

# UNA PRUEBA DE RACHAS PARA ALTERNATIVA DE TENDENCIA CON MUESTRAS PEQUEÑAS

por

Jimmy Corzo & Emilse Gómez Torres<sup>1</sup>

## Resumen

**Corzo J. & Gómez Torres E.:** Una prueba de rachas para alternativa de tendencia con muestras pequeñas. Rev. Acad. Colomb. Cienc. 24(92): 417-426, 2000. ISSN 0370-3908.

Se presenta un estudio para estimar la potencia de una prueba de rachas para la alternativa de tendencia monótona, se compara con otras pruebas para la misma alternativa, empleando métodos de simulación para observaciones provenientes de distribuciones doble exponencial, logística y normal. Se presenta además un método para construir la distribución exacta de la estadística de prueba.

**Palabras clave:** Prueba de rachas, tendencias, métodos de simulación.

## Abstract

A study is presented to estimate the power of a run test for the alternative of monotonous tendency and to compare it with other tests for the same alternative, using simulation methods for observations coming from distributions double exponential, logistics and normal. A method to build the exact distribution of the test statistic is presented.

**Key words:** Run tests, trends, simulation methods.

## 1. Introducción

Para una sucesión de variables aleatorias independientes  $Y_1, Y_2, \dots, Y_T$ , con funciones de distribución continuas  $F_1, F_2, \dots, F_T$ , respectivamente, una forma de realizar un análisis de tendencia es examinando la hipótesis<sup>2</sup>:

$$H_1 : F_1 = F_2 = \dots = F_T$$

Para la cual la alternativa general se puede formular como sigue:  $K_1$ : existen  $t_1 \neq t_2 \in \{1, 2, \dots, T\}$  tales que  $F_{t_1} \neq F_{t_2}$ .

<sup>1</sup> Departamento de Matemáticas y Estadística, Universidad Nacional de Colombia, Bogotá. Direcciones electrónicas: Emilse Gómez: emilse@unete.com y Jimmy Corzo: jcorzo@matematicas.unal.edu.co

<sup>2</sup> Lehman y D'Abbrera (1975), págs. 289, 313.

Dado que la construcción de pruebas para una clase tan amplia de alternativas puede conducir a pruebas ineficientes para muchas o todas las alternativas consideradas<sup>3</sup>, se estudia una clase de alternativas más restringida, que represente una tendencia monótona. Esta alternativa asume una sucesión de variables aleatorias independientes, de naturaleza continua, ordenadas en el tiempo, con distribuciones idénticas en forma pero diferentes en localización. Esto suele expresarse de la forma

$$y_t = f(t) + \varepsilon_t, \quad t = 1, 2, \dots, T \quad (1.1)$$

donde  $f(t)$  es una función monótona en  $t$  tal que  $f(t) = \beta_T a(t) + b$ ,  $\beta_T$  es un valor constante para las sucesiones de tamaño  $T$ ,  $a(t)$  es una función monótona en  $t$ , continua en  $[1, T]$  y derivable en cada punto de  $(1, T)$  y  $b$  es un número real, constante. El estudio se limitará al caso en que  $b$  es cero.  $\varepsilon_t, t = 1, 2, \dots, T$ , son variables aleatorias independientes idénticamente distribuidas con función de distribución continua  $F$ , de media cero y varianza constante, igual a uno.

De (1.1)  $Y_1, Y_2, \dots, Y_T$ , es una sucesión de variables aleatorias independientes con función de distribución continua desconocida  $F(y - \beta_T a(t))$ ,  $t = 1, 2, \dots, T$ ; por tanto la hipótesis sobre la igualdad de las distribuciones se puede formular en términos de  $a(t)$  como sigue:  $H_2 : a'(t) = 0$ , para todo  $t \in (1, T)$ .

Las alternativas de tendencia consideradas para esta hipótesis son las siguientes:

$$K_2 : a'(t) > 0$$

para todo  $t \in (1, T)$ , tendencia monótona creciente;

$$K_3 : a'(t) < 0$$

para todo  $t \in (1, T)$  tendencia monótona decreciente.

Obsérvese que la función  $a'(t)$  es responsable de la tendencia en las alternativas consideradas. Esto permite esperar que las pruebas para tales alternativas de tendencia sean más potentes en la medida que la forma funcional  $a'(t)$  presente una mayor rapidez de crecimiento, según sea la alternativa  $K_2$  o  $K_3$ .

## 2. Construcción de la prueba de rachas<sup>4</sup> para variables continuas

La construcción de la estadística de prueba requiere la transformación de las observaciones  $y_1, y_2, \dots, y_T$  en una

sucesión dicotómica<sup>5</sup>  $\eta_1, \eta_2, \dots, \eta_N$ , la cual será obtenida aplicando los siguientes dos métodos:

1. Dicotomización de signos sucesivos:

$$\eta_t = \begin{cases} 1 & \text{si } y_t < y_{t+1} \\ 0 & \text{si } y_t > y_{t+1} \end{cases} \quad (2.1)$$

para  $t = 1, 2, \dots, T - 1$ .

