



UNIVERSIDAD DE BUENOS AIRES
Facultad de Ciencias Exactas y Naturales
Departamento de Computación

ESTUDIO EXPERIMENTAL
DE DISCREPANCIA DE NORMALIDAD

Tesis presentada para optar al título de Licenciado de la Universidad de Buenos Aires
en el área de Ciencias de la Computación

Alejandro Alfonso
L.U. 050/07

Directora de Tesis: *Dra. Verónica Becher*

Buenos Aires, Diciembre 2012.

Índice general

Índice general	I
1. Introducción	1
1.1. Objetivo de esta tesis	2
2. Definiciones y teoremas sobre normalidad y discrepancia	4
2.1. Notación	4
2.2. Definiciones	4
2.3. Teoremas de normalidad y discrepancia	6
3. Casos de estudio	8
4. Experimentos	11
4.1. Especificación de los experimentos	11
4.2. Experimentos sobre secuencias individuales	12
4.2.1. Champernowne en base 2 y en base 3	12
4.2.2. Ford extendida en base 2 y en base 3	34
4.2.3. Otras de Bruijn extendidas en base 3	47
4.2.4. Ehrenfeucht-Mycielski	57
4.3. Resultados de comparación	64
4.3.1. Todas las variantes de Champernowne tienen la misma discrepancia?	64
4.3.2. Todas las de Bruijn extendidas tienen la misma discrepancia?	66
4.3.3. Las de Bruijn extendidas y Champernowne tienen la misma discrepancia?	69
4.3.4. Cuánta discrepancia tiene la secuencia de Ehrenfeucht-Mycielski?	75
4.3.5. Todas en el mismo gráfico	79
5. Conclusiones	86
5.1. Variantes de secuencias de Bruijn	86
5.2. Variantes de Champernowne	86
5.3. Relación entre los diferentes tipos de secuencia	87
5.4. Conjeturas sobre las secuencias estudiadas	87
Bibliografía	88

Capítulo 1

Introducción

Hace ya más de cien años, Émile Borel definió la propiedad de *normalidad* [5] para números reales. Un número real es normal en una base dada si, en su expansión en esa base, todos los bloques de dígitos de igual longitud ocurren con la misma frecuencia asintótica. Un número real es absolutamente normal si es normal en toda base entera mayor o igual que dos. La absoluta normalidad es una versión débil de aleatoriedad, ya que asegura que, a la larga, ninguna secuencia de dígitos estará favorecida o perjudicada. Borel demostró que casi todos los números reales son absolutamente normales (el conjunto tiene medida de Lebesgue igual a uno), y planteó el problema de dar ejemplos concretos. Hasta el día de hoy sigue siendo un problema muy difícil. Ha sido mucho más fácil conjeturar la propiedad de normalidad que demostrarla. Por ejemplo, sigue abierta la conjetura de que las constantes matemáticas fundamentales como π y e , $\sqrt{2}$ son normales en todas las bases. El mismo Borel en 1950 conjeturó que todos los números reales algebraicos son normales. David Bailey, Borwein y Richard Crandall en [1, 6] llevan adelante una investigación matemática asistida por computadora que ellos llaman *matemática experimental*. En particular, exploran las propiedades de aleatoriedad de las constantes fundamentales, y las expansiones binarias de números algebraicos. Sus experimentos sostienen las conjeturas de normalidad. Una presentación cuidadosa de los avances logrados sobre secuencias normales se encuentra en la última edición del libro de Kuipers y Niederreiter [14] de 2006 y en el libro de Bugeaud de 2012 [24].

Los únicos ejemplos conocidos de números absolutamente normales están dados por definiciones o construcciones que no permiten exhibir los números. Las primeras fueron de Henri Lebesgue [15] y de Waclaw Sierpiński [22] en 1917. Se sabe que hay números absolutamente normales que son computables, pero hasta el momento hay solamente dos algoritmos: Uno es una reformulación computable del trabajo de Sierpiński de 1917 [22], el otro es la reconstrucción del algoritmo dado por Alan Turing [18, 2]. Estos dos algoritmos no resultan satisfactorios porque tienen una complejidad computacional doblemente exponencial: para arrojar el n -ésimo dígito de la expansión del número el algoritmo debe realizar una cantidad de operaciones doblemente exponencial en n .

Las investigaciones sobre números normales en bases particulares han sido mucho más satisfactorias y hay muchos ejemplos. Sin embargo, en todos los casos solamente se sabe que el número definido es normal en la base particular utilizada, pero no se sabe si es además normal en otra base multiplicativamente independiente. El primer ejemplo de número normal en base 10 (pero puede definirse en cualquier otra base) fue dado en 1933 [8], y sigue siendo el más simple, y el más famoso : *es el número de Champernowne*

0,1234567891011121341516...

Es decir, es la concatenación de todos los números naturales en orden creciente. Unos años después Erdos y Copeland [10] generalizaron la construcción de Champernowne y, por ejemplo, consiguieron mostrar que la concatenación de los números primos también es normal en la base considerada. La otra familia de números normales que nos interesa considerar es la que se define por medio de las *secuencias de Brujijn*. Se las conoce por este nombre por el trabajo del matemático Nicolaas Govert de Brujijn de 1946 [7], pero fueron descubiertas y redescubiertas varias veces (por ejemplo por I.J. Good en 1946 [13], y anteriormente por Flye Sainte-Marie en 1894 [19]). La historia de las secuencias de Brujijn está muy bien presentada en el artículo de Jan Berstel y

Dominique Perrin de 2007 [4]. Karl Popper, en su libro “The Logic of Scientific Discovery” las llama “shortest random-like sequences”, que se puede traducir como las secuencias más cortas del azar. Popper las nombra así porque son secuencias que tienen una cantidad mínima de subsecuencias repetidas. Sin embargo, no son exactamente lo que esperaríamos como secuencias finitas aleatorias de la teoría de aleatoriedad (complejidad de Chaitin-Kolmogorov), principalmente porque para cada longitud considerada, las secuencias de Bruijn son solamente una pequeña proporción del total de secuencias de dicha longitud. Las secuencias de Bruijn han sido y continúan siendo estudiadas, tanto por sus propiedades combinatorias como por sus aplicaciones prácticas. Se usan en aplicaciones ingenieriles, por ejemplo, para representar y razonar sobre redes de comunicación, y también circuitos integrados. En el campo de la bioinformática, las secuencias de Bruijn son la base de algoritmos de ensamblado de las lecturas de ADN que realizan los secuenciadores modernos. Las secuencias de Bruijn dan el límite teórico de la compresión de datos.

Una secuencia de Bruijn de orden n en un alfabeto de k símbolos es una secuencia de longitud $k^n + n - 1$ tal que cada secuencia de longitud n ocurre exactamente una vez. Las secuencias de Bruijn de orden n pueden ser extendidas a secuencias de orden $n + 1$ cuando el alfabeto tiene más de dos símbolos. Y no pueden extenderse a orden $n + 1$ pero sí a orden $n + 2$ cuando el alfabeto tiene exactamente dos símbolos [3]. En base a este resultado, las secuencias de Bruijn se pueden extender infinitamente, y se ha demostrado que estas secuencias son normales en el alfabeto considerado [23, 24].

Las secuencias de Bruijn extendidas son una subfamilia propia de una familia de secuencias que se define golosamente en base la medida de complejidad desarrollada por V.Becher y P.Heiber, a la que llamaron I -complejidad. A grandes rasgos, esta es una medida que clasifica a las secuencias finitas de forma inversamente proporcional a la cantidad de repeticiones en las secuencias. Las secuencias más complejas son las que tienen menos repeticiones y son precisamente las secuencias de Bruijn. Las menos complejas son las que utilizan un único símbolo. Esta medida satisface propiedades fundamentales de la complejidad de Chaitin-Kolmogorov y es computable en tiempo y espacio lineal.

Becher y Heiber conjeturan que las secuencias que se definen maximizando golosamente la I -complejidad de sus segmentos iniciales, a las que llaman *secuencias I -golosas*, son normales. Las secuencias I -golosas incluyen propiamente a las secuencias de Bruijn extendidas, y también incluyen a una secuencia de ceros y unos famosa, sobre la que se sabe muy poco: *la secuencia de Ehrenfeucht–Mycielski* [11]. Acerca de la secuencia de Ehrenfeucht–Mycielski solamente se sabe que toda secuencia finita aparece infinitas veces en su expansión (esta propiedad se conoce con el nombre de *disjuntividad*). En su artículo [11], que ya tiene más de diez años, Ehrenfeucht–Mycielski definen la secuencia constructivamente de manera muy simple, y preguntan si su secuencia tiene la misma frecuencia asintótica de ceros que de unos. Esta es una condición necesaria para la normalidad.

Cuando tenemos un caso de una secuencia normal nos interesa saber a qué velocidad cumple la propiedad de normalidad. A este indicador se lo llama *discrepancia de normalidad*. Para cada bloque de símbolos, la discrepancia es la diferencia entre la frecuencia esperada del bloque y su frecuencia real, en cada posición de la secuencia. Se sabe que la discrepancia de normalidad de las secuencias de Bruijn extendidas infinitamente está acotada superiormente por $O\left(\sqrt{\frac{\log \log n}{\log n}}\right)$ [24]. Sin embargo, no se sabe si esta cota superior es ajustada. Tampoco están identificadas qué secuencias de esta familia de Bruijn tienen máxima discrepancia. Becher y Heiber conjeturan que la máxima discrepancia está en la secuencia de Bruijn extendida lexicográficamente mínima.

1.1. Objetivo de esta tesis

En esta tesis desarrollamos un estudio experimental de estos problemas abiertos:

1. ¿La discrepancia de normalidad de las secuencias de Bruijn extendidas es menor que la de Champernowne?
2. ¿La máxima discrepancia de las secuencias de Bruijn extendidas es la de la lexicográficamente mínima?
3. ¿Cuán grande es la discrepancia de la secuencia Ehrenfeucht–Mycielski?

La tesis hace un estudio experimental para juntar evidencias sobre estas preguntas. Ilustra de qué manera se presentan las discrepancias en segmentos iniciales de las secuencias analizadas. Nos concentramos exclusiva-

mente en alfabetos de dos símbolos y de tres símbolos. En cada caso damos evidencias de si hay o no simetría en discrepancia por arriba y por abajo, y exhibimos los testigos que determinan la discrepancia en cada base, y longitud de bloque. Los experimentos realizados muestran la forma que toma la discrepancia en cada caso estudiado, y permiten predecir como sigue cada gráfico de discrepancia al agrandar el tamaño de segmento.

Capítulo 2

Definiciones y teoremas sobre normalidad y discrepancia

2.1. Notación

De aquí en adelante, A denota un alfabeto, es decir, un conjunto finito de dos o más símbolos. A^n es el conjunto de cadenas de n símbolos de A , $A^{<n} = \bigcup_{i=0}^{n-1} A^i$ es el conjunto de cadenas de longitud estrictamente menor que n , A^* es el conjunto de cadenas finitas de cualquier longitud y A^ω es el conjunto de secuencias infinitas de símbolos de A . $|A|$ es la cardinalidad de A . Notar que $|A^n| = |A|^n$. λ es la cadena vacía, $|s|$ es la longitud de la cadena s , $s[i]$ es el símbolo en la posición i -ésima de s , para $1 \leq i \leq |s|$, y $s[i..i+k-1]$ es la cadena de k símbolos consecutivos de s comenzando desde la posición i , para $1 \leq i \leq |s| - k + 1$.

2.2. Definiciones

Émile Borel en 1909 define la propiedad de normalidad para números reales como una forma de azar. Un número real es normal en una base si cada bloque de dígitos de la misma longitud aparece con la misma frecuencia en el límite en la expresión decimal del número en esa base. Por ejemplo, si un número es normal en base 2, ambos dígitos 0 y 1 ocurren la mitad de las veces en el límite. Cada uno de los bloques 00, 01, 10, y 11 ocurre un cuarto de las veces y así sucesivamente. Un número real que es normal en todas las bases se dice *absolutamente normal*.

Dado que en este trabajo nos dedicaremos exclusivamente a la propiedad de normalidad en una base dada, presentamos la definición de normalidad para secuencias infinitas de un alfabeto A en lugar de presentarla para números reales.

Definición 1 (Normalidad). Sea $occ(x, s) = |\{i : x = s[i..i + |x| - 1]\}|$ la cantidad de ocurrencias del bloque x en la cadena s . Una secuencia $\alpha \in A^\omega$ es normal para el alfabeto A si cada bloque de símbolos de A tiene la misma frecuencia asintótica en α :

$$\forall x \in A^*, \quad \lim_{k \rightarrow \infty} \frac{occ(x, \alpha[1..k])}{k} = \frac{1}{|A|^{|x|}}.$$

Por ejemplo la secuencia 0,1010010000100000100000... no cumple la condición para bloques de tamaño 1. Por contraste, 0,10101010101010101010101... cumple la condición para bloques de tamaño 1 pero no para bloques de tamaño 2.

Cabe aclarar que hay varias formulaciones de la propiedad de normalidad que se demuestran equivalentes. Las demostraciones de equivalencia se pueden leer en [14] o en [24].

La velocidad de convergencia del límite en la definición de normalidad puede variar significativamente. Para estudiarla se usa la noción de *discrepancia*.

Definición 2 (Discrepancia). *La discrepancia de normalidad de una secuencia infinita $\alpha \in A^\omega$ para la posición n y tamaño de bloque ℓ es la mayor diferencia con el valor esperado de la frecuencia con la que ocurren los bloques de tamaño ℓ como subcadenas de α .*

$$D(\alpha, n, \ell) = \max_{x \in A^\ell} \left| \frac{\text{occ}(x, \alpha[1..n])}{n} - \frac{1}{|A|^\ell} \right|$$

donde $\text{occ}(x, s) = |\{i : x = s[i..i + |x| - 1]\}|$ es la cantidad de ocurrencias del símbolo x en la cadena s .

Podemos expresar $D(\alpha, n, \ell)$ como $O(f(n))$, donde f depende de α y la constante oculta depende del tamaño de bloque ℓ . Podemos entonces ignorar el parámetro ℓ cuando hablamos de cotas de discrepancia y decir que $D(\alpha, n)$ es $O(f(n))$.

Definición 3 (Secuencia de Bruijn). *Una secuencia de Bruijn (no cíclica) de orden n en un alfabeto de k símbolos es una secuencia de longitud $k^n + n - 1$ tal que cada secuencia de longitud n ocurre exactamente una vez como subcadena[7].*

Si el alfabeto A tiene k símbolos, entonces hay $(k!)^{k^{n-1}}/k^n$ secuencias de Bruijn distintas de orden n [7]. Entre ellas está la secuencia de Ford, que cumple las siguiente propiedades:

1. Es la secuencia de Bruijn lexicográficamente menor.
2. Es el resultado de aplicar el algoritmo goloso *least-first* para construir una secuencia de Bruijn, comenzando con 1^k .
3. Es el resultado de concatenar todas las *Lyndon words* de cada longitud en orden lexicográfico.
4. Es la secuencia de Bruijn generada por un *shift register* cuya tabla de verdad tiene peso mínimo.

Definición 4 (Grafo de Bruijn). *El Grafo de Bruijn de orden n para un alfabeto A es el grafo dirigido que tiene como nodos los elementos de A^n y cuyo conjunto de ejes está definido por*

$$x \rightarrow y \quad \text{sii} \quad x[2..n] = y[1..n - 1].$$

Teorema 5. *El grafo de línea de un Grafo de Bruijn de orden n es un Grafo de Bruijn de orden $n + 1$ para el mismo alfabeto.*

Teorema 6. *Sea G un Grafo de Bruijn de orden n . Si rotulamos los ejes $x \rightarrow y$ con el símbolo $y[n]$, un ciclo Hamiltoniano en G es una secuencia de Bruijn de orden n y un ciclo Euleriano en G es una secuencia de Bruijn de orden $n + 1$.*

Teorema 7.

1. *Toda secuencia de Bruijn de orden n en un alfabeto de al menos tres símbolos puede ser extendida a una secuencia de Bruijn de orden $n + 1$.*
2. *Toda secuencia de Bruijn de orden n en un alfabeto de dos símbolos no puede ser extendida a una secuencia de Bruijn de orden $n + 1$, pero sí a una de orden $n + 2$.*

Corolario Existen secuencias de Bruijn infinitas para cualquier alfabeto. Una demostración completa y elemental de la existencia de secuencias de Bruijn extendidas se puede leer de [3].

Definición 8 (Secuencia de Bruijn infinita). *Una secuencia infinita es una secuencia de Bruijn infinita si es el límite inductivo de la sucesión de secuencias de Bruijn donde cada secuencia es una extensión de la anterior.*

Definición 9 (Secuencia Champernowne). *La secuencia de Champernowne en una base b es la concatenación de todos los números enteros en orden consecutivo escritos en su expresión en base b . Para $b = 10$, la secuencia de Champernowne es*

012345678910111213141516171819202122232425262728 ...

Definición 10 (Secuencia I-golosa). Decimos que $\alpha \in A^\omega$ es I-golosa si es el límite inductivo de la sucesión de cadenas $(\alpha[1..n])_{n \geq 1}$ donde $\alpha[1..n] = \alpha[1..(n-1)]b$, con el símbolo b tal que para todo $c \in A$,

$$I(\alpha[1..(n-1)]b) \geq I(\alpha[1..(n-1)]c).$$

El símbolo b no siempre es único. Dependiendo de la estrategia utilizada para desempatar se definen diferentes secuencias. Algunas estrategias de desempate posibles son: lexicográficamente, según la cantidad de ocurrencias del símbolo a a agregar en la cadena hasta el momento, según la posición del mayor sufijo que se repite, entre otras.

Teorema 11. En un alfabeto de al menos tres símbolos, cada secuencia infinita de Bruijn es I-Golosa.

Es un problema abierto identificar una estrategia lineal de desempate que permita distinguir cuales de las secuencias I-Golosas son de Bruijn extendidas.

Para alfabetos de dos símbolos, las secuencias I-golosas no incluyen a las de Bruijn extendidas. Sin embargo, hay una estrategia de desempate que define a la secuencia Ehrenfeucht-Mycielski [11].

Definición 12 (Secuencia Ehrenfeucht-Mycielski). La secuencia Ehrenfeucht-Mycielski es una secuencia infinita sobre el alfabeto $0, 1$ construida con el siguiente algoritmo:

1. Comenzar con $u_1 = 0$.
2. Suponer que los primeros n bits $U_n = u_1 u_2 \dots u_n$ ya han sido elegidos.
3. Encontrar el sufijo más largo v de U_n que aparece en U_{n-1} .
4. Si b es el bit que le sigue a v en su última ocurrencia en U_{n-1} , entonces $u_{n+1} = \bar{b}$, el complemento de b .
5. Si no existe ninguna ocurrencia previa de ningún sufijo no vacío, el último bit de la secuencia es alternado.

2.3. Teoremas de normalidad y discrepancia

Borel demostró que casi todos los números reales son absolutamente normales, y fue a partir de este teorema que pidió un ejemplo explícito.

Teorema 13 (Borel [5]). El conjunto de números normales tiene medida de Lebesgue 1.

La discrepancia de una secuencia infinita α no puede ser muy chica y satisface

$$1/n < D(\alpha, n) \leq 1.$$

La parte izquierda de la inecuación puede ser mejorada sustancialmente para valores arbitrariamente grandes de n , como demostró W. M. Schmidt [21].

Teorema 14 (Schmidt [21]). Para toda secuencia infinita α , existen enteros n arbitrariamente grandes tales que $D(\alpha, n) \geq \log n / 25n$. Otra manera de decirlo es $\limsup_{n \rightarrow \infty} D(\alpha, n)n / \log n > 0$.

Teorema 15. Gál and Gál [12] probaron que para casi todos los números reales α , en cada base, la discrepancia $D(\alpha, n)$ es $O(\sqrt{(\log \log n)/n})$.

M. Levin [16] definió números absolutamente normales cuya discrepancia es $O(\sqrt{(\log n)^6/n})$. En otro trabajo [17] Levin da números reales que son normales para una base particular con una discrepancia de $O((\log n)^2/n)$. Teniendo en cuenta el Teorema 14, esto es cercano de ser lo mejor posible.

Los números de Stoneham $\sum_{j \geq 1} 1/c^j b^{c^j}$, y los números de Korobov $\sum_{j \geq 0} 1/c^{d^j} b^{c^{d^j}}$, donde b y c son enteros coprimos ambos mayores o iguales que 2, y d también es un entero mayor o igual que 2, son normales en la base b . Su discrepancia de normalidad es $O((\log n)^2/\sqrt{n})$ [1]. Fue demostrado que no son absolutamente normales [24].

Teorema 16 (Champernowne [8]). *La secuencia de Champernowne es normal en el alfabeto considerado.*

Teorema 17 (Schiffer [20]). *La secuencia de Champernowne tiene discrepancia $O(1/\log n)$.*

Teorema 18 (Ugalde [23]). *Las secuencias de Bruijn extendidas infinitamente tienen discrepancia $O(\sqrt{\log \log n / \log n})$.*

Ugalde conjetura que esta estimación no es la mejor posible, porque sus experimentos numéricos indican que la discrepancia debería ser $O(1/\log n)$, al menos para la gran mayoría de las secuencias de Bruijn extendidas.

Toda secuencia de Bruijn tiene una serie de propiedades que le dan carácter aleatorio: cada palabra de longitud apropiada aparece como subsecuencia, las cantidades de bloques de cada longitud están balanceadas, la cantidad de 0 es igual a la cantidad de 1, etc. Sin embargo, la secuencia de Ford extendida se construye mediante un algoritmo goloso que usa 0 antes que 1 cuando es posible. Es natural sospechar que la secuencia Ford tiene exceso de 0 antes de equilibrarlos. Es decir, la diferencia entre la cantidad de ceros y unos es grande en los segmentos iniciales.

Teorema 19 (Cooper y Heitsch [9]). *Sea el alfabeto de dos símbolos. La diferencia de ceros y unos de la secuencia de Ford de longitud n es $\Theta(\log \log n / \log n)$. Es decir, para la secuencia Ford de orden $\log n$, la diferencia entre la frecuencia de 0 y la de 1 está acotada asintóticamente por arriba y por debajo por un factor constante por $(\log \log n / \log n)$.*

Este teorema solamente indica la discrepancia de un solo dígito. Es aún un problema abierto demostrar la discrepancia de bloques de tamaño mayor. También queda por demostrar si esta cota superior se mantiene para alfabetos más grandes.

Es posible que la diferencia de ceros y unos de la secuencia Ford se propague en la secuencia Ford extendida infinitamente. Hasat ahora el único resultado conocido es la cota dada en el Teorema 18 que aplica a todas las secuencias de Bruijn extendidas infinitamente.

Capítulo 3

Casos de estudio

Para este trabajo, decidimos usar alfabetos de 2 y 3 símbolos. La elección particular de estas bases fue por simpleza y costo computacional. De manera indistinta hablamos de alfabetos de dos símbolos o de base 2; y de alfabetos de tres símbolos o base 3.

Analizamos dos familias de secuencias: la familia de Bruijn extendidas y variantes de la secuencia de Champernowne. Además, para el caso de alfabeto de dos símbolos la secuencia de Ehrenfeucht-Mycielski. Específicamente consideramos las siguientes secuencias. Explicitamos a continuación un segmento inicial de cada una.

En Base 2

1. Ford Extendida: la secuencia de Bruijn extendida lexicográficamente menor
010001110100101011001101111100000100000011000010100001111001000100100110010110101010011100010111001...
2. Champernowne original
01101110010111011110001001101010111001101111011111000010001100101001110100101011011010111100011001...
3. Ehrenfeucht-Mycielski
01001101011100010000111101100101001001110100011000001011011110011001111101010110001110010001010100...

En Base 3

1. Ford Extendida: la secuencia de Bruijn extendida lexicográficamente menor
0120022110001011121020212220100110102100200002111102200121120112202012212120222210120101000001110010...
2. De Bruijn extendida eligiendo al azar cuando hay posibilidades con 25% de probabilidad de elegir la opción lexicográficamente menor (ver algoritmo)
0121102200100021222101112020112212012202211212102021111012002222010102100200001100120222021202101222...
3. De Bruijn extendida eligiendo al azar cuando hay posibilidades con 50% de probabilidad de elegir la opción lexicográficamente menor (ver algoritmo)
0120221100010111222002121020110222212210121201021111010022020210011211200002001220120111021011001020...
4. Variante de Champernowne, orden lexicográfico
0120001021011122021220000010020100110120200210221001011021101111121201211222002012022102112122202212...
5. Variante de Champernowne, orden aleatorio
0210022201221011110020121110201221200012222112012021021101120020210101002001212200222210001012102120...
6. Variante de Champernowne, orden que demora un dígito lo más posible
0120001101102122021220000010100111001011101110020120200211021121201212002012102112022120221222202212...

Algoritmo para generar las variantes de Champernowne

La idea es ordenar todos los bloques del mismo tamaño bajo algún criterio y luego colocar primero los bloques de tamaño 1, luego los de tamaño 2, etc. Recordemos que escribimos 2^L para indicar el bloque que consiste en la concatenación de L veces el símbolo 2.

$L := 1$

Repetir:

Imprimir todas las cadenas de largo L una atrás de la otra (*)

$L = L + 1$

Dependiendo de cómo se ordenen las secuencias en el paso (*) obtenemos las distintas variantes:

1. La primer variante usa el orden lexicográfico, obteniendo una secuencia similar a la secuencia de Champernowne original.
2. La segunda variante ordena de manera aleatoria.
3. La tercer variante usa el siguiente criterio de comparación:
 - Primero se compara por la cantidad de apariciones de la cadena 2^L .
 - En caso de empate, por la cantidad de apariciones de la cadena 2^{L-1} .
 - En caso de empate, cantidad de apariciones de 2^{L-2} .
 - ...
 - En caso de empate, cantidad de apariciones de $2^1 = 2$.
 - En caso de empate, lexicográficamente.

Algoritmo para generar de Bruijn extendida

Básicamente, el algoritmo consiste en mapear una secuencia de Bruijn de orden i en un camino Hamiltoniano del grafo de Bruijn de orden i y extenderlo a un ciclo Euleriano, obteniendo así una secuencia de Bruijn de orden $i + 1$. Este proceso garantiza que la secuencia original de orden i es prefijo de la secuencia obtenida de orden $i + 1$. Escribimos G_i para indicar el grafo de Bruijn de orden i . El mapeo utilizado asigna a la cadena s el camino en G_i que comienza en el nodo $s[1..i]$ y continúa siguiendo los rótulos de los ejes. En cada iteración, al finalizar el cuerpo del ciclo s es la secuencia de Bruijn de orden $i + 1$.

Empezar con $i = 1, s = 01\dots|A| - 1$, la secuencia de Bruijn de orden 1

Repetir mientras s no alcance la longitud deseada:

Mapear en G_i el camino Hamiltoniano que representa a s

Extender este camino a un ciclo Euleriano en G_i . (*)

$s =$ secuencia que representa este ciclo Euleriano

$i = i + 1$

Hay que hacer una salvedad para $|A| = 2$ dado que, como vimos en el capítulo 2, una secuencia de orden n no puede extenderse a una de orden $n + 1$ pero sí a una de orden $n + 2$. En este caso, i debe incrementarse no en 1 sino en 2. El camino que describe la vieja secuencia en el grafo de Bruijn dos órdenes más grandes no representa un camino Hamiltoniano porque solamente visita la mitad de los nodos. Sin embargo, es un camino que no pasa dos veces por un mismo nodo, por lo cual el resto del algoritmo aplica de todos modos.

La secuencia obtenida depende de la forma de extender el camino en el paso (*). Para obtener la secuencia lexicográficamente menor, se intenta siempre extender el camino agregando el eje con el menor rótulo entre las opciones posibles. Si se llega a una situación donde el camino no puede ser extendido sin haberse cerrado el ciclo Euleriano, se hace backtracking tomando la otra ruta.

El caso $|A| = 3$ presenta una situación particular. Al mapear una secuencia de Bruijn sobre el camino Hamiltoniano que representa en el grafo de Bruijn del mismo orden, todos los nodos tienen exactamente dos ejes salientes no utilizados por el camino, con excepción del último. Es decir, cuando intentemos extender el camino, cada vez que se visite un nodo por primera vez (segunda vez si se cuenta el mapeo sobre el ciclo Hamiltoniano) hay exactamente dos extensiones posibles.

Podemos entonces, en lugar de favorecer la alternativa lexicográficamente menor en todo momento, favorecerla con probabilidad $1-p$. Tenemos entonces una probabilidad p de diverger de la secuencia lexicográficamente menor. Para este trabajo, utilizamos las siguientes variantes:

1. $p = 0$, obteniendo la secuencia lexicográficamente menor.
2. $p = 0,5$, sin favorecer ninguna alternativa al momento de extender el camino, y
3. $p = 0,25$, favoreciendo la alternativa lexicográficamente menor, pero no estrictamente.

Capítulo 4

Experimentos

4.1. Especificación de los experimentos

Para cada una de las secuencias previamente enumeradas, computamos su primer millón de dígitos y calculamos la discrepancia de normalidad para diferentes tamaños de bloque. La elección del número 1 000 000 ha sido arbitraria en base al tiempo de cómputo. Consideramos todas las posibles longitudes de bloque hasta $\lceil \log_b 1\,000\,000 \rceil$, donde b es la base (o tamaño del alfabeto). Es, decir, consideramos longitudes de bloque hasta 20 para base 2 y hasta 13 para base 3. La elección de esta máxima longitud de bloque es bastante más holgada que lo que sería estadísticamente significativo, ya que hasta la posición un millón, a lo sumo, pueden ocurrir todos los bloques de longitud $\lfloor \log_b 1\,000\,000 \rfloor$, pero esto no es posible para bloques un dígito más largos.

Hicimos dos tipos de análisis:

1. Cómputo de discrepancia por exceso y por defecto, para cada posición $n = 1 \dots 1\,000\,000$.

Para una secuencia α , una posición n y tamaño de bloque ℓ , decimos que:

$$D_{exceso}(\alpha, n, \ell) = \max_{x \in A^\ell} \frac{occ(x, \alpha[1..n])}{n} - \frac{1}{|A|^\ell}$$

$$D_{defecto}(\alpha, n, \ell) = \min_{x \in A^\ell} \frac{occ(x, \alpha[1..n])}{n} - \frac{1}{|A|^\ell}$$

donde $occ(x, s) = |\{i : x = s[i..i + |x| - 1]\}|$ es la cantidad de ocurrencias del bloque x en la cadena s .

Observar que:

$$D_{defecto}(\alpha, n, \ell) \leq 0 \leq D_{exceso}(\alpha, n, \ell)$$

Y además:

$$D(\alpha, n, \ell) = \max\{D_{exceso}(\alpha, n, \ell), |D_{defecto}(\alpha, n, \ell)|\}$$

2. Cómputo de bloque responsable (o testigo) de la discrepancia por exceso y por defecto para cada posición $n = 1 \dots 1\,000\,000$. En caso de empate, elegimos el lexicográficamente menor.

$$T_{exceso}(\alpha, n, \ell) = \min\{x \in A^\ell \mid occ(x, \alpha[1..n]) \text{ es máximo}\}$$

$$T_{defecto}(\alpha, n, \ell) = \min\{x \in A^\ell \mid occ(x, \alpha[1..n]) \text{ es mínimo}\}$$

Para evitar la propagación de errores de precisión aritmética, trabajamos siempre con números enteros hasta el momento de hacer el plot de los gráficos. El cálculo de las secuencias fue realizado por programas escritos en C++ o Python según conveniencia y disponibilidad de código existente. Luego, para cada secuencia y tamaño de bloque, generamos un archivo de texto con la cantidad de apariciones del bloque más y menos frecuente

hasta el momento para cada posición, junto con la codificación en base 10 de cuáles son esos bloques. Con esta información realizamos todos los gráficos utilizando el paquete *GNU Plot*.

