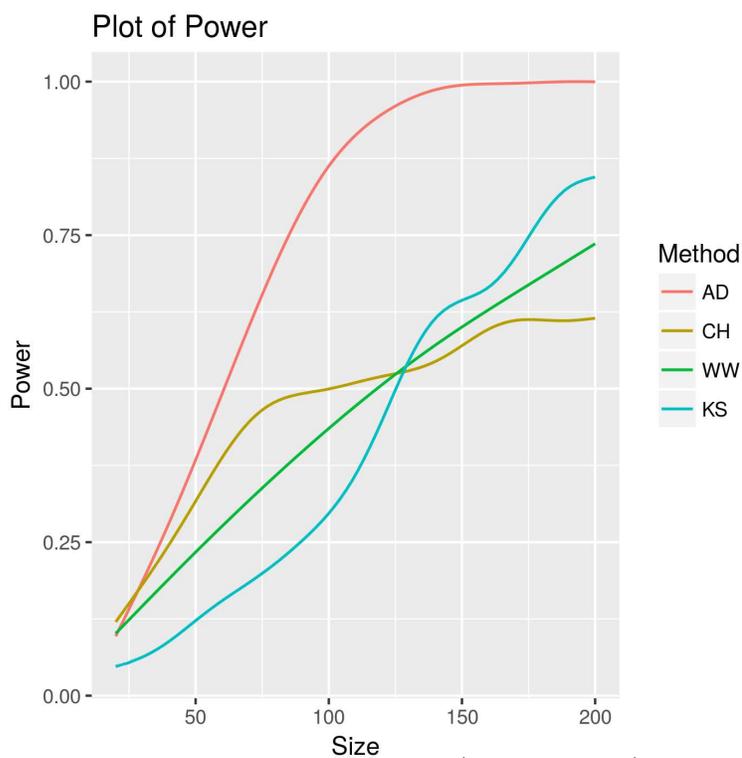


Figura 4-1: Pruebas más potentes para $N(\mu = 0, \sigma = 1)$ contra $t(df = 1)$.

- $N(\mu = 0, \sigma = 1)$ contra $t(df = 2)$

La Tabla 4-2 muestra que es necesario muestras de mayor tamaño superior a 200 para que la prueba de Chi-Cuadrado tenga un poder de por lo menos 0.83, le sigue la prueba de Anderson-Darlin con una potencia de 0.67 para muestras de tamaño 200. La Figura 4-2 muestra las cuatro pruebas con el mejor rendimiento.

n	CH	AD	WW	KS	WT	PT
20	0.07	0.07	0.03	0.04	0.06	0.05
30	0.12	0.08	0.09	0.03	0.06	0.05
40	0.16	0.08	0.06	0.03	0.05	0.05
50	0.24	0.13	0.11	0.06	0.05	0.06
60	0.29	0.14	0.10	0.07	0.05	0.05
70	0.35	0.18	0.11	0.06	0.07	0.06
80	0.42	0.21	0.07	0.06	0.05	0.06
90	0.46	0.22	0.11	0.06	0.04	0.05
100	0.51	0.27	0.17	0.08	0.05	0.05
120	0.59	0.37	0.13	0.08	0.05	0.04
140	0.68	0.44	0.13	0.13	0.05	0.06
160	0.73	0.53	0.18	0.11	0.05	0.05
180	0.79	0.56	0.21	0.14	0.04	0.03
200	0.83	0.67	0.23	0.15	0.05	0.05

Tabla 4-2: Estimación del poder en función del tamaño de muestra. $N(\mu = 0, \sigma = 1)$ contra $t(df = 2)$.

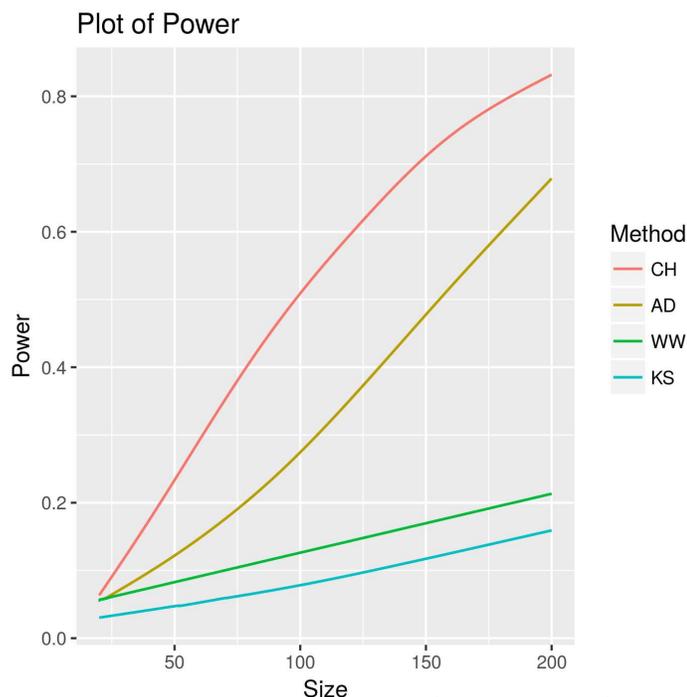


Figura 4-2: Pruebas más potentes para $N(\mu = 0, \sigma = 1)$ contra $t(df = 2)$.

- $N(\mu = 0, \sigma = 1)$ contra $t(df = 10)$

Dado que los grados de libertad aumentan, la distribución t se aproxima a la distribución normal, la potencia de las pruebas disminuye sustancialmente. La Tabla 4-3 y La figura 4-3 muestra que ninguna prueba es relativamente buena, (inclusive para tamaños de muestra grande) bajo esta hipótesis alterna.

n	CH	AD	WW	KS	WT	PT
20	0.63	0.11	0.05	0.06	0.06	0.07
20	0.29	0.06	0.03	0.04	0.05	0.05
50	0.79	0.36	0.26	0.11	0.05	0.05
60	0.91	0.50	0.27	0.19	0.07	0.05
60	0.57	0.14	0.09	0.08	0.06	0.05
100	0.73	0.27	0.17	0.07	0.05	0.05
140	0.83	0.44	0.14	0.12	0.06	0.06
180	0.95	0.63	0.18	0.16	0.06	0.06
220	0.96	0.77	0.24	0.20	0.05	0.04
260	0.97	0.86	0.27	0.27	0.06	0.06
300	0.98	0.92	0.32	0.32	0.05	0.05
340	0.98	0.97	0.34	0.41	0.05	0.05
380	1.00	0.98	0.35	0.44	0.05	0.06
400	0.99	0.99	0.42	0.52	0.05	0.04

Tabla 4-3: Estimación del poder en función del tamaño de muestra. $N(\mu = 0, \sigma = 1)$ contra $t(df = 10)$.

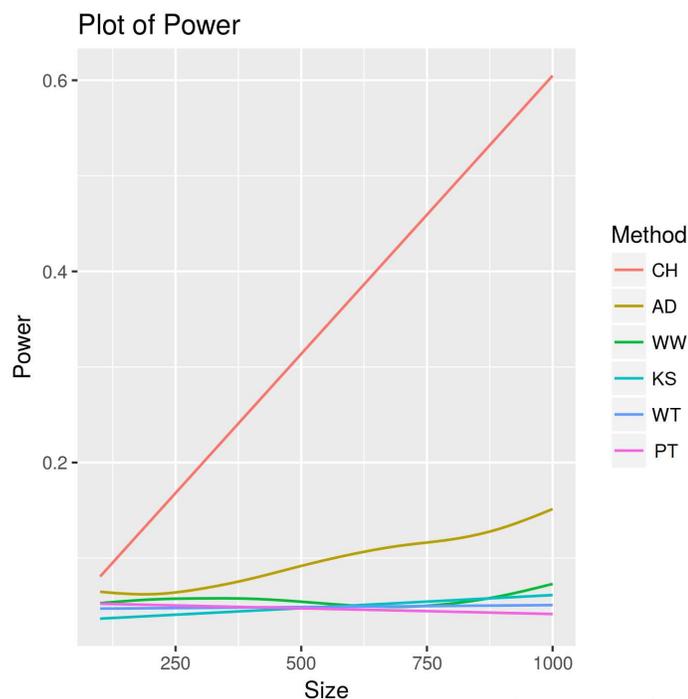


Figura 4-3: Potencias para $N(\mu = 0, \sigma = 1)$ contra $t(df = 10)$.

- $N(\mu = 0, \sigma = 1)$ contra $N(\mu \in [-1, 1], \sigma = 1)$, ($m = n = 50$)

La Tabla 4-4 muestra que las potencias bajo esta hipótesis alterna son muy similares para las pruebas de Permutación, Anderson-Darling y Wilcoxon, la prueba de Kolmogorov-Smirnov es conservadora; es decir, la probabilidad real de rechazar la hipótesis nula cuando es cierta es menor que el nivel de significación $\alpha = 0.05$. La Figura 4-4 no evidencia diferencias significativas del poder entre las mejores pruebas. La peor prueba bajo esta alternativa es la prueba de Wald-Wolfowitz.

μ	PT	WT	AD	KS	CH	WW
-1.00	1.00	1.00	1.00	0.99	0.93	0.50
-0.83	0.99	0.98	0.98	0.94	0.84	0.30
-0.67	0.90	0.88	0.88	0.79	0.63	0.19
-0.50	0.69	0.68	0.68	0.53	0.37	0.09
-0.25	0.22	0.20	0.20	0.15	0.11	0.07
-0.08	0.07	0.07	0.07	0.06	0.07	0.05
0.00	0.04	0.04	0.04	0.03	0.04	0.05
0.08	0.07	0.07	0.07	0.05	0.07	0.06
0.25	0.22	0.21	0.22	0.16	0.11	0.07
0.58	0.82	0.81	0.80	0.67	0.51	0.12
0.83	0.98	0.98	0.98	0.94	0.84	0.32
1.00	1.00	1.00	1.00	0.98	0.93	0.52

Tabla 4-4: Estimación del poder en función de μ . $N(\mu = 0, \sigma = 1)$ contra $N(\mu \in [-1, 1], \sigma = 1)$.

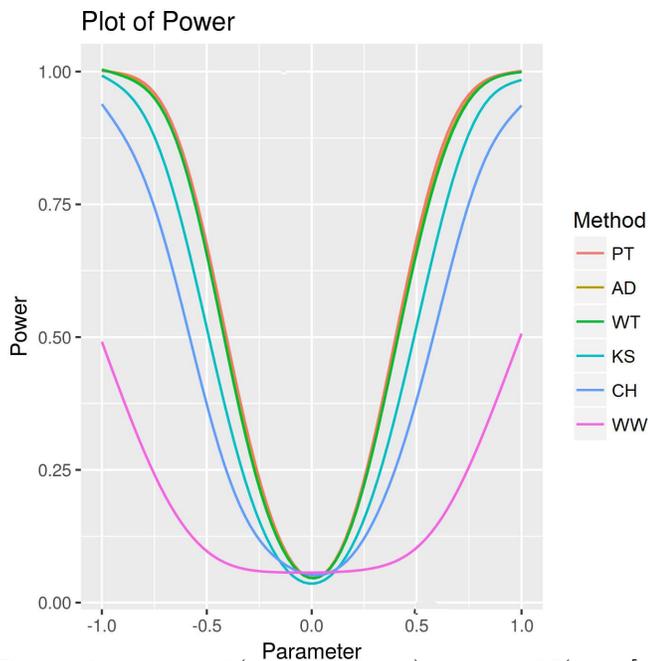


Figura 4-4: Potencias para $N(\mu = 0, \sigma = 1)$ contra $N(\mu \in [-1, 1], \sigma = 1)$.

- $N(\mu = 0, \sigma = 1)$ contra $N(\mu = 0, \sigma \in (0, 2])$, ($m = n = 50$)

En la Tabla 5-5 y Figura 4-5 vemos que la mejor prueba que más detecta diferencias entre dos poblaciones normales con igual media es la prueba de Chi Cuadrado, seguido por la prueba de Anderson-Darling.

σ	CH	AD	WW	KS	WT	PT
0.10	1.00	1.00	1.00	1.00	0.09	0.04
0.26	1.00	1.00	0.98	0.98	0.05	0.04
0.42	0.97	0.94	0.62	0.61	0.06	0.04
0.57	0.69	0.49	0.24	0.21	0.06	0.06
0.73	0.26	0.12	0.09	0.08	0.04	0.04
0.89	0.07	0.06	0.06	0.05	0.05	0.05
0.97	0.05	0.06	0.06	0.05	0.06	0.06
1.05	0.06	0.06	0.06	0.04	0.06	0.06
1.13	0.07	0.06	0.04	0.05	0.06	0.06
1.21	0.13	0.08	0.07	0.05	0.05	0.05
1.37	0.27	0.13	0.08	0.09	0.05	0.05
1.52	0.45	0.26	0.12	0.14	0.06	0.04
1.68	0.63	0.43	0.21	0.20	0.05	0.06
1.84	0.78	0.62	0.30	0.29	0.06	0.05
2.00	0.86	0.76	0.40	0.39	0.06	0.06

Tabla 4-5: Estimación del poder en función de σ . $N(\mu = 0, \sigma = 1)$ contra $N(\mu = 0, \sigma \in (0, 2])$.