2. Dicotomización de la prueba del rango

$$\xi_{ij} = \begin{cases} 1 & \text{si } y_i < y_j \\ 0 & \text{si } y_i > y_j \end{cases}, \quad (2.2)$$

para  $j > i$ , con  $i = 1, 2, \dots, T - 1$  y  $j = 2, 3, \dots, T$ . La sucesión dicotomizada  $\eta_1, \eta_2, \dots, \eta_{T(T-1)/2}$  es la sucesión de los  $\xi_{ij}$  arreglados de manera que  $i = 1, \dots, T - 1$  y  $j = i + 1, \dots, T$  en ese orden. Así:  $\eta_1, \eta_2, \dots, \eta_{T-1}$  corresponde a  $\xi_{12}, \xi_{13}, \dots, \xi_{1T}$ ;  $\eta_T, \eta_{T+1}, \dots, \eta_{2T-3}$ , corresponde a  $\xi_{23}, \xi_{24}, \dots, \xi_{2T}$ , y así sucesivamente, hasta tener que  $\eta_{T(T-1)/2}$  es  $\xi_{T-1,T}$ . Todos los  $\xi_{ij}$  se pueden ver como un arreglo triangular superior, de modo que la sucesión  $\eta_1, \eta_2, \dots, \eta_{T(T-1)/2}$  se obtiene al leer el arreglo por filas, es decir cuando se toman primero los elementos de la primera fila, luego los de la segunda y así sucesivamente hasta llegar a la fila  $T - 1$ , como se ve en el siguiente esquema:

$$\begin{array}{ccccccc} \xi_{12} & \xi_{13} & \cdots & \xi_{1T} & \rightarrow & \eta_1 & \eta_2 \cdots & \eta_{(T-1)} \\ & \xi_{23} & \cdots & \xi_{2T} & \rightarrow & & \eta_T \cdots & \eta_{(T-1)+(T-2)} \\ & & \ddots & & & & & \vdots \\ & & & \xi_{T-1,T} & \rightarrow & & & \eta_{T(T-1)/2} \end{array}$$

Bajo la hipótesis nula se espera que la sucesión dicotómica  $\eta_1, \eta_2, \dots, \eta_N$  no presente "grandes" agrupamientos de elementos del mismo tipo en ninguna parte<sup>6</sup>.

Para la construcción de las estadísticas de prueba, a partir de una sucesión dicotomizada  $\eta_1, \eta_2, \dots, \eta_N$ , se procede como sigue :

<sup>3</sup>Op. Cit., pág. 105

<sup>4</sup>Dada una secuencia de dos o más tipos de símbolos, una racha se define como una sucesión de uno o más símbolos idénticos que son seguidos o precedidos por un símbolo diferente o por ningún símbolo (Gibbons (1971, pág. 50))

<sup>5</sup>Este proceso se denomina aquí dicotomización.

<sup>6</sup>Esta prueba fue propuesta por Corzo (1996, págs. 11-15).

- Definir una variable indicadora de permanencia en un estado:

$$\begin{aligned} \bar{I}_1 &= 1 \\ \bar{I}_t &= \begin{cases} 1 & \text{si } \eta_t = \eta_{t-1} \\ 0 & \text{si } \eta_t \neq \eta_{t-1} \end{cases} \end{aligned}$$

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

- Definir un indicador de la tendencia de la sucesión  $y_1, y_2, \dots, y_T$  a permanecer en algún estado hasta el  $t$ -ésimo instante:

$$\bar{r}_t = \sum_{i=1}^t \bar{I}_i \quad \text{para } t = 1, 2, \dots, N$$

que es el número de antirachas hasta el  $t$ -ésimo elemento de la sucesión  $\eta_1, \eta_2, \dots, \eta_N$ .<sup>7</sup>

Con estos elementos se define al estadístico de prueba:

$$\bar{C} = \frac{1}{\bar{r}_N} \sum_{t=1}^N \delta_t \bar{r}_t,$$

donde la función:

$$\delta_t = \begin{cases} 1 & \text{si } \eta_t = 1 \\ -1 & \text{si } \eta_t = 0 \end{cases} \quad \text{para } t = 1, 2, \dots, N,$$

sirve para indicar el estado del proceso en el  $t$ -ésimo instante.

Dada esta construcción,  $\bar{C}$  contiene información sobre la trayectoria de la sucesión  $y_1, y_2, \dots, y_T$  en términos de su tendencia a permanecer en un estado durante el período observado.

Un valor "grande" (positivo) del estadístico  $\bar{C}$  refleja una tendencia a permanecer en un estado al darse muchos agrupamientos de unos. Un valor "pequeño" (negativo) del estadístico  $\bar{C}$  indica una tendencia a permanecer en el otro estado pues se tienen muchos agrupamientos de ceros. La hipótesis nula se rechaza en favor de la alternativa de tendencia a permanecer en algún estado cuando el estadístico toma valores absolutos "grandes".

<sup>7</sup>Nótese que, si se define  $I_1 = 1$ ,  $I_t = 1 - \bar{I}_t$ , para  $t = 2, \dots, N$ , se obtiene

$$r_t = \sum_{k=1}^t I_k,$$

que corresponde al número de rachas hasta el  $t$ -ésimo instante en la sucesión dicotomizada  $\eta_1, \eta_2, \dots, \eta_N$ . Además, se puede mostrar que:

$$\bar{r}_t = t - r_t + 1;$$

por esta razón,  $\bar{r}_t$  es un estadístico de rachas, y de ahí su nombre número de antirachas.