Presentamos un gráfico para discrepancia por exceso y por defecto. En el eje horizontal mostramos las posiciones de la secuencia en escala logarítmica. En el eje vertical usamos escala lineal. Son gráficos con dos curvas, una roja por defecto, y una verde por exceso.

4.2. Experimentos sobre secuencias individuales

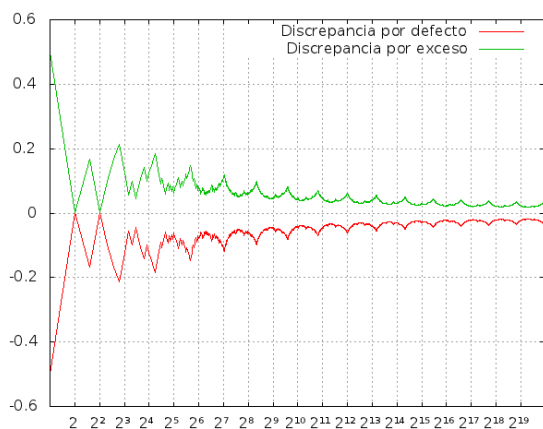
En esta sección presentamos los resultados de los experimentos, volcados en dos tipos de gráficos. Los gráficos de discrepancia de normalidad por exceso y por defecto para cada longitud de bloque, para el primer millón de posiciones de cada una de las secuencias estudiadas. Se discrimina la discrepancia por exceso (verde) y por defecto (rojo). El eje x representa las posiciones en la secuencia en escala logarítmica y el eje y indica el valor de la discrepancia para la posición correspondiente.

Los gráficos de testigos muestran, para cada longitud de bloque, cuales son los que exhiben máxima discrepancia. Se discrimina la discrepancia por exceso (verde) y por defecto (rojo). El eje x representa las posiciones en la secuencia en escala logarítmica y el eje y contiene todos los bloques existentes para el tamaño dado, con $y = 0$ siendo el bloque 0^k y el valor máximo siendo el bloque b^k (donde b es 2 o 3 según el caso). En caso de que haya dos bloques con la misma discrepancia máxima, el gráfico muestra el lexicográficamente menor. Es importante tener en cuenta que para tamaños de bloque chicos, el eje y contiene pocos valores y es esperable que se alcancen pocos valores.

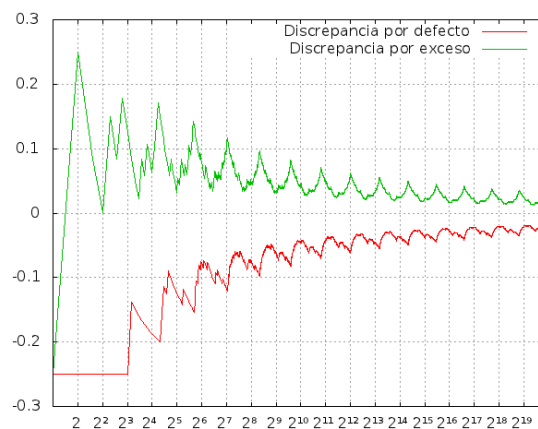
4.2.1. Champernowne en base 2 y en base 3

Base 2

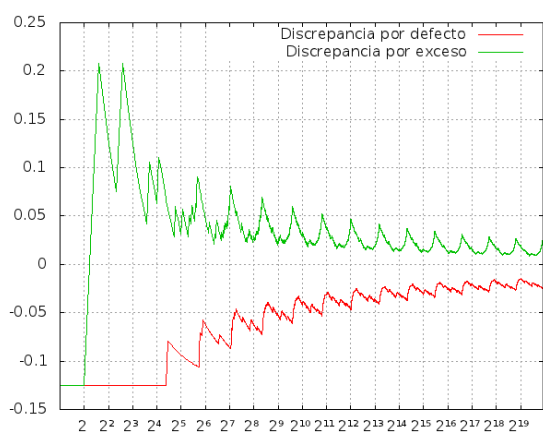
Gráficos de Discrepancia Champerowne en base 2. Son 20 gráficos para longitud de bloque de 1 a 20. Presentan un patrón regular, con forma de serrucho con picos sucesivamente más chicos tanto por exceso (verde) como por defecto (rojo), aunque la discrepancia por defecto se acerca a 0 mucho más lentamente que la discrepancia por exceso a medida que crece el tamaño del bloque. Esta tendencia a formar patrones (graficados sobre una escala logarítmica en x) es algo que veremos en casi todas las secuencias estudiadas.



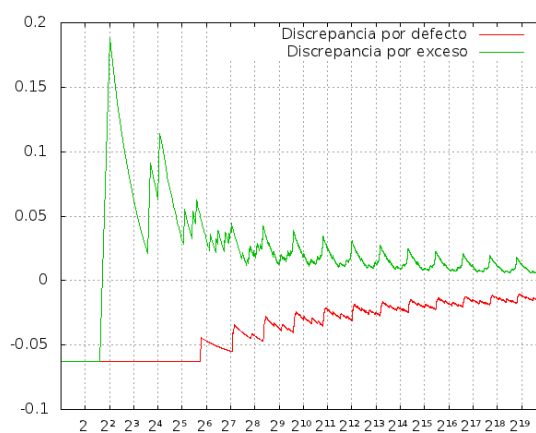
Discrepancia de Champerowne (Base 2)
Bloques de tamaño 1



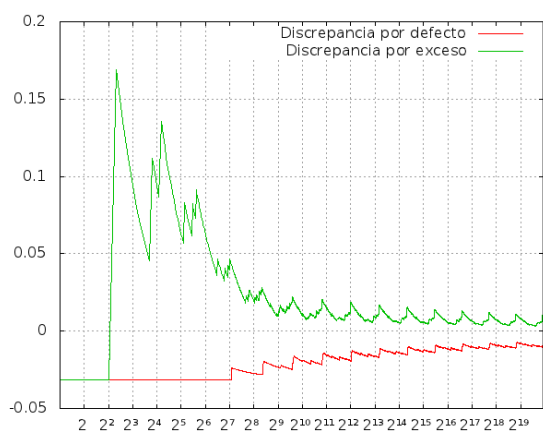
Discrepancia de Champerowne (Base 2)
Bloques de tamaño 2



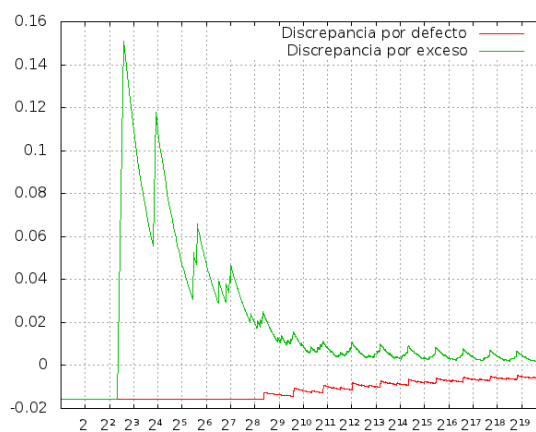
Discrepancia de Champernowne (Base 2)
Bloques de tamaño 3



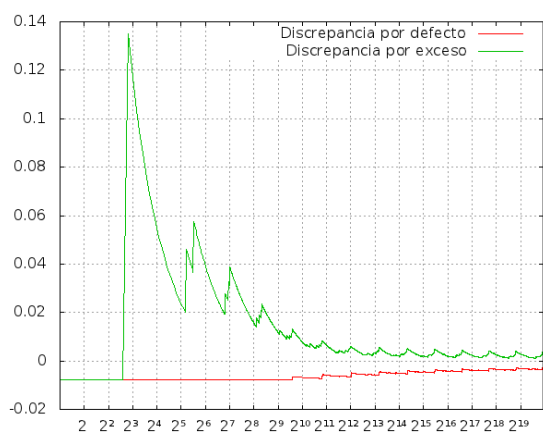
Discrepancia de Champernowne (Base 2)
Bloques de tamaño 4



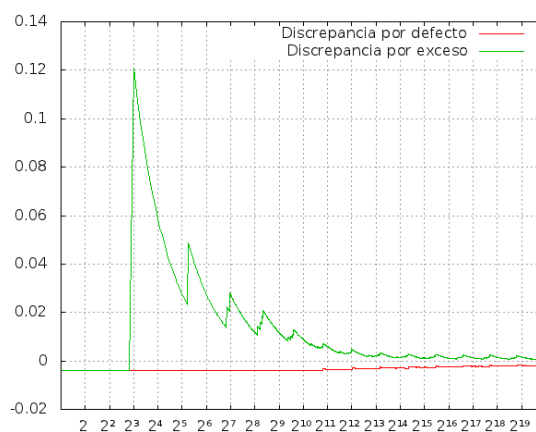
Discrepancia de Champernowne (Base 2)
Bloques de tamaño 5



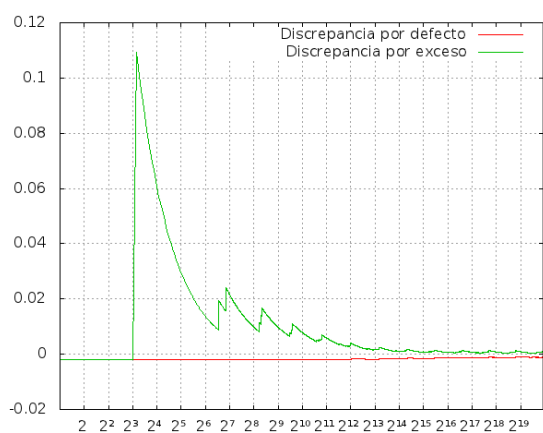
Discrepancia de Champernowne (Base 2)
Bloques de tamaño 6



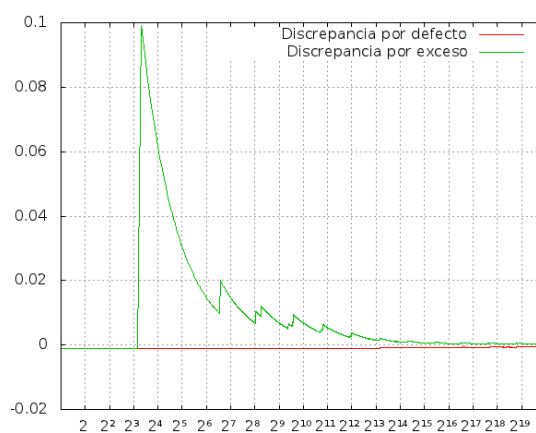
Discrepancia de Champernowne (Base 2)
Bloques de tamaño 7



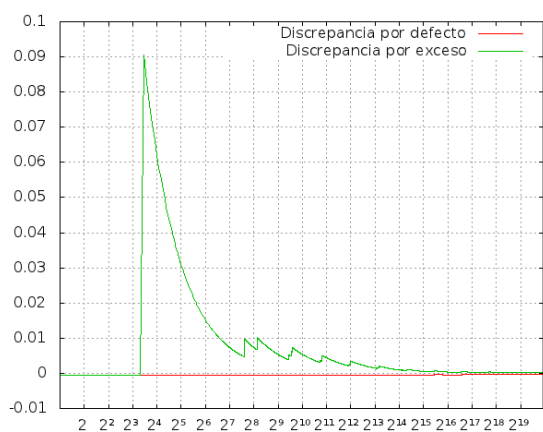
Discrepancia de Champernowne (Base 2)
Bloques de tamaño 8



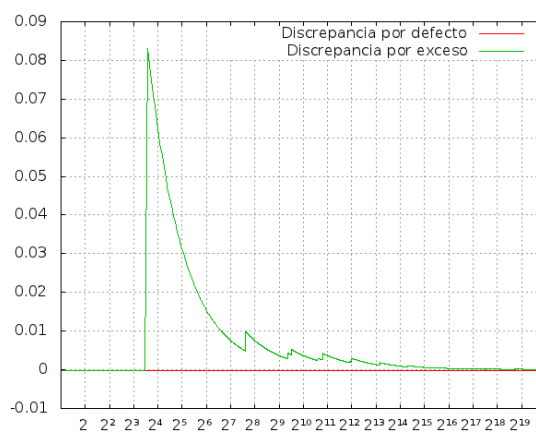
Discrepancia de Champernowne (Base 2)
Bloques de tamaño 9



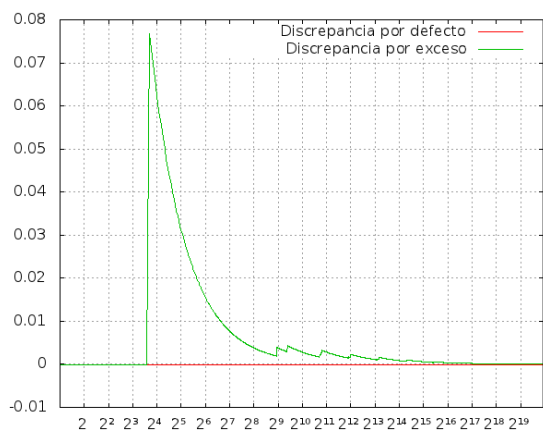
Discrepancia de Champernowne (Base 2)
Bloques de tamaño 10



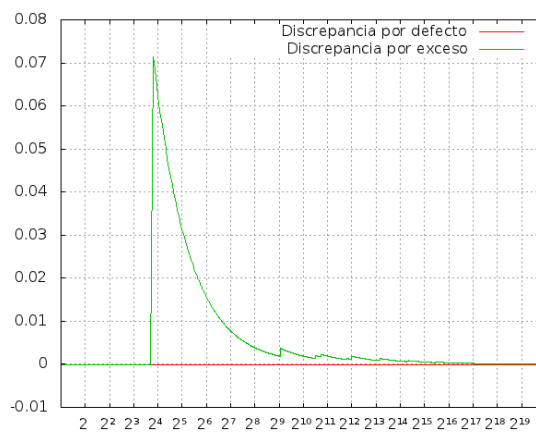
Discrepancia de Champernowne (Base 2)
Bloques de tamaño 11



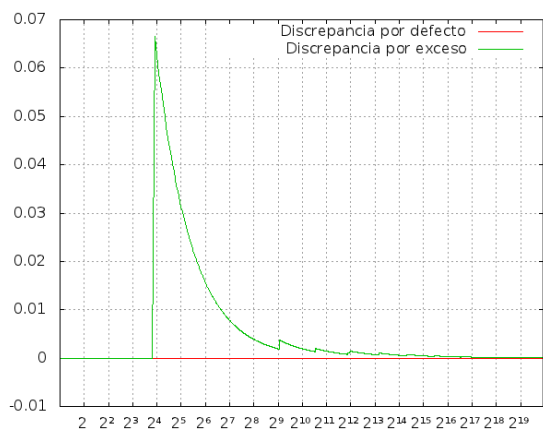
Discrepancia de Champernowne (Base 2)
Bloques de tamaño 12



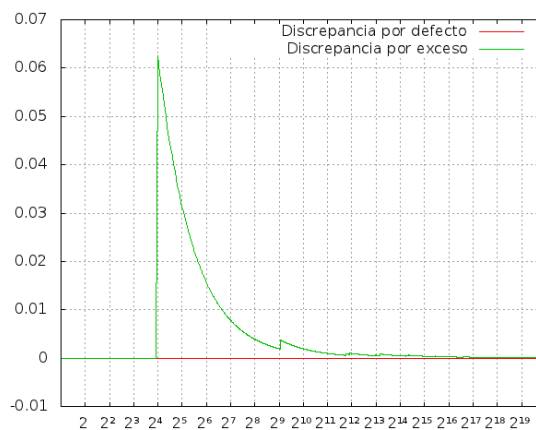
Discrepancia de Champernowne (Base 2)
Bloques de tamaño 13



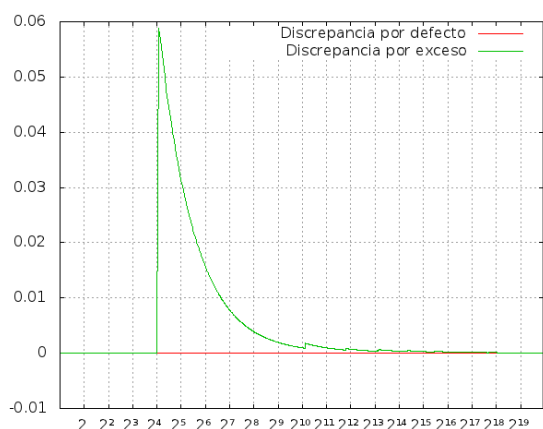
Discrepancia de Champernowne (Base 2)
Bloques de tamaño 14



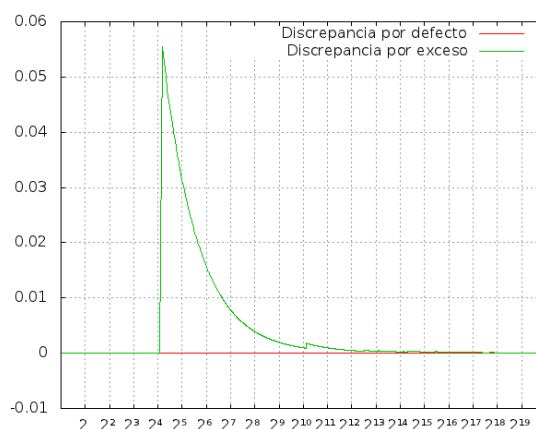
Discrepancia de Champernowne (Base 2)
Bloques de tamaño 15



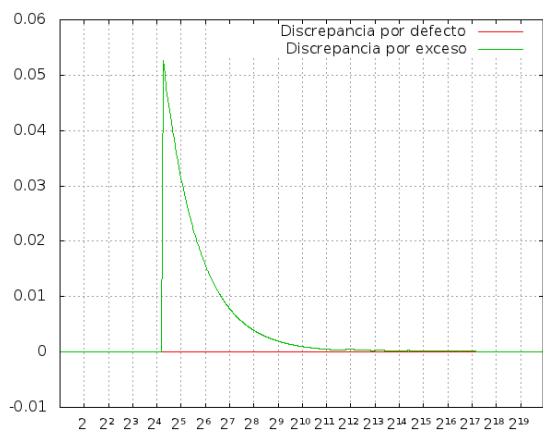
Discrepancia de Champernowne (Base 2)
Bloques de tamaño 16



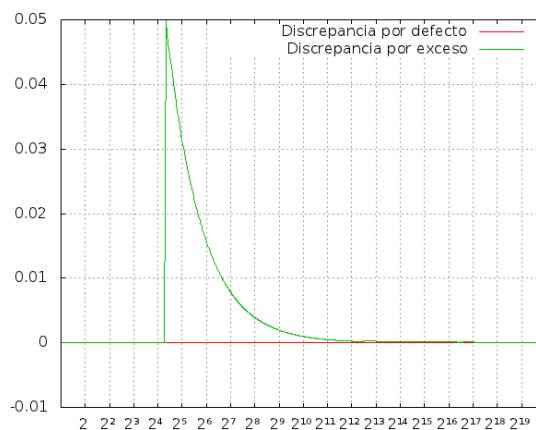
Discrepancia de Champernowne (Base 2)
Bloques de tamaño 17



Discrepancia de Champernowne (Base 2)
Bloques de tamaño 18

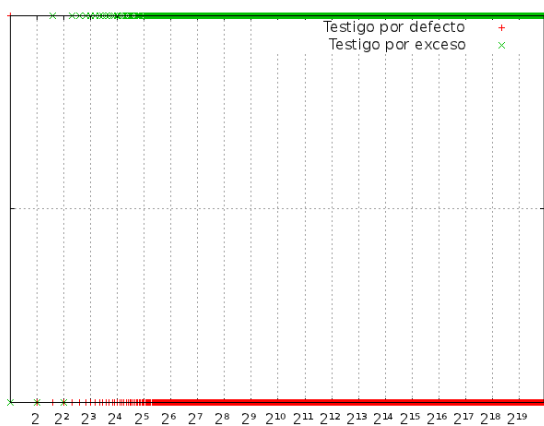


Discrepancia de Champernowne (Base 2)
Bloques de tamaño 19

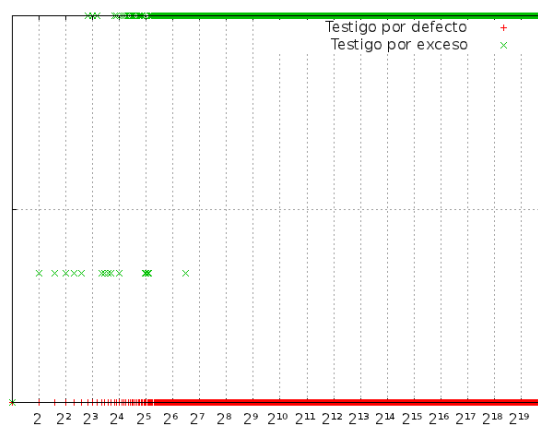


Discrepancia de Champernowne (Base 2)
Bloques de tamaño 20

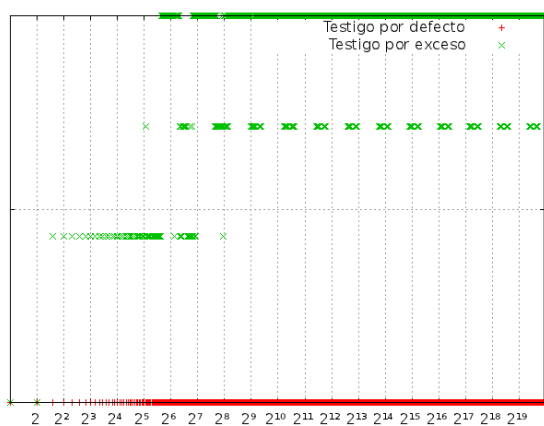
Gráficos de Testigos. En la secuencia de Champernowne en base 2 los bloques testigo de la discrepancia por exceso (verde) varían relativamente poco, pero el testigo por defecto (rojo) es siempre 0^i . Esto es de esperarse, pues el bloque 0^i recién aparece cuando la “cuenta” llega hasta el número 2^i .



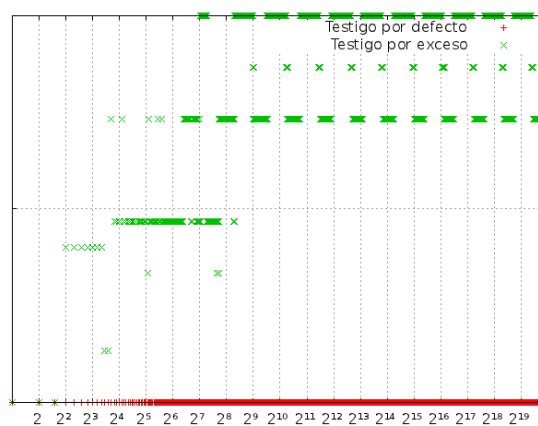
Testigos de Champernowne (Base 2)
Bloques de tamaño 1



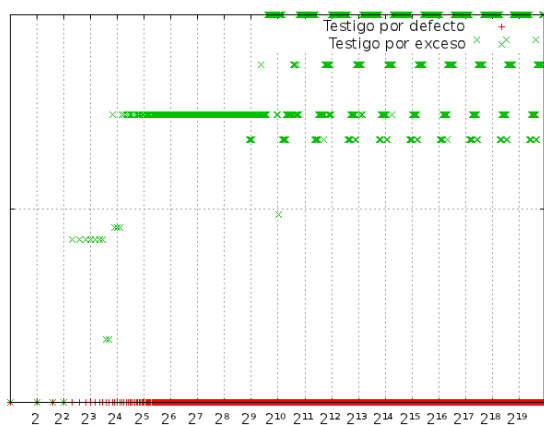
Testigos de Champernowne (Base 2)
Bloques de tamaño 2



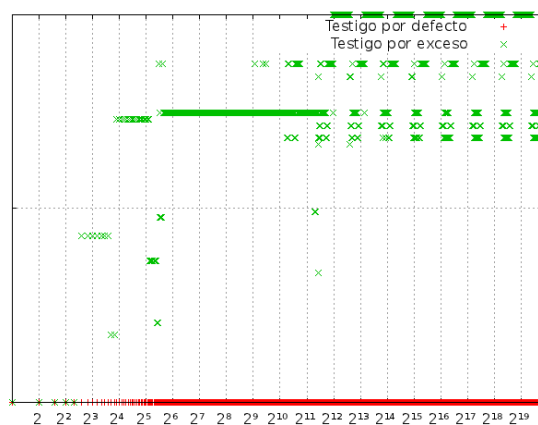
Testigos de Champernowne (Base 2)
Bloques de tamaño 3



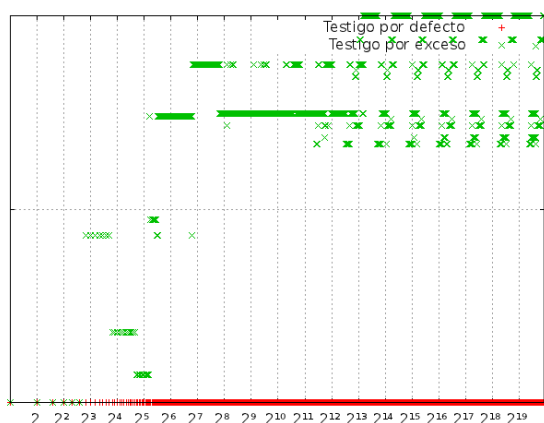
Testigos de Champernowne (Base 2)
Bloques de tamaño 4



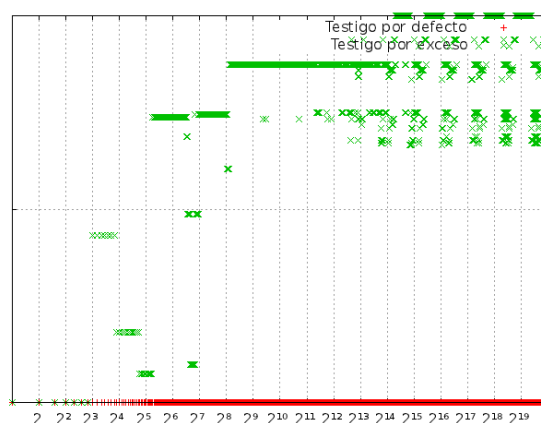
Testigos de Champernowne (Base 2)
Bloques de tamaño 5



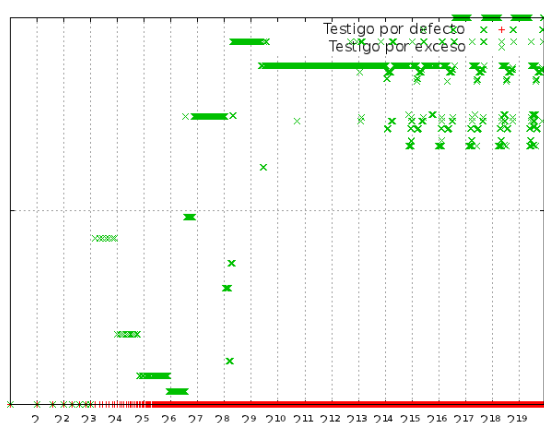
Testigos de Champernowne (Base 2)
Bloques de tamaño 6



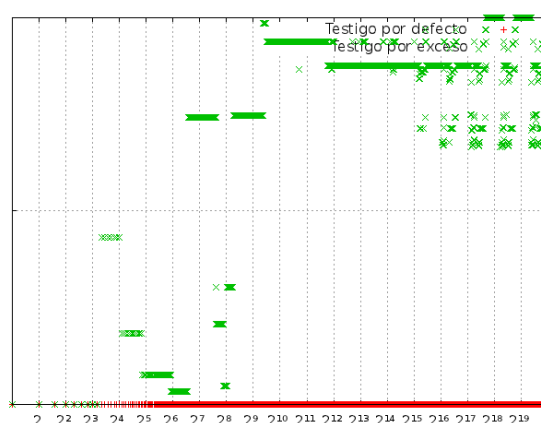
Testigos de Champernowne (Base 2)
Bloques de tamaño 7



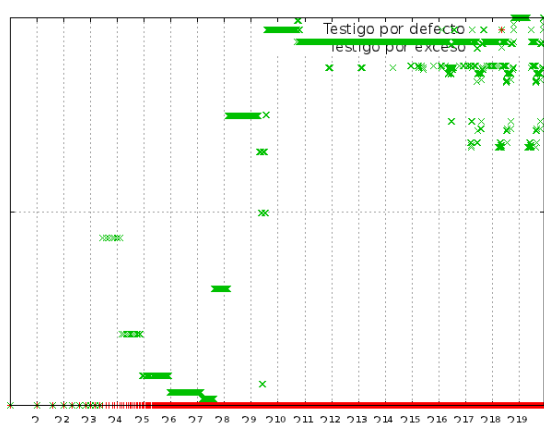
Testigos de Champernowne (Base 2)
Bloques de tamaño 8



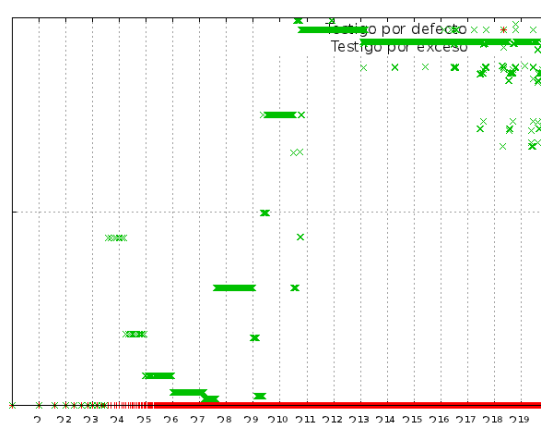
Testigos de Champernowne (Base 2)
Bloques de tamaño 9



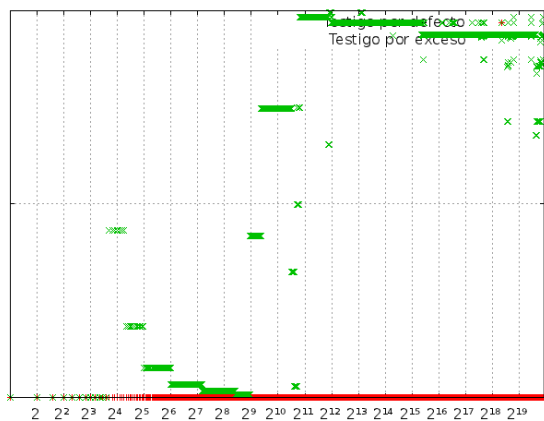
Testigos de Champernowne (Base 2)
Bloques de tamaño 10



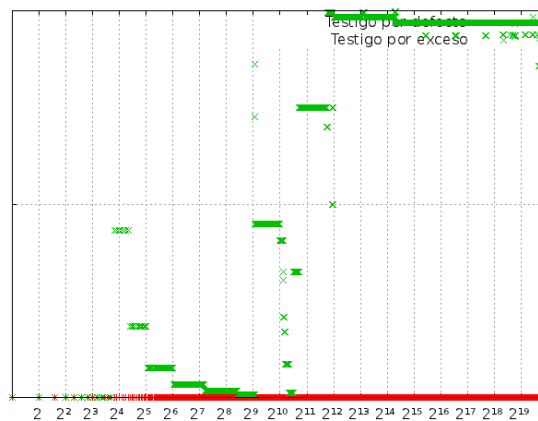
Testigos de Champernowne (Base 2)
Bloques de tamaño 11