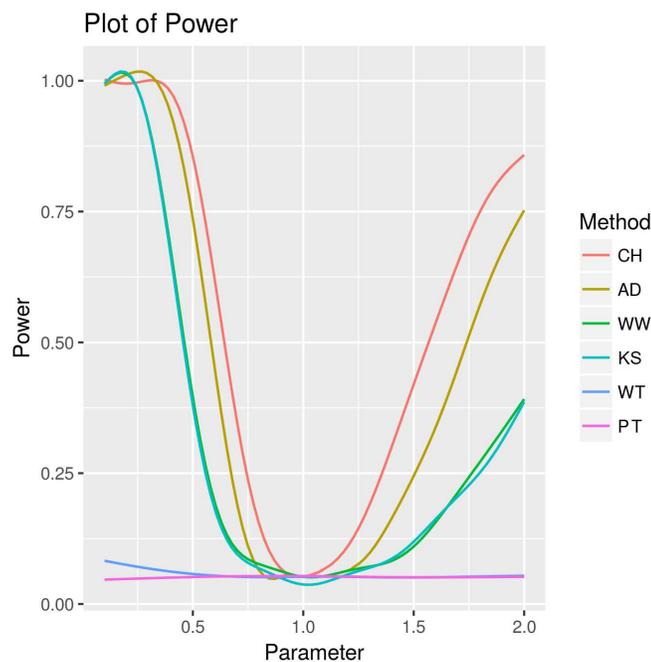


Figura 4-5: Potencias para $N(\mu = 0, \sigma = 1)$ contra $N(\mu = 0, \sigma \in (0, 2])$.

Muestras de diferentes tamaños

- $N(\mu = 0, \sigma = 1, n \in [100, 1000])$ contra $t(df = 1, m = 1000)$

La Tabla 4-6 muestra que las pruebas basadas en la función de densidad empírica son las más potentes cuando se comparan muestras de distintos tamaños. El poder de la prueba de Chi-Cuadrado tiene un comportamiento aproximadamente lineal como se observa en la Figura 4-6.

n	AD	KS	WW	CH	WT	PT
100	1.00	0.82	0.17	0.01	0.02	0.00
200	1.00	1.00	0.42	0.07	0.03	0.00
300	1.00	1.00	0.71	0.15	0.02	0.00
400	1.00	1.00	0.85	0.26	0.04	0.00
500	1.00	1.00	0.94	0.36	0.04	0.00
600	1.00	1.00	0.97	0.41	0.04	0.00
700	1.00	1.00	0.99	0.51	0.05	0.01
800	1.00	1.00	1.00	0.58	0.04	0.01
900	1.00	1.00	1.00	0.67	0.04	0.03
1000	1.00	1.00	1.00	0.73	0.05	0.06

Tabla 4-6: Estimación del poder en función del tamaño de muestra. $N(\mu = 0, \sigma = 1, n \in [100, 1000])$ contra $t(df = 1, m = 1000)$.

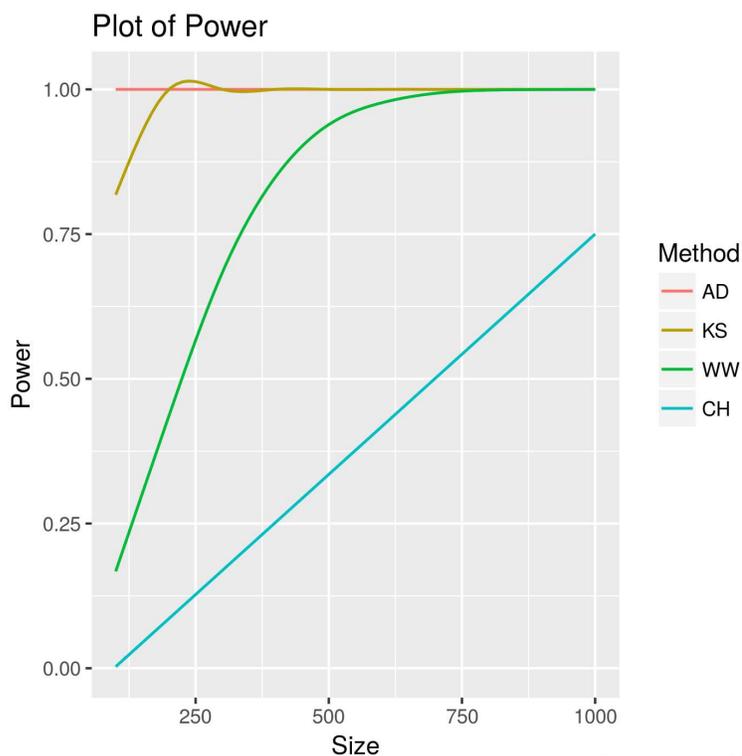


Figura 4-6: Pruebas más potentes. $N(\mu = 0, \sigma = 1, n \in [100, 1000])$ contra $t(df = 1, m = 1000)$.

- $N(\mu = 0, \sigma = 1, n \in [100, 1000])$ contra $t(df = 10, m = 1000)$

Dado que los grados de libertad aumentan, la distribución t se aproxima a la distribución normal, la potencia de las pruebas disminuye sustancialmente. La Tabla 4-7 y la Figura 4-7 muestra que la prueba Chi-Cuadrado supera en gran medida a las otras pruebas aunque con muestras de tamaños 1000 tienen una potencia que no supera al 60 %.

n	CH	AD	KS	WW	WT	PT
100	0.02	0.04	0.05	0.04	0.04	0.04
200	0.05	0.06	0.04	0.05	0.04	0.03
300	0.11	0.07	0.05	0.05	0.04	0.04
400	0.17	0.08	0.05	0.05	0.04	0.04
500	0.27	0.10	0.06	0.04	0.05	0.05
600	0.35	0.11	0.04	0.06	0.05	0.04
700	0.42	0.13	0.07	0.06	0.07	0.06
800	0.48	0.14	0.06	0.05	0.05	0.05
900	0.56	0.14	0.06	0.06	0.05	0.03
1000	0.60	0.15	0.06	0.07	0.05	0.05

Tabla 4-7: Estimación del poder en función del tamaño de muestra. $N(\mu = 0, \sigma = 1, n \in [100, 1000])$ contra $(df = 10, m = 1000)$.

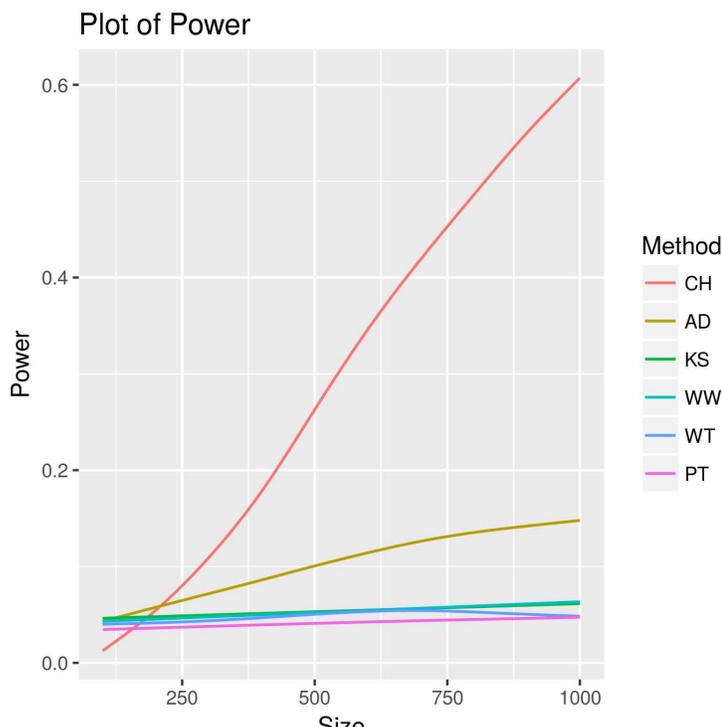


Figura 4-7: Pruebas más potentes. $N(\mu = 0, \sigma = 1, n \in [100, 1000])$ contra $(df = 10, m = 1000)$.

- $N(\mu \in [-1, 1], \sigma = 1, n = 100)$ contra $N(\mu = 0, \sigma = 1, m = 1000)$

En la Figura 4-8 se observa un comportamiento similar (en cuanto al orden de las mejores pruebas) al contrastar $N(\mu \in [-1, 1], \sigma = 1, n = 100)$ contra $N(\mu = 0, \sigma = 1, m = 1000)$, cuando los tamaños de muestras son 100 y 1000 respectivamente a cuando los tamaños de muestra son 50 para ambas muestras (ver Figura 4-4). Bajo estas condiciones de tamaños de muestra, el poder de las pruebas crece significativamente comparado con el poder de estas mismas muestras pero con tamaños de 50. Por ejemplo, ahora el poder aproximado de las mejores pruebas, (PT, AD y WT) cuando $\mu = 0.5$ es aproximadamente 1, en cambio al usar tamaños de muestra iguales a 50, su poder es aproximadamente 0.7. Ver Tabla 4-8.

μ	PT	WT	AD	KS	CH	WW
-1.00	1.00	1.00	1.00	1.00	1.00	0.86
-0.89	1.00	1.00	1.00	1.00	1.00	0.66
-0.78	1.00	1.00	1.00	1.00	1.00	0.48
-0.67	1.00	1.00	1.00	1.00	1.00	0.32
-0.56	1.00	1.00	1.00	1.00	0.99	0.20
-0.44	0.99	0.99	0.99	0.95	0.88	0.13
-0.33	0.88	0.87	0.86	0.76	0.60	0.08
-0.22	0.57	0.56	0.54	0.41	0.30	0.06
-0.11	0.18	0.19	0.17	0.15	0.09	0.04
0.00	0.06	0.06	0.06	0.05	0.06	0.06
0.11	0.18	0.18	0.18	0.14	0.09	0.05
0.22	0.57	0.56	0.54	0.41	0.31	0.07
0.33	0.90	0.88	0.87	0.78	0.64	0.07
0.44	0.99	0.98	0.98	0.94	0.89	0.12
0.56	1.00	1.00	1.00	1.00	0.98	0.18
0.67	1.00	1.00	1.00	1.00	1.00	0.30
0.78	1.00	1.00	1.00	1.00	1.00	0.50
0.89	1.00	1.00	1.00	1.00	1.00	0.66
1.00	1.00	1.00	1.00	1.00	1.00	0.83

Tabla 4-8: Estimación del poder en función μ . $N(\mu \in [-1, 1], \sigma = 1, n = 100)$ contra $N(\mu = 0, \sigma = 1, m = 1000)$.

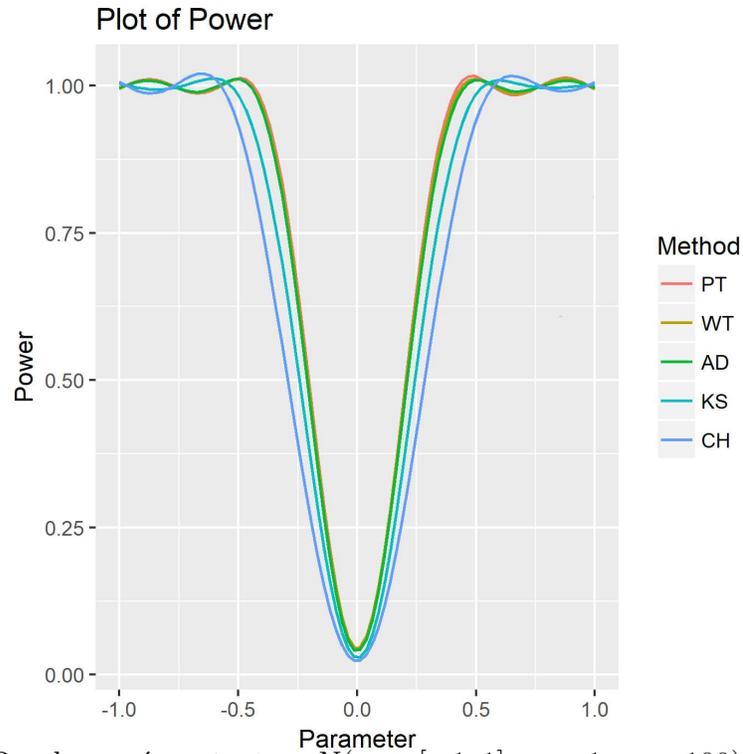


Figura 4-8: Pruebas más potentes. $N(\mu \in [-1, 1], \sigma = 1, n = 100)$ contra $N(\mu = 0, \sigma = 1, m = 1000)$.