### 3. Distribución exacta del estadístico $\bar{C}$ bajo la hipótesis nula y tabla de valores críticos

Como las dos dicotomizaciones utilizadas identifican el crecimiento (decrecimiento) de acuerdo con la magnitud relativa de las observaciones, para dicotomizar una muestra aleatoria proveniente de una distribución continua basta tomar la sucesión de sus  $T$  rangos. Sea  $v_1, v_2, \dots, v_T$  la sucesión de rangos de las  $T$  observaciones, donde  $v_i$  es el rango de  $y_i$  en la sucesión  $\{y_1, y_2, \dots, y_T\}$ . Nótese que la estructura de crecimiento se conserva (cuando  $y_i < y_j, v_i < v_j$  y cuando  $y_i > y_j, v_i > v_j$ ); de manera que los métodos de dicotomización (2.1) y (2.2) se pueden expresar en términos de los rangos así:

1. Dicotomización de signos sucesivos:

$$\eta_t = \begin{cases} 1 & \text{si } v_t < v_{t+1} \\ 0 & \text{si } v_t > v_{t+1} \end{cases} \quad (3.1)$$

para  $t = 1, 2, \dots, T - 1$ .

2. Dicotomización de la prueba del rango

$$\xi_{ij} = \begin{cases} 1 & \text{si } v_i < v_j \\ 0 & \text{si } v_i > v_j \end{cases}, \quad (3.2)$$

para  $j > i$ , con  $i = 1, 2, \dots, T - 1$  y  $j = 2, 3, \dots, T$ .

Sea  $\Omega_T$  el conjunto de todas las permutaciones de los  $T$  rangos y denotando  $v = \{v_1, v_2, \dots, v_T\}$ ; para un arreglo cualquiera  $\varpi \in \Omega_T$ ,  $\varpi = \{\varpi_1, \varpi_2, \dots, \varpi_T\}$ , donde  $\varpi_i \in \{1, 2, \dots, T\}$ :

$$P(v = \varpi) = P(v_1 = \varpi_1, v_2 = \varpi_2, \dots, v_T = \varpi_T) = \frac{1}{T!}$$

Como cada arreglo de  $v$  produce una sucesión dicotómica, también genera un valor del estadístico  $\bar{C}$ , por tanto los valores posibles del estadístico  $\bar{C}$  se obtienen a partir de las  $T!$  permutaciones de los enteros de 1 a  $T$ . Denotando como  $\#\{\varpi \in \Omega_T : \bar{C} = c\}$  al número de arreglos  $\varpi \in \Omega_T$  tales que el valor de  $\bar{C}$  es  $c$ , se tiene que la probabilidad del valor  $c$  es:

$$P(\bar{C} = c) = \frac{\#\{\varpi \in \Omega_T : \bar{C} = c\}}{T!}$$

En las siguientes tablas se ilustra el proceso para obtener la distribución exacta de los estadísticos  $\bar{C}_1$ , que usa

la dicotomización de signos sucesivos, y  $-C_2$ , que usa la dicotomización del rango, para una muestra de 4 observaciones. La tabla 1 contiene las  $4!$  permutaciones de los enteros de 1 a 4, las sucesiones dicotómicas obtenidas y el valor de los estadísticos  $-C_1$  y  $-C_2$ ; además se incluyen los estadísticos de las pruebas de Mann ( $\tau^*$ ) y de Daniels ( $\rho^*$ ), correspondientes a modificaciones en los coeficientes de correlación de Kendall ( $\tau$ ) y de Spearman ( $\rho$ )<sup>8</sup>.

En la tabla 1 se observan varios aspectos:

- $-C_2$  es más sensible que  $-C_1$  a los cambios que presenta la sucesión de rangos, pues produce una mayor cantidad de valores del estadístico, lo cual permite detectar con mayor efectividad desviaciones de la hipótesis nula.

- $\tau^*$  considera el número de incrementos sucesivos en la sucesión, tomando sólo un valor más que  $-C_1$ .

- $\rho^*$  refiere los incrementos o decrementos de acuerdo con la magnitud relativa de la observación dentro de la sucesión, dando un rango de variación más amplio que  $-C_1$ , pero menor que  $-C_2$ .

Tabla 1. Valores de los estadísticos para una muestra de 4 observaciones

t-ésimo No. t-ésimo obs	t-ésimo Rangos $v = \{v_1, v_2, \dots, v_T\}$	t-ésimo Dicotomización con t-ésimo signos sucesivos		t-ésimo Dicotomización de la t-ésimo prueba del rango		t-ésimo Otros estadís- t-ésimos	
		t-ésimo Sucesión t-ésimo Dicotómica	t-ésimo Valor de $C_1$	t-ésimo Sucesión t-ésimo Dicotómica	t-ésimo Valor de $C_2$	$\tau^*$	$\rho^*$
1	4 3 2 1	0 0 0	-2	0 0 0 0 0	-3,5	-1	-1
2	3 4 2 1	1 0 0	-1	1 0 0 0 0	-2,8	-0,667	-0,8
3	4 2 3 1	0 1 0	-1	0 0 0 1 0	-2,5	-0,667	-0,8
4	2 4 3 1	1 0 0	-1	1 1 0 0 0	-2,2	-0,333	-0,4
5	4 3 1 2	0 0 1	-0,5	0 0 0 0 1	-2	-0,667	-0,8
6	3 2 4 1	0 1 0	-1	0 1 0 1 0	-1,5	-0,333	-0,4
7	3 4 1 2	1 0 1	1	1 0 0 0 1	-1,25	-0,333	-0,6
8	1 4 3 2	1 0 0	-1	1 1 1 0 0	-1,2	0	0,2
9	4 1 3 2	0 1 0	-1	0 0 0 1 1	-0,75	-0,333	-0,4
10	2 3 4 1	1 1 0	0,5	1 1 0 1 0	-0,667	0	-0,2
11	4 2 1 3	0 0 1	-0,5	0 0 0 1 1	-0,2	-0,333	-0,4
12	2 4 1 3	1 0 1	1	1 0 1 0 0	0	0	0
13	3 1 4 2	0 1 0	-1	0 1 0 1 1	0	0	0
14	1 3 4 2	1 1 0	0,5	1 1 1 1 0	0,2	-0,3333	0,4
15	3 2 1 4	0 0 1	-0,5	0 0 1 0 1	0,667	0	0,2
16	1 4 2 3	1 0 1	1	1 1 1 0 0	0,75	0,3333	0,4
17	4 1 2 3	0 1 1	1	0 0 0 1 1	1,2	0	-0,2
18	2 1 4 3	0 1 0	-1	0 1 1 1 0	1,25	0,3333	0,6
19	2 3 1 4	1 0 1	1	1 0 1 0 1	1,5	0,3333	0,4
20	1 2 4 3	1 1 0	0,5	1 1 1 1 0	2	0,6667	0,8
21	3 1 2 4	0 1 1	1	0 0 1 1 1	2,2	0,3333	0,4
22	1 3 2 4	1 0 1	1	1 1 1 0 1	2,5	0,6667	0,8
23	2 1 3 4	0 1 1	1	0 1 1 1 1	2,8	0,6667	0,8
24	1 2 3 4	1 1 1	2	1 1 1 1 1	3,5	1	1