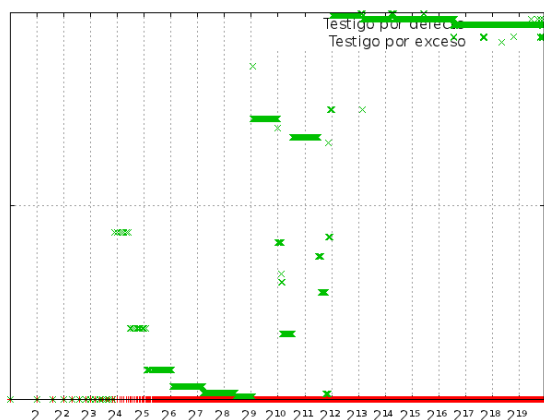
Testigos de Champernowne (Base 2)
Bloques de tamaño 12



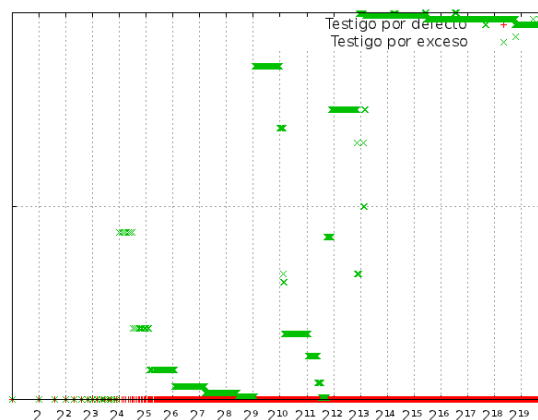
Testigos de Champernowne (Base 2)
Bloques de tamaño 13



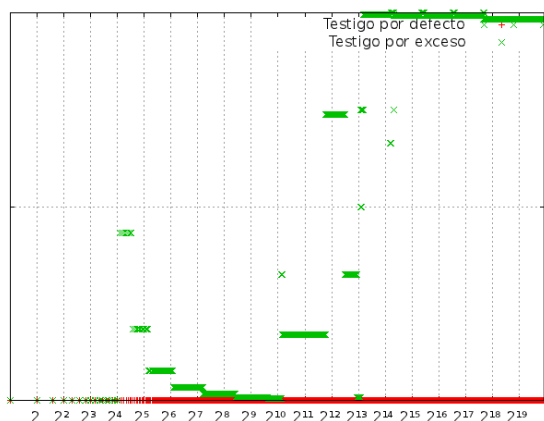
Testigos de Champernowne (Base 2)
Bloques de tamaño 14



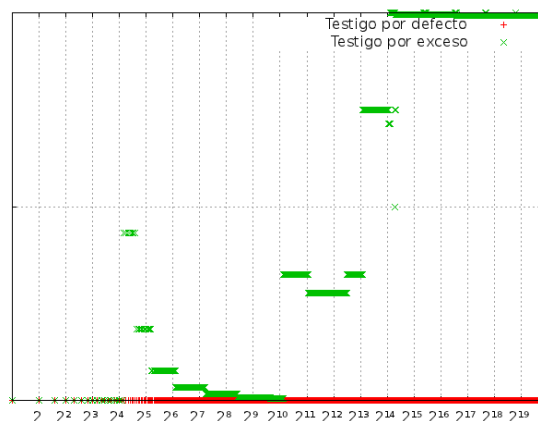
Testigos de Champernowne (Base 2)
Bloques de tamaño 15



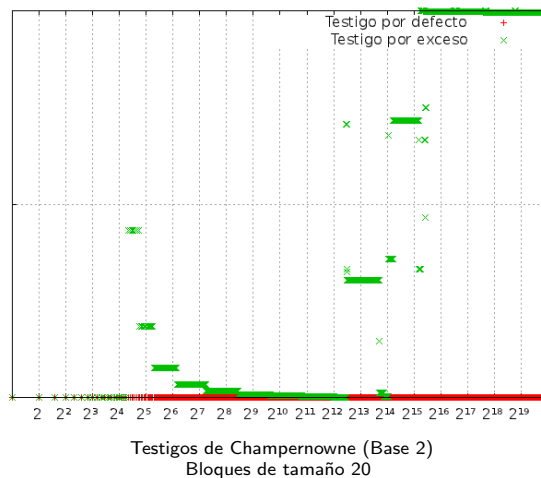
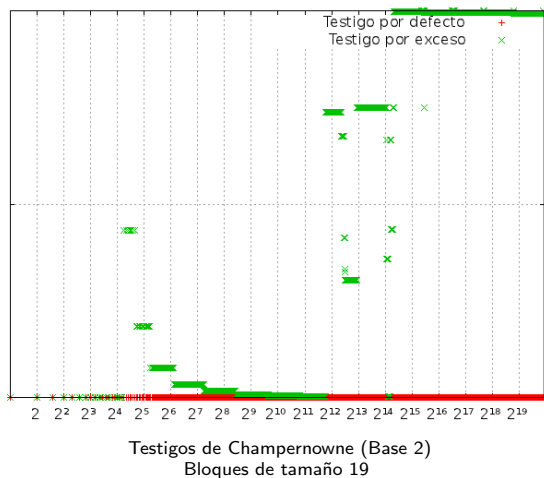
Testigos de Champernowne (Base 2)
Bloques de tamaño 16



Testigos de Champernowne (Base 2)
Bloques de tamaño 17

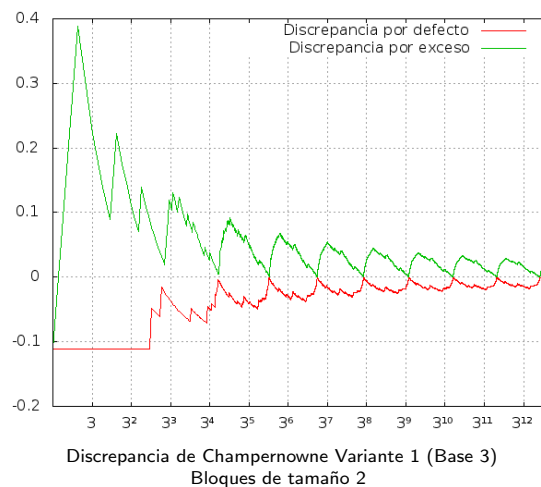
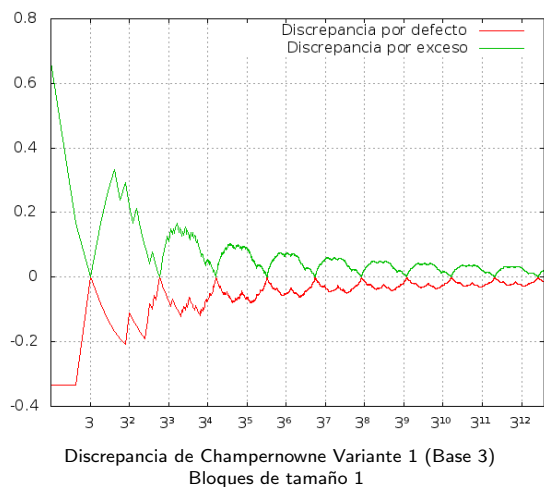


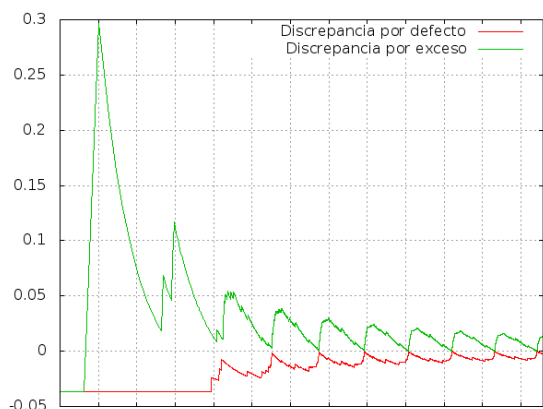
Testigos de Champernowne (Base 2)
Bloques de tamaño 18



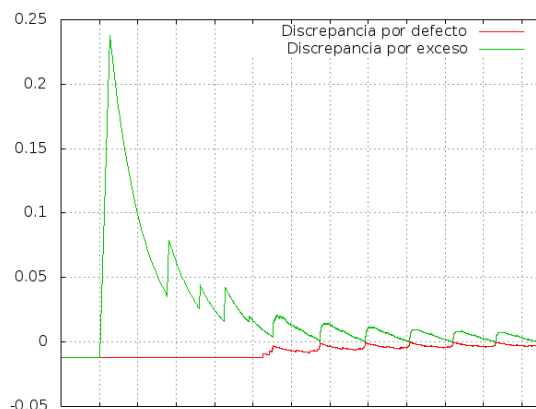
Base 3

Discrepancia de Champernowne Variante 1 en base 3. Se ve claramente un patrón de arcos que se repite a lo largo del eje x . Recordemos que para construir estas variantes, se ordenan todos los bloques de tamaño n antes de mostrar los de tamaño $n + 1$, lo cual explica esta regularidad.

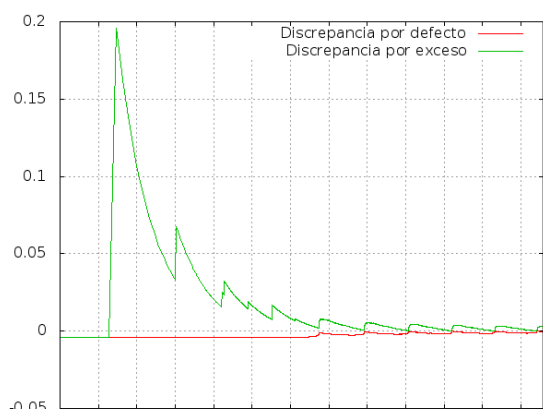




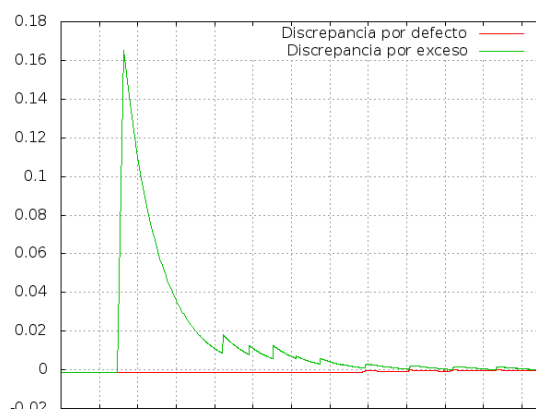
Discrepancia de Champernowne Variante 1 (Base 3)
Bloques de tamaño 3



Discrepancia de Champernowne Variante 1 (Base 3)
Bloques de tamaño 4



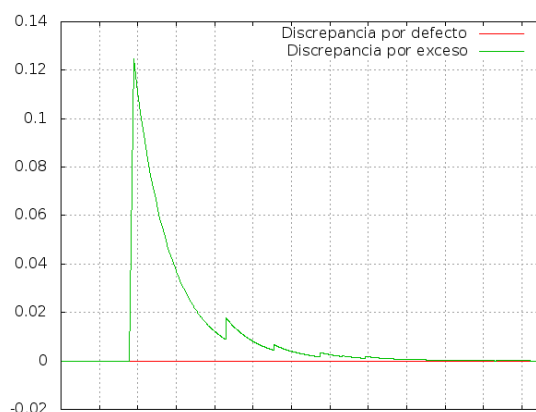
Discrepancia de Champernowne Variante 1 (Base 3)
Bloques de tamaño 5



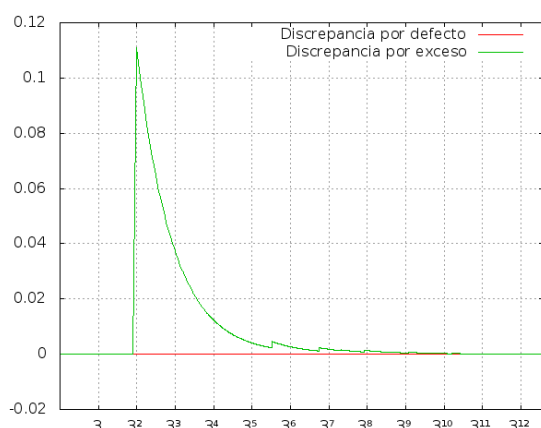
Discrepancia de Champernowne Variante 1 (Base 3)
Bloques de tamaño 6



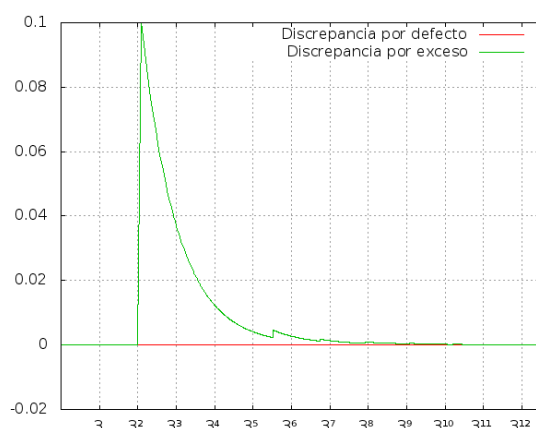
Discrepancia de Champernowne Variante 1 (Base 3)
Bloques de tamaño 7



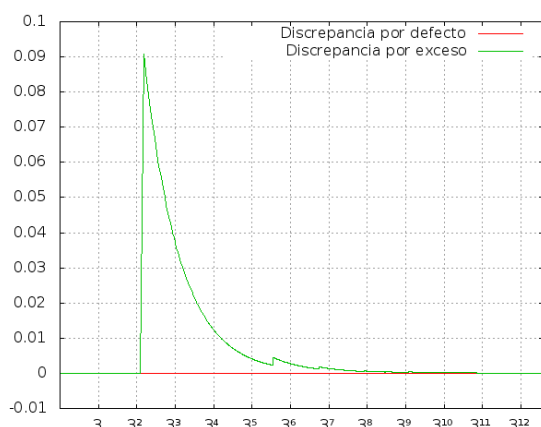
Discrepancia de Champernowne Variante 1 (Base 3)
Bloques de tamaño 8



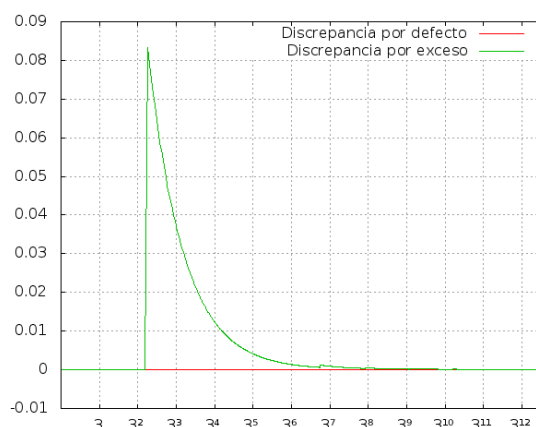
Discrepancia de Champernowne Variante 1 (Base 3)
Bloques de tamaño 9



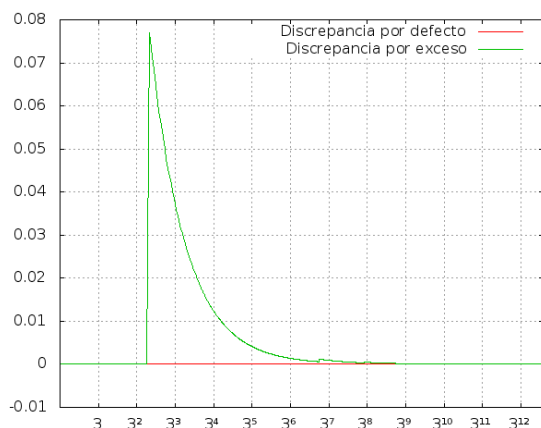
Discrepancia de Champernowne Variante 1 (Base 3)
Bloques de tamaño 10



Discrepancia de Champernowne Variante 1 (Base 3)
Bloques de tamaño 11



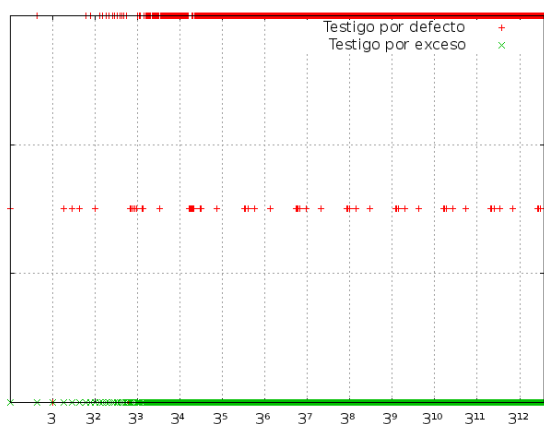
Discrepancia de Champernowne Variante 1 (Base 3)
Bloques de tamaño 12



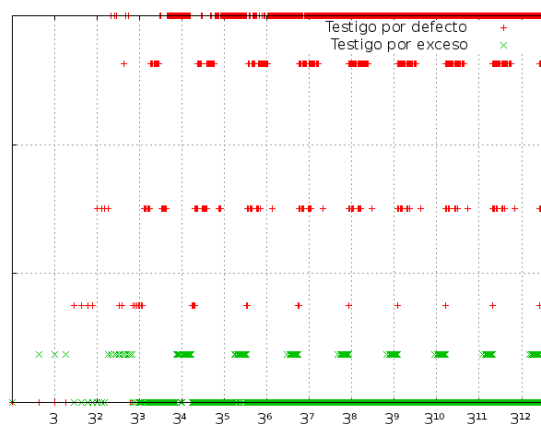
Discrepancia de Champernowne Variante 1 (Base 3)
Bloques de tamaño 13

Testigos de Champernowne Variante 1 en base 3. Esta variante presenta gráficos de testigos curiosos. El testigo por exceso (verde) es casi siempre 0^i , pero el testigo por defecto (rojo) muestra un patrón donde comienza desde 0^i cambiando en cada paso haciendo una barrida de los bloques que comienzan con 0 (el tercio de abajo). Luego se saltea los bloques que comienzan con 1 (el tercio del medio) y continúa con un barrido más

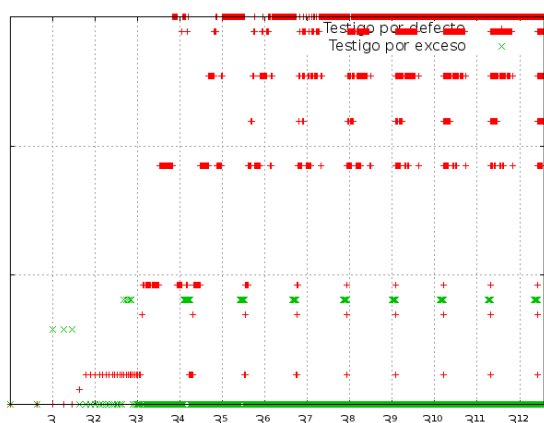
errático de los bloques que comienzan con 2 (el tercio de arriba).



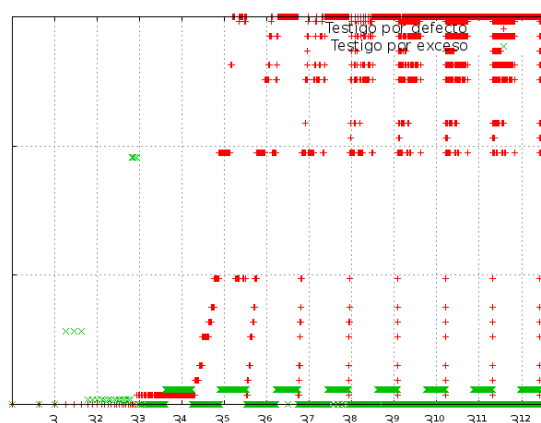
Testigos de Champernowne Variante 1 (Base 3)
Bloques de tamaño 1



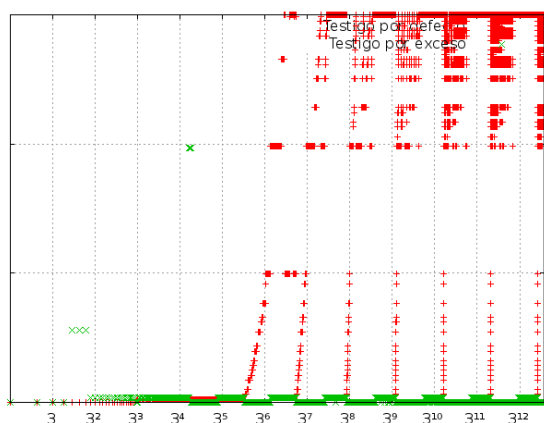
Testigos de Champernowne Variante 1 (Base 3)
Bloques de tamaño 2



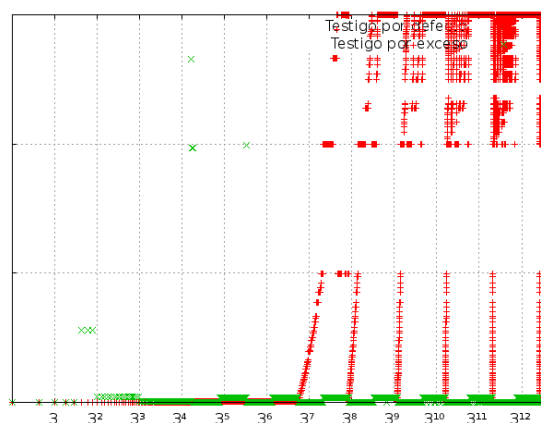
Testigos de Champernowne Variante 1 (Base 3)
Bloques de tamaño 3



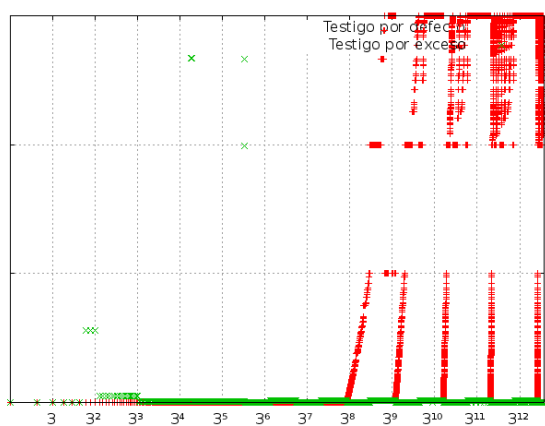
Testigos de Champernowne Variante 1 (Base 3)
Bloques de tamaño 4



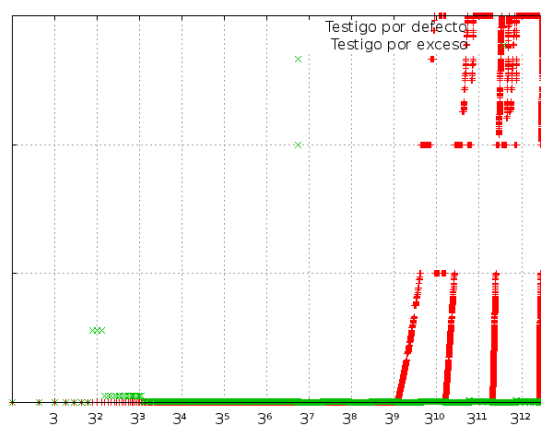
Testigos de Champernowne Variante 1 (Base 3)
Bloques de tamaño 5



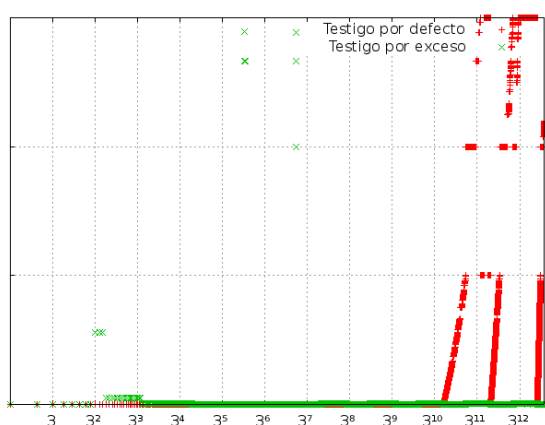
Testigos de Champernowne Variante 1 (Base 3)
Bloques de tamaño 6



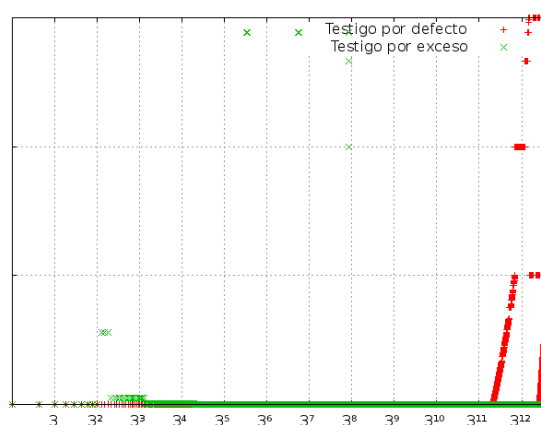
Testigos de Champernowne Variante 1 (Base 3)
Bloques de tamaño 7



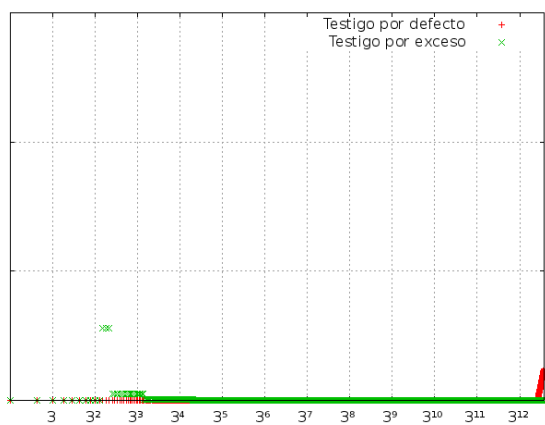
Testigos de Champernowne Variante 1 (Base 3)
Bloques de tamaño 8



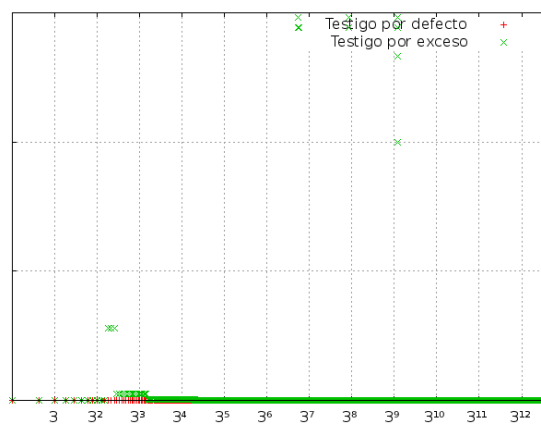
Testigos de Champernowne Variante 1 (Base 3)
Bloques de tamaño 9



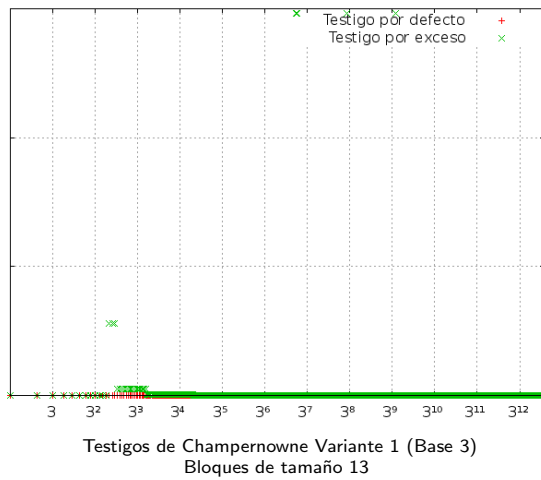
Testigos de Champernowne Variante 1 (Base 3)
Bloques de tamaño 10



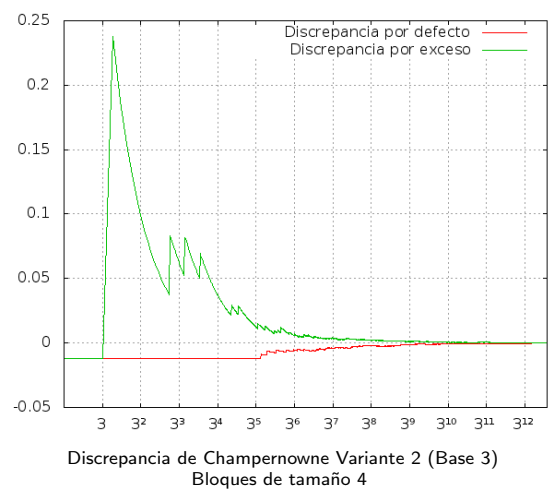
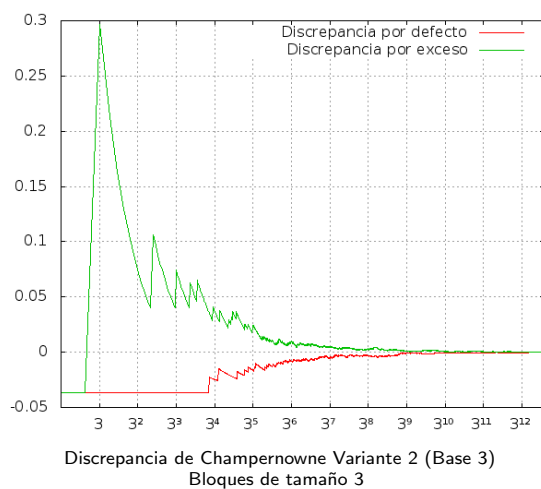
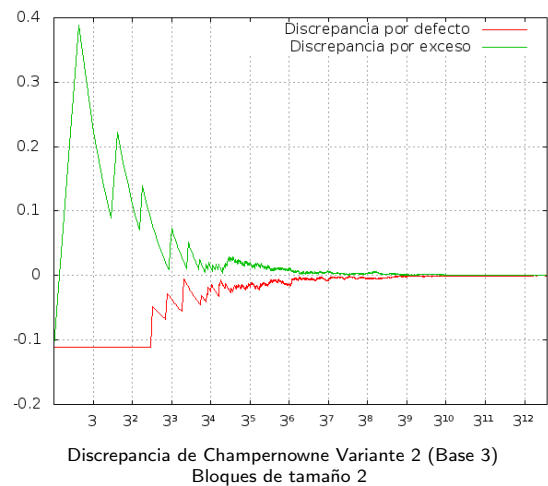
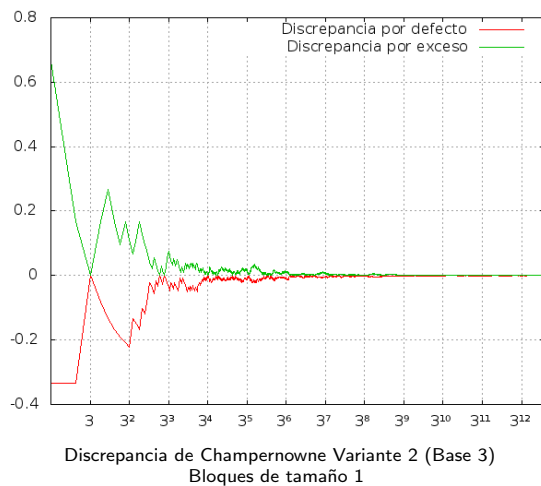
Testigos de Champernowne Variante 1 (Base 3)
Bloques de tamaño 11

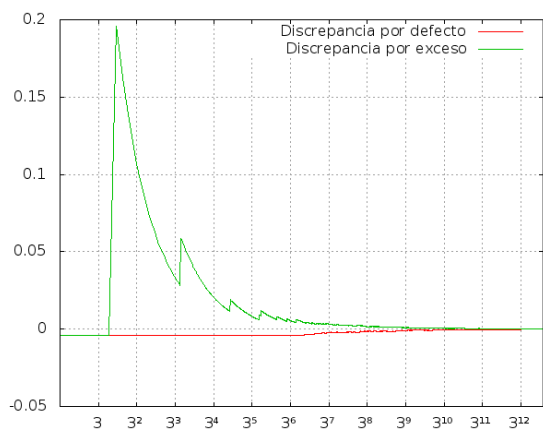


Testigos de Champernowne Variante 1 (Base 3)
Bloques de tamaño 12

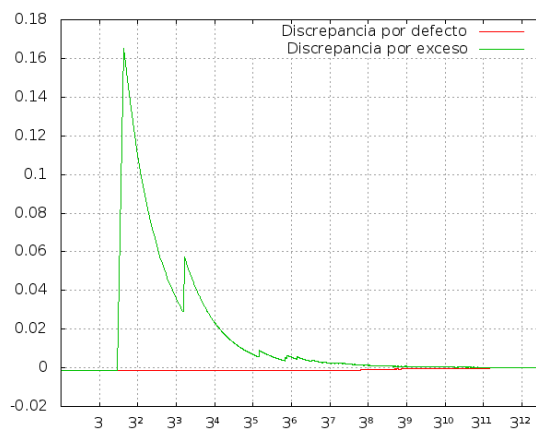


Discrepancia de Champernowne Variante 2 en base 3. La discrepancia de esta variante converge muy rápidamente a 0 tanto por exceso como por defecto, incluso para tamaños de bloque chicos. Recordemos que esta es la variante con orden aleatorio.