- $N(\mu = 0, \sigma \in (0, 2), n = 100)$ contra $N(\mu = 0, \sigma = 1, m = 1000)$

En la Figura 4-9 y Tabla 4-9 se observa que las pruebas que mejor desempeño tienen son las pruebas de Anderson-Darling y Chi.Cuadrado, note que en distribuciones alternas en donde el parámetro de escala varía, estas pruebas resultan con un mayor poder.

σ	AD	CH	KS	WW	PT	WT
0.10	1.00	0.94	1.00	0.99	0.00	0.00
0.23	1.00	0.84	1.00	0.65	0.00	0.00
0.35	1.00	0.72	1.00	0.28	0.00	0.00
0.48	1.00	0.61	0.86	0.15	0.00	0.00
0.61	0.65	0.37	0.37	0.07	0.00	0.00
0.73	0.15	0.15	0.10	0.04	0.01	0.02
0.86	0.07	0.05	0.06	0.04	0.03	0.04
0.99	0.04	0.04	0.03	0.02	0.04	0.04
1.11	0.10	0.15	0.07	0.04	0.08	0.07
1.24	0.26	0.40	0.11	0.07	0.09	0.06
1.37	0.50	0.66	0.18	0.13	0.15	0.08
1.49	0.75	0.84	0.29	0.25	0.17	0.10
1.62	0.90	0.93	0.43	0.39	0.20	0.11
1.75	0.96	0.98	0.56	0.54	0.21	0.11
1.87	0.99	1.00	0.70	0.77	0.27	0.13
2.00	0.99	1.00	0.81	0.86	0.27	0.14

Tabla 4-9: Estimación del poder en función de σ . $N(\mu = 0, \sigma \in (0, 2), n = 100)$ contra $N(\mu = 0, \sigma = 1, m = 1000)$.

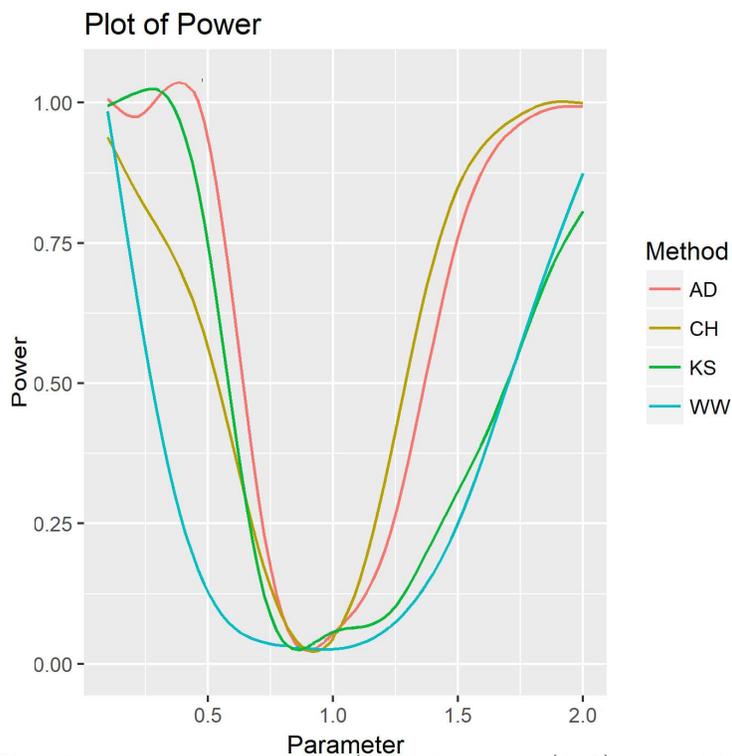


Figura 4-9: Pruebas más potentes. $N(\mu = 0, \sigma \in (0, 2), n = 100)$ contra $N(\mu = 0, \sigma = 1, m = 1000)$.

4.0.2. Soporte semi infinito

Muestras de igual tamaño

- $Exp(\lambda = 1)$ contra $F(df_1 = 2, df_2 = 5)$

La Tabla 4-10 muestra que las pruebas tienen, en general, un bajo rendimiento. La prueba más sobresaliente es la prueba de permutación; no obstante, necesita una muestra de tamaño 200 para alcanzar una potencia de 95 %. En la Figura 4-10, las mejores tres pruebas bajo la hipótesis alterna considerada.

n	PT	CH	AD	WT	KS	WW
60	0.47	0.19	0.19	0.13	0.13	0.08
70	0.49	0.21	0.22	0.16	0.10	0.06
80	0.56	0.27	0.25	0.16	0.12	0.05
90	0.65	0.32	0.29	0.18	0.14	0.06
100	0.71	0.38	0.34	0.19	0.16	0.08
110	0.73	0.42	0.38	0.23	0.20	0.09
120	0.77	0.48	0.36	0.21	0.19	0.07
140	0.83	0.56	0.43	0.25	0.28	0.09
160	0.91	0.63	0.53	0.30	0.31	0.09
180	0.93	0.68	0.57	0.32	0.37	0.10
200	0.95	0.76	0.61	0.35	0.36	0.10

Tabla 4-10: Estimación del poder en función del tamaño de muestra. $Exp(\lambda = 1)$ contra $F(df_1 = 2, df_2 = 5)$.

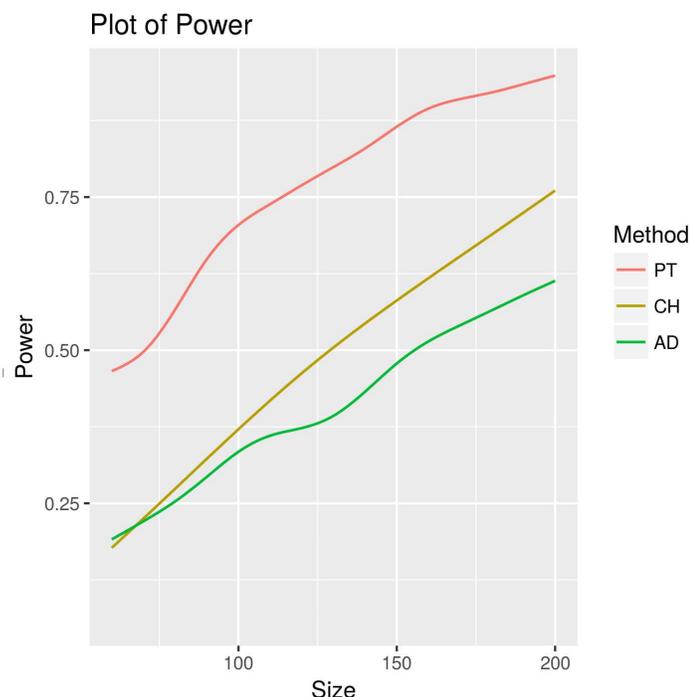


Figura 4-10: Potencias para $Exp(\lambda = 1)$ contra $F(df_1 = 2, df_2 = 5)$.

- $Exp(\lambda = 1)$ contra $Weibull(shape = 1, scale \in [1, 2], shift = 0)$. $m = n = 50$

La mayoría de las pruebas, bajo la hipótesis alterna considerada tienen un bajo rendimiento (el tamaño de ambas muestra es de 50). Solo en la prueba de Permutación, el 92 % de las veces, en promedio identifica una diferencia significativa en muestras de tamaño 50 cuyas distribuciones son $Weibull(shape = 1, scale = 1, shift = 0)$ y $Weibull(shape = 1, scale = 2, shift = 0)$. ver Tabla 4-11 y Figura 4-11.

<i>scale</i>	PT	AD	WT	KS	CH	WW
1.00	0.05	0.04	0.04	0.03	0.04	0.05
1.08	0.06	0.05	0.06	0.04	0.04	0.05
1.17	0.10	0.10	0.10	0.08	0.07	0.06
1.25	0.19	0.16	0.15	0.10	0.10	0.07
1.33	0.33	0.25	0.26	0.18	0.13	0.06
1.42	0.39	0.32	0.31	0.23	0.16	0.07
1.50	0.50	0.43	0.43	0.32	0.23	0.07
1.58	0.61	0.52	0.51	0.40	0.29	0.08
1.67	0.72	0.62	0.59	0.46	0.39	0.11
1.75	0.77	0.68	0.65	0.52	0.43	0.12
1.83	0.84	0.76	0.73	0.61	0.52	0.15
1.92	0.89	0.81	0.78	0.67	0.56	0.16
2.00	0.92	0.85	0.83	0.75	0.65	0.18

Tabla 4-11: Estimación del poder en función de *scale*. $Exp(\lambda = 1)$ contra $Weibull(shape = 1, scale \in [1, 2], shift = 0)$.

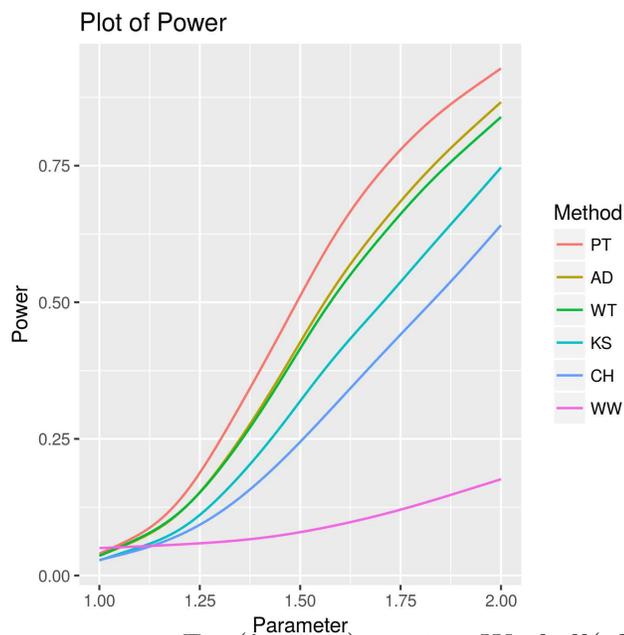


Figura 4-11: Potencias para $Exp(\lambda = 1)$ contra $Weibull(shape = 1, scale \in [1, 2], shift = 0)$.

- $Exp(\lambda = 1)$ contra $Exp(\lambda \in (0, 2])$. $m = n = 50$

Como se observa en la Tabla 4-12 y Figura 4-12, la prueba que más distingue entre las hipótesis alternas es la prueba de Permutación seguida de la prueba de Anderson-Darling.

λ	PT	AD	WT	KS	CH	WW
0.10	1.00	1.00	1.00	1.00	0.94	1.00
0.20	1.00	1.00	1.00	1.00	0.98	0.94
0.30	1.00	1.00	1.00	1.00	0.95	0.67
0.40	1.00	0.98	0.98	0.94	0.87	0.37
0.50	0.93	0.88	0.85	0.76	0.65	0.17
0.60	0.69	0.61	0.59	0.46	0.38	0.10
0.80	0.19	0.16	0.15	0.13	0.10	0.06
1.00	0.05	0.06	0.06	0.04	0.03	0.04
1.20	0.13	0.11	0.11	0.10	0.07	0.05
1.40	0.38	0.31	0.30	0.22	0.17	0.07
1.50	0.51	0.42	0.40	0.30	0.24	0.08
1.60	0.65	0.54	0.54	0.41	0.30	0.10
1.70	0.73	0.64	0.62	0.50	0.40	0.10
1.80	0.80	0.71	0.69	0.55	0.46	0.12
1.90	0.88	0.80	0.78	0.68	0.56	0.15
2.00	0.92	0.86	0.83	0.75	0.65	0.19

Tabla 4-12: Estimación del poder en función de λ . $Exp(\lambda = 1)$ contra $Exp(\lambda \in (0, 2])$.

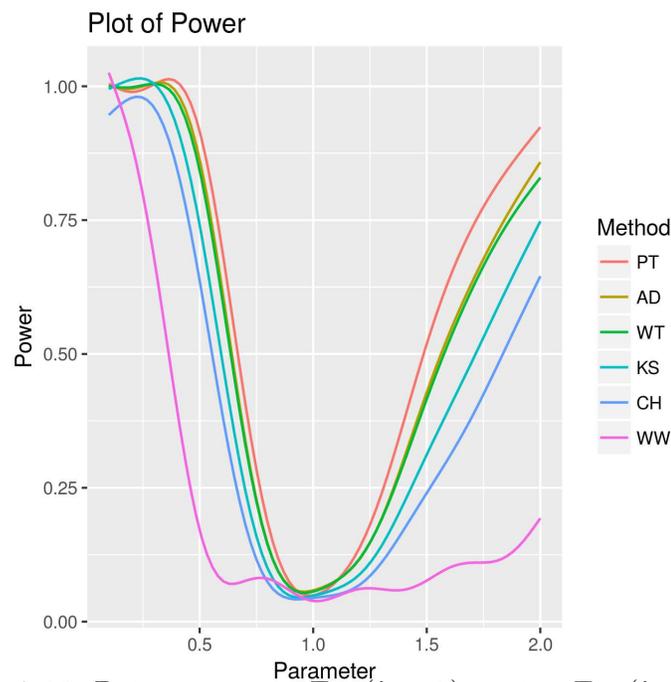


Figura 4-12: Potencias para $Exp(\lambda = 1)$ contra $Exp(\lambda \in (0, 2])$.