De la tabla 1 se construyen las tablas 2 y 3 que tienen las distribuciones de frecuencias acumuladas de  $-C_1$  y  $-C_2$ , de las cuales se extraen los valores críticos correspondientes al nivel de significancia deseado, que son los valores de los estadísticos con probabilidades más cercanas a éste.

Es importante resaltar que en ambos casos la distribución exacta del estadístico es simétrica con respecto a cero; en Corzo (1993) se pueden consultar los detalles acerca de la demostración de la simetría alrededor de cero.

### Tabla de valores críticos de $-C$

Las tablas de valores críticos para  $-C_1$  y  $-C_2$  (tablas 4 y 5), se construyen a partir de las distribuciones acumuladas exactas, seleccionando de estas últimas, aquellos valores de la estadística correspondientes a los niveles de significancia más cercanos a 0,5%, 1%, 2,5% y 5%.

En estas tablas se tienen los valores de  $-C_1$  y  $-C_2$  cuyo nivel de significancia natural,  $\alpha^*$ , es más cercano por defecto ( $\alpha^* \leq \alpha$ , primer renglón de la celda) y por exceso ( $\alpha^* > \alpha$ , segundo renglón de la celda) al deseado,  $\alpha$ .

<sup>8</sup>Conover (1980, págs. 252-261)

Tabla 2. Distribución exacta de  $^{-}C_1$  para muestras de 4.

$c$	-2	-1	-0.5	0.5	1	2
Distribución de probabilidades $P(^{-}C = c)$	0.042	0.333	0.125	0.125	0.333	0.042
Distribución acumulada $P(^{-}C \leq c)$	0.042	0.375	0.5	0.625	0.958	1

Tabla 3. Distribución exacta de  $^{-}C_2$  para muestras de 4.

En esta tabla no se incluye la distribución de probabilidades, pues al tomar tantos valores diferentes se tiene que  $P(^{-}C = c) = 0,04167$  para  $c \neq 0$  y  $P(^{-}C = 0) = 0,08333$ .

$c$	-3.5	-2.8	-2.5	-2.2	-2	-1.5	-1.25	-1.2
Distribución acumulada $P(^{-}C \leq c)$	0.042	0.083	0.125	0.167	0.208	0.250	0.292	0.333

$c$	-0.75	-0.667	-0.2	0	0.2	0.667	0.75	1.2
Distribución acumulada $P(^{-}C \leq c)$	0.375	0.417	0.458	0.542	0.583	0.625	0.667	0.708

$c$	1.25	1.5	2	2.2	2.5	2.8	-3.5
Distribución acumulada $P(^{-}C \leq c)$	0.750	0.792	0.833	0.875	0.917	0.958	1.000

Tabla 4. Valores críticos de  $^{-}C_1$  para muestras de tamaño 4 a 11.

Los valores entre paréntesis son los niveles de significancia naturales, esto es  $\alpha^* = P[C \geq c]$ <sup>9</sup>

Tamaño Sucesión	$\alpha$			
	5%	2.5%	1%	0.5%
4	2 (0.04167)	*	*	*
	1 (0.37500)	2 (0.04167)	2 (0.04167)	2 (0.04167)
5	1.667 (0.04167)	2.5 (0.00833)	2.5 (0.00833)	*
	1.5 (0.19167)	1.667 (0.04167)	1.667 (0.04167)	2.5 (0.00833)
6	2.25 (0.00833)	2.25 (0.00833)	2.25 (0.00833)	3 (0.00139)
	2 (0.07361)	2 (0.07361)	2 (0.07361)	2.5 (0.00833)
7	1.75 (0.03691)	2.5 (0.02282)	2.8 (0.00139)	2.8 (0.00139)
	1.667 (0.08472)	2.2 (0.02579)	2.5 (0.02282)	2.5 (0.02282)
8	2.25 (0.02510)	2.4 (0.00878)	2.4 (0.00878)	3.333 (0.00020)
	2 (0.08968)	2.25 (0.02510)	2.25 (0.02510)	3 (0.00598)
9	2 (0.04994)	2.667 (0.00733)	2.667 (0.00733)	3 (0.00183)
	1.833 (0.05039)	2.5 (0.03481)	2.5 (0.03481)	2.8 (0.00637)
10	2.143 (0.04306)	2.286 (0.02221)	3.286 (0.00158)	3.286 (0.00158)
	2 (0.09803)	2.25 (0.04218)	3 (0.01196)	3 (0.01196)
11	2.222 (0.04731)	2.571 (0.01726)	2.857 (0.00620)	3.125 (0.00424)
	2.2 (0.05862)	2.5 (0.04302)	2.8 (0.01379)	3 (0.00601)

<sup>9</sup>El estadístico no tiene valores con probabilidad inferior al nivel de significancia deseado.