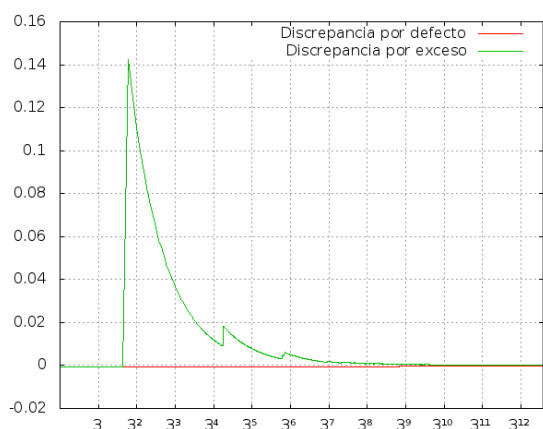




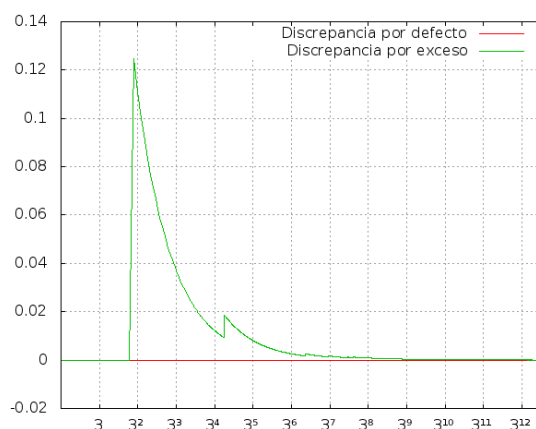
Discrepancia de Champernowne Variante 2 (Base 3)
Bloques de tamaño 5



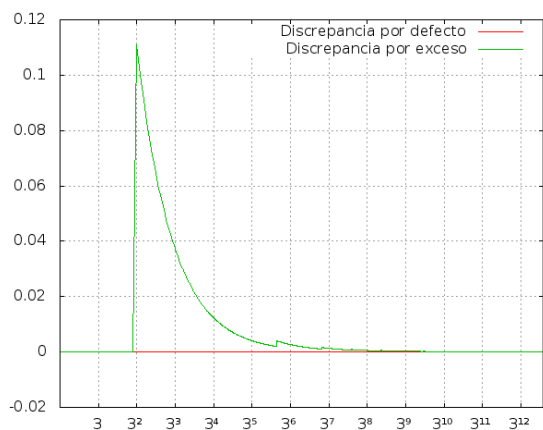
Discrepancia de Champernowne Variante 2 (Base 3)
Bloques de tamaño 6



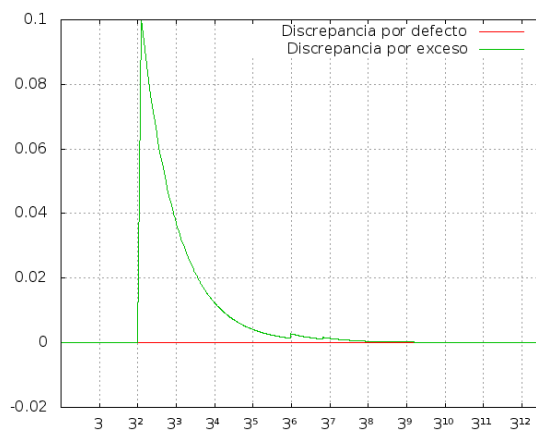
Discrepancia de Champernowne Variante 2 (Base 3)
Bloques de tamaño 7



Discrepancia de Champernowne Variante 2 (Base 3)
Bloques de tamaño 8



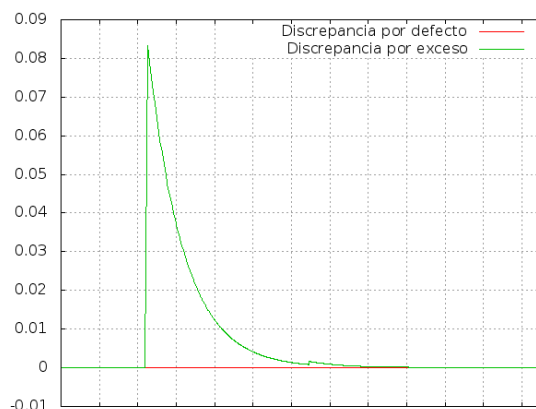
Discrepancia de Champernowne Variante 2 (Base 3)
Bloques de tamaño 9



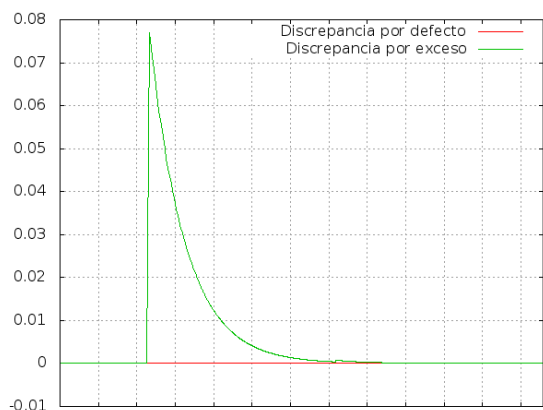
Discrepancia de Champernowne Variante 2 (Base 3)
Bloques de tamaño 10



Discrepancia de Champernowne Variante 2 (Base 3)
Bloques de tamaño 11

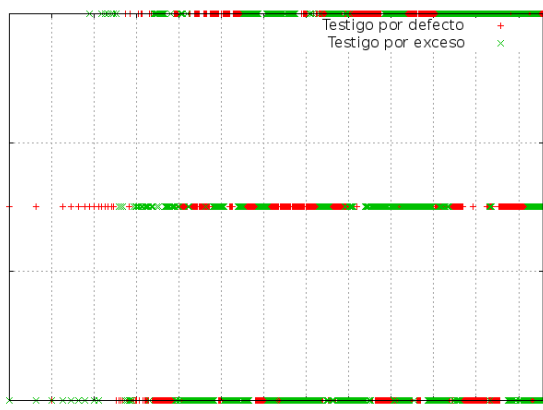


Discrepancia de Champernowne Variante 2 (Base 3)
Bloques de tamaño 12

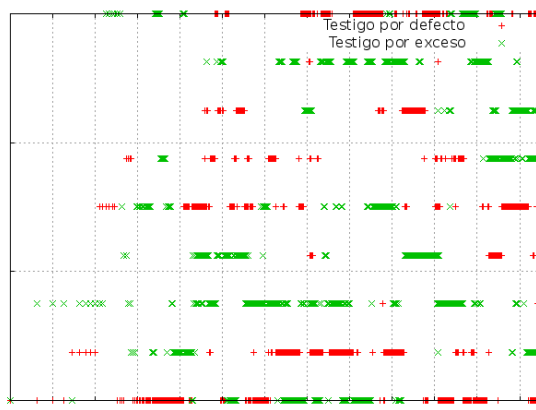


Discrepancia de Champernowne Variante 2 (Base 3)
Bloques de tamaño 13

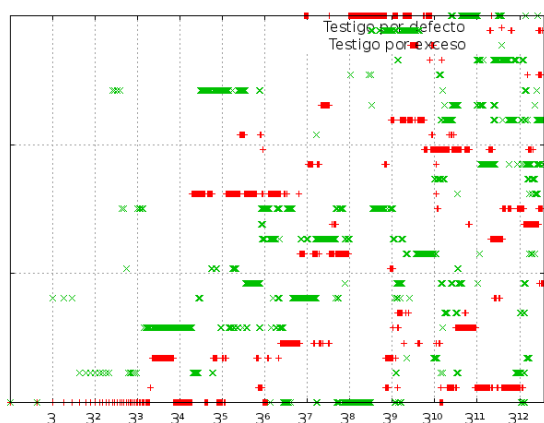
Testigos de Champernowne Variante 2. Los testigos para esta variante cambian constantemente, como se esperaría por su construcción fuertemente aleatoria. Hay puntos verdes (exceso) y rojos (defecto) saltando de renglón en renglón en lugar de mantenerse constante.



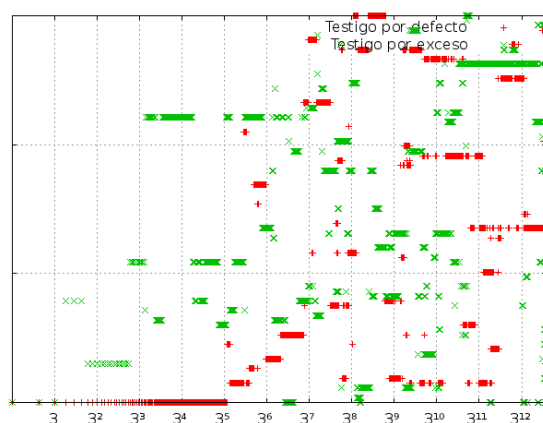
Testigos de Champernowne Variante 2 (Base 3)
Bloques de tamaño 1



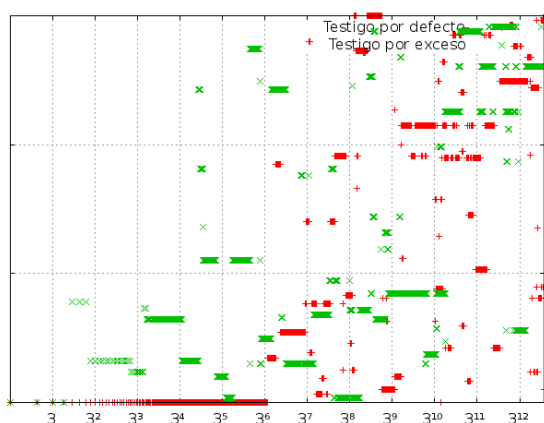
Testigos de Champernowne Variante 2 (Base 3)
Bloques de tamaño 2



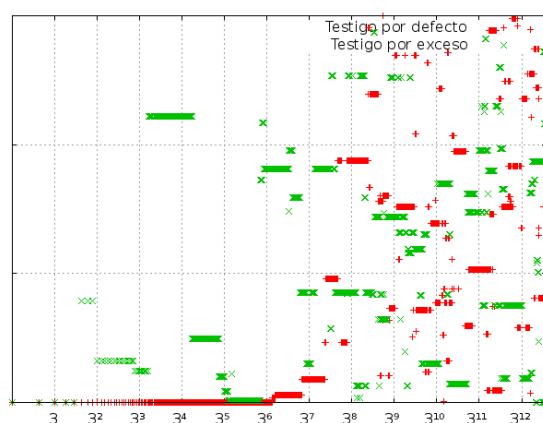
Testigos de Champernowne Variante 2 (Base 3)
Bloques de tamaño 3



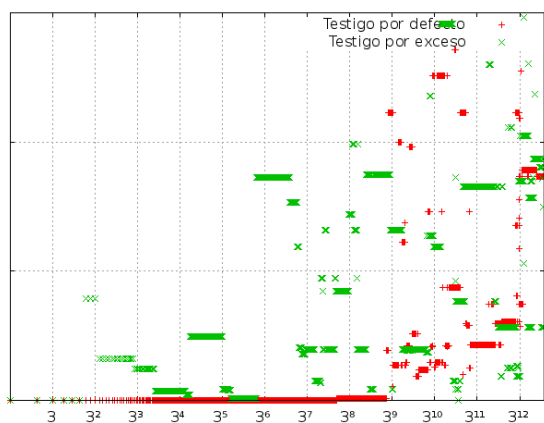
Testigos de Champernowne Variante 2 (Base 3)
Bloques de tamaño 4



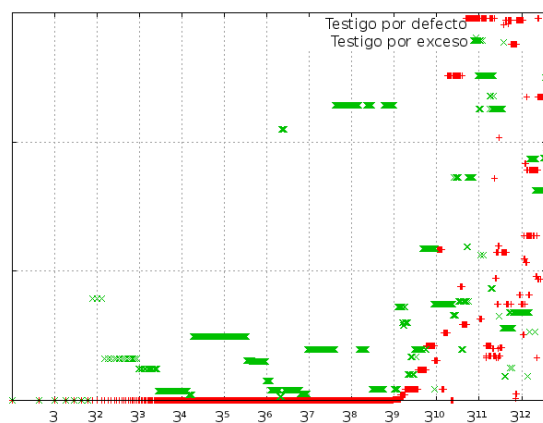
Testigos de Champernowne Variante 2 (Base 3)
Bloques de tamaño 5



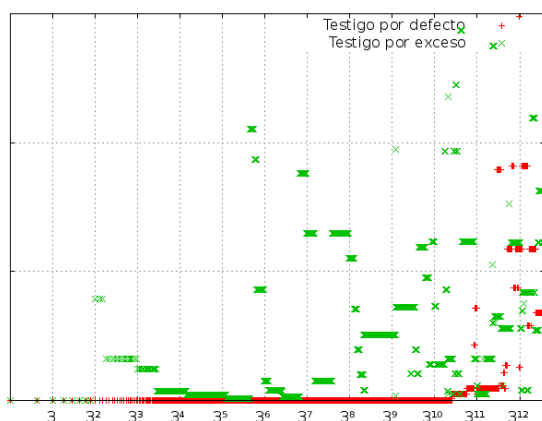
Testigos de Champernowne Variante 2 (Base 3)
Bloques de tamaño 6



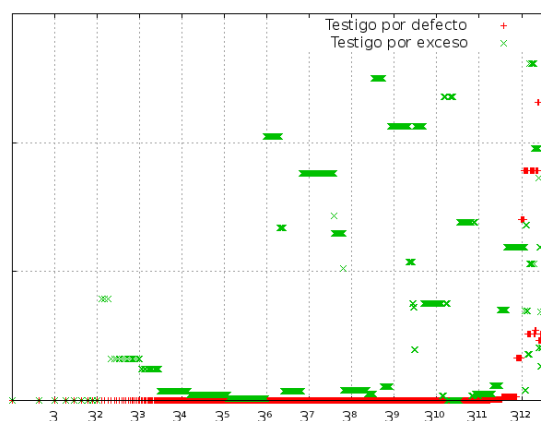
Testigos de Champernowne Variante 2 (Base 3)
Bloques de tamaño 7



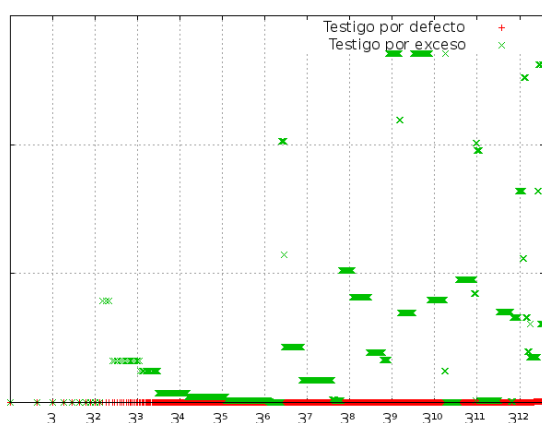
Testigos de Champernowne Variante 2 (Base 3)
Bloques de tamaño 8



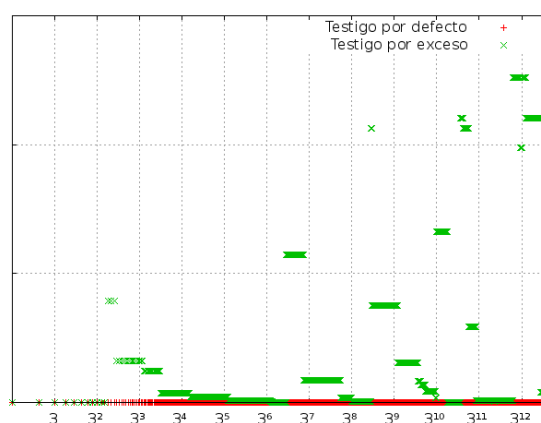
Testigos de Champernowne Variante 2 (Base 3)
Bloques de tamaño 9



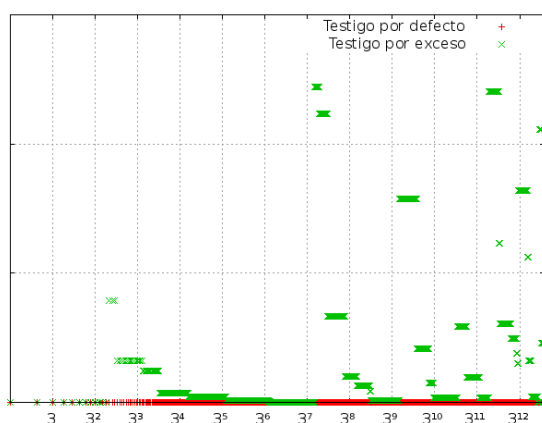
Testigos de Champernowne Variante 2 (Base 3)
Bloques de tamaño 10



Testigos de Champernowne Variante 2 (Base 3)
Bloques de tamaño 11

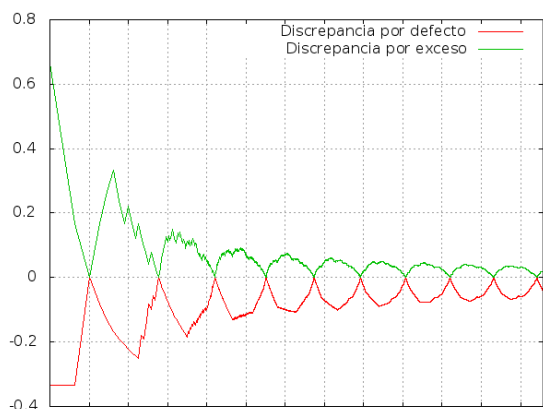


Testigos de Champernowne Variante 2 (Base 3)
Bloques de tamaño 12

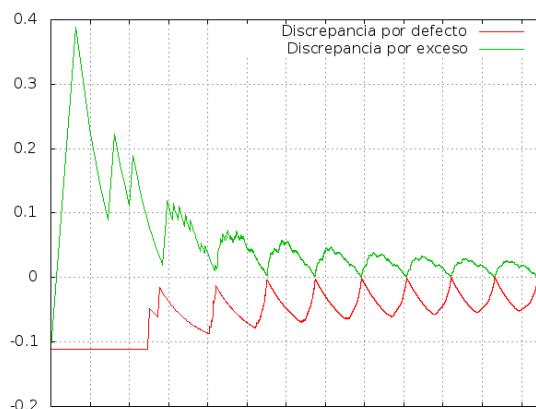


Testigos de Champernowne Variante 2 (Base 3)
Bloques de tamaño 13

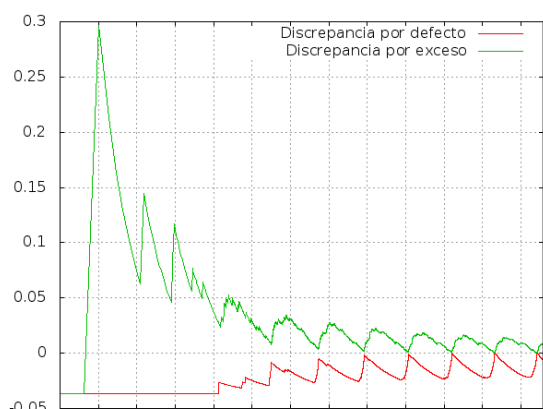
Discrepancia de Champernowne Variante 3 en base 3. Esta variante tiene una discrepancia similar a la variante 1, con un patrón que se repite a lo largo del eje x .



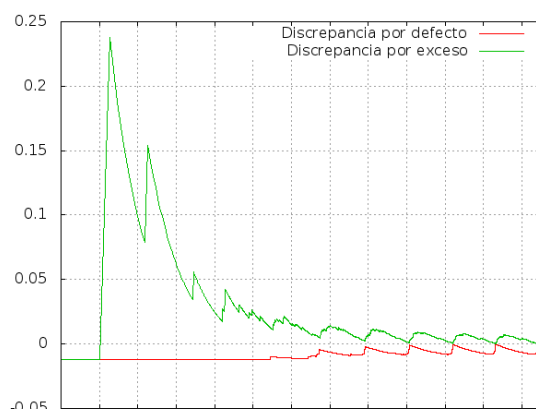
Discrepancia de Champernowne Variante 3 (Base 3)
Bloques de tamaño 1



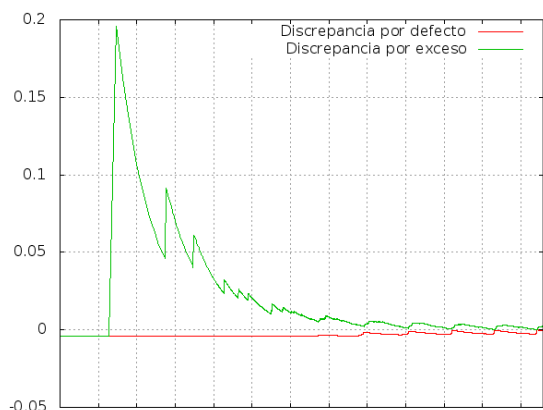
Discrepancia de Champernowne Variante 3 (Base 3)
Bloques de tamaño 2



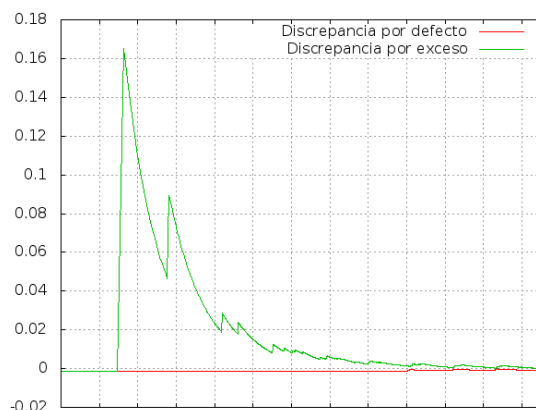
Discrepancia de Champernowne Variante 3 (Base 3)
Bloques de tamaño 3



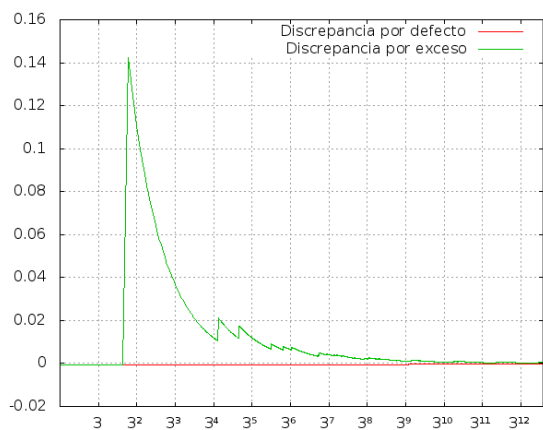
Discrepancia de Champernowne Variante 3 (Base 3)
Bloques de tamaño 4



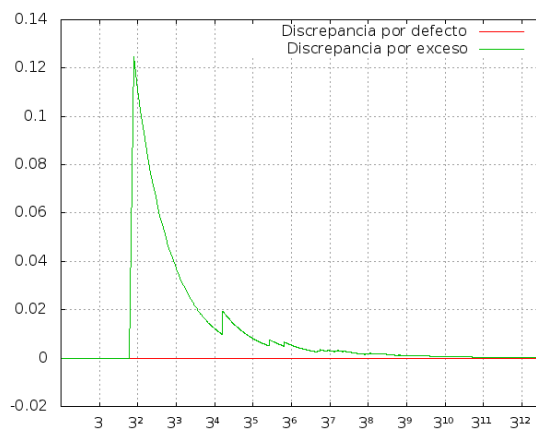
Discrepancia de Champernowne Variante 3 (Base 3)
Bloques de tamaño 5



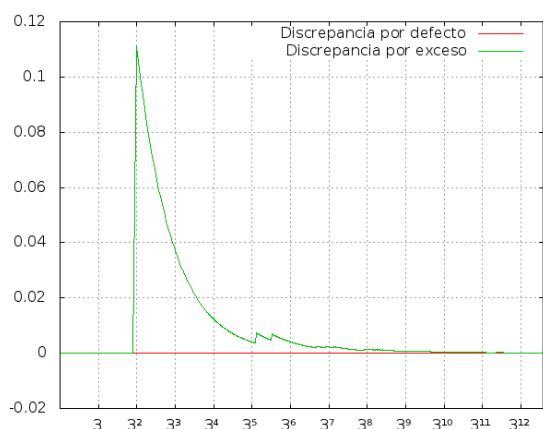
Discrepancia de Champernowne Variante 3 (Base 3)
Bloques de tamaño 6



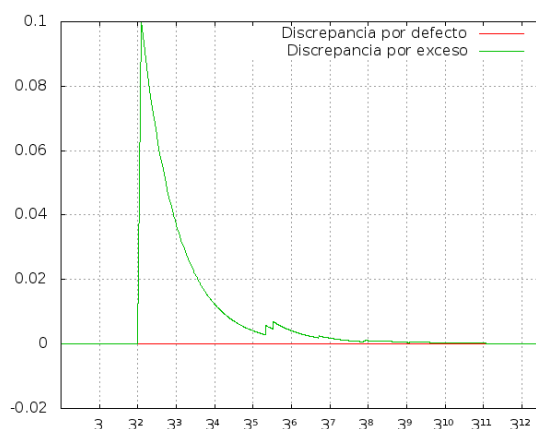
Discrepancia de Champernowne Variante 3 (Base 3)
Bloques de tamaño 7



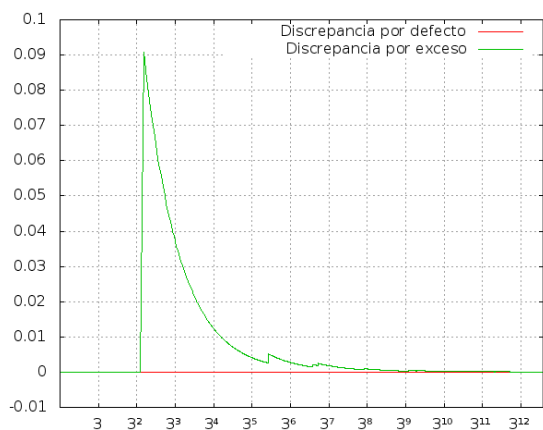
Discrepancia de Champernowne Variante 3 (Base 3)
Bloques de tamaño 8



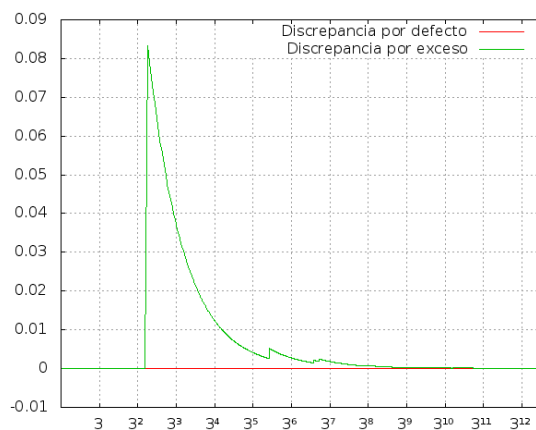
Discrepancia de Champernowne Variante 3 (Base 3)
Bloques de tamaño 9



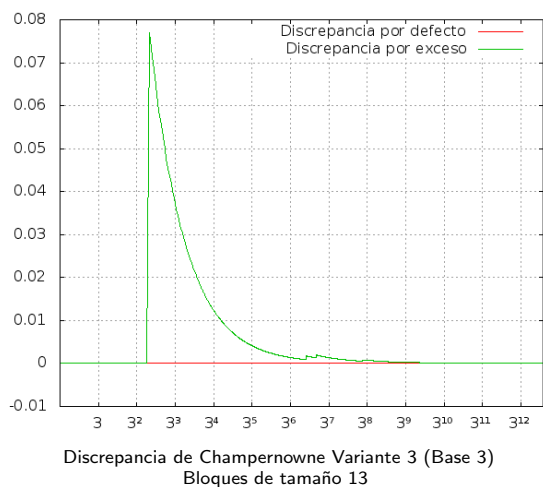
Discrepancia de Champernowne Variante 3 (Base 3)
Bloques de tamaño 10



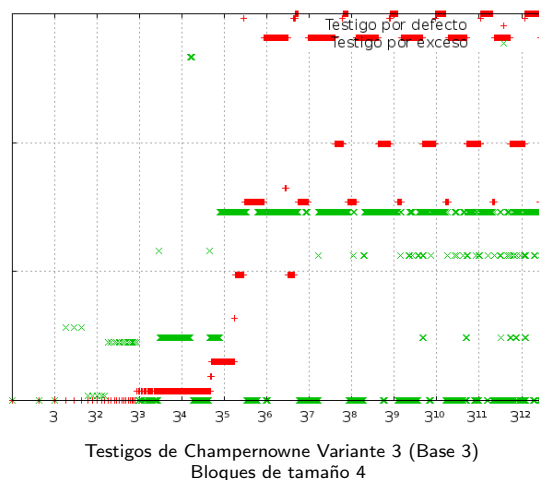
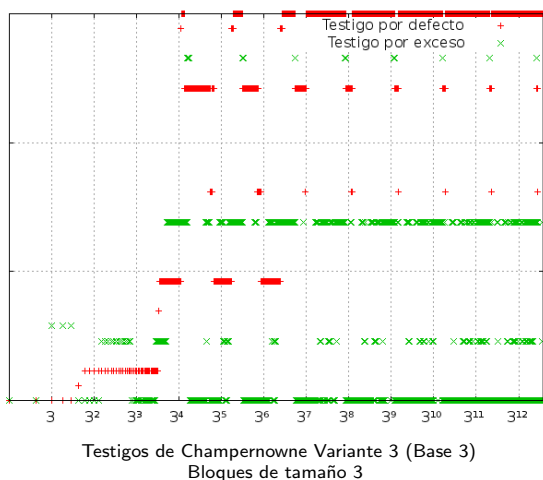
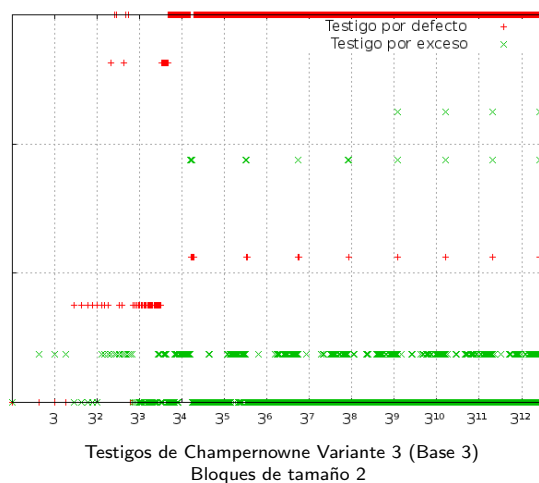
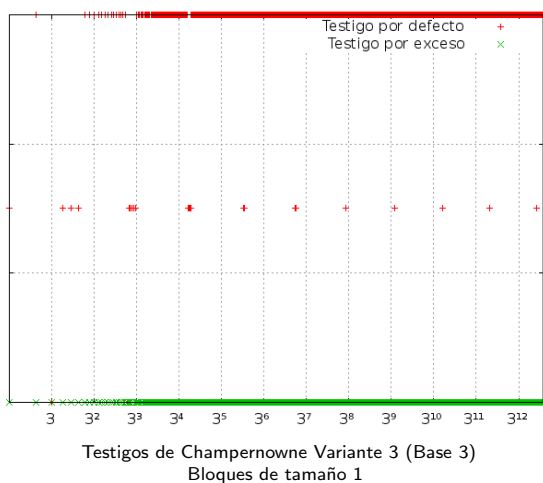
Discrepancia de Champernowne Variante 3 (Base 3)
Bloques de tamaño 11

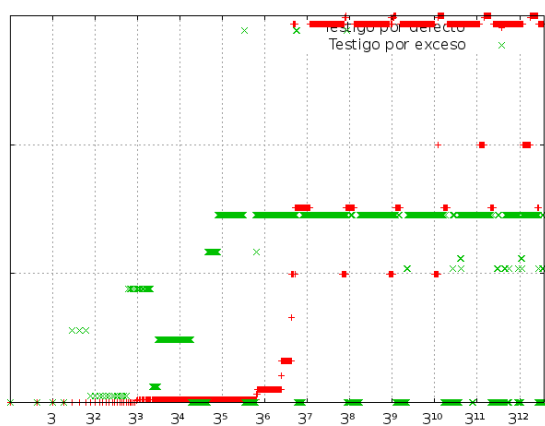


Discrepancia de Champernowne Variante 3 (Base 3)
Bloques de tamaño 12

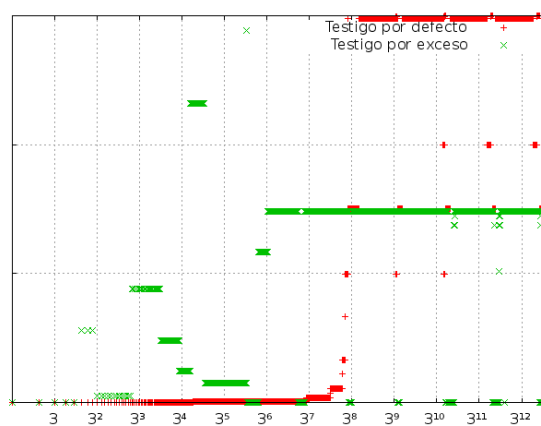


Testigos de Champernowne Variante 3 en base 3. Esta variante es menos errática que la anterior en sus testigos, favoreciendo en general a bloques de la forma x^i , aunque también presenta bastantes saltos.