• $Exp(\lambda = 1)$ contra $Gamma(shape \in [0.25, 1.75])$. $m = n = 50$ En la Tabla 4-13 y Figura 4-13 vemos que a medida que $shape$ se acerca a 1, las potencias de las pruebas tienden al nivel de significancia $\alpha = 0.05$. Este comportamiento se debe a que $gamma(shape = 1, rate = 1)$ corresponde a una distribución exponencial con $\lambda = 1$. Note que las pruebas de Anderson-Darling y de Wilcoxon tiene aproximadamente el mismo poder y superan al resto de pruebas significativamente.

$shape$	AD	WT	PT	KS	CH	WW
0.25	1.00	1.00	1.00	1.00	0.47	0.94
0.42	1.00	1.00	0.95	0.98	0.40	0.48
0.58	0.84	0.84	0.68	0.73	0.25	0.15
0.75	0.36	0.35	0.29	0.26	0.11	0.08
0.92	0.08	0.07	0.07	0.06	0.05	0.05
1.00	0.05	0.05	0.06	0.04	0.03	0.06
1.08	0.07	0.06	0.08	0.05	0.04	0.05
1.25	0.29	0.29	0.21	0.21	0.09	0.05
1.42	0.56	0.56	0.46	0.41	0.17	0.08
1.58	0.83	0.83	0.72	0.71	0.33	0.17
1.75	0.96	0.95	0.90	0.89	0.48	0.26

Tabla 4-13: Estimación del poder en función de $shape$. $Exp(\lambda = 1)$ contra $Gamma(shape \in [0.25, 1.75])$.

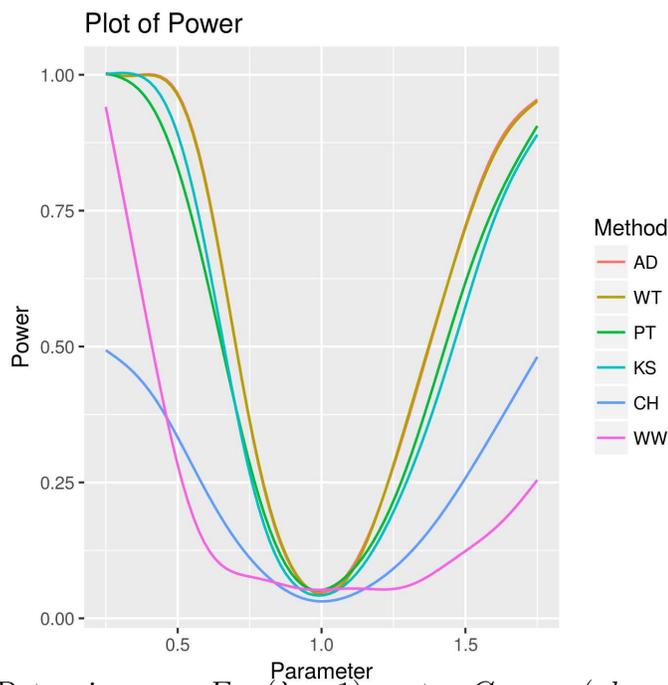


Figura 4-13: Potencias para $Exp(\lambda = 1)$ contra $Gamma(shape \in [0.25, 1.75])$.

Muestras de diferentes tamaños

- $Exp(\lambda = 1, n = 100)$ contra $Gamma(shape \in [0.25, 1.75], rate = 1, m = 1000)$

En la Tabla 4-14 y Figura 4-14 muestra un comportamiento del poder similar al tener muestras de tamaños 50. La prueba de Anderson-Darling tiene el mismo poder que una prueba de Wilcoxon bajo estas hipótesis alternas.

<i>shape</i>	AD	WT	KS	PT	CH	WW
0.25	1.00	1.00	1.00	1.00	0.79	1.00
0.46	1.00	1.00	1.00	1.00	0.64	0.90
0.68	0.99	0.98	0.96	0.90	0.21	0.19
0.89	0.25	0.23	0.19	0.17	0.04	0.04
1.00	0.06	0.05	0.05	0.05	0.05	0.04
1.11	0.21	0.22	0.18	0.17	0.10	0.04
1.32	0.92	0.92	0.85	0.82	0.52	0.08
1.54	1.00	1.00	1.00	1.00	0.88	0.18
1.75	1.00	1.00	1.00	1.00	0.99	0.33

Tabla 4-14: Estimación del poder en función de *shape*. $Exp(\lambda = 1, n = 100)$ contra $Gamma(shape \in [0.25, 1.75], m = 1000)$.

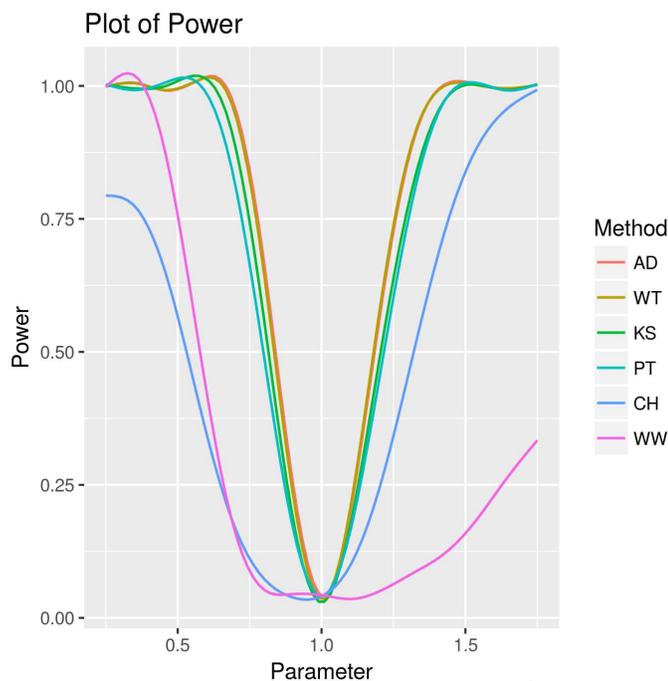


Figura 4-14: Potencias para $Exp(\lambda = 1, n = 100)$ contra $Gamma(shape \in [0.25, 1.75], m = 1000)$.

- $Exp(\lambda = 1, n = 1000)$ contra $Weibull(shape = 1, scale \in [1, 2], shift = 0, m = 100)$

Al aumentar el tamaño de una muestra, vemos en la Tabla 4-15 y Figura 4-15 que las potencias mejoran, note que la prueba de Permutación tiene una buena ventaja en comparación con las demás. La prueba de rachas de Wald-Wolfowitz es la menos potente; En todas las pruebas se alcanza un potencia aproximadamente de 1 cuando $scale = 1.75$, en cambio Wald-Wolfowitz alcanza un poder de un poco más del 0.25.

<i>scale</i>	PT	AD	CH	WT	KS	WW
1.00	0.04	0.04	0.06	0.04	0.02	0.04
1.17	0.36	0.28	0.27	0.24	0.20	0.05
1.33	0.81	0.68	0.64	0.65	0.56	0.09
1.42	0.95	0.87	0.83	0.82	0.72	0.11
1.50	0.98	0.94	0.90	0.90	0.84	0.18
1.58	1.00	0.98	0.96	0.96	0.93	0.18
1.67	1.00	0.99	0.98	0.98	0.95	0.26
1.83	1.00	1.00	1.00	0.99	0.99	0.40
2.00	1.00	1.00	1.00	1.00	0.99	0.55

Tabla 4-15: Estimación del poder en función de $scale$. $Exp(\lambda = 1, n = 1000)$ contra $Weibull(shape = 1, scale \in [1, 2], shift = 0, m = 100)$.

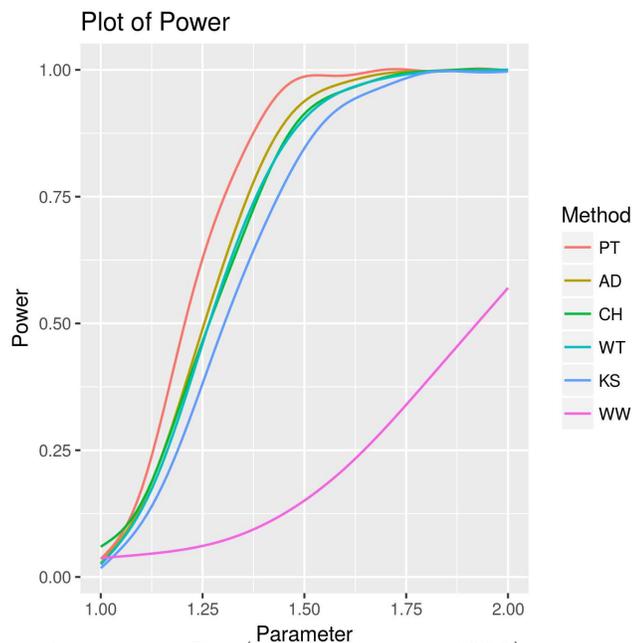


Figura 4-15: Potencias para $Exp(\lambda = 1, n = 1000)$ contra $Weibull(shape = 1, scale \in [1, 2], shift = 0, m = 100)$.

4.0.3. Soporte finito (0,1)

muestras de igual tamaño

- $U(a = 0, b = 1)$ contra $Beta(\alpha \in (0, 2], \beta = 1)$. $m = n = 50$

Las pruebas de Anderson-Darling, Permutación y Wilcoxon son las más óptimas con poderes muy similares como lo muestra la Figura 4-16. Note que la prueba de rachas de Wald-Wolfowitz tiene poco poder por que el comportamiento de las funciones de probabilidad hace que el número de rachas en la sucesión dicotomizada de la muestra combinada ordenada tenga muchas rachas a pesar de la diferencia entre las distribuciones. Ver Tabla 4-16.

α	AD	PT	WT	KS	CH	WW
0.10	1.00	1.00	1.00	1.00	0.74	1.00
0.20	1.00	1.00	1.00	1.00	0.82	0.94
0.30	1.00	1.00	1.00	1.00	0.80	0.67
0.40	0.98	0.95	0.97	0.93	0.67	0.36
0.50	0.85	0.78	0.82	0.72	0.49	0.17
0.60	0.63	0.58	0.60	0.48	0.29	0.10
0.70	0.34	0.31	0.33	0.25	0.19	0.07
0.80	0.15	0.14	0.15	0.11	0.08	0.05
0.90	0.07	0.07	0.07	0.05	0.06	0.04
1.00	0.06	0.05	0.05	0.04	0.04	0.07
1.10	0.07	0.07	0.08	0.06	0.05	0.06
1.20	0.13	0.13	0.13	0.10	0.09	0.07
1.30	0.22	0.22	0.21	0.15	0.12	0.07
1.40	0.30	0.31	0.29	0.20	0.19	0.06
1.50	0.41	0.42	0.40	0.31	0.27	0.07
1.60	0.53	0.54	0.51	0.40	0.33	0.08
1.70	0.66	0.69	0.65	0.53	0.46	0.12
1.80	0.72	0.74	0.69	0.59	0.52	0.14
1.90	0.80	0.81	0.78	0.67	0.63	0.14
2.00	0.85	0.87	0.83	0.70	0.67	0.18

Tabla 4-16: Estimación del poder en función de α . $U(a = 0, b = 1)$ contra $Beta(\alpha \in (0, 2], \beta = 1)$.