Obsérvese que, como  $-C_2$  toma una mayor cantidad de valores que  $-C_1$ , los niveles de significancia naturales de los valores críticos son más cercanos al deseado, reflejándose en una mayor sensibilidad a desviaciones de la hipótesis nula de la prueba basada en  $-C_2$  con respecto a la prueba basada en  $-C_1$ .

**Tabla 5. Valores críticos de  $-C_2$  para muestras de tamaño 4 a 11.**  
Los valores entre paréntesis son los niveles de significancia naturales, esto es  $\alpha^* = P[C \geq c]$ .

Tamaño Sucesión	$\alpha$			
	5%	2.5%	1%	0.5%
4	3.5 (0.041667)	*	*	*
	2.8 (0.083333)	3.5 (0.04167)	3.5 (0.04167)	3.5 (0.04167)
5	4.5 (0.04167)	4.556 (0.02500)	5.5 (0.00833)	*
	4 (0.05833)	4.5 (0.04167)	4.889 (0.01667)	5.5 (0.00833)
6	5.1 (0.05)	6 (0.025)	6.857 (0.00972)	7.2143 (0.00417)
	5.091 (0.05139)	5.923 (0.02639)	6.5 (0.01111)	7 (0.00833)
7	6.25 (0.05)	7.222 (0.02480)	8.278 (0.00912)	9.25 (0.00357)
	6.235 (0.05020)	7.214 (0.02559)	8.25 (0.01111)	9 (0.00556)
8	7.476 (0.04988)	8.68 (0.02500)	10.038 (0.00960)	10.870 (0.00494)
	7.474 (0.05007)	8.667 (0.02547)	10 (0.01014)	10.864 (0.00501)
9	8.731 (0.04998)	10.219 (0.02499)	11.828 (0.00997)	12.852 (0.00498)
	8.727 (0.05001)	10.217 (0.02502)	11.826 (0.01006)	12.846 (0.00501)
10	10.455 (0.04998)	11.805 (0.02496)	13.718 (0.00999)	14.967 (0.00498)
	10.435 (0.050021)	11.8 (0.02500)	13.714 (0.01000)	14.966 (0.00500)
11	11.431 (0.04996)	13.448 (0.02499)	15.683 (0.00999)	17.127 (0.004997)
	11.429 (0.05000)	13.447 (0.02501)	15.682 (0.01000)	17.125 (0.005002)

giendo  $\beta_T = 1$ , para tendencia creciente, y  $\beta_T = -1$ , para tendencia decreciente, y eligiendo un caso de  $a^j(t)$  (tabla 7), según se quiera tendencia lenta (casos 1, 2 y 3), moderada (casos 4 y 5) o rápida (casos 6, 7 y 8) y estableciendo un tamaño para la sucesión,  $T = 5, 6, \dots, 11$ .

#### 4.1. Algoritmo para la estimación de la potencia

Esta estimación es un procedimiento iterativo, donde cada iteración  $i$  está dada por la realización de los siguientes pasos:

Paso 1: Simular una sucesión de  $T$  observaciones con el comportamiento de alguna de las funciones de tendencia escogidas (funciones utilizadas por Aiyar, Guillier y Albers (1979)) (Tabla 7), procediendo de la siguiente manera:

- *Parte aleatoria:* generar un número aleatorio  $u_t$  de la distribución uniforme (0, 1), y transformarlo

#### 4. Estudio de potencia

La estimación de las potencias de las pruebas se efectúa para cada función de tendencia específica con la forma definida en la ecuación (1.1), la cual se obtiene seleccionando una distribución para  $\varepsilon_t$  (tabla 6), esco-

en una observación de la distribución seleccionada para  $\varepsilon_t$ ,  $t = 1, 2, \dots, T$ , usando su función percentil (funciones utilizadas por Aiyar, Guillier y Albers (1979)) (tabla 6).

- *Parte determinística:* calcular  $f(t) = \beta_T a^j(t)$ ,  $t = 1, 2, \dots, T$ .
- Obtener la  $t$ -ésima observación en la sucesión como  $y_t = f(t) + \varepsilon_t$

Paso 2: Calcular los estadísticos de prueba para la sucesión de observaciones generadas.

Paso 3: Decidir si se rechaza o no la hipótesis nula con base en los estadísticos calculados.

Paso 4: Replicar los pasos anteriores  $n$  veces, donde  $n_i = 2n_{i-1}$  ( $n_1 = 100$ ).

Paso 5: Estimar la potencia: calcular el porcentaje de rechazos de la hipótesis nula al unir las sucesiones de

la iteración anterior con las generadas en la iteración actual.

Paso 6: Determinar la estabilidad de la estimación de la potencia: comparar las potencias estimadas en las dos últimas iteraciones, si la diferencia entre ellas es menor o igual que  $0,01^{10}$  finaliza el proceso, en otro caso se realiza la siguiente iteración,  $i + 1$ .