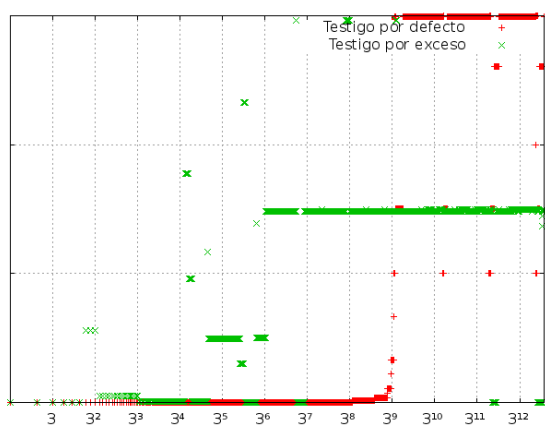




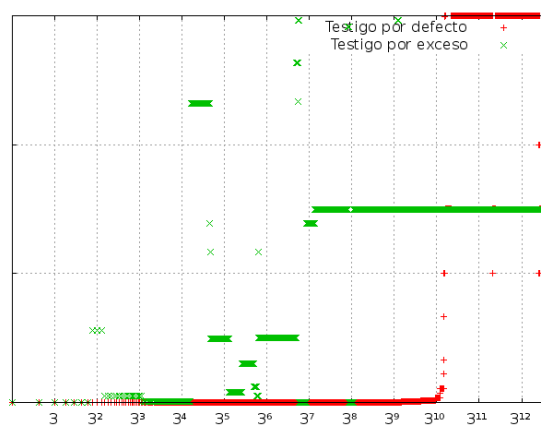
Testigos de Champernowne Variante 3 (Base 3)
Bloques de tamaño 5



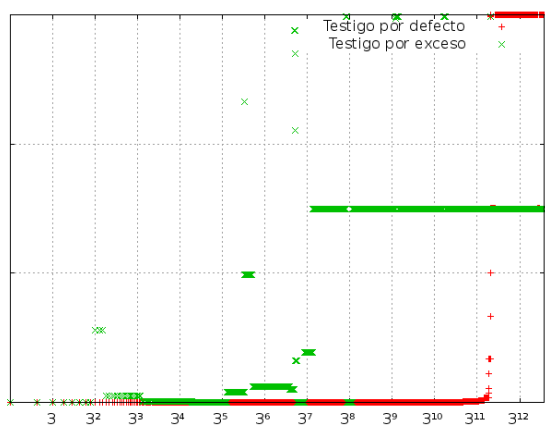
Testigos de Champernowne Variante 3 (Base 3)
Bloques de tamaño 6



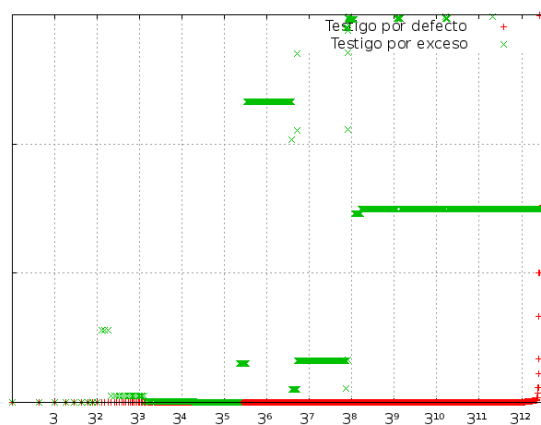
Testigos de Champernowne Variante 3 (Base 3)
Bloques de tamaño 7



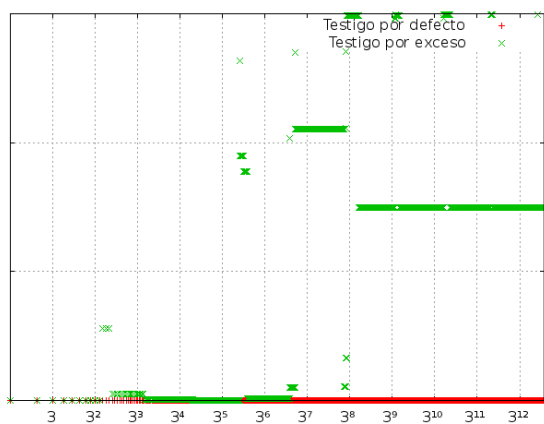
Testigos de Champernowne Variante 3 (Base 3)
Bloques de tamaño 8



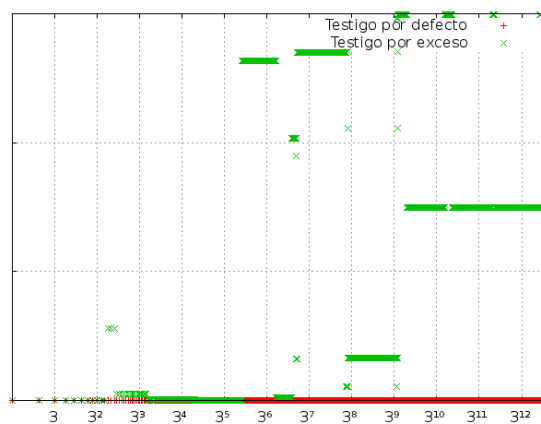
Testigos de Champernowne Variante 3 (Base 3)
Bloques de tamaño 9



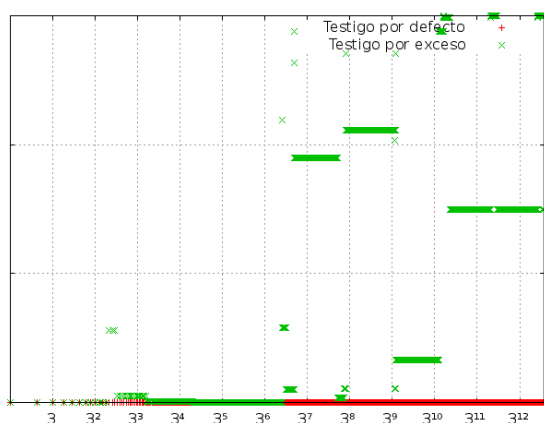
Testigos de Champernowne Variante 3 (Base 3)
Bloques de tamaño 10



Testigos de Champernowne Variante 3 (Base 3)
Bloques de tamaño 11



Testigos de Champernowne Variante 3 (Base 3)
Bloques de tamaño 12

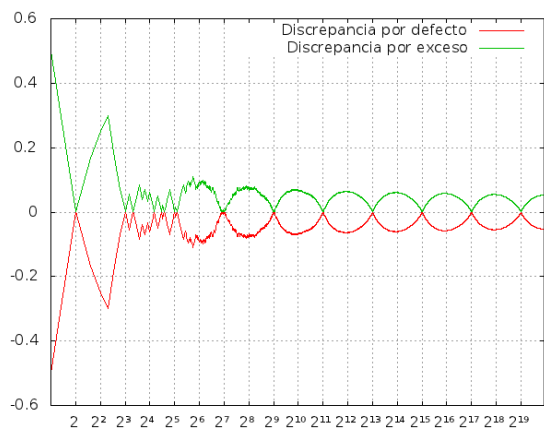


Testigos de Champernowne Variante 3 (Base 3)
Bloques de tamaño 13

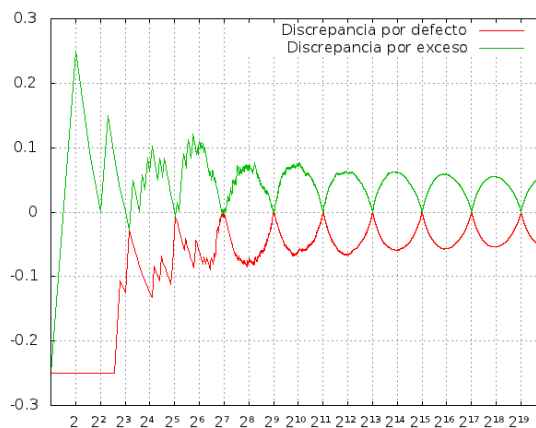
4.2.2. Ford extendida en base 2 y en base 3

Base 2

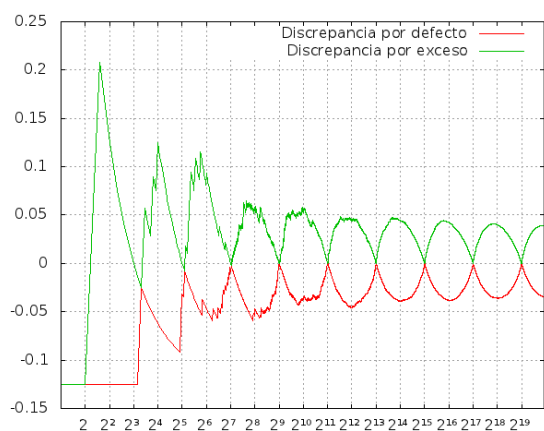
Discrepancia de Ford Extendida. En el caso de la secuencia de Ford Extendida, la regularidad en la curva es aún mayor que en las otras secuencias. La curva es mayormente suave, casi sin picos que le den una apariencia dentada como otras secuencias. Por definición, la discrepancia llega a 0 en todos los puntos de la forma 2^{n+1} por lo que el período de regularidad es siempre el mismo.



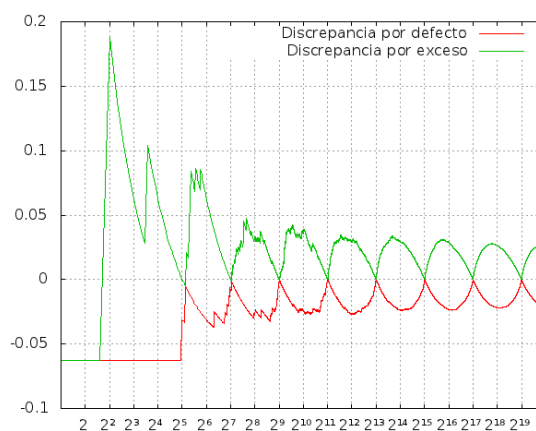
Discrepancia de Ford Extendida (Base 2)
Bloques de tamaño 1



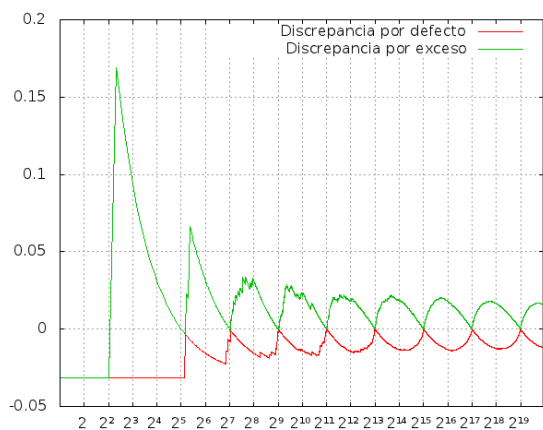
Discrepancia de Ford Extendida (Base 2)
Bloques de tamaño 2



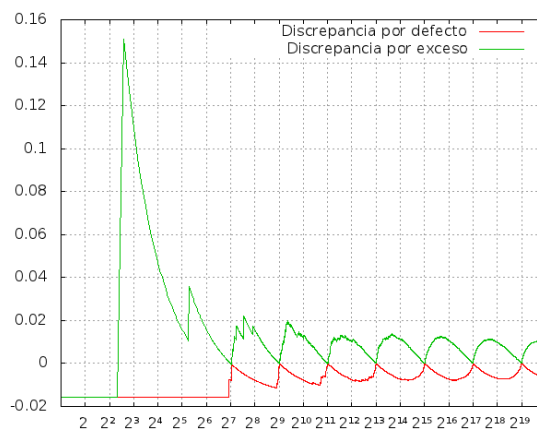
Discrepancia de Ford Extendida (Base 2)
Bloques de tamaño 3



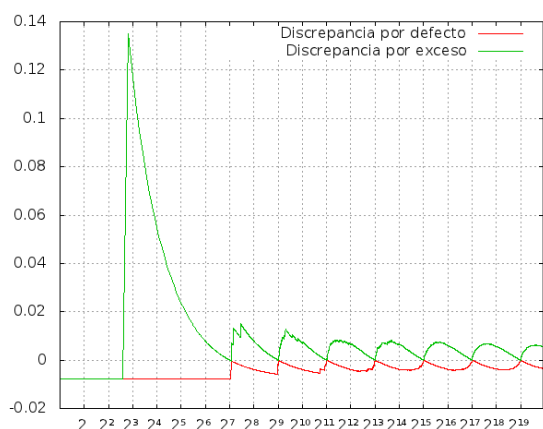
Discrepancia de Ford Extendida (Base 2)
Bloques de tamaño 4



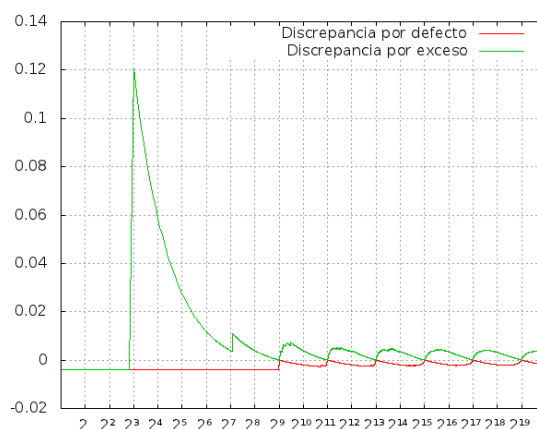
Discrepancia de Ford Extendida (Base 2)
Bloques de tamaño 5



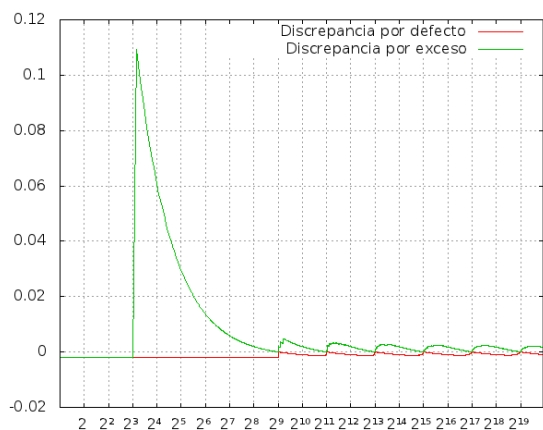
Discrepancia de Ford Extendida (Base 2)
Bloques de tamaño 6



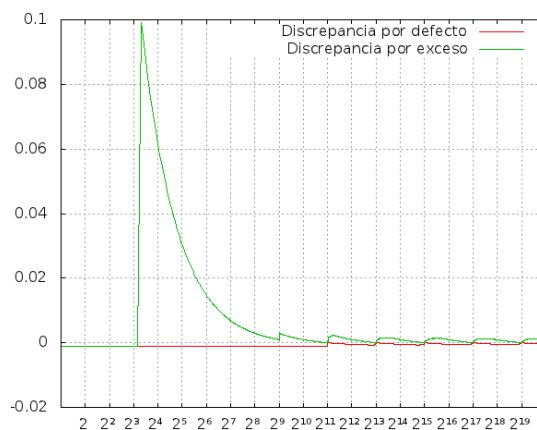
Discrepancia de Ford Extendida (Base 2)
Bloques de tamaño 7



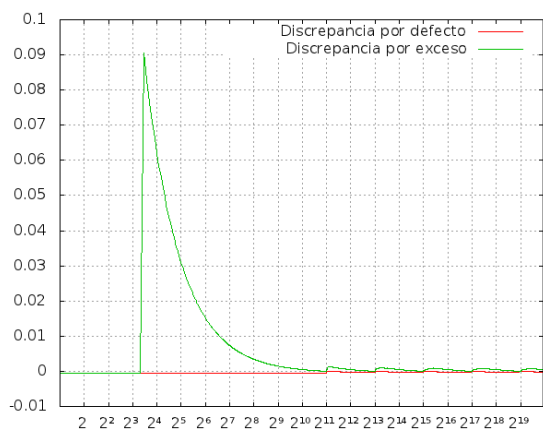
Discrepancia de Ford Extendida (Base 2)
Bloques de tamaño 8



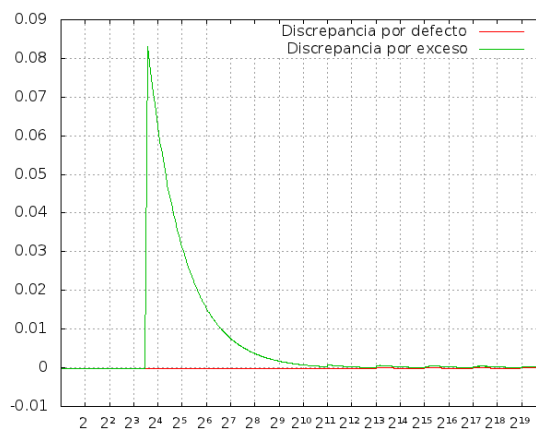
Discrepancia de Ford Extendida (Base 2)
Bloques de tamaño 9



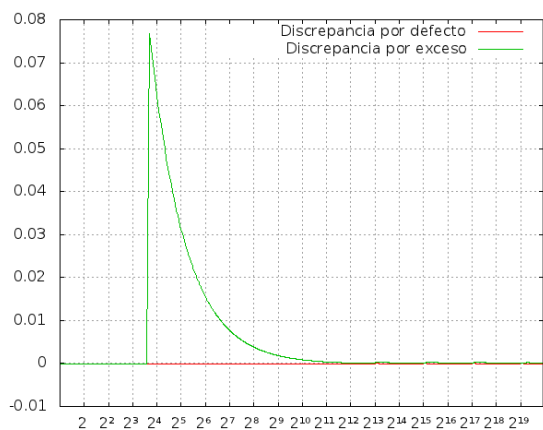
Discrepancia de Ford Extendida (Base 2)
Bloques de tamaño 10



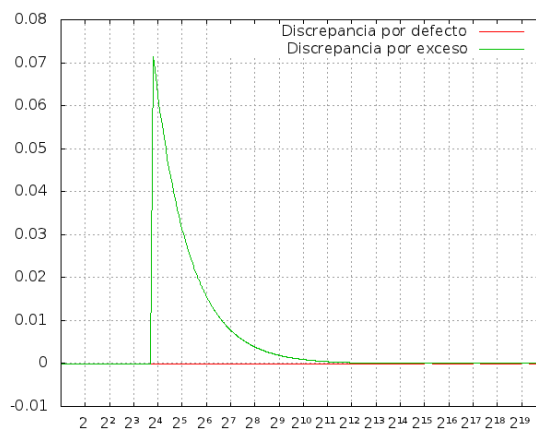
Discrepancia de Ford Extendida (Base 2)
Bloques de tamaño 11



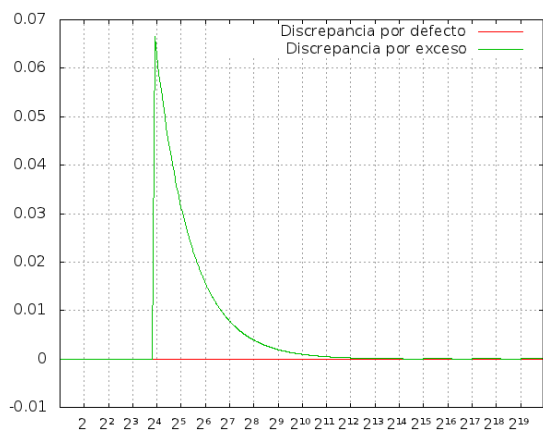
Discrepancia de Ford Extendida (Base 2)
Bloques de tamaño 12



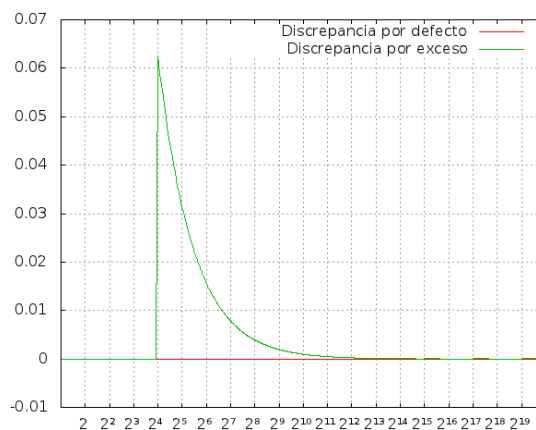
Discrepancia de Ford Extendida (Base 2)
Bloques de tamaño 13



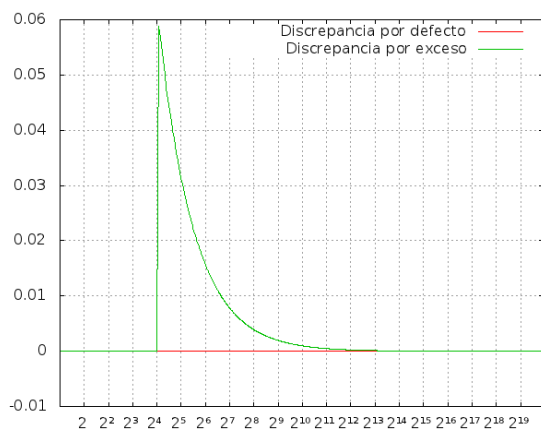
Discrepancia de Ford Extendida (Base 2)
Bloques de tamaño 14



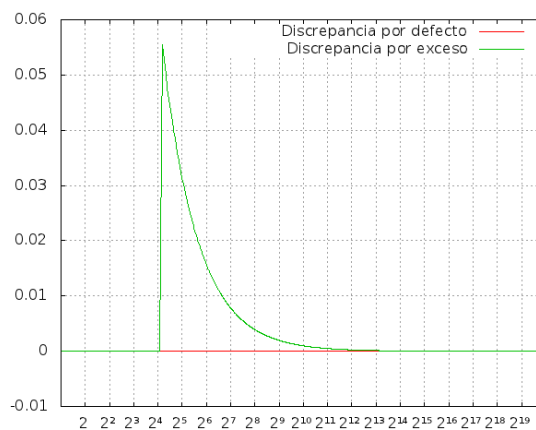
Discrepancia de Ford Extendida (Base 2)
Bloques de tamaño 15



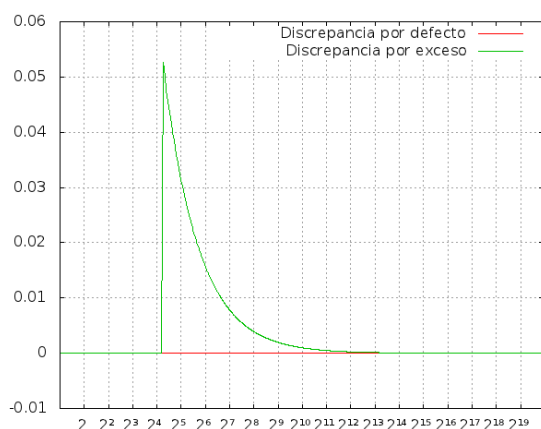
Discrepancia de Ford Extendida (Base 2)
Bloques de tamaño 16



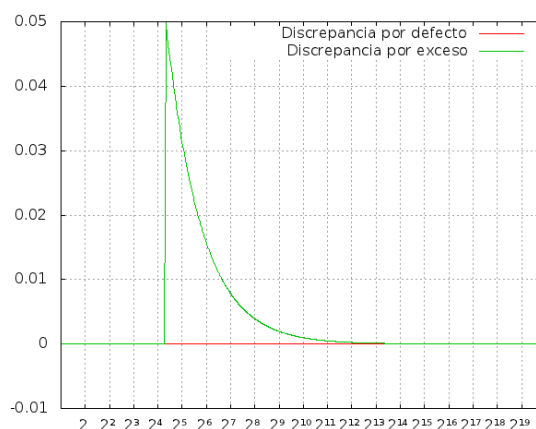
Discrepancia de Ford Extendida (Base 2)
Bloques de tamaño 17



Discrepancia de Ford Extendida (Base 2)
Bloques de tamaño 18

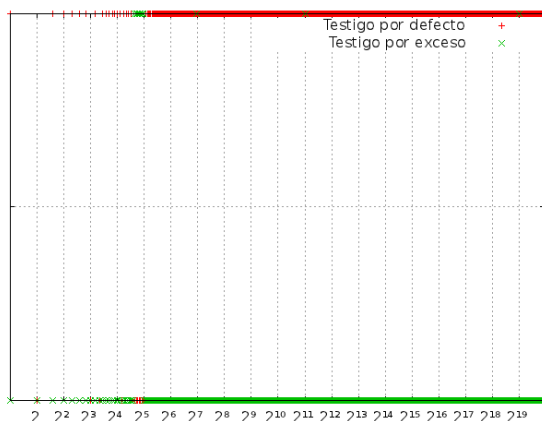


Discrepancia de Ford Extendida (Base 2)
Bloques de tamaño 19

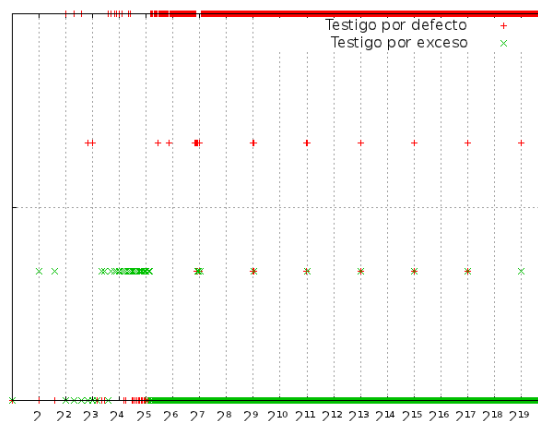


Discrepancia de Ford Extendida (Base 2)
Bloques de tamaño 20

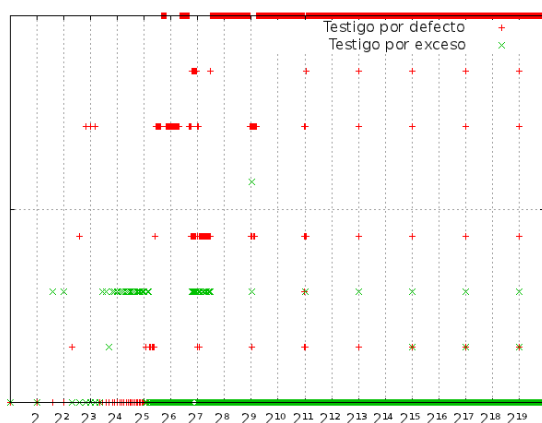
Testigos de Ford Extendida. El testigo por exceso (verde) es el bloque 0^i en casi todos los puntos, por lo que surge la pregunta de qué tan exactamente podemos aproximar la discrepancia de la secuencia de Ford Extendida si miramos las ocurrencias del bloque 0^i exclusivamente, en lugar de considerar todos los bloques de tamaño i . Por otro lado, los testigos por defecto (rojo) siguen el mismo comportamiento en cada período $(2^{n-1}, 2^{n+1})$: comienza en el bloque 0^i y va cambiando a bloques cada vez mayores lexicográficamente hasta alcanzar el mayor de los bloques, 1^i . El comportamiento regular y sin saltos bruscos de los testigos por ambos lados explica la suavidad de la curva de discrepancia.