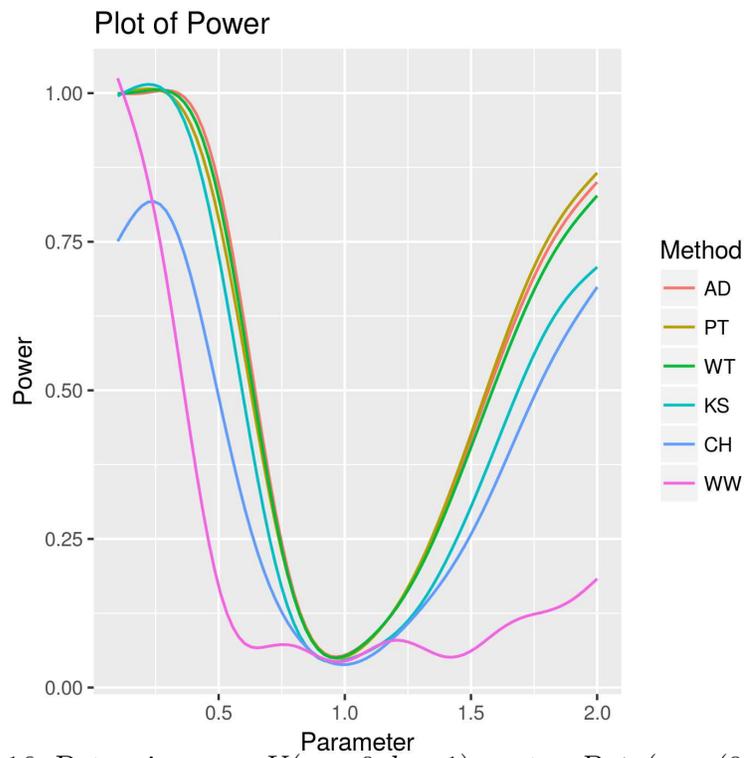


Figura 4-16: Potencias para $U(a = 0, b = 1)$ contra $Beta(\alpha \in (0, 2], \beta = 1)$.

- $U(a = 0, b = 1)$ contra $Lineal(Intercepto = 2)$.

Todas las pruebas alcanzan un poder mayor a 95% para muestras mayores a 100, excepto la prueba de rachas de Wald-Wolfowitz. Ver Figura 4-17 y Tabla 4-17.

n	PT	AD	WT	KS	CH	WW
20	0.47	0.45	0.44	0.33	0.23	0.10
30	0.68	0.65	0.65	0.49	0.41	0.12
40	0.78	0.75	0.74	0.57	0.54	0.17
50	0.86	0.85	0.83	0.73	0.68	0.18
60	0.94	0.93	0.91	0.84	0.77	0.20
70	0.96	0.95	0.93	0.86	0.83	0.21
80	0.98	0.98	0.97	0.92	0.88	0.26
90	0.99	0.99	0.98	0.94	0.92	0.24
100	1.00	0.99	0.99	0.96	0.95	0.28
120	1.00	0.99	0.99	0.99	0.98	0.34
140	1.00	1.00	1.00	1.00	1.00	0.38
160	1.00	1.00	1.00	1.00	1.00	0.40
180	1.00	1.00	1.00	1.00	1.00	0.49
200	1.00	1.00	1.00	1.00	1.00	0.51

Tabla 4-17: Estimación del poder en función del tamaño de muestra. $U(a = 0, b = 1)$ contra $Lineal(Intercepto = 2)$

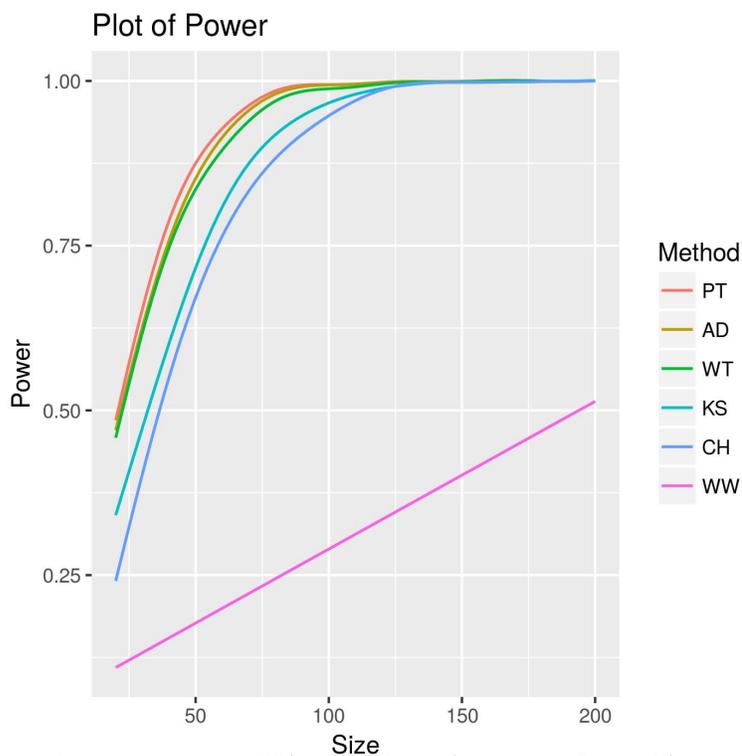


Figura 4-17: Potencias para $U(a = 0, b = 1)$ contra $Lineal(Intercepto = 2)$.

- $U(a = 0, b = 1)$ contra $Linear(Intercepto \in [0, 2])$

Nuevamente las pruebas tienen un comportamiento similar a los anteriores dos casos. Note la simetría respecto a 1, a medida que varía el parámetro del intercepto, la función de probabilidad lineal gira al rededor de la media 1/2. Ver Figura 4-18 y Tabla 4-18.

<i>Intercepto</i>	PT	AD	WT	KS	CH	WW
0.00	0.87	0.86	0.84	0.73	0.67	0.16
0.17	0.72	0.70	0.69	0.56	0.45	0.12
0.33	0.50	0.48	0.48	0.38	0.28	0.08
0.50	0.30	0.28	0.30	0.22	0.18	0.08
0.67	0.14	0.14	0.15	0.11	0.09	0.07
0.83	0.09	0.08	0.09	0.06	0.06	0.06
1.00	0.05	0.06	0.06	0.04	0.05	0.06
1.17	0.07	0.06	0.06	0.05	0.05	0.06
1.33	0.14	0.13	0.14	0.11	0.09	0.05
1.50	0.29	0.28	0.28	0.21	0.18	0.06
1.67	0.53	0.51	0.51	0.41	0.32	0.09
1.83	0.71	0.70	0.68	0.56	0.47	0.12
2.00	0.87	0.85	0.84	0.74	0.67	0.16

Tabla 4-18: Estimación del poder en función del intercepto. $U(a = 0, b = 1)$ contra $Linear(Intercepto \in [0, 2])$

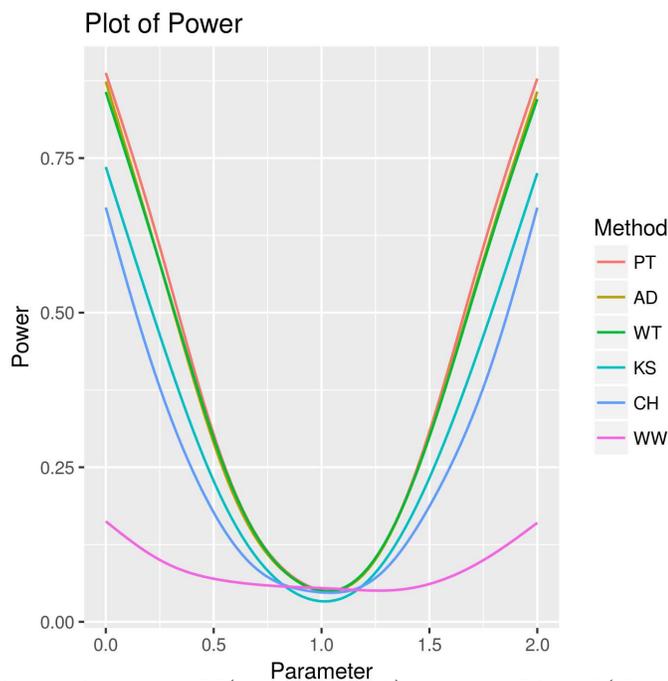


Figura 4-18: Potencias para $U(a = 0, b = 1)$ contra $Linear(Intercepto \in [0, 2])$.

- *Exponencial truncada*($min = 0, max = 1, \lambda = 1$) contra *Lineal*(*Intercepto* = 2)

La Tabla 4-19 muestra que en general, las pruebas no tienen buen poder. Se necesitan muestras de tamaño 200 para un poder superior al 90% en las pruebas de Anderson-Darling y la prueba de Permutación. En la Figura 4-19 se observa que los poderes crecen a medida que aumentan los tamaños muestrales en una forma aproximadamente lineal.

n	PT	AD	WT	CH	KS	WW
20	0.18	0.17	0.16	0.09	0.10	0.04
40	0.30	0.26	0.25	0.20	0.15	0.09
60	0.43	0.39	0.34	0.30	0.31	0.09
80	0.51	0.48	0.41	0.37	0.32	0.07
100	0.63	0.60	0.53	0.50	0.41	0.09
120	0.71	0.69	0.60	0.59	0.49	0.08
140	0.78	0.76	0.67	0.65	0.60	0.09
160	0.83	0.82	0.73	0.71	0.64	0.10
180	0.89	0.88	0.79	0.77	0.75	0.10
200	0.90	0.90	0.82	0.82	0.73	0.12

Tabla 4-19: Estimación del poder en función del tamaño de muestra. *Exponencial truncada*($min = 0, max = 1, \lambda = 1$) contra *Lineal*(*Intercepto* = 2)

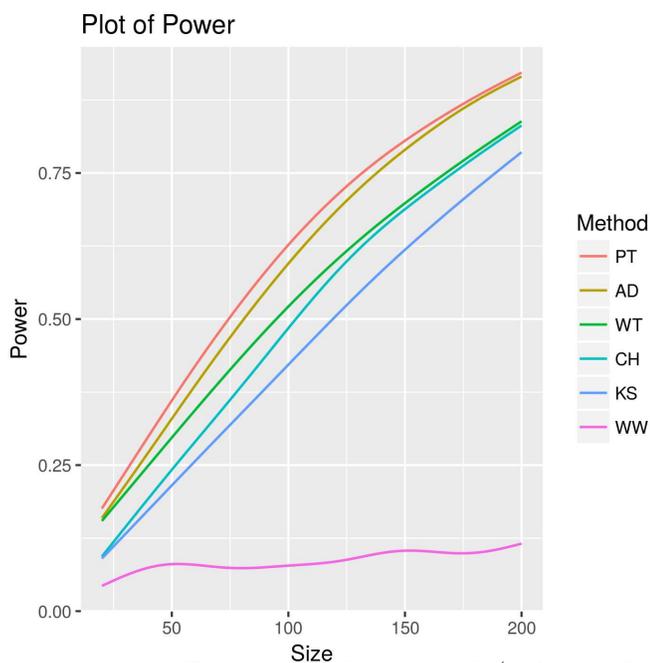


Figura 4-19: Potencias para *Exponencial truncada*($min = 0, max = 1, \lambda = 1$) contra *Lineal*(*Intercepto* = 2).

- *Exponencial truncada*($min = 0, max = 1, \lambda = 1$) contra *Lineal*(*Intercepto* = 1.5)

Como se observa en la Tabla 4-20, inclusive para tamaños de muestra bastante grandes, las pruebas tienen poco poder. Este comportamiento se debe a la gran similitud entre las dos distribuciones de la hipótesis alterna.

n	CH	AD	KS	PT	WW	WT
1000	0.07	0.06	0.07	0.06	0.05	0.05
5000	0.28	0.19	0.13	0.08	0.05	0.04
10000	0.43	0.24	0.22	0.08	0.05	0.04

Tabla 4-20: Estimación del poder en función del tamaño de muestra. *Exponencial truncada*($min = 0, max = 1, \lambda = 1$) contra *Lineal*(*Intercepto* = 1.5)

muestras de diferentes tamaños

- $U(a = 0, b = 1, n = 1000)$ contra $Beta(\alpha \in (0, 2], \beta = 1, m = 100)$.

Note que el comportamiento de las potencias de las pruebas es similar cuando tenemos muestras de tamaños 50 (ver Figura 4-16). Como se observa en la Figura 4-20, al considerar tamaños de 100 y 1000, solo aumenta las potencias. Una comparación entre la Tabla 4-21 y la tabla 4-16 muestra que, por ejemplo, la potencia de la prueba de Anderson-Darling cuando el parámetro $\beta = 1.50$ aumenta más del doble (de 0.40 a 0.91) cuando tenemos tamaños de 100 y 1000.