#### 4.2 Resultados del estudio de potencia

La comparación de pruebas se efectúa a través del cotejo de las potencias empíricas de las pruebas en consideración, cuando se evalúan bajo las mismas condiciones o, en algunos casos, condiciones muy similares. En este estudio, se comparan las potencias de las pruebas que utilizan los estadísticos propuestos  $^{-}C_1$  y  $^{-}C_2$  con las de Mann basada en  $\tau^*$  y de Daniels basada en  $\rho^*$ .

En la Plancha 1 se presentan las tablas y gráficos de las potencias estimadas de las pruebas al examinar la

hipótesis de una cola al 5% con 12800 muestras, cuando  $\beta_T = 1$  (tendencia creciente) y la distribución de  $\varepsilon_t$  es exponencial doble; se ha estructurado de manera que facilite la exploración de las potencias estimadas al aumentar la rapidez de la tendencia según las funciones seleccionadas, presentando 7 secciones, cada una con una tabla y un gráfico asociado a ésta, utilizando la siguiente notación:

- $T$ : tamaño de la muestra; varía entre 5 y 11
- $NS$  nominal: nivel de significancia nominal de cada prueba
- Caso: modalidad de la tendencia ensayada, va de 1 a 8 (tabla 7)
- $PC_1$ : potencia estimada de la prueba con  $^{-}C_1$  (dicotomización de signos sucesivos)
- $PC_2$ : potencia estimada de la prueba con  $^{-}C_2$  (dicotomización de la prueba del rango)
- $P_{tau}$ : potencia estimada de la prueba de Mann (con  $\tau^*$ )
- $P_{rho}$ : potencia estimada de la prueba de Daniels (con  $\rho^*$ )

Tabla 6. Función percentil de distribuciones estandarizadas.

Distribución	Función percentil de distribución estándar (número aleatorio)
Exponencial doble	$\varepsilon_t = \begin{cases} \frac{1}{\sqrt{2}} \log_e (2u_t) & \text{si } 0 < u_t \leq \frac{1}{2} \\ -\frac{1}{\sqrt{2}} \log_e (2(1 - u_t)) & \text{si } \frac{1}{2} < u_t < 1 \end{cases}$
Logística	$\varepsilon_t = -\frac{\sqrt{3}}{\pi} \log_e \left( \frac{1}{u_t} - 1 \right)$
Normal	<p>Con dos números aleatorios de uniforme (0,1), <math>u_{t_1}, u_{t_2}</math>, se tiene:</p> $\varepsilon_{t_1} = \sqrt{-2 \log_e (u_{t_1})} \cos (2\pi u_{t_2})$ $\varepsilon_{t_2} = \sqrt{-2 \log_e (u_{t_1})} \sin (2\pi u_{t_2})$

En la lectura de los resultados se tiene en cuenta que la desigualdad de las condiciones bajo las cuales se aplican las pruebas incide en las potencias, en este caso se

tiene que los niveles de significancia nominales no son iguales. Nótese que los niveles de significancia de la prueba con  $^{-}C_1$  son muy bajos con respecto al deseado,

<sup>10</sup>Corzo (1984)

a excepción de  $T = 9$  y  $11$ . Para la prueba con  ${}^{-}C_2$  son muy cercanos al 5%, excepto cuando  $T = 5$ . Para la prueba de Mann siempre está por debajo de 4.2%. Para la prueba de Daniels es muy cercano al 5% a partir de  $T = 8$ .

En este sentido se observa que para  $T = 6$ , las potencias son más bajas que con otros tamaños de muestra, empezando a crecer en el caso 5, pasando a su máximo valor desde el caso 6.

Esta disminución se puede atribuir al hecho de que para este tamaño de muestra el nivel de significancia natural de la prueba es 0,8333% (ver tabla 5), el cual

es bastante menor que el nivel de significancia teórico, 5%; de modo que la probabilidad de rechazo es más baja que con otros tamaños de muestra, cuyos niveles de significancia exactos son más cercanos al teórico.

Para  $T = 5$  y  $6$ , con excepción de la potencia de la prueba con  ${}^{-}C_1$ , las potencias de las otras pruebas son bajas en los tres primeros casos, empiezan a crecer en el caso 4, aumentan un poco en el caso 5 y se maximizan desde el caso 6.

A partir de  $T = 7$  el aumento en el caso 5 es más pronunciado, siendo las potencias muy cercanas al valor máximo, excepto la de la prueba con  ${}^{-}C_1$ .

Tabla 7. Casos de tendencia utilizados<sup>11</sup>

	Caso $j$	$a^j(t)$	$a^j(t)'$ : Rapidez de crecimiento
Lenta ( $a^j(t)' < \frac{1}{4}$ )	Caso 1	$a^1(t) = -t^{-1/8}$	$a^1(t)' = (1/8)t^{-9/8}$
	Caso 2	$a^2(t) = -t^{-1/4}$	$a^2(t)' = (1/4)t^{-5/4}$
	Caso 3	$a^3(t) = t^{1/4}$	$a^3(t)' = (1/4)t^{-3/4}$
Moderada ( $\frac{1}{4} \leq a^j(t)' \leq 1$ )	Caso 4	$a^4(t) = \log_e t$	$a^4(t)' =  t ^{-1}$
	Caso 5	$a^5(t) = t$	$a^5(t)' = 1$
Rápida ( $a^j(t)' > 1$ )	Caso 6	$a^6(t) = t^2$	$a^6(t)' = 2t$
	Caso 7	$a^7(t) = t^3$	$a^7(t)' = 3t^2$
	Caso 8	$a^8(t) = t^4$	$a^8(t)' = 4t^3$

Las potencias de la prueba con  ${}^{-}C_1$  siempre son menores que las potencias de las otras pruebas, siendo parecidas a las de Mann y de Daniels ante tendencias lentas, menores ante moderadas e iguales ante rápidas.