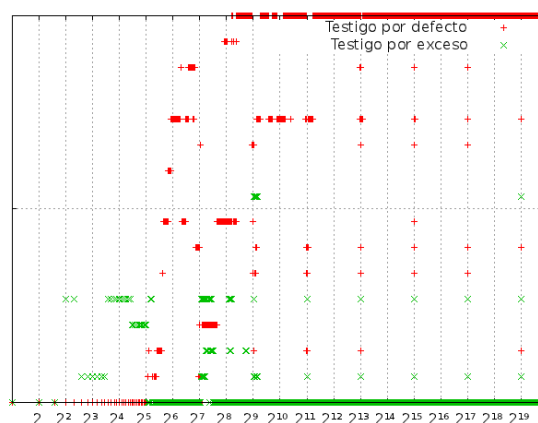
Testigos de Ford Extendida (Base 2)
Bloques de tamaño 1



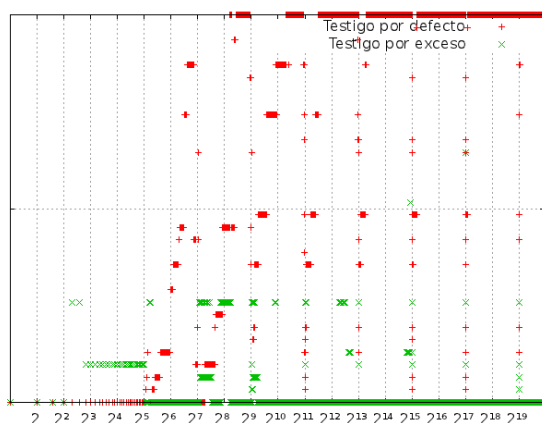
Testigos de Ford Extendida (Base 2)
Bloques de tamaño 2



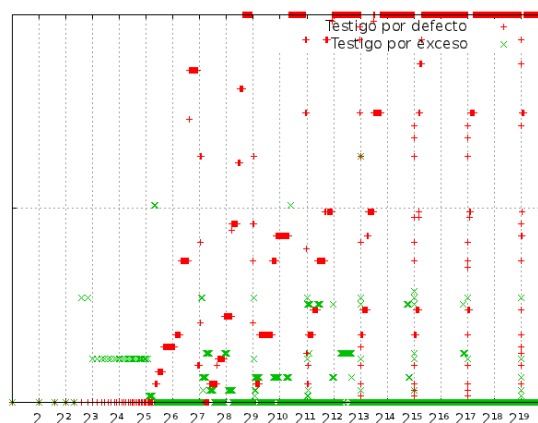
Testigos de Ford Extendida (Base 2)
Bloques de tamaño 3



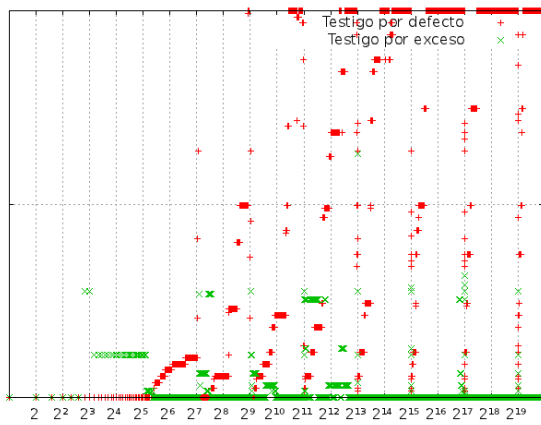
Testigos de Ford Extendida (Base 2)
Bloques de tamaño 4



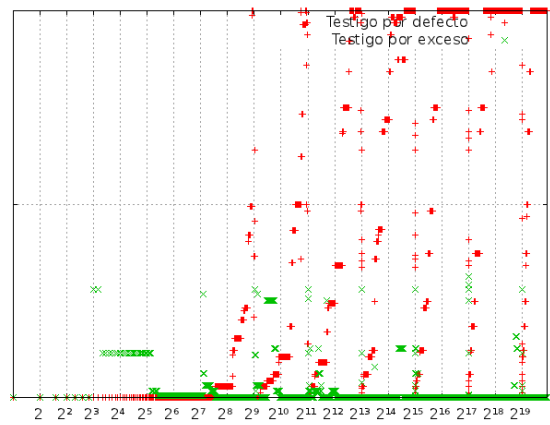
Testigos de Ford Extendida (Base 2)
Bloques de tamaño 5



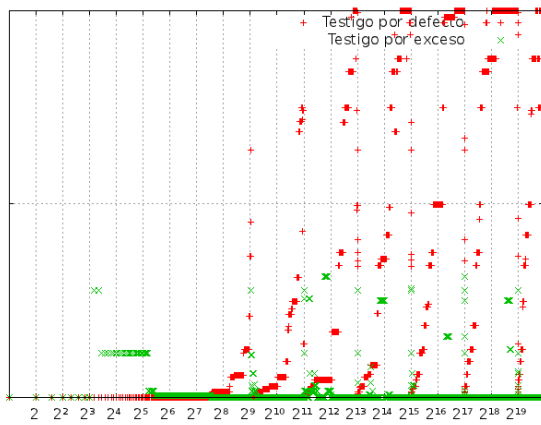
Testigos de Ford Extendida (Base 2)
Bloques de tamaño 6



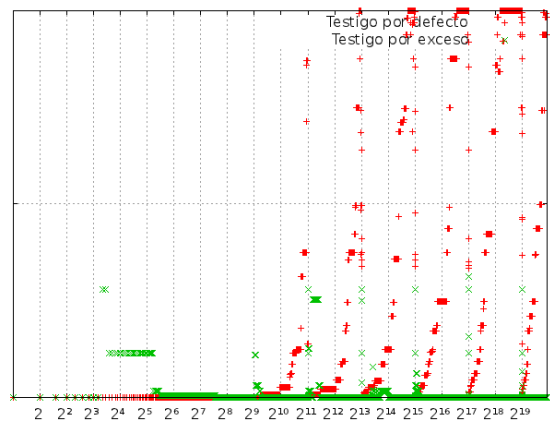
Testigos de Ford Extendida (Base 2)
Bloques de tamaño 7



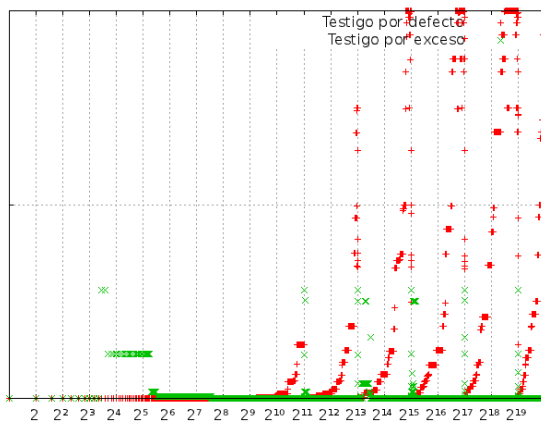
Testigos de Ford Extendida (Base 2)
Bloques de tamaño 8



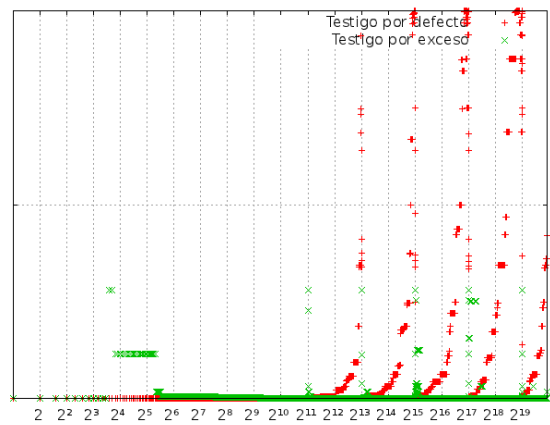
Testigos de Ford Extendida (Base 2)
Bloques de tamaño 9



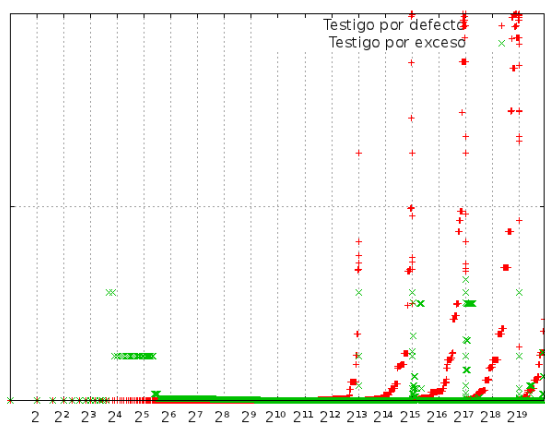
Testigos de Ford Extendida (Base 2)
Bloques de tamaño 10



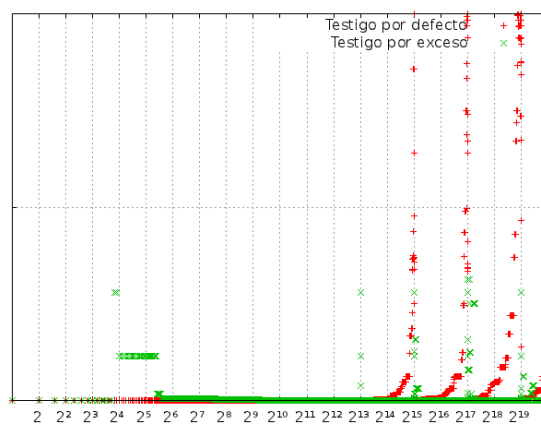
Testigos de Ford Extendida (Base 2)
Bloques de tamaño 11



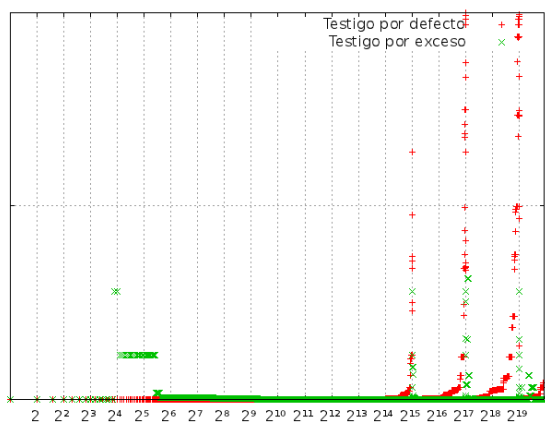
Testigos de Ford Extendida (Base 2)
Bloques de tamaño 12



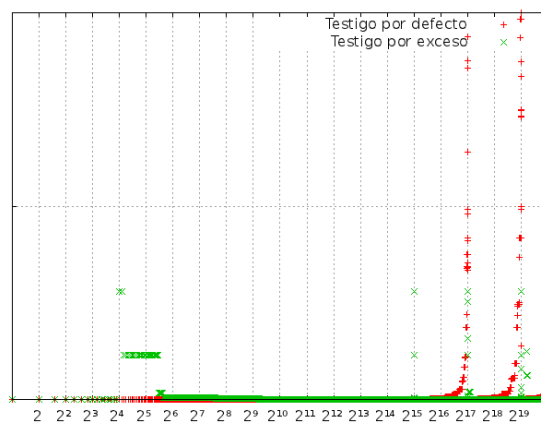
Testigos de Ford Extendida (Base 2)
Bloques de tamaño 13



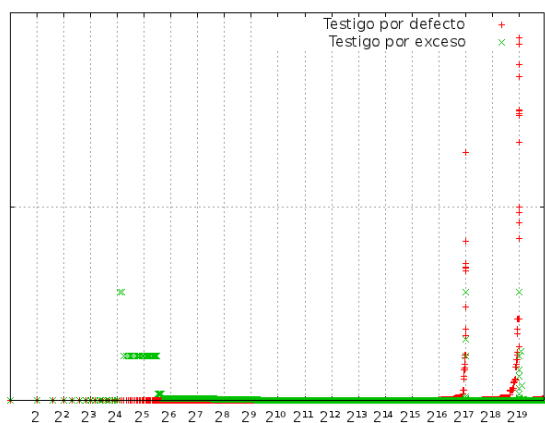
Testigos de Ford Extendida (Base 2)
Bloques de tamaño 14



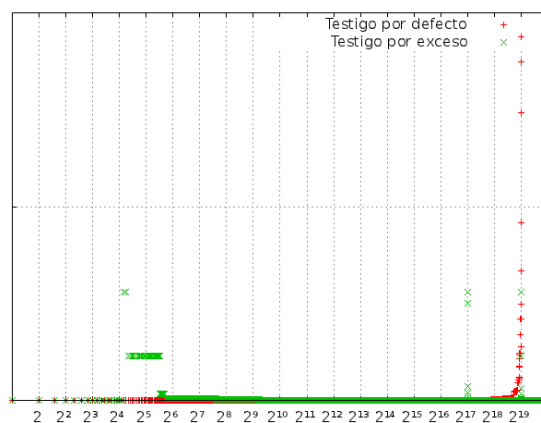
Testigos de Ford Extendida (Base 2)
Bloques de tamaño 15



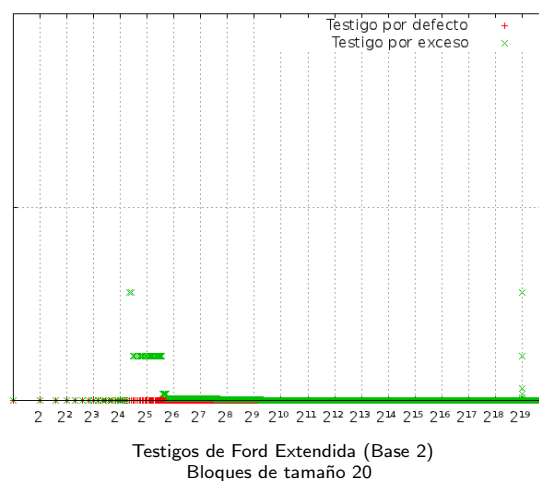
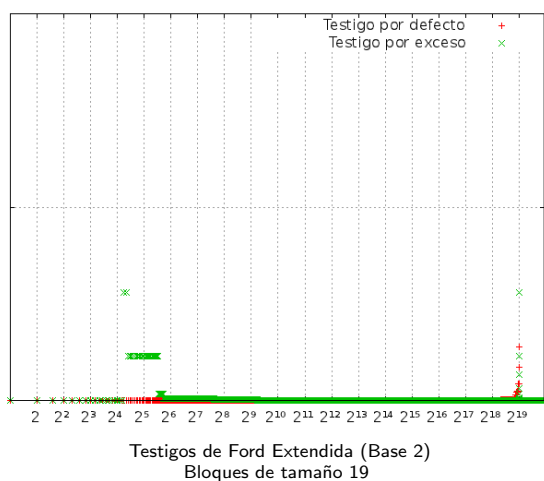
Testigos de Ford Extendida (Base 2)
Bloques de tamaño 16



Testigos de Ford Extendida (Base 2)
Bloques de tamaño 17

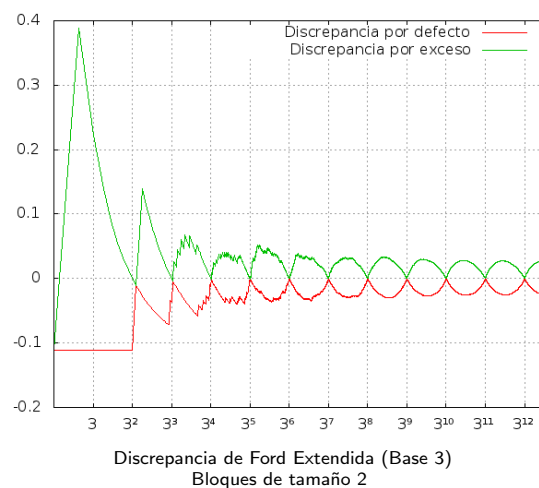
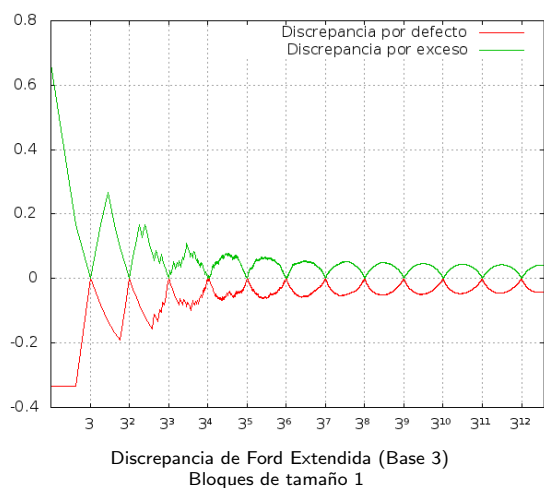


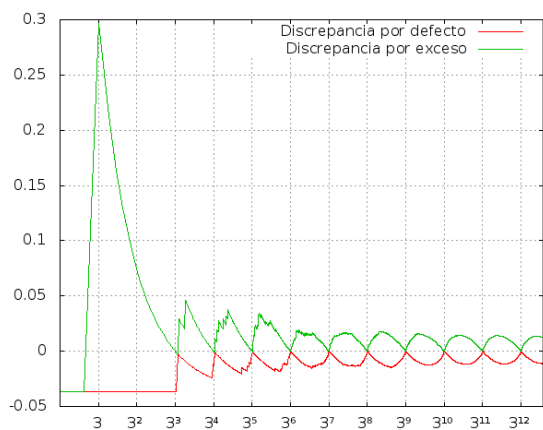
Testigos de Ford Extendida (Base 2)
Bloques de tamaño 18



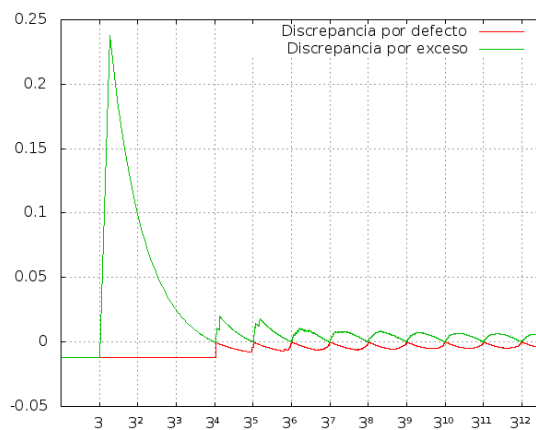
Base 3

Discrepancia de Ford Extendida en base 3. Para alfabeto de tamaño 3, la secuencia de Ford Extendida se comporta igual que para el alfabeto de tamaño 2, con una curva sin cambios bruscos. La única salvedad es que la regularidad se presenta entre cada potencia de 3, sin saltar ninguna. Esto era de esperarse, debido a que en un alfabeto de tamaño 3, podemos extender una secuencia de Bruijn a otra de orden inmediatamente superior, a diferencia de un alfabeto de tamaño 2 donde debemos saltar un orden.

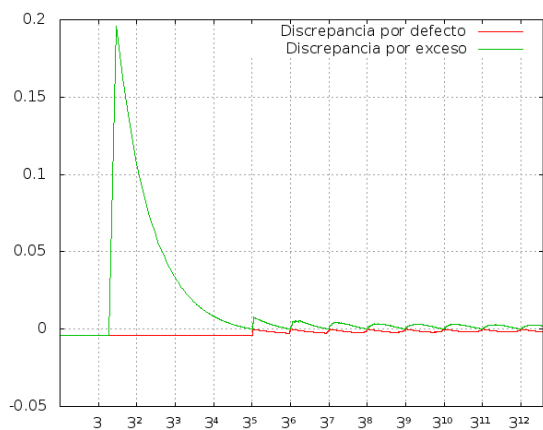




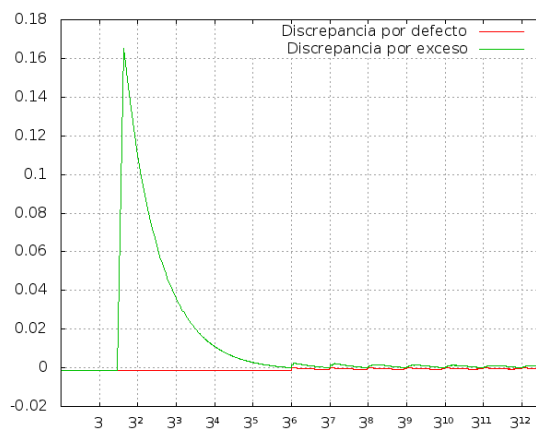
Discrepancia de Ford Extendida (Base 3)
Bloques de tamaño 3



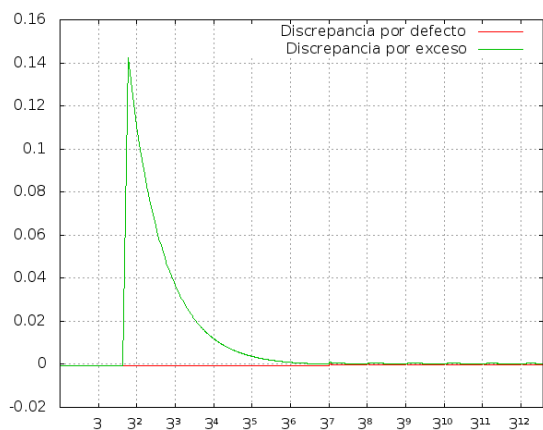
Discrepancia de Ford Extendida (Base 3)
Bloques de tamaño 4



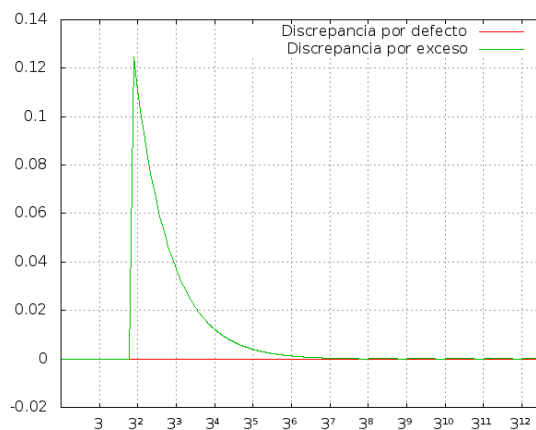
Discrepancia de Ford Extendida (Base 3)
Bloques de tamaño 5



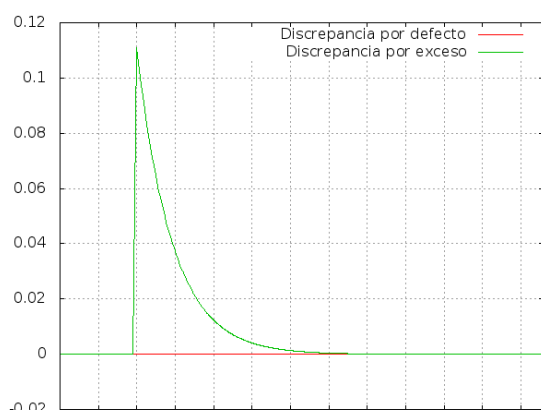
Discrepancia de Ford Extendida (Base 3)
Bloques de tamaño 6



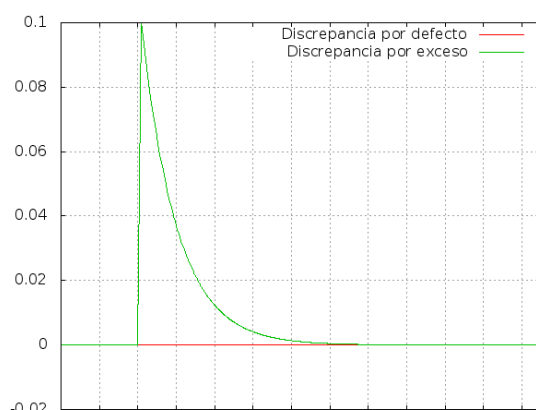
Discrepancia de Ford Extendida (Base 3)
Bloques de tamaño 7



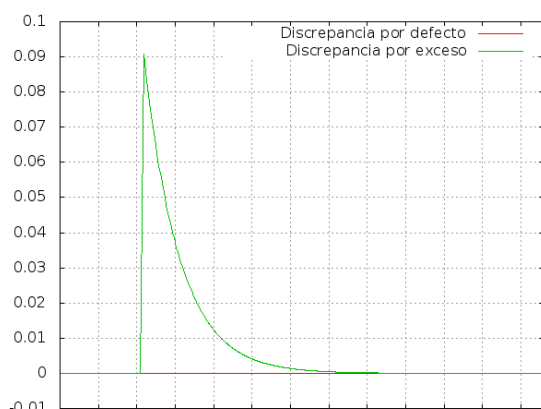
Discrepancia de Ford Extendida (Base 3)
Bloques de tamaño 8



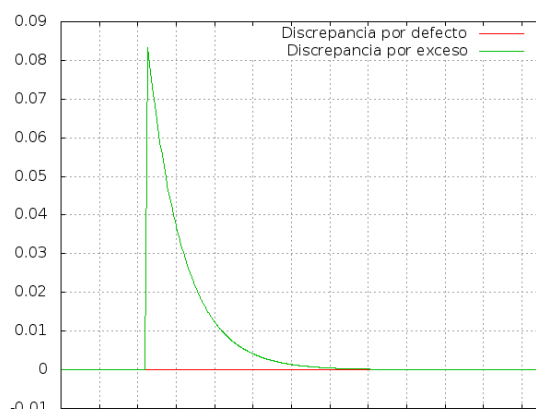
Discrepancia de Ford Extendida (Base 3)
Bloques de tamaño 9



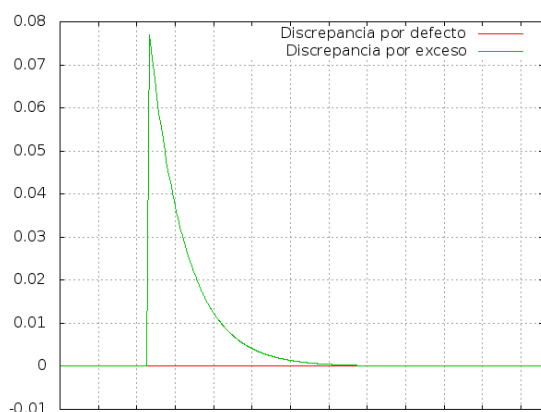
Discrepancia de Ford Extendida (Base 3)
Bloques de tamaño 10



Discrepancia de Ford Extendida (Base 3)
Bloques de tamaño 11

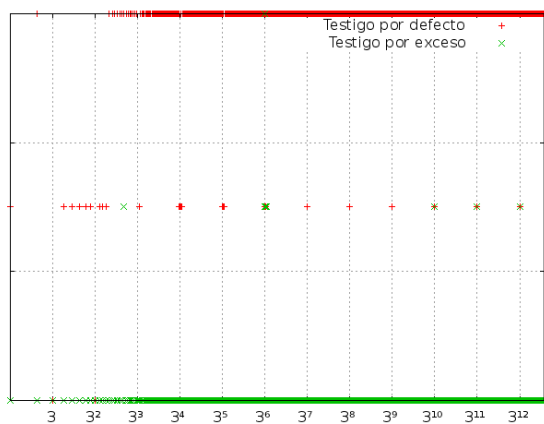


Discrepancia de Ford Extendida (Base 3)
Bloques de tamaño 12

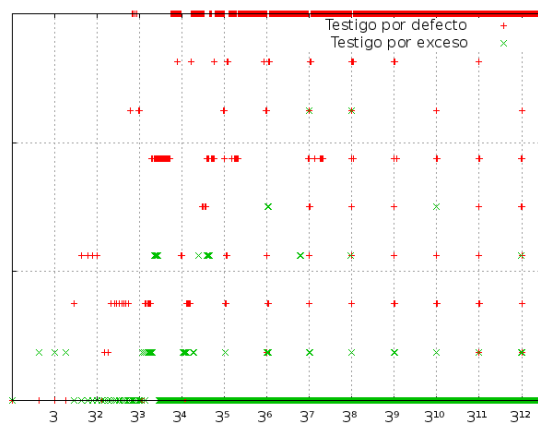


Discrepancia de Ford Extendida (Base 3)
Bloques de tamaño 13

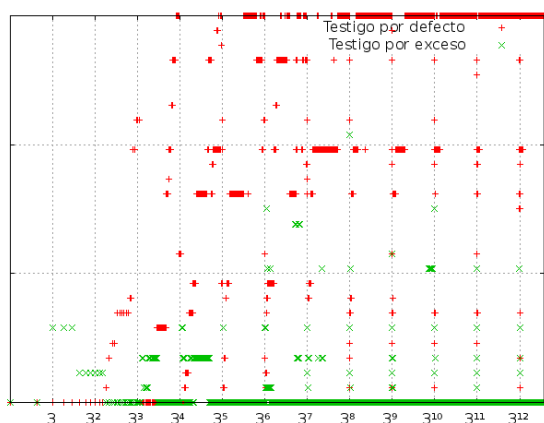
Testigos de Ford Extendida en base 3. Al igual que con alfabeto bináreo, el testigo por exceso (verde) es casi siempre el bloque 0^i y el testigo por defecto (rojo) respeta un patrón que se va repitiendo: Comienza en el bloque 0^i y va aumentando lexicográficamente hasta alcanzar el bloque 1^i .