α	AD	PT	WT	KS	CH	WW
0.10	1.00	1.00	1.00	1.00	1.00	1.00
0.21	1.00	1.00	1.00	1.00	1.00	1.00
0.31	1.00	1.00	1.00	1.00	1.00	1.00
0.42	1.00	1.00	1.00	1.00	1.00	0.90
0.52	1.00	1.00	1.00	1.00	0.99	0.54
0.63	0.99	0.96	0.96	0.94	0.87	0.23
0.73	0.75	0.71	0.72	0.65	0.51	0.08
0.84	0.30	0.27	0.27	0.20	0.14	0.04
0.94	0.07	0.07	0.07	0.06	0.06	0.04
1.05	0.09	0.09	0.09	0.08	0.06	0.06
1.16	0.21	0.21	0.21	0.16	0.12	0.04
1.26	0.49	0.51	0.50	0.40	0.30	0.05
1.37	0.76	0.76	0.75	0.64	0.54	0.06
1.47	0.91	0.91	0.91	0.83	0.73	0.08
1.58	0.98	0.98	0.97	0.94	0.91	0.08
1.68	0.99	0.99	0.99	0.97	0.95	0.08
1.79	1.00	1.00	1.00	0.99	0.98	0.14
1.89	1.00	1.00	1.00	1.00	1.00	0.15
2.00	1.00	1.00	1.00	1.00	1.00	0.17

Tabla 4-21: Estimación del poder en función de α . $U(a = 0, b = 1, n = 1000)$ contra $Beta(\alpha \in (0, 2], \beta = 1, m = 100)$.

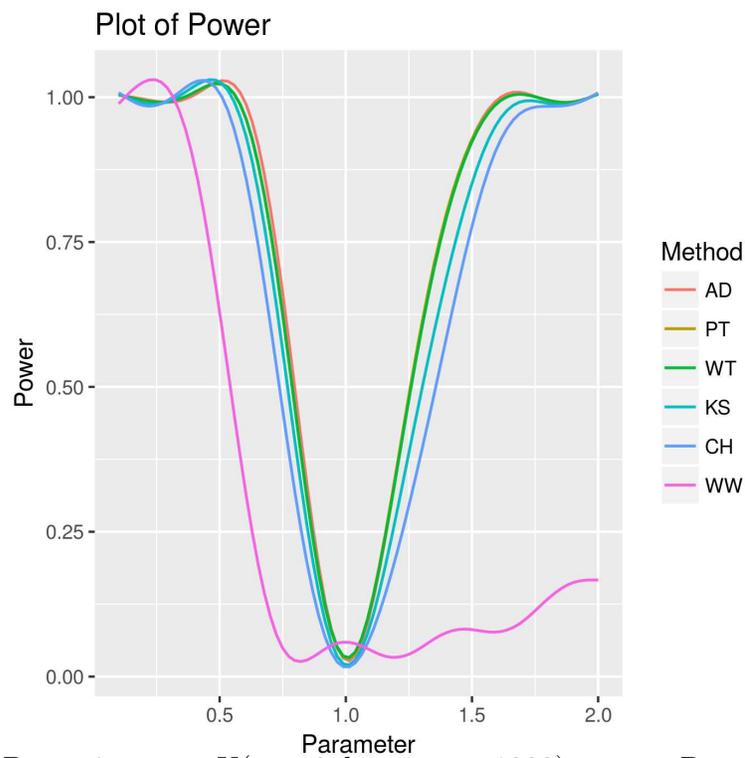


Figura 4-20: Potencias para $U(a = 0, b = 1, n = 1000)$ contra $Beta(\alpha \in (0, 2], \beta = 1, m = 100)$.

- $U(a = 0, b = 1, n = 1000)$ contra $Lineal(intercepto = 2, m \in [100, 1000])$.

La Tabla 4-21 muestra que en todas las pruebas excepto la prueba de rachas de Wald-Wolfowitz se tiene un poder de 100%.

m	PT	KS	WT	AD	CH	WW
100	1.00	1.00	1.00	1.00	1.00	0.20
200	1.00	1.00	1.00	1.00	1.00	0.39
300	1.00	1.00	1.00	1.00	1.00	0.63
400	1.00	1.00	1.00	1.00	1.00	0.76
500	1.00	1.00	1.00	1.00	1.00	0.90
600	1.00	1.00	1.00	1.00	1.00	0.86
700	1.00	1.00	1.00	1.00	1.00	0.97
800	1.00	1.00	1.00	1.00	1.00	0.99
900	1.00	1.00	1.00	1.00	1.00	0.98
1000	1.00	1.00	1.00	1.00	1.00	1.00

Tabla 4-22: Estimación del poder en función de $U(a = 0, b = 1, n = 1000)$ contra $Lineal(intercepto = 2, m \in [100, 1000])$.

- $U(a = 0, b = 1, n = 1000)$ contra $Lineal(intercepto \in [0, 2], m = 100)$.

Como se observa en la Figura 4-21 y en la Tabla 4-21 las pruebas de Anderson-Darling, Permutación y Wilcoxon tienen un poder muy similar. Nuevamente, la prueba de rachas de Wald-Wolfowitz no es eficiente en términos del poder.

$intercepto$	PT	WT	AD	KS	CH	WW
0.00	1.00	1.00	1.00	1.00	1.00	0.16
0.14	1.00	1.00	1.00	1.00	0.99	0.13
0.29	0.98	0.98	0.98	0.96	0.91	0.09
0.43	0.87	0.86	0.86	0.81	0.68	0.08
0.57	0.66	0.65	0.64	0.57	0.41	0.06
0.71	0.36	0.35	0.35	0.29	0.19	0.04
0.86	0.11	0.11	0.10	0.09	0.07	0.04
1.00	0.06	0.06	0.04	0.06	0.05	0.05
1.14	0.12	0.13	0.13	0.10	0.08	0.05
1.29	0.37	0.36	0.35	0.28	0.18	0.05
1.43	0.67	0.66	0.64	0.56	0.41	0.06
1.57	0.90	0.90	0.89	0.83	0.72	0.06
1.71	0.98	0.98	0.98	0.95	0.91	0.09
1.86	1.00	1.00	1.00	1.00	0.99	0.12
2.00	1.00	1.00	1.00	1.00	1.00	0.17

Tabla 4-23: Estimación del poder en función de $U(a = 0, b = 1, n = 1000)$ contra $Lineal(intercepto \in [0, 2], m = 100)$.

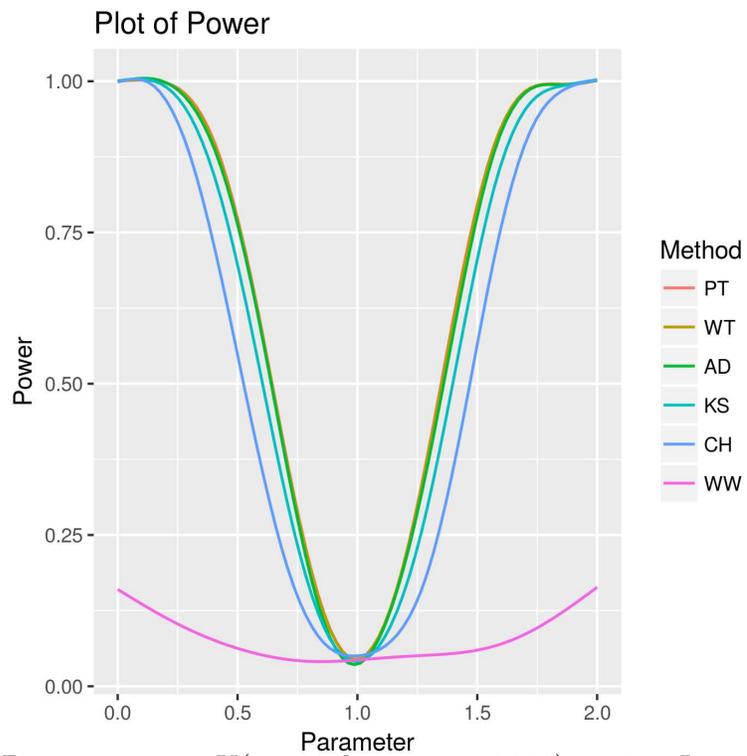
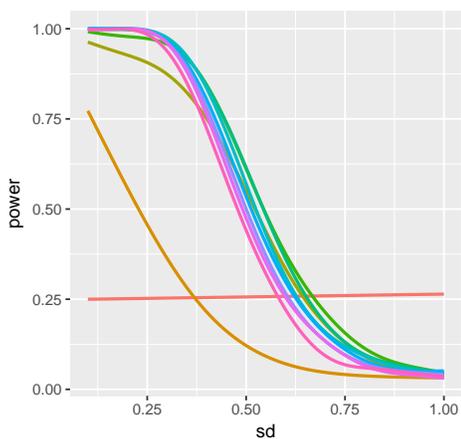


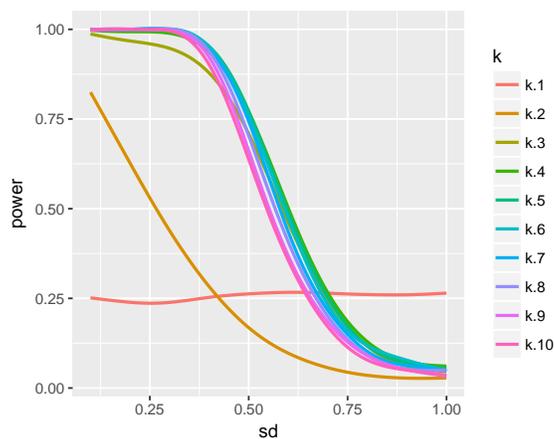
Figura 4-21: Potencias para $U(a = 0, b = 1, n = 1000)$ contra $Lineal(intercepto \in [0, 2], m = 100)$.

4.0.4. Elección del mejor número de intervalos (k) en la prueba de Chi-Cuadrado

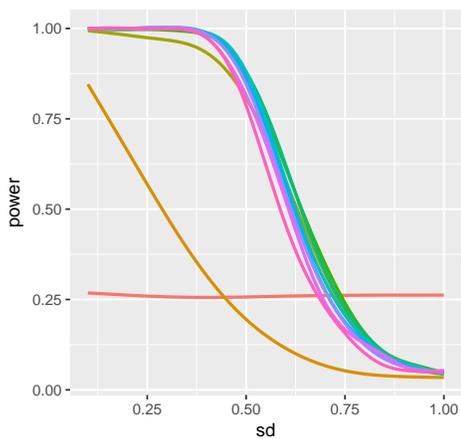
Como se mencionó en el Capítulo 2, en el uso de la prueba de Chi-Cuadrado, surge un problema en cómo discretizar los datos, es decir, cómo elegir el número de clases k y dónde ubicar los límites de dichas clases (equiprobable, equiespaciado o alguna otra configuración). En este estudio, se usan clases o intervalos equiespaciados. Queda por determinar el número de clases dependiendo del tamaño muestral, ya que el poder de la prueba se afecta por la cantidad de intervalos a considerar. Para determinar el mejor número de clases en función del tamaño muestral se considera lo siguiente: Se generan 1,000 muestras pseudoaleatorias para cada muestra: Normal estándar ($N(\mu = 0, \sigma = 1)$) y muestras normales con media 0, variando la desviación estándar de 0 hasta 1 ($N(\mu = 0, \sigma \in (0, 1])$). Para cada muestras del del mismo tamaño (30, 40, 50, 60, 70, 80, 90, 500) se estima el poder (tasa de error), considerando de 1 a 10 clases en la prueba de Chi-Cuadrado. Se tienen los siguientes resultados:



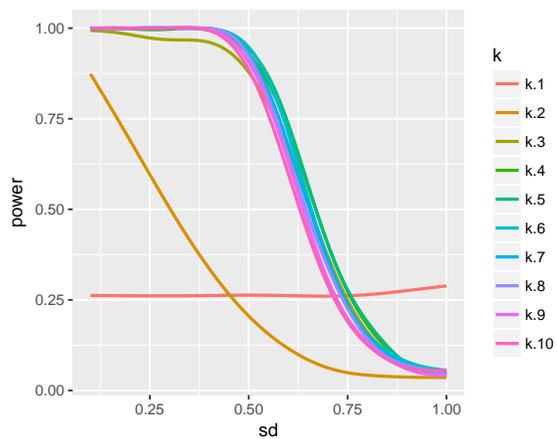
Muestras de tamaño 30



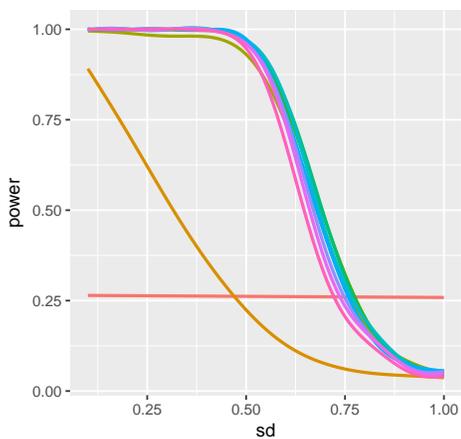
Muestras de tamaño 40



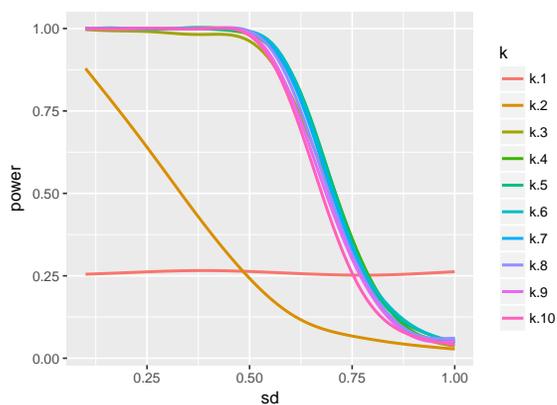
Muestras de tamaño 50



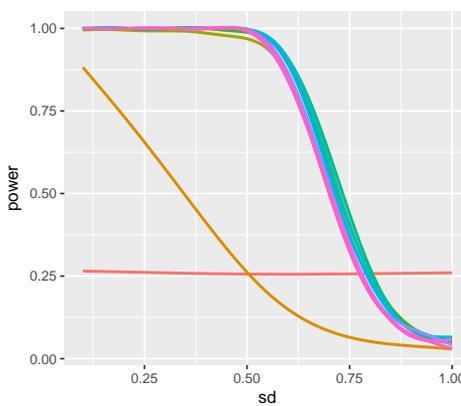
Muestras de tamaño 60



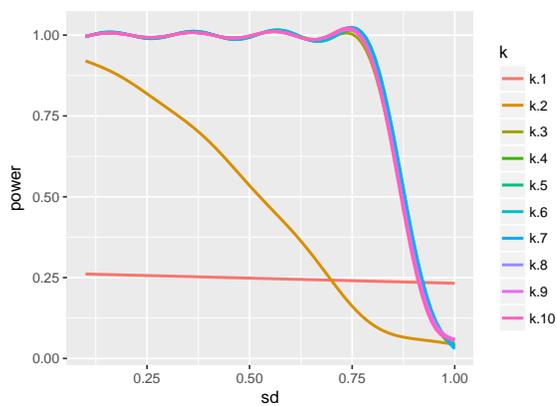
Muestras de tamaño 70



Muestras de tamaño 80



Muestras de tamaño 90



Muestras de tamaño 500

De las gráficas anteriores se tienen los mejores k para cada tamaño muestral:

Tamaños de muestra	Mejor k en orden descendente de acuerdo al poder									
30	4	5	6	7	8	3	9	10	1	2
40	4	5	6	7	8	3	9	10	1	2
50	4	5	6	7	8	3	9	10	2	1
60	5	4	6	7	8	3	9	10	2	1
70	5	6	4	7	8	3	9	10	2	1
80	5	6	4	7	8	3	9	10	2	1
90	5	6	4	7	8	3	9	10	2	1
500	6	7	5	4	8	9	3	10	2	1

Tabla 4-24: Elección del mejor k que resulta en una prueba más potente.

En la literatura estadística existen varias fórmulas para seleccionar el número de clases, entre las que se basan en el tamaño muestral se destacan: La Fórmula de Sturges, La Regla de Rice y la Raíz Cuadrada. A continuación una comparación de estas fórmulas con los mejores k de acuerdo a la Tabla 4-24.

Tamaño	Regla de Sturges $k = 1 + \log_2 n$	Raíz Cuadrada $k = \sqrt{n}$	Regla de Rice $k = 2n^{\frac{1}{3}}$	Rice Modificada $k = 2n^{\frac{1}{5}}$	k óptimo
30	6	5	6	4	4
40	6	6	7	4	4
50	7	7	7	4	4
60	7	8	8	5	5
70	7	8	8	5	5
80	7	9	9	5	5
90	7	9	9	5	5
500	10	22	16	7	6

Tabla 4-25: Comparación de distintas fórmulas para seleccionar el número de clases (se aproxima k al entero más cercano)

Como ve observa en la Tabla 4-25, la fórmula de Rice modificada es la que mejor de aproxima al k óptimo dado en las simulaciones anteriores. Note en las gráficas anteriores, que a medida que el tamaño de muestras aumenta, el poder de la prueba converge a un mismo punto indistintamente si el número de clases es superior a 3.

Capítulo 5

LA APLICACIÓN WEB INTERACTIVA

[Shiny](#) es un paquete de R que facilita la creación de aplicaciones web interactivas directamente desde R. Se pueden alojar aplicaciones independientes en una página web, insertarlas en documentos de R Markdown o crear dashboards.

En este capítulo haremos una descripción de la aplicación usada en el Capítulo 4, la cual está escrita en código R usando el paquete de Shiny. La aplicación se encuentra disponible en https://server-cristian.shinyapps.io/two_sample_problem/.

La aplicación tiene por nombre “Two Sample Problem”. En la vista principal tenemos dos opciones para escoger:

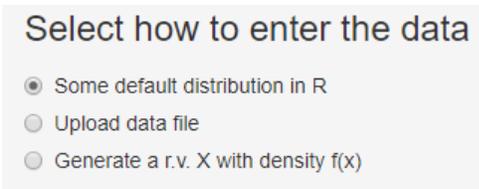
1. Do you have your own data?
2. Do you want to study the power?

En la primera opción podemos realizar las 6 pruebas descritas en el Capítulo 2 para comparar dos muestras. En la segunda opción podemos realizar un estudio del poder en las diferentes pruebas. A continuación una descripción en detalle:

5.1. Do you have your own data?

En esta parte podemos ingresar los dos conjuntos de datos para calcular los diferentes p-values correspondientes a las pruebas.

Tenemos tres formas de ingresar los datos en el panel “Select how to enter the data”:



⊙ Some default distribution in R

En el panel derecho se despliega “Select Test”, se seleccionan las pruebas que deseamos ejecutar. Seguido, aparecen las opciones para simular los dos conjuntos de datos a los cuales queremos aplicar las pruebas: “Select Data Set1” y “Select Data Set2”. Podemos simular datos que provengan de una distribución normal estándar, uniforme con soporte $(0, 1)$, exponencial con parámetro $\lambda = 1$ o una distribución t con un grado de libertad.

⊙ Upload data file

En esta opción es posible importar los datos desde un archivo (.csv) o un archivo (.xlsx). Tenemos dos formas de importarlos:

“**A single file, two columns per variable**”. Como un archivo único con las dos muestras (organizadas por columnas).

“**Two files, one for each variable**”. Dos archivos separados, uno para cada muestra

⊙ Generate a r.v. X with density f(x)

Esta opción nos permite escribir una función de densidad de probabilidad para generar una muestra aleatoria. La función debe estar definida en un intervalo acotado y ser integrable. En las barras deslizadoras “Sample1 size” y “Sample2 size” especificamos los tamaños de muestras respectivos. Por defecto, los tamaños de muestras son de 50 para cada muestra.

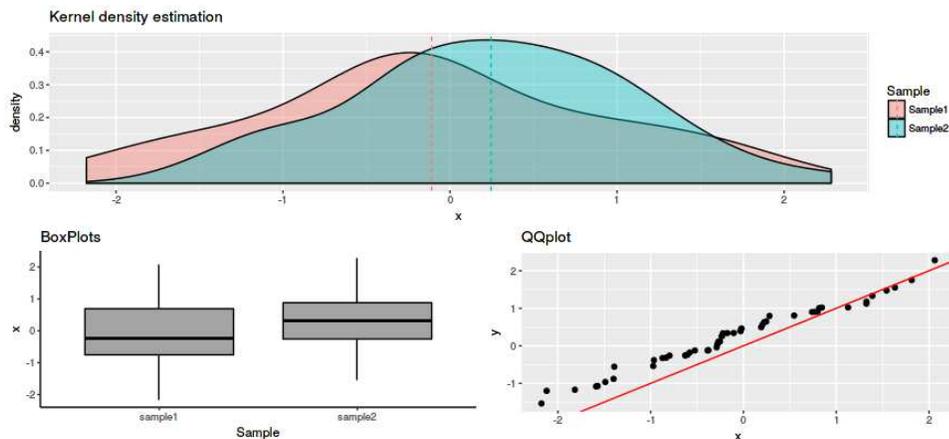
Luego de seleccionar los conjuntos de datos, en el panel derecho se mostrará las diferentes salidas: Table, Plot, Summary y Display data, que describimos a continuación.

Table: Aquí se muestran los p-values de las pruebas seleccionadas.

Method	Pvalue	Signif.codes
Permutation Test (Monte Carlo)	0.0054	**
Kolmogorov Smirnov	0	***
Wilcoxon Rank Sum	0.00064	***
Anderson-Darling	1e-05	***
Wald-Wolfowitz Runs Test	1e-05	***
Chi-squared	5e-04	***

```
[1] "Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1"
```

Plot: En esta sección se muestran tres gráficos: Las densidades estimadas no paramétricas, el diagrama de cajas y un gráfico cuantil-cuantil.



Summary: Muestra un resumen de los estadísticos más comunes para ambas muestras.

Table	Plot	Summary	Display data
		<pre> sample1 nbr.val 50.000 nbr.null 0.000 nbr.na 0.000 min -2.177 max 2.062 range 4.239 sum -5.351 median -0.225 mean -0.107 SE.mean 0.146 CI.mean.0.95 0.293 var 1.066 std.dev 1.032 coef.var -9.646 </pre>	<pre> sample2 nbr.val 50.000 nbr.null 0.000 nbr.na 0.000 min -1.534 max 2.282 range 3.816 sum 12.215 median 0.302 mean 0.244 SE.mean 0.118 CI.mean.0.95 0.237 var 0.695 std.dev 0.833 coef.var 3.411 </pre>

Display Data: Muestra los conjuntos de datos.

Table	Plot	Summary	Display data
		data1	data2
		-0.628406108540218	0.40997218694294
		1.32550525021688	-0.554998315736747
		-0.221051158510925	-0.114591088723284
		-0.972706972137183	-0.259408399646196
		-0.280615618496004	1.02627784426671
		1.12760659942923	1.0055985981907

5.2. Do you want to study the power?

Cuando seleccionamos este apartado, tenemos opciones General case, Specific case y Precalculated powers que describimos a continuación.

The image shows a light gray rounded rectangle containing four radio button options. The first two options are stacked vertically: 'Do you have your own data?' (unselected) and 'Do you want to study the power?' (selected). Below these, there are two options on the same line: 'General case' (selected) and 'Specific case' (unselected). The final option, 'Precalculated powers', is on a separate line below the others and is unselected.

⊙ General case

Aquí calculamos los poderes de las distintas pruebas cuando contrastamos las dos muestras.

Primero se debe especificar el número de simulaciones en “Number of simulations”, esto es, seleccionamos cuántas veces queremos repetir las pruebas para la proporción de veces en que se rechaza la hipótesis nula cuando la hipótesis alternativa es verdadera (poder de la prueba). Luego, “ α ” seleccionamos el nivel de significancia de las pruebas: (0.01, 0.02, 0.03, 0.05, 0.1).

Se puede seleccionar una gran variedad de distribuciones, para generar las respectivas muestras aleatorias.

Beta	Cauchy
Chi-Cuadrado	Exponencial
F	Gamma
Log Normal	Normal
t de Student	Uniforme
Weibull	Exponencial truncada en (0,1)
Distribución lineal	

También es posible visualizar las funciones de densidad de probabilidad de ambas muestras en el panel central.

Existe la opción de estimar el tiempo que tardará en ejecutarse todas las simulaciones en las distintas pruebas. (Este tiempo se especifica en “Check Time?”). Al seleccionar el tiempo, inmediatamente saldrá en pantalla un estimado del tiempo en horas, minutos y segundos de cada prueba y el tiempo global.

Después de seleccionar las distribuciones con sus respectivos parámetros y tamaños, el nivel de significancia y el número de simulaciones se oprime el botón . Se mostrará una barra indicadora en la pantalla, contabilizando el progreso de las simulaciones.

Al finalizar las simulaciones, aparecerá una tabla con el poder de cada prueba:

Method	Power
Permutation Test (Monte Carlo)	0.33
Kolmogorov Smirnov	0.20
Wilcoxon Rank Sum	0.33
Anderson-Darling	0.32
Wald-Wolfowitz Runs Test	0.07
Chi-squared	0.15

⊙ Specific case

En este apartado, podemos hacer un estudio detallado del poder de las pruebas bajos distintos escenarios; variando parámetros y tamaños de muestras. Tenemos las siguientes opciones:

What do you want to variation?

- variations of any one parameter
- variations of size (one sample)
- variations of both sample sizes

Variations of any one parameter

Al seleccionar la primera opción “variations of any one parameter”, se muestran dos paneles. En el panel de la izquierda se puede escoger la distribución de la muestra que se quiere simular. Se escoge el parámetro correspondiente a dicha distribución que deseamos variar, desde un extremo (mínimo) hasta el otro extremo (máximo) y en cuántos puntos en ese rango deseamos generar los datos:

Min	Max	Steps
0	1	25

En la parte inferior del mismo panel se fija el valor de los otros parámetros y el tamaño de la muestra.