Para  $T = 5$ , la potencia de la prueba con  ${}^{-}C_2$  empieza a crecer en el caso 4, aumenta un poco en el caso 5 y se maximiza a partir del caso 6. Para  $T = 6$ , el comportamiento de la potencia es análogo, pero a partir de  $T = 7$  el aumento en el caso 5 es más pronunciado, siendo la potencia muy cercana al máximo.

La prueba con  ${}^{-}C_2$  es competitiva frente a las de Mann y Daniels, pues tiene valores muy cercanos en tendencias lentas, levemente menores en el caso 4, similares en el caso 5 e idénticos en rápidas. Cabe resaltar que para  $T = 6$  la potencia de la prueba con  ${}^{-}C_2$  está por encima de las otras pruebas.

Los resultados bajo las otras distribuciones (logística y normal) presentan estas mismas características. Para alternativa de una cola al 1% son análogos a las descriptas, excepto porque en  $T = 5$  las potencias son iguales,

<sup>11</sup>Ibid

## Plancha 1. Potencias estimadas de pruebas a una cola

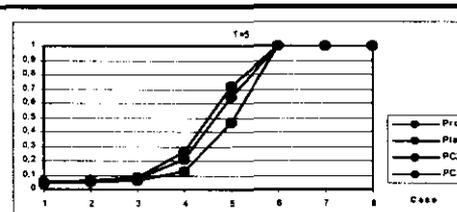
Nivel de significancia deseado: 5%

N= 12800

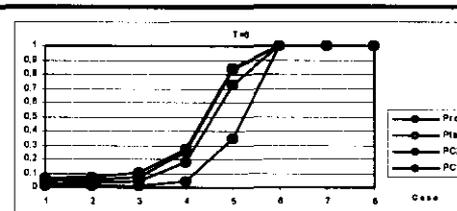
Beta= 1

Distribución Exponencial Doble

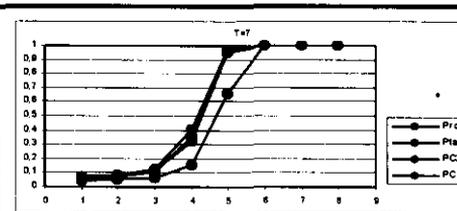
T=		PC1	PC2	Ptau	Pro
5	NS nominal	0,041667	0,041667	0,042	0,0417
	Caso				
	1	0,046094	0,050781	0,052656	0,052656
	2	0,051328	0,061875	0,066016	0,066016
	3	0,06125	0,078516	0,084922	0,084922
	4	0,123828	0,211094	0,259922	0,259922
	5	0,460703	0,642656	0,717188	0,717188
	6	0,997969	1	1	1
	7	1	1	1	1
	8	1	1	1	1



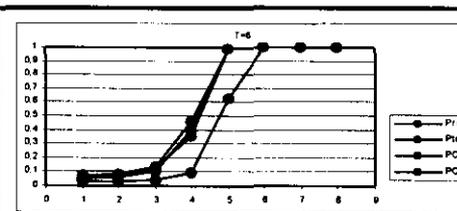
T=		PC1	PC2	Ptau	Pro
6	NS nominal	0,008333	0,05	0,028	0,0292
	Caso				
	1	0,009297	0,065781	0,039063	0,023828
	2	0,01125	0,078125	0,050781	0,031953
	3	0,015391	0,102266	0,07	0,046328
	4	0,04375	0,269297	0,252344	0,174063
	5	0,344844	0,84	0,834141	0,728828
	6	0,997969	1	1	1
	7	1	1	1	1
	8	1	1	1	1



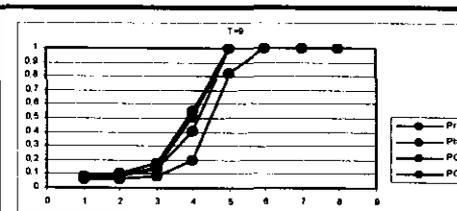
T=		PC1	PC2	Ptau	Pro
7	NS nominal	0,036905	0,05	0,035	0,044
	Caso				
	1	0,043984	0,067734	0,052266	0,065
	2	0,049219	0,0825	0,067969	0,084141
	3	0,062344	0,1175	0,103281	0,124375
	4	0,153594	0,319766	0,358047	0,403047
	5	0,6525	0,945313	0,962734	0,969453
	6	1	1	1	1
	7	1	1	1	1
	8	1	1	1	1



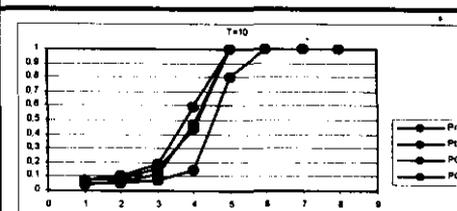
T=		PC1	PC2	Ptau	Pro
8	NS nominal	0,025099	0,049876	0,031	0,0481
	Caso				
	1	0,027891	0,068359	0,047891	0,063672
	2	0,031016	0,083203	0,062891	0,083906
	3	0,041172	0,126953	0,105703	0,136953
	4	0,096484	0,3575	0,400781	0,454219
	5	0,628906	0,984844	0,990234	0,992109
	6	1	1	1	1
	7	1	1	1	1
	8	1	1	1	1