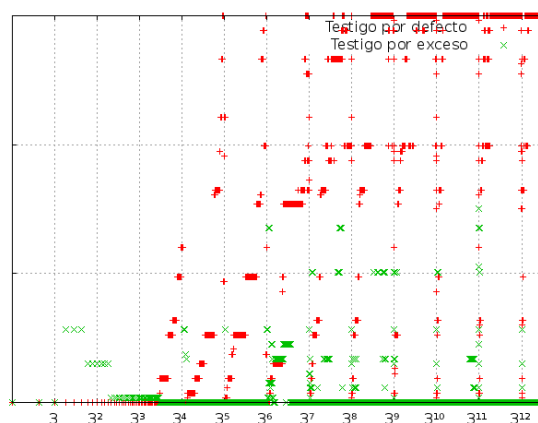
Testigos de Ford Extendida (Base 3)
Bloques de tamaño 1



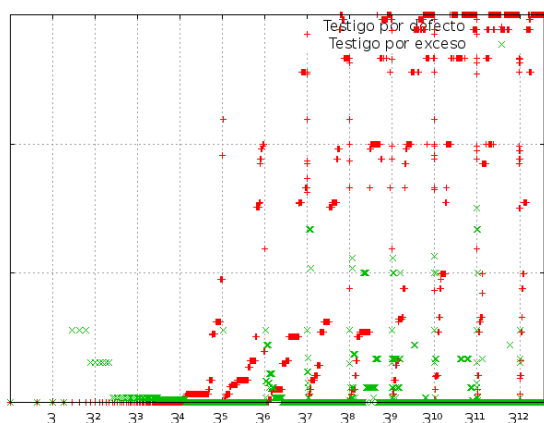
Testigos de Ford Extendida (Base 3)
Bloques de tamaño 2



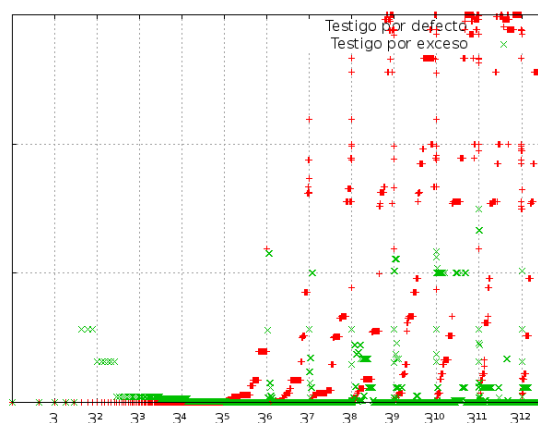
Testigos de Ford Extendida (Base 3)
Bloques de tamaño 3



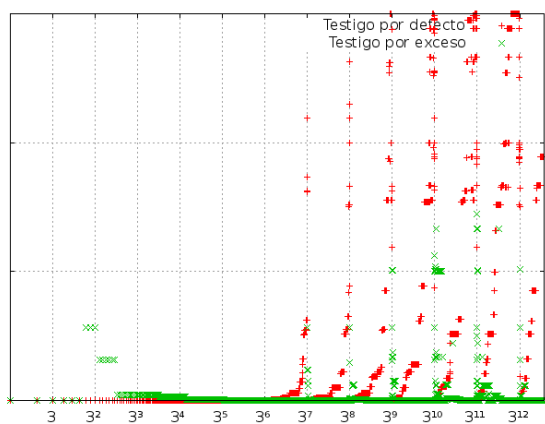
Testigos de Ford Extendida (Base 3)
Bloques de tamaño 4



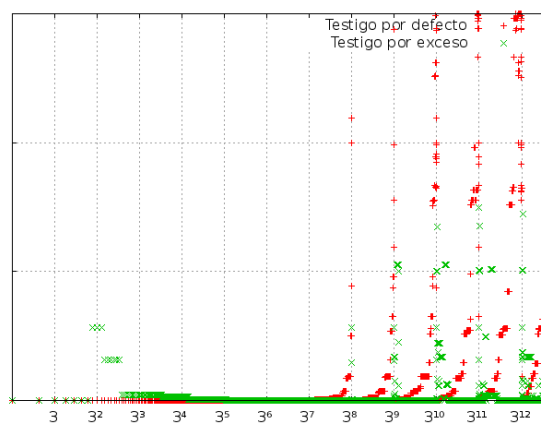
Testigos de Ford Extendida (Base 3)
Bloques de tamaño 5



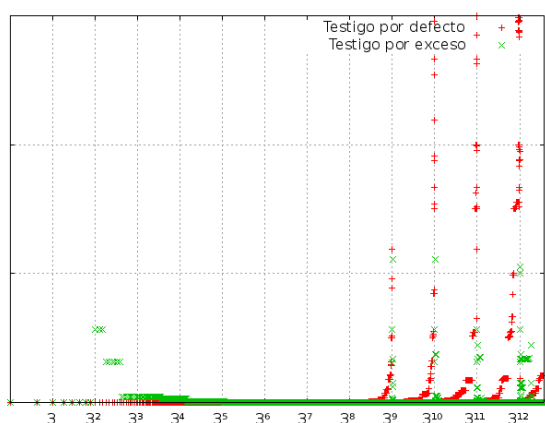
Testigos de Ford Extendida (Base 3)
Bloques de tamaño 6



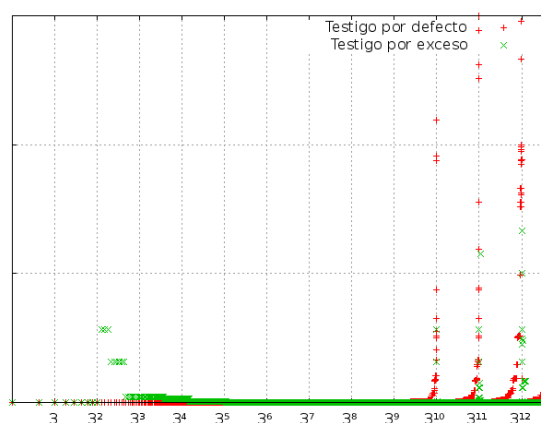
Testigos de Ford Extendida (Base 3)
Bloques de tamaño 7



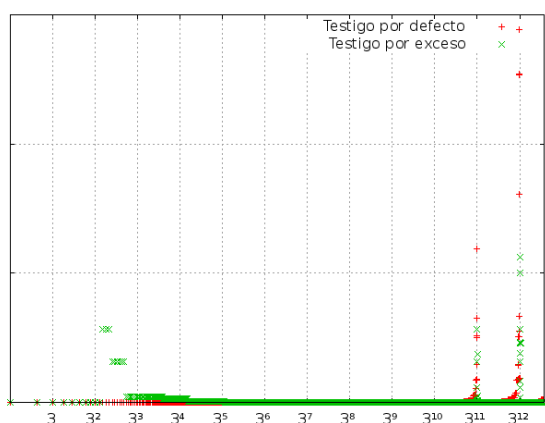
Testigos de Ford Extendida (Base 3)
Bloques de tamaño 8



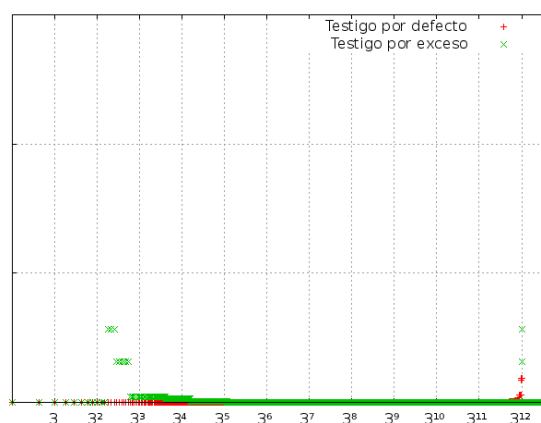
Testigos de Ford Extendida (Base 3)
Bloques de tamaño 9



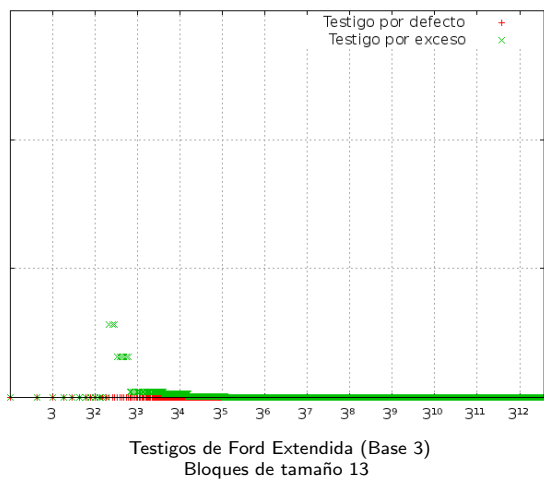
Testigos de Ford Extendida (Base 3)
Bloques de tamaño 10



Testigos de Ford Extendida (Base 3)
Bloques de tamaño 11

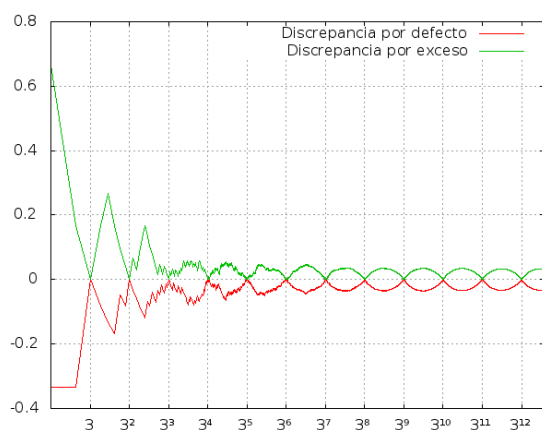


Testigos de Ford Extendida (Base 3)
Bloques de tamaño 12

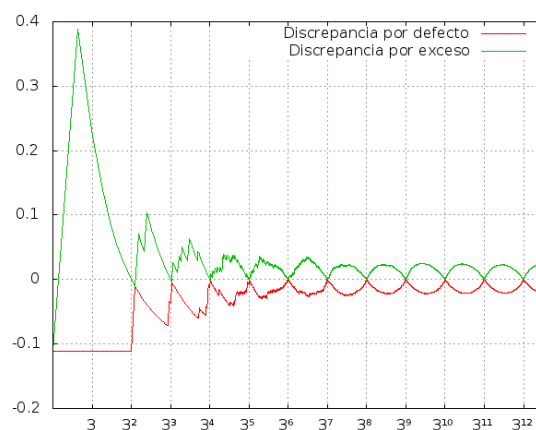


4.2.3. Otras de Bruijn extendidas en base 3

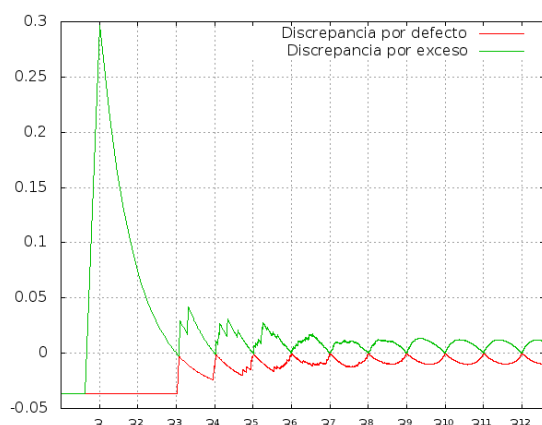
Discrepancia de Bruijn Variante 25 %. Esta variante tiene un comportamiento similar a la lexicográficamente menor, pero un poco menos regular. También presenta un patrón de arcos entre las potencias de 3, pero algunos son más pronunciados o más chatos que sus vecinos.



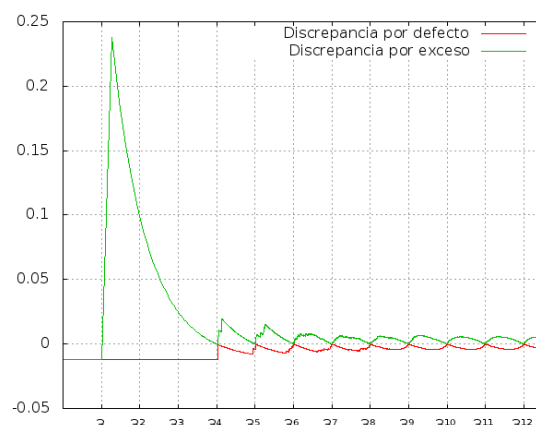
Discrepancia de Bruijn Variante 25 % (Base 3)
Bloques de tamaño 1



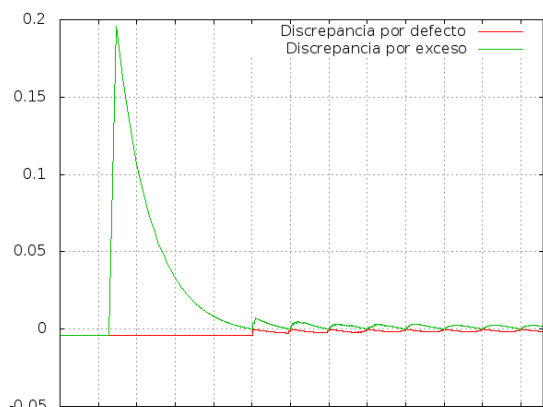
Discrepancia de Bruijn Variante 25 % (Base 3)
Bloques de tamaño 2



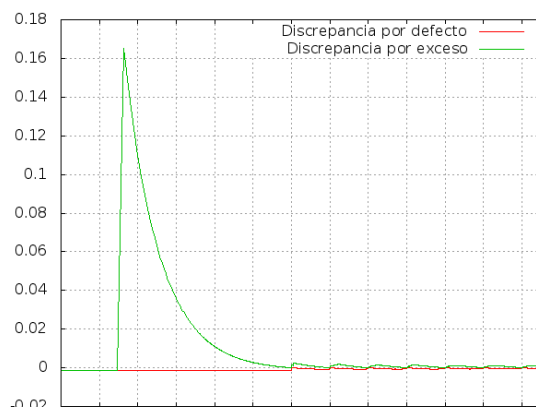
Discrepancia de Bruijn Variante 25 % (Base 3)
Bloques de tamaño 3



Discrepancia de Bruijn Variante 25 % (Base 3)
Bloques de tamaño 4



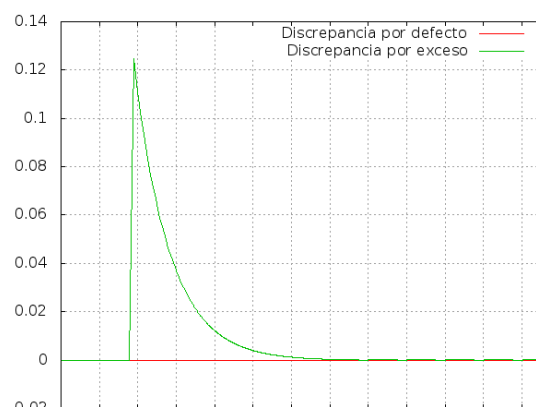
Discrepancia de Bruijn Variante 25 % (Base 3)
Bloques de tamaño 5



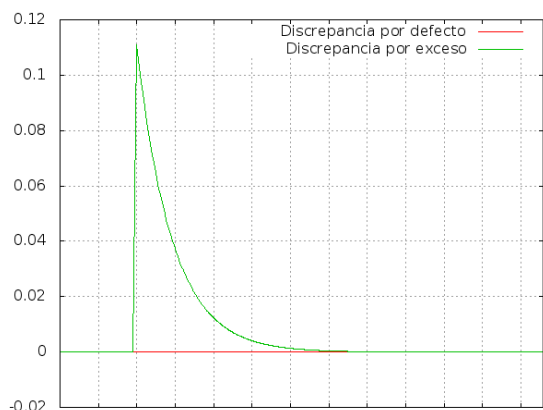
Discrepancia de Bruijn Variante 25 % (Base 3)
Bloques de tamaño 6



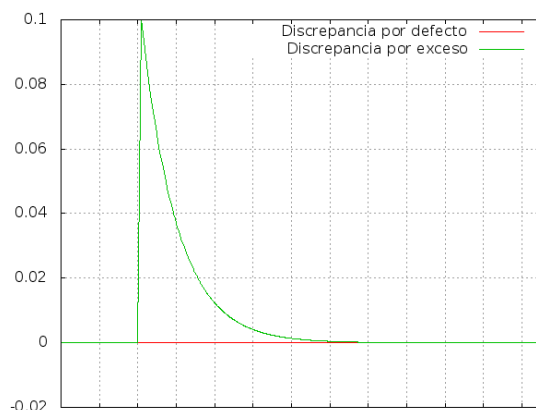
Discrepancia de Bruijn Variante 25 % (Base 3)
Bloques de tamaño 7



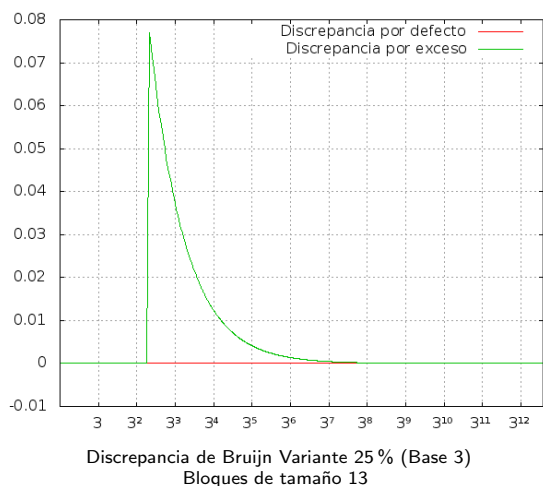
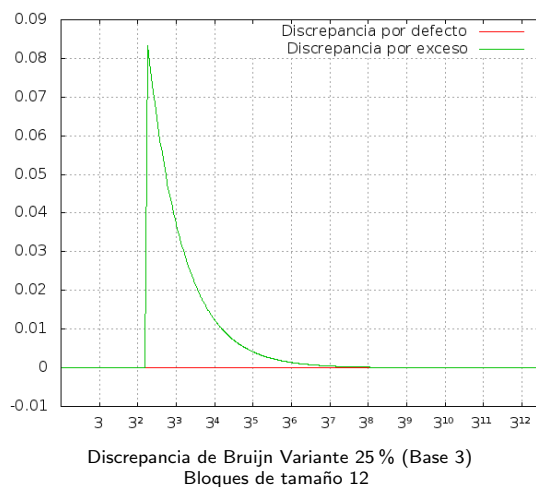
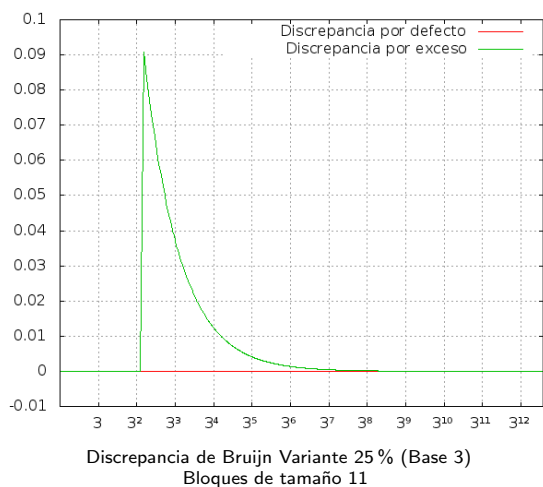
Discrepancia de Bruijn Variante 25 % (Base 3)
Bloques de tamaño 8



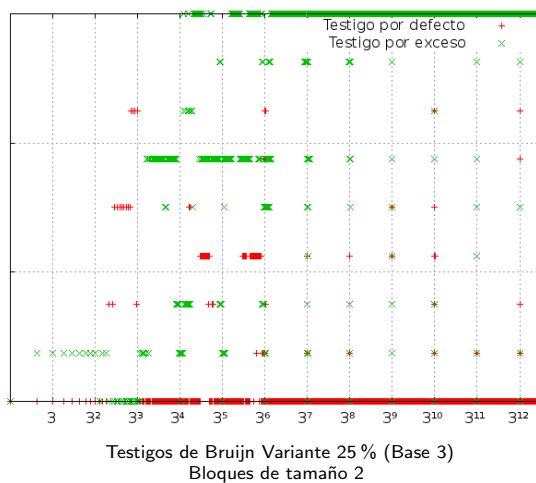
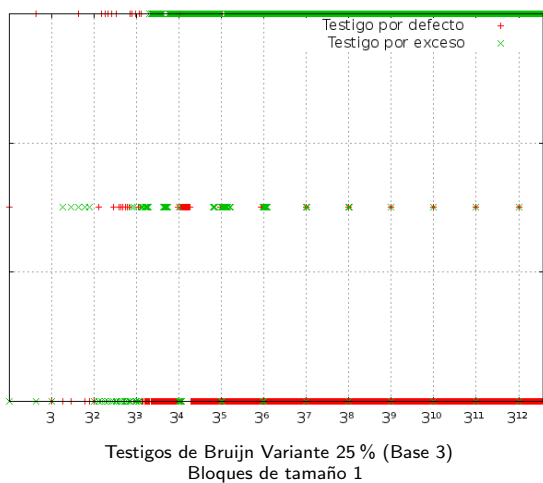
Discrepancia de Bruijn Variante 25 % (Base 3)
Bloques de tamaño 9

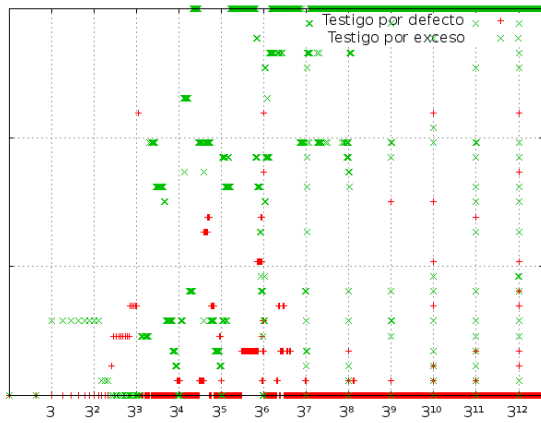


Discrepancia de Bruijn Variante 25 % (Base 3)
Bloques de tamaño 10

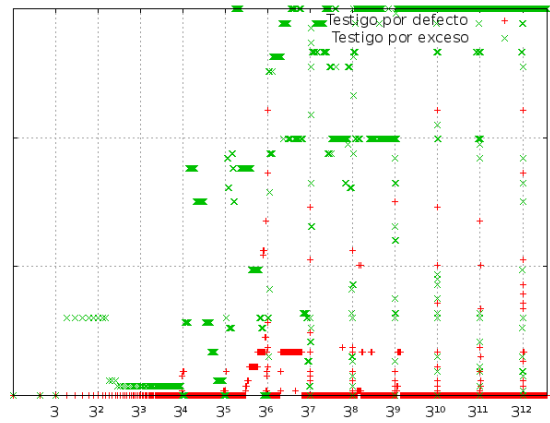


Testigos de Bruijn Variante 25 %. En este caso son los testigos por defecto (rojo) los que se mantienen relativamente constantes, con los testigos por exceso (verde) saltando erráticamente entre varios valores.

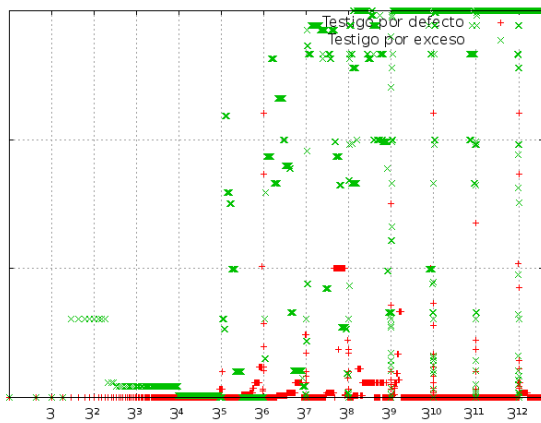




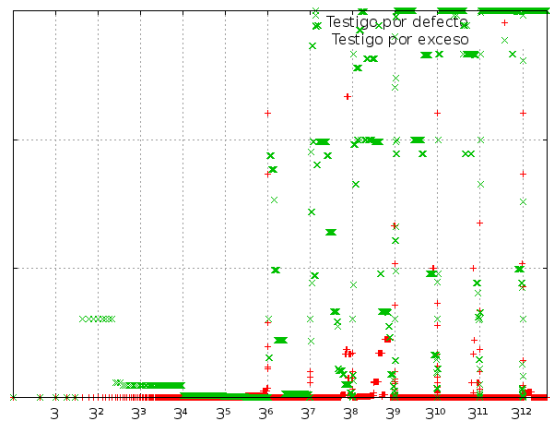
Testigos de Bruijn Variante 25% (Base 3)
Bloques de tamaño 3



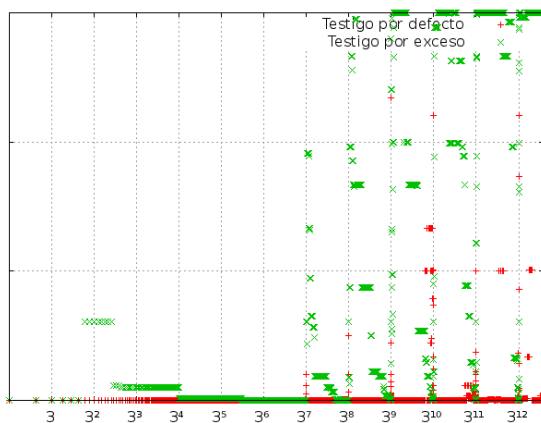
Testigos de Bruijn Variante 25% (Base 3)
Bloques de tamaño 4



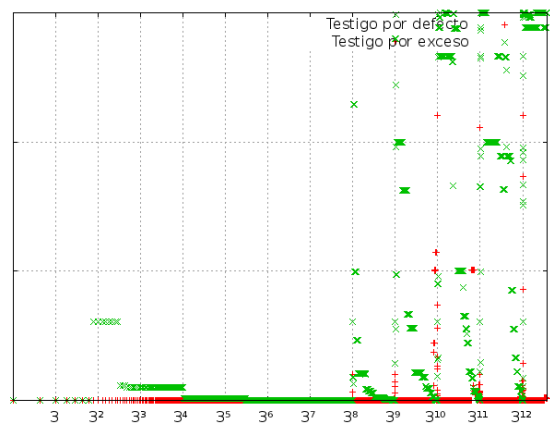
Testigos de Bruijn Variante 25% (Base 3)
Bloques de tamaño 5



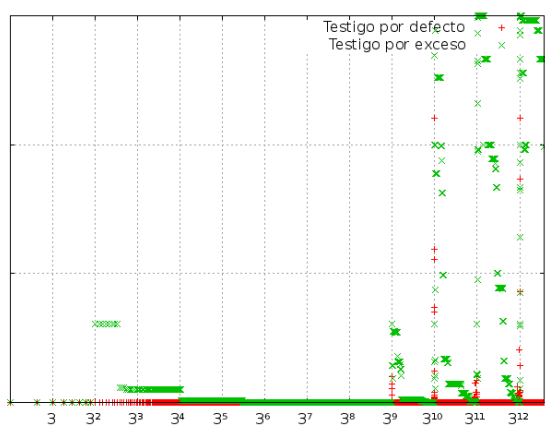
Testigos de Bruijn Variante 25% (Base 3)
Bloques de tamaño 6



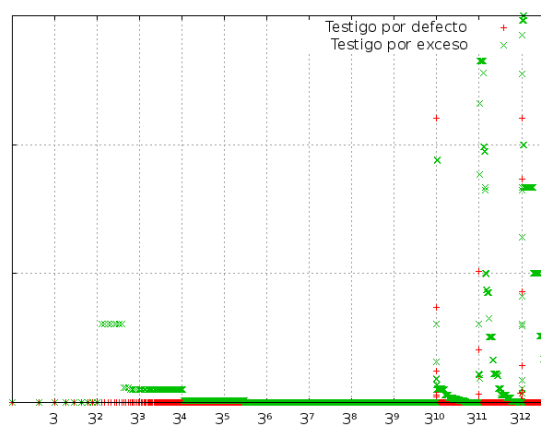
Testigos de Bruijn Variante 25% (Base 3)
Bloques de tamaño 7



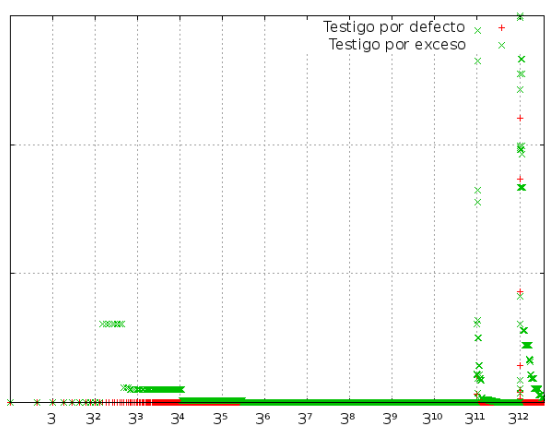
Testigos de Bruijn Variante 25% (Base 3)
Bloques de tamaño 8



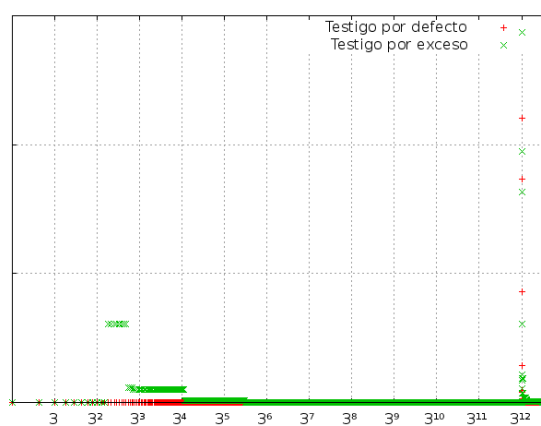
Testigos de Bruijn Variante 25% (Base 3)
Bloques de tamaño 9



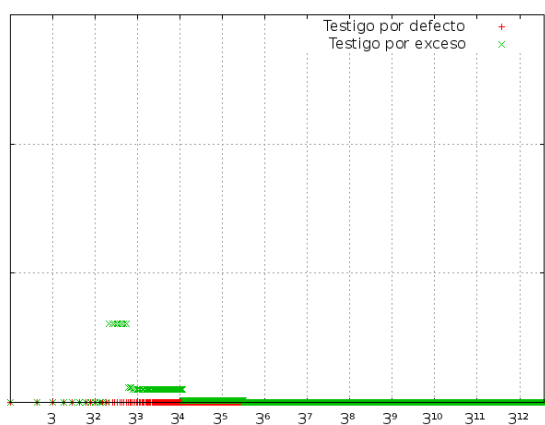
Testigos de Bruijn Variante 25% (Base 3)
Bloques de tamaño 10



Testigos de Bruijn Variante 25% (Base 3)
Bloques de tamaño 11

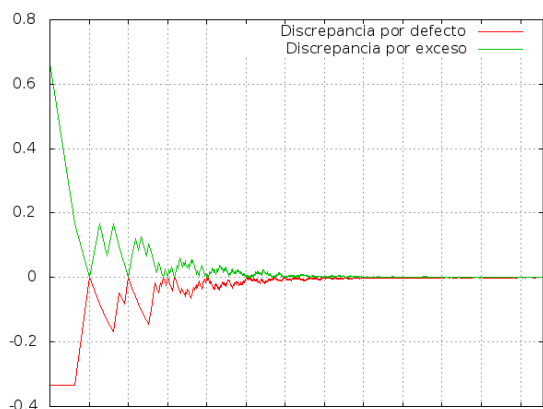


Testigos de Bruijn Variante 25% (Base 3)
Bloques de tamaño 12

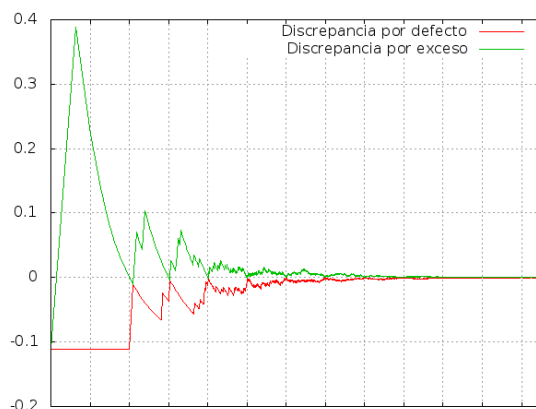


Testigos de Bruijn Variante 25% (Base 3)
Bloques de tamaño 13

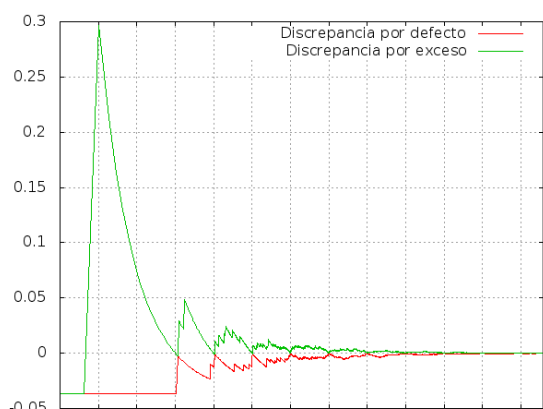
Discrepancia de Bruijn Variante 50 %. Esta variante sigue respetando el patrón de arcos entre potencias de 3, pero mucho más pequeños e irregulares, con muchas variaciones bruscas que le dan una apariencia dentada.