En el segundo panel (a la derecha) se escoge la distribución de la otra muestra con su respectivos parámetros y tamaño de muestra. También se puede estimar el tiempo en que tardará en ejecutarse todas las simulaciones. La opción “Smooth curve?” por defecto está activa. Esto mostrará las gráficas del poder con las curvas suavizadas para una mejor visualización. Esta opción se puede deshabilitar y mostrar las curvas del poder con las fluctuaciones características de una simulación.

Con el botón , se ejecutan las simulaciones. Aparecerá en pantalla una barra indicadora que contabilizando el progreso de las calculaciones. Al finalizar se reporta una gráfica con los distintos poderes de las pruebas y la respectiva tabla con los valores específicos. Podemos descargar el gráfico en “Download plot”.

Variations of size (one sample)

Si se quiere comparar el poder de las pruebas para dos muestras variando solo el tamaño en una muestra, se selecciona la opción “Variations of size (one sample)”: Tal como en el caso anterior se especifican las distribuciones y parámetros en los distintos paneles.

Variations of both sample sizes

Aquí ambos tamaños de muestras varían concomitantemente, manteniendo fijos los parámetros en ambas muestras.

⊙ Precalculated powers

En este apartado se presentan algunas simulaciones ya predeterminadas. En cada simulación se describe los parámetros que se variaron y se presentan sus respectivas gráficas del poder.

Capítulo 6

CONCLUSIONES Y TRABAJOS FUTUROS

Dadas dos muestras aleatorias (que provienen de distribuciones continuas), determinar si existen evidencias (estadísticamente significativas) para concluir que ambas muestras no provienen de la misma distribución es una tarea difícil. Hemos visto que no existe una prueba superior a las demás para todas las posibles hipótesis alternas, sin embargo, la prueba de Anderson-Darling regularmente funciona mejor en comparación con las otras cinco pruebas tratadas aquí. En contraste la prueba de rachas de Wald-Wolfowitz es la de menor desempeño.

Para muestras de tamaño 50 o más, en hipótesis alternas con soporte finito, cuando las distribuciones mantienen la misma forma pero solo se diferencian en un parámetro de localización θ (desconocido), es decir la primera muestra viene de una distribución $F(x)$ y la segunda, de una distribución $F(x - \theta)$, la prueba de Anderson-Darling, Permutación y Wilcoxon tienen un buen rendimiento. Cuando las muestras provienen de distribuciones que solo se distinguen por un parámetro de escala $\varsigma \neq 0$ (desconocido), es decir la primera muestra viene de una distribución $F(x)$ y la segunda, de una distribución $F(x/\varsigma)$, las pruebas de Anderson-Darling es una de las más poderosas; solamente superada por la prueba de Chi-Cuadrado que, aunque para esta prueba los datos se discretizan perdiendo información, resulta ser la mejor prueba. En las distribuciones con soporte semi infinito las pruebas de Permutación y Anderson-Darling muestran una potencia mayor que las demás pruebas. Un comportamiento similar se observa en las distribuciones con soporte finito donde las mejores pruebas resultan ser estas mismas.

No se observa diferencia significativa en cuanto al orden de las mejores pruebas cuando se contrastan muestras con igual tamaño y muestras de tamaños muy dispares. Solamente se observa un incremento natural en las potencias de las pruebas al aumentar el tamaño de una de las muestras.

Al determinar el número óptimo de intervalos equidistantes a considerar en la prueba de Chi-Cuadrado, se observa que es irrelevante el número de intervalos a considerar (siempre y cuando sea mayor a dos) para muestras de tamaño grande.

Como trabajo futuro se podría hacer uso de herramientas de aprendizaje automático, y minería de datos para construir una prueba en base al error de clasificación. Razonando de esta manera: Si ambas muestras provienen de la misma distribución, entonces se espera que un método de clasificación, por ejemplo arboles de decisión, tenga un error de mala clasificación bastante alto. En las pruebas tratadas aquí sólo consideramos vectores de datos unidimensionales, por lo que en trabajos futuros se podría considerar extender el problema de dos muestras a varias variables, aplicando por ejemplo, pruebas como Kolmogorov Smirnov multidimensional, o reducir la dimensionalidad usando técnicas como análisis de componentes principales para luego aplicar pruebas univariadas. En un problema típico de bondad de ajuste se requiere, por ejemplo, contrastar $H_0 : X \sim N(0, 1)$ vs $H_1 : X \not\sim N(0, 1)$. Estas hipótesis pueden reformularse como un problema general de dos muestras: simulando un conjunto de datos Y de una distribución $N(0, 1)$ y contrastar $H_0 : F(X) = G(Y)$ vs $H_1 : F(X) \neq G(Y)$. En trabajos futuros se podría estudiar bajo qué condiciones es posible reemplazar un problema de bondad de ajuste en un problema de dos muestras de tal forma que ambos métodos sean estadísticamente iguales.

Dada la importancia de las distribuciones normales, como trabajo futuro se podrá realizar una comparacin entre el poder de las pruebas aqu estudiadas contra la prueba t para dos muestras independientes. Cohen estudia el tamaño del efecto para medir la magnitud de un efecto de tratamiento y lo define como $d = (\mu_1 - \mu_2)/\sigma$,

donde σ es una desviación estándar basada en una o ambas poblaciones (Cohen, 1988). Dependiendo de la desviación estándar, d se ve afectada y a su vez la potencia de la prueba. Un estudio consiste en comparar las potencias de las pruebas dependiendo el tamaño de la desviación estándar.

Extender la adición de bandas de confianza en los gráficos de Cuantil-Cuantil, requiere que las hipótesis nulas estén bien definidas. Una posible solución para implementar este método en hipótesis generales ($H_0 : F(x) = G(x)$) para el problema de dos muestras es tratar de estimar las funciones de distribución acumulada implementando métodos como densidad de núcleo y/o ajustar un modelo polinomial.

APENDICES

Apéndice A

PROPOSICIONES

Teorema A.0.1. *Sea X_1, X_2, \dots, X_n una muestra aleatoria de una distribución discreta con función de probabilidad $f_X(x_i) = p_i$, donde $x_1 < x_2 < \dots$ son los posibles valores de X en orden ascendente. Definamos*

$$P_0 = 0, P_1 = p_1, P_2 = p_1 + p_2, \dots, P_i = p_1 + p_2 + \dots + p_i, \dots$$

Sea $X_{(1)}, \dots, X_{(n)}$ las estadísticas del orden de la muestra. Entonces

$$P(X_{(j)} \leq x_i) = \sum_{k=j}^n \binom{n}{k} P_i^k (1 - P_i)^{n-k}. \quad (\text{A.1})$$

Demostración. Fijemos i , y sea Y una variable aleatoria que contabiliza el número de X_1, \dots, X_n que son menores o iguales a x_i . Para cada X_1, \dots, X_n , llamemos al evento $\{X_j \leq x_i\}$ “éxito” y al evento $\{X_j > x_i\}$ “fracaso”. Entonces, es claro que, Y es el número de éxitos en n intentos. Dado que X_1, \dots, X_n son *i.i.d.*, la probabilidad de cada éxito $P_i = P(X_j \leq x_i)$ es la misma para cada intento. Los éxitos o el fracaso de cada ensayo son independientes del resultado de cualquier otro ensayo dado que los X_j s son independientes. Luego $Y \sim \text{Binomial}(n, P_i)$.

Note que $X_{(j)} \leq x_i$ si y solo si al menos j observaciones de X_1, X_2, \dots, X_n es menor o igual a x_i , por tanto $P(X_{(j)} \leq x_i) = P(Y \geq j) = \sum_{k=j}^n \binom{n}{k} P_i^k (1 - P_i)^{n-k} \quad \square$

Teorema A.0.2. *Sea $X_{(1)}, \dots, X_{(n)}$ los estadísticos del orden de una muestra aleatoria X_1, \dots, X_n , de una distribución continua con función de distribución acumulada $F_X(x)$ y función de densidad de probabilidad $f_X(x)$. Entonces la función de densidad*

de probabilidad de $X_{(j)}$ es :

$$f_{X_{(j)}} = \frac{n!}{(j-1)!(n-j)!} f_X(x) [F_X(x)]^{j-1} [1 - F_X(x)]^{n-j}. \quad (\text{A.2})$$

Demostración. Primero encontremos la función de distribución acumulada de $X_{(j)}$ y luego diferenciarla con respecto a x . Tal como se definió en el teorema A.0.1, sea Y una variable aleatoria que contabiliza el número de X_1, \dots, X_n menores o iguales a x . Definimos el “éxito” al evento $\{X_j \leq x\}$. Note que podemos escribir $P_i = F_X(x_i)$ en el teorema A.0.1, entonces $Y \sim \text{Binomial}(n, F_X(x))$. Luego

$$F_{X_{(j)}}(x) = P(Y \geq j) = \sum_{k=j}^n \binom{n}{k} [F_X(x)]^k [1 - F_X(x)]^{n-k},$$

derivando $F_{X_{(j)}}(x)$ con respecto a x ,

$$\begin{aligned} f_{X_{(j)}}(x) &= \frac{d}{dx} F_{X_{(j)}}(x) \\ &= \sum_{k=j}^n \binom{n}{k} (k[F_X(x)]^{k-1} [1 - F_X(x)]^{n-k} f_X(x) \\ &\quad - (n-k)[F_X(x)]^k [1 - F_X(x)]^{n-k-1} f_X(x)) \\ &= \binom{n}{j} j f_X(x) [F_X(x)]^{j-1} [1 - F_X(x)]^{n-j} \\ &\quad + \sum_{k=j+1}^n \binom{n}{k} k [F_X(x)]^{k-1} [1 - F_X(x)]^{n-k} f_X(x) \\ &\quad - \sum_{k=j}^{n-1} \binom{n}{k} (n-k) [F_X(x)]^k [1 - F_X(x)]^{n-k-1} f_X(x) \\ &= \frac{n!}{(j-1)!(n-j)!} f_X(x) [F_X(x)]^{j-1} [1 - F_X(x)]^{n-j} \\ &\quad + \sum_{k=j}^{n-1} \binom{n}{k+1} (k+1) [F_X(x)]^k [1 - F_X(x)]^{n-k-1} f_X(x) \\ &\quad - \sum_{k=j}^{n-1} \binom{n}{k} (n-k) [F_X(x)]^k [1 - F_X(x)]^{n-k-1} f_X(x) \end{aligned}$$

Note que $\binom{n}{k+1}(k+1) = \frac{n!}{k!(n-k-1)!} = \binom{n}{k}(n-k)$ y los últimos dos términos cancelan, entonces

$$f_{X_j} = \frac{n!}{(j-1)!(n-j)!} f_X(x) [F_X(x)]^{j-1} [1 - F_X(x)]^{n-j}$$

□

Bibliografía

- Cohen, J. (1988). Statistical power analysis for the behavioral sciences. *NJ: Erlbaum*.
- Darling, D. (1957). The Kolmogorov-Smirnov, Cramer-von Mises tests. *Annals of Mathematical Statistics*, 28:823– 838.
- Gibbons, J. (2003). *Nonparametric Statistical Inference*. Marcel Dekker, fourth edition.
- Gnedenko, B. (1954). Tests of homogeneity of probability distributions in two independent samples (in Russian). *Doklody Akademii Nauk SSSR*, 80:525–528.
- Hoeffding, W. (1952). The large sample power of test based on permutations of observations. *Annals of Mathematical Statistics*, 23:169–192.
- Korolyuk, V. S. (1961). On the discrepancy of empiric distributions for the case of two independent samples. *Selected Translations in Mathematical Statistics and Probability*, 1:105–121.
- Makarov, A. y Simonova, G. I. (2017). Studying the power of th two-sample Anderson-Darling test in the case of contamination of one sample. *Mathematical Sciences*, 221:580– 507.
- Noiman, S. et al. (2013). The Power to See: A New Graphical Test of Normality. *The American Statistician*, 67.
- Pittitt, A. N. (1976). A Two-Sample Anderson-Darling Rank Statistic. *Biometrika*, 63:161– 168.
- Révész, P. (1968). *The Laws of Large Numbers*. Academic Press, first edition.
- Rolke, W. An extension of the normal probability plot. *Unpublished manuscript*.
- Stanford, J. y Vardeman, S. (1994). *Statistical Methods for Physical Science*. Academic Press, first edition.

- Thas, O. (2010). *Comparing Distributions*. Springer Series in Statistics, first edition.
- Wald, A. y Wolfowitz, J. (1940). On a test whether two samples are from the same population. *Ann Math*, 11:147– 162.
- Wilcoxon, F. (1954). Individual Comparisons by Ranking Methods. *Biometric*, 6:80–83.