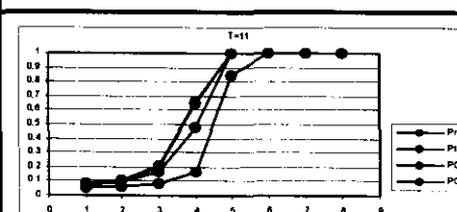
T=		PC1	PC2	Ptau	Pro
9	NS nominal	0,049942	0,049983	0,038	0,0484
	Caso				
	1	0,058516	0,070313	0,061719	0,079844
	2	0,066328	0,088047	0,083906	0,104141
	3	0,085078	0,139531	0,145156	0,177734
	4	0,2025	0,405625	0,509688	0,555781
	5	0,821563	0,996875	0,999375	0,999609
	6	1	1	1	1
	7	1	1	1	1
	8	1	1	1	1



T=		PC1	PC2	Ptau	Pro
10	NS nominal	0,043061	0,04998	0,036	0,0481
	Caso				
	1	0,048281	0,070938	0,042969	0,075234
	2	0,052813	0,087188	0,058828	0,102578
	3	0,068203	0,151172	0,112813	0,182578
	4	0,140781	0,4425	0,472422	0,591094
	5	0,801875	0,999609	0,999922	0,999922
	6	1	1	1	1
	7	1	1	1	1
	8	1	1	1	1



T=		PC1	PC2	Ptau	Pro
11	NS nominal	0,047309	0,049958	0,043	0,047
	Caso				
	1	0,055781	0,071094	0,074375	0,079531
	2	0,062344	0,088906	0,101641	0,108516
	3	0,080547	0,163203	0,195781	0,210859
	4	0,167344	0,474063	0,645234	0,652813
	5	0,84625	0,999922	1	1
	6	1	1	1	1
	7	1	1	1	1
	8	1	1	1	1



cuando  $T = 6$  la potencia de la prueba con  $^{-}C_1$  es ligeramente mayor que las de las pruebas de Mann y de Daniels y cuando  $T = 6$  y  $7$  la potencia de la prueba con  $^{-}C_2$  supera a las demás. Para alternativas de dos colas son similares a los anteriores, resaltando que con  $\alpha = 1\%$  cuando  $T = 5$  la prueba no se realiza debido a la ausencia de puntos críticos al nivel de significancia deseado y cuando  $T = 6$  las potencias de las pruebas con  $^{-}C_1$ , de Mann y de Daniels son iguales<sup>12</sup>.

### Conclusiones

Las potencias de las pruebas con  $^{-}C_1$  y  $^{-}C_2$  son competitivas frente a las pruebas de Mann y de Daniels, ya que en muchos casos toman valores cercanos. Las potencias de las pruebas tienen el comportamiento esperado, pues crecen con la rapidez de la tendencia, ya que tienen bajas potencias en tendencias lentas, mejoran en tendencias moderadas y presentan su mejor desempeño cuando las tendencias son rápidas.

Por razones prácticas, resulta más conveniente el uso de  $^{-}C_1$ , en lugar de  $^{-}C_2$ , cuando los tamaños de muestra son grandes y la tendencia es rápida o moderada, pues tiene un buen desempeño y su cálculo es sencillo. Por otra parte, el uso de  $^{-}C_2$  es más adecuado cuando la tendencia es moderada.

El resultado más importante es haber encontrado una prueba tan potente como las de Mann y de Daniels, pero que admite el manejo de datos categóricos ordenables, ya que las rachas se pueden definir a partir de éste tipo de datos, y además se puede aplicar en datos longitudinales, pues para el cálculo de las rachas no se necesita igual espaciamiento de las observaciones.

### Otras posibilidades de investigación

Este estudio se puede continuar desarrollando en diferentes sentidos relacionados con las funciones que caracterizan el modelo, ampliando la gama de funciones

de tendencia moderadas en la parte determinística y la cobertura de otros tipos de funciones de distribución en la parte aleatoria. El estadístico  $^{-}C$  tiene la propiedad de cubrir diversos tipos de alternativas de acuerdo con el mecanismo de dicotomización utilizado, así se pueden abordar alternativas de autocorrelación (utilizando dicotomizaciones con las magnitudes relativas de las observaciones con respecto a valores fijos o parámetros de centralidad) o tendencia en variables categóricas ordenables (como las tradicionales de rachas, que manejan el número y el tamaño), entre otras.

### Referencias bibliográficas

1. Aiyar, R; Guillier, C. & Albers, W. (1979). *Asymptotic Relative Efficiencies of Rank Tests for Trend Alternatives*. Journal of American Statistician Association. **74**, 226-231.
2. Conover, W. J. (1980). *Practical nonparametric statistics*. John Wiley & Sons. New York.
3. Corzo, J. A. (1984). *Una prueba de dispersión basada en secuencias*. Tesis de Maestría. Universidad Nacional de Colombia.
4. — (1993). *Algunas propiedades distribucionales de estadísticas de rachas*. Reporte interno No. 25. Departamento de Matemáticas y Estadística. Universidad Nacional de Colombia, Santa Fe de Bogotá.
5. — (1996). *Análisis de datos a través de rachas*. Coloquio Distrital de Matemáticas y Estadística, Santa Fe de Bogotá.
6. Gibbons, J. D. (1971). *Nonparametric statistical inference*. McGraw-Hill. Tokyo.
7. Gómez, E. (1999) *Estimación de la potencia de una prueba de rachas para alternativa de tendencia*. Trabajo de grado. Universidad Nacional de Colombia.
8. Lehmann, E. L. & D'Abbrera (1975). *Nonparametrics statistical methods based on ranks*. McGra-Hill. San Francisco.

<sup>12</sup>Estos resultados se pueden consultar en Gómez (1999)