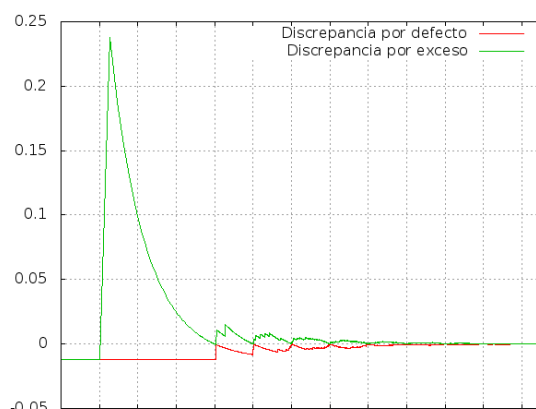
Discrepancia de Bruijn Variante 50% (Base 3)
Bloques de tamaño 1



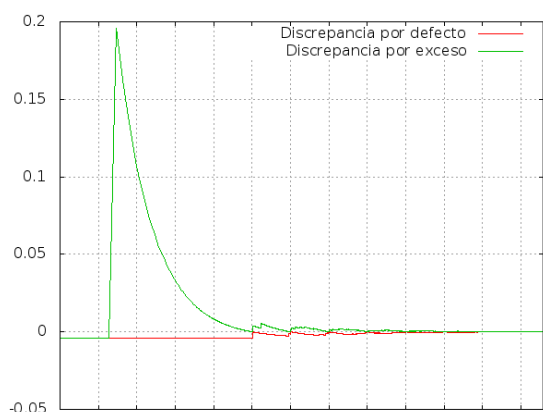
Discrepancia de Bruijn Variante 50% (Base 3)
Bloques de tamaño 2



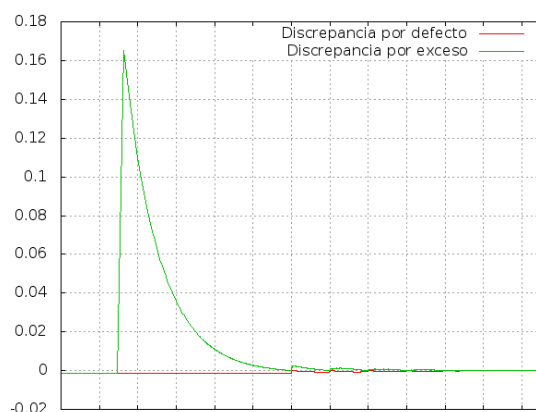
Discrepancia de Bruijn Variante 50% (Base 3)
Bloques de tamaño 3



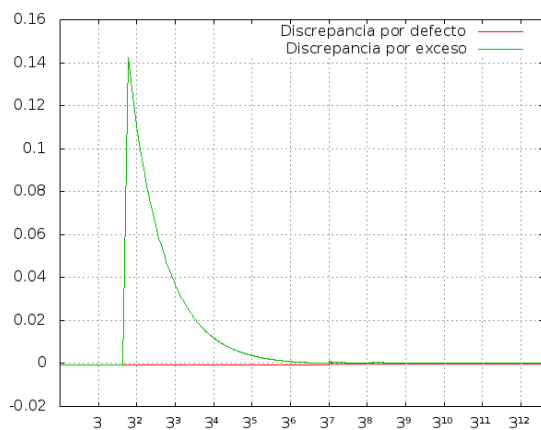
Discrepancia de Bruijn Variante 50% (Base 3)
Bloques de tamaño 4



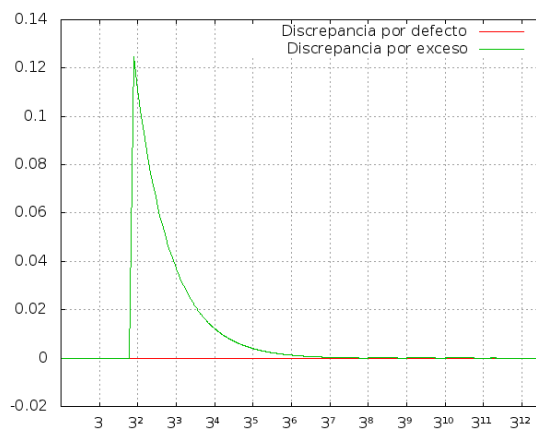
Discrepancia de Bruijn Variante 50% (Base 3)
Bloques de tamaño 5



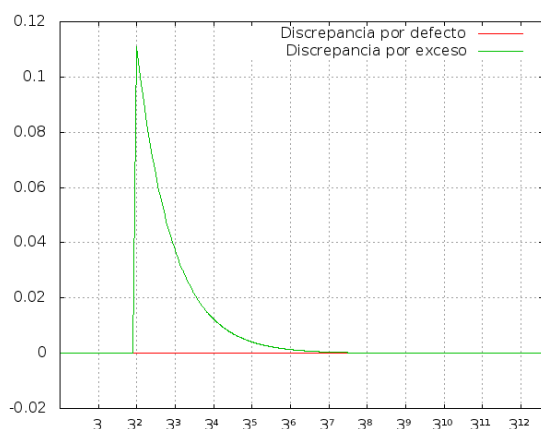
Discrepancia de Bruijn Variante 50% (Base 3)
Bloques de tamaño 6



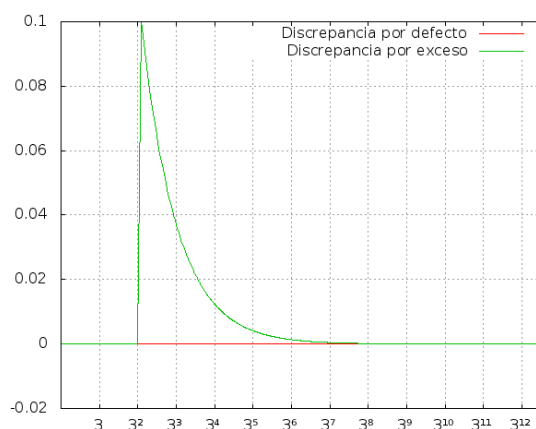
Discrepancia de Bruijn Variante 50% (Base 3)
Bloques de tamaño 7



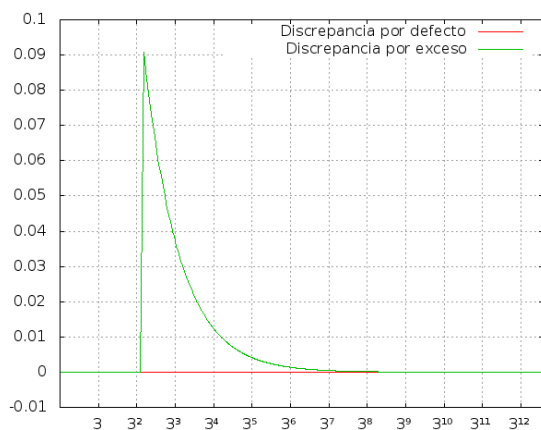
Discrepancia de Bruijn Variante 50% (Base 3)
Bloques de tamaño 8



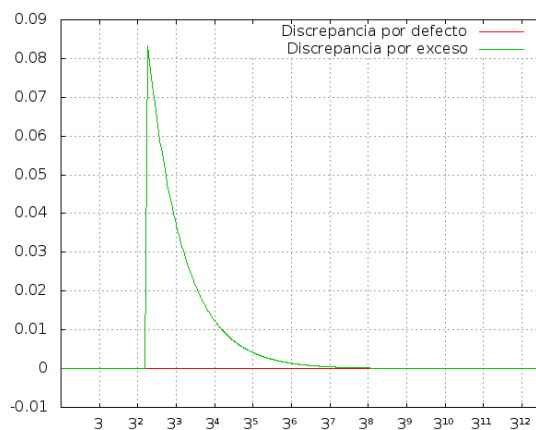
Discrepancia de Bruijn Variante 50% (Base 3)
Bloques de tamaño 9



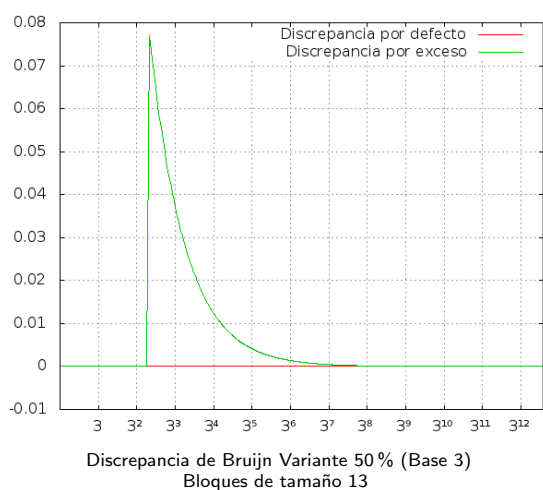
Discrepancia de Bruijn Variante 50% (Base 3)
Bloques de tamaño 10



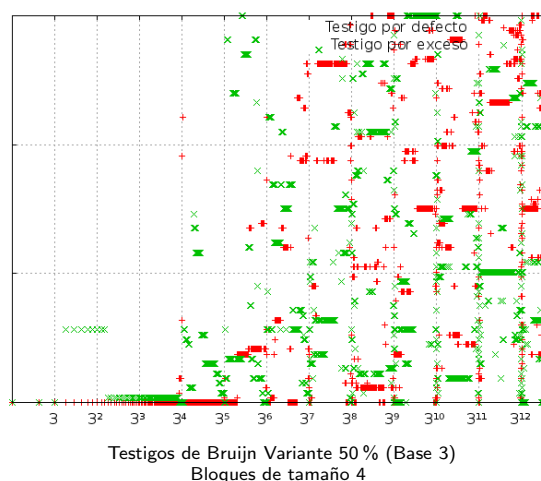
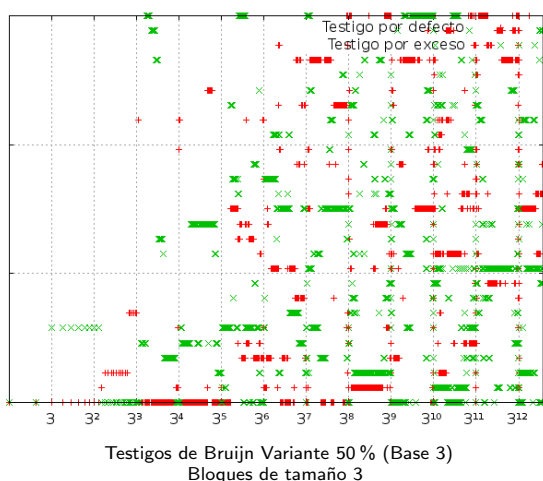
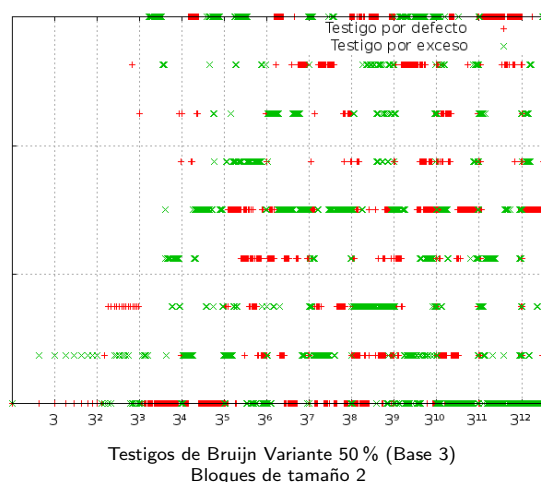
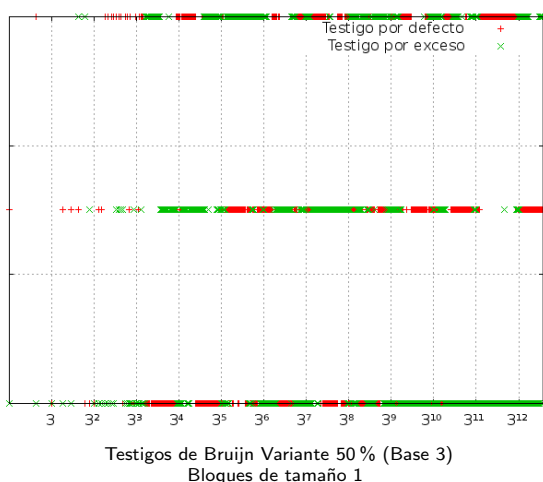
Discrepancia de Bruijn Variante 50% (Base 3)
Bloques de tamaño 11

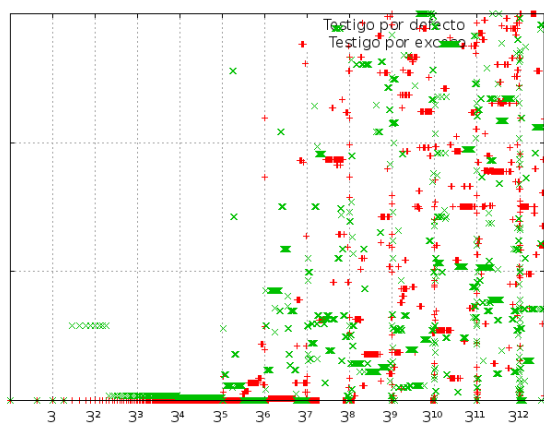


Discrepancia de Bruijn Variante 50% (Base 3)
Bloques de tamaño 12

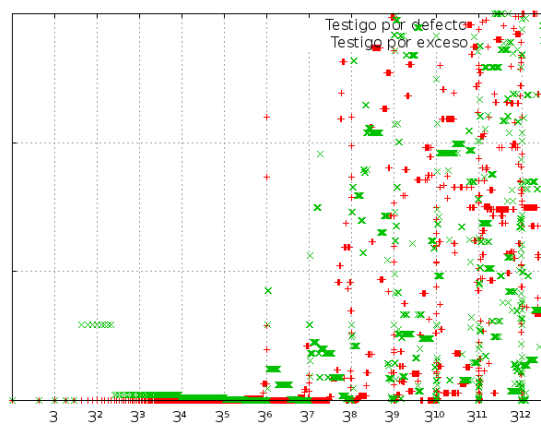


Testigos de Bruijn Variante 50 %. Como era de esperarse mirando la apariencia dentada de la discrepancia, los testigos no siguen ningún tipo de patrón. El plot de los testigos parece un salpicado de pintura, reflejando el factor aleatorio en la generación de esta secuencia.

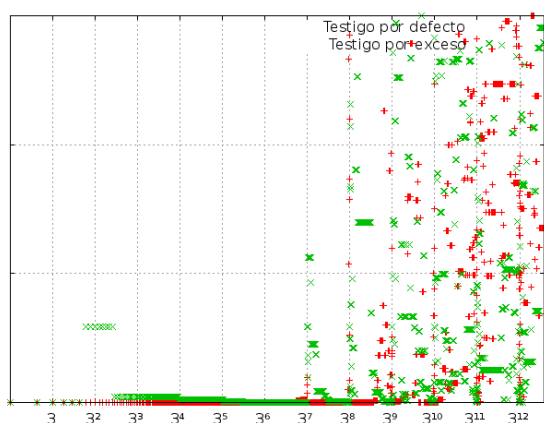




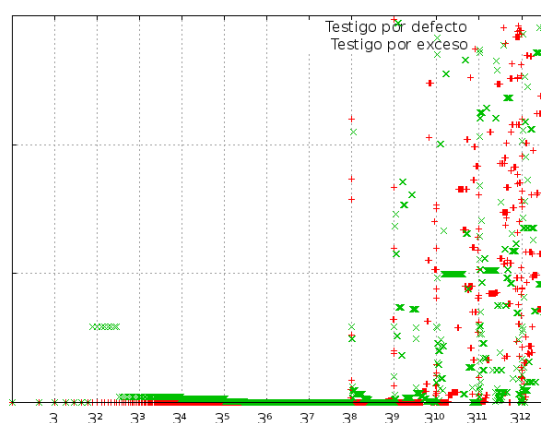
Testigos de Bruijn Variante 50% (Base 3)
Bloques de tamaño 5



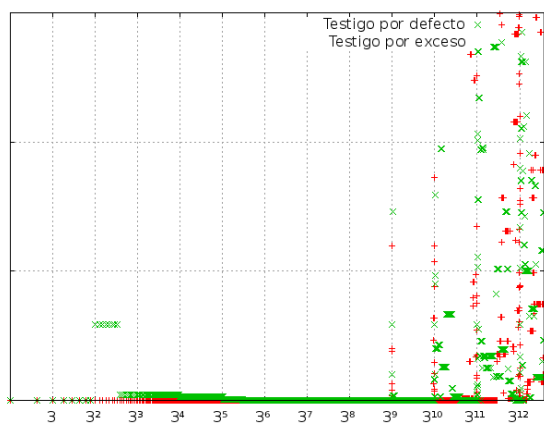
Testigos de Bruijn Variante 50% (Base 3)
Bloques de tamaño 6



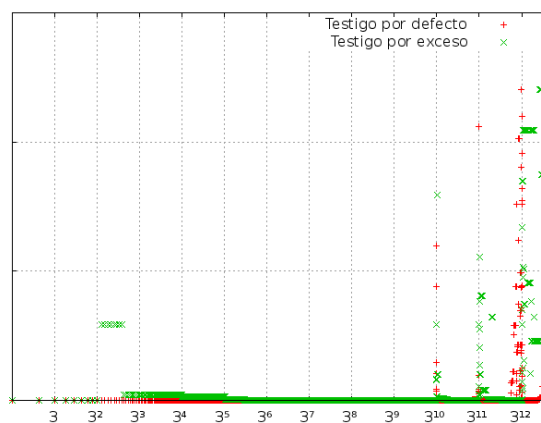
Testigos de Bruijn Variante 50% (Base 3)
Bloques de tamaño 7



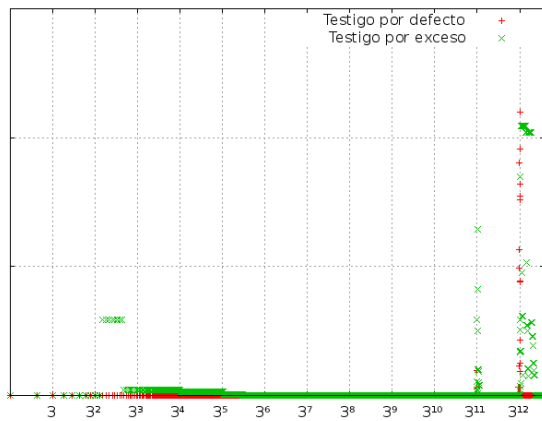
Testigos de Bruijn Variante 50% (Base 3)
Bloques de tamaño 8



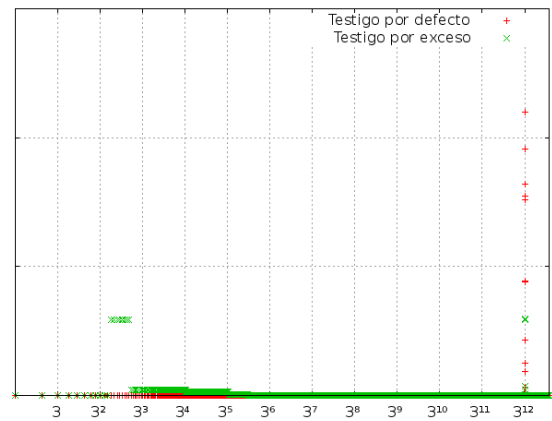
Testigos de Bruijn Variante 50% (Base 3)
Bloques de tamaño 9



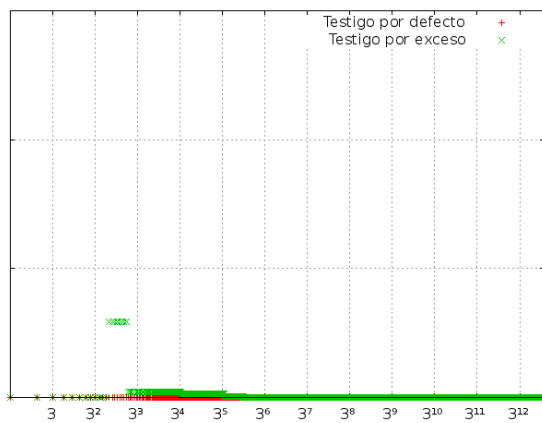
Testigos de Bruijn Variante 50% (Base 3)
Bloques de tamaño 10



Testigos de Bruijn Variante 50% (Base 3)
Bloques de tamaño 11



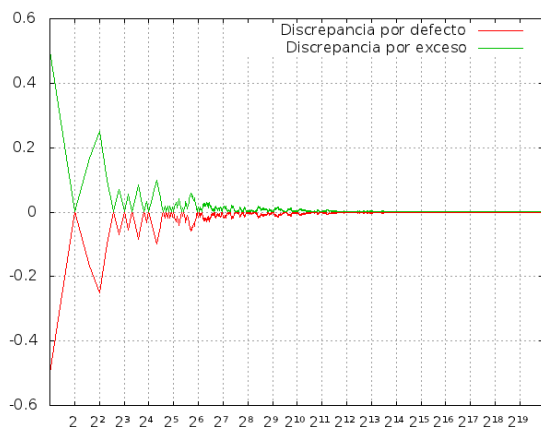
Testigos de Bruijn Variante 50% (Base 3)
Bloques de tamaño 12



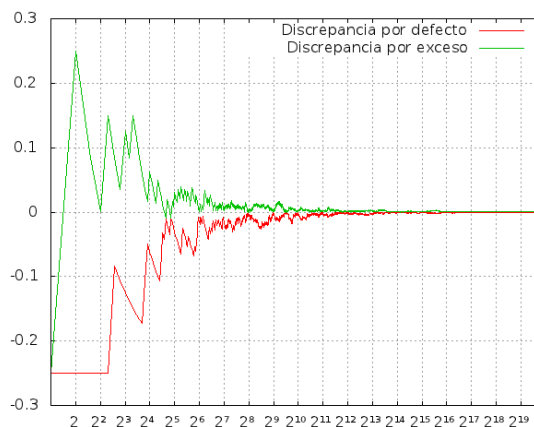
Testigos de Bruijn Variante 50% (Base 3)
Bloques de tamaño 13

4.2.4. Ehrenfeucht-Mycielski

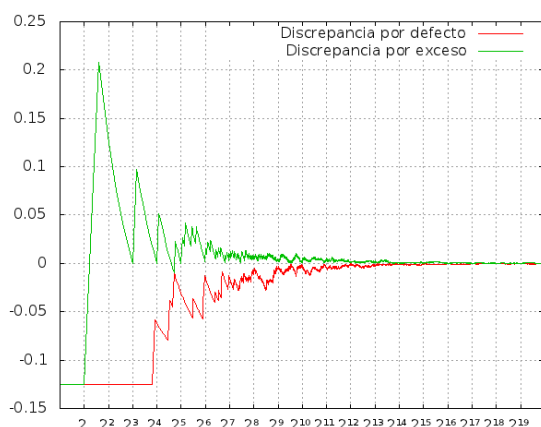
Discrepancia de Ehrenfeucht-Mycielski. La secuencia de Ehrenfeucht-Mycielski es una de las pocas entre las estudiadas que no sigue un patrón en su gráfico. En cambio, converge rápidamente por ambos extremos y presenta cambios bruscos que le hacen que la curva tenga forma de serrucho.



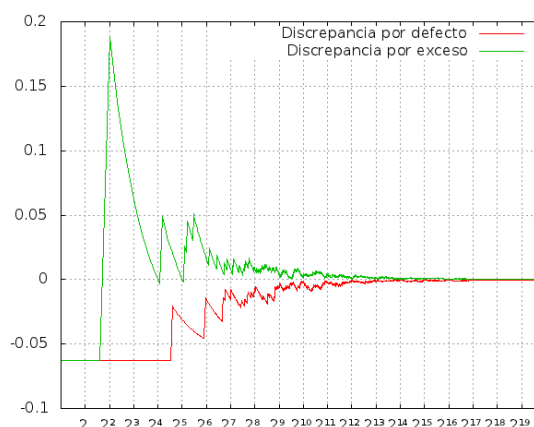
Discrepancia de Ehrenfeucht-Mycielski (Base 2)
Bloques de tamaño 1



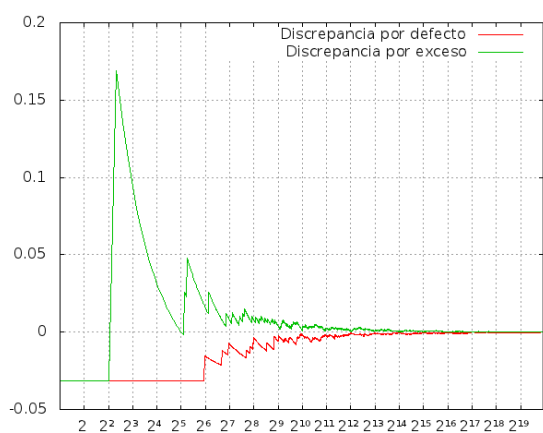
Discrepancia de Ehrenfeucht-Mycielski (Base 2)
Bloques de tamaño 2



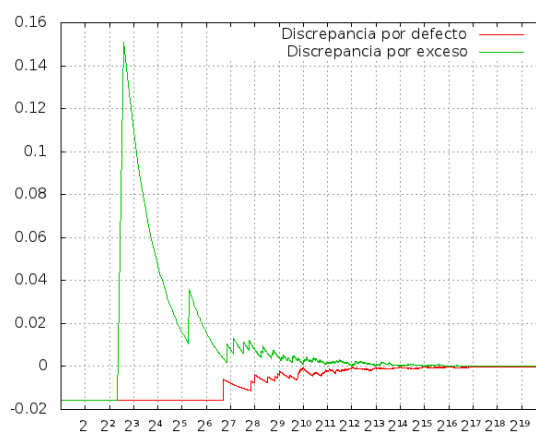
Discrepancia de Ehrenfeucht-Mycielski (Base 2)
Bloques de tamaño 3



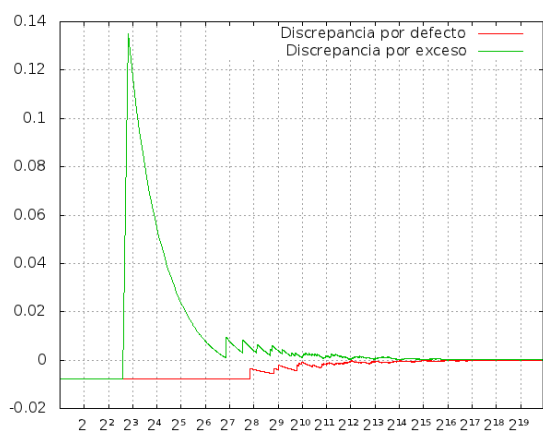
Discrepancia de Ehrenfeucht-Mycielski (Base 2)
Bloques de tamaño 4



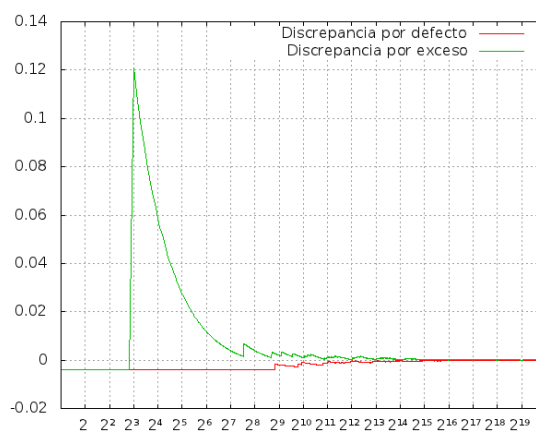
Discrepancia de Ehrenfeucht-Mycielski (Base 2)
Bloques de tamaño 5



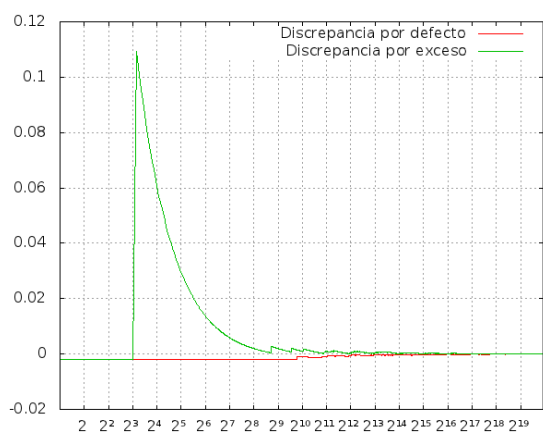
Discrepancia de Ehrenfeucht-Mycielski (Base 2)
Bloques de tamaño 6



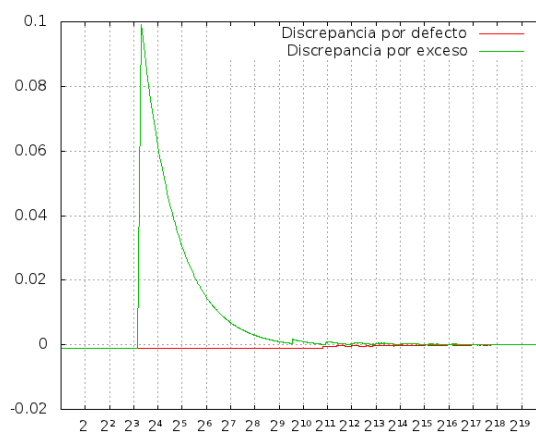
Discrepancia de Ehrenfeucht-Mycielski (Base 2)
Bloques de tamaño 7



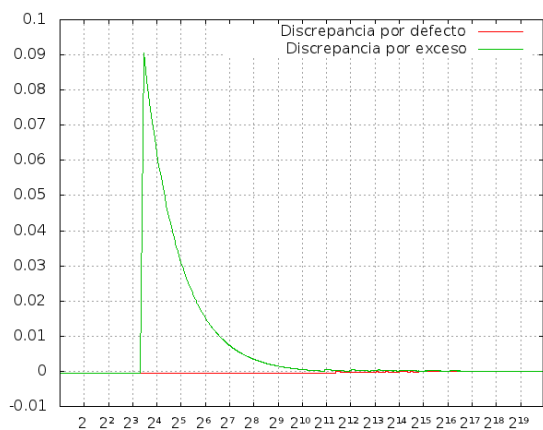
Discrepancia de Ehrenfeucht-Mycielski (Base 2)
Bloques de tamaño 8



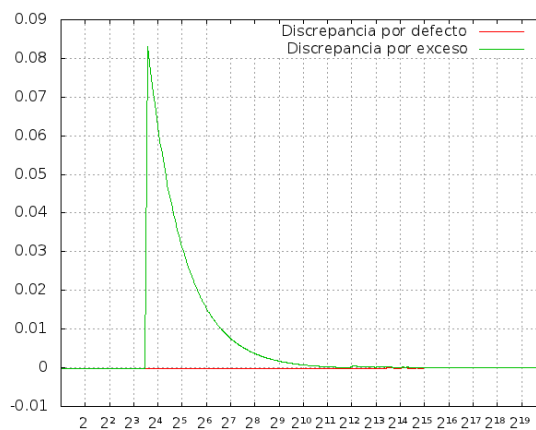
Discrepancia de Ehrenfeucht-Mycielski (Base 2)
Bloques de tamaño 9



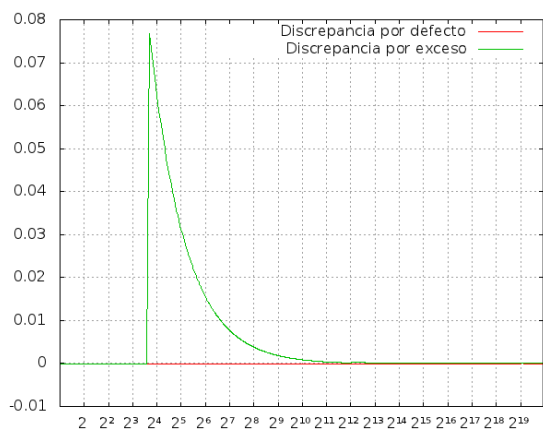
Discrepancia de Ehrenfeucht-Mycielski (Base 2)
Bloques de tamaño 10



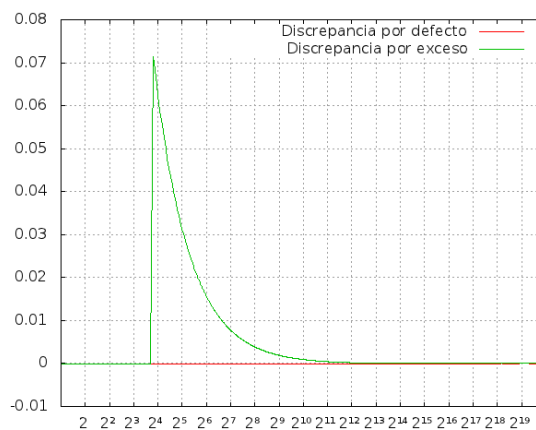
Discrepancia de Ehrenfeucht-Mycielski (Base 2)
Bloques de tamaño 11



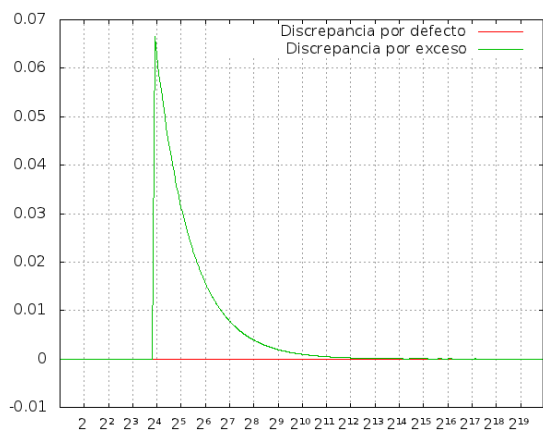
Discrepancia de Ehrenfeucht-Mycielski (Base 2)
Bloques de tamaño 12



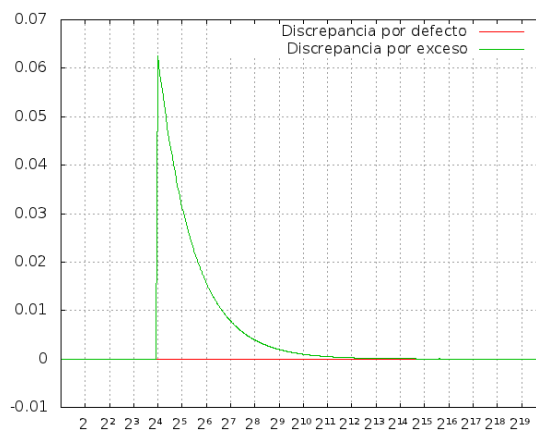
Discrepancia de Ehrenfeucht-Mycielski (Base 2)
Bloques de tamaño 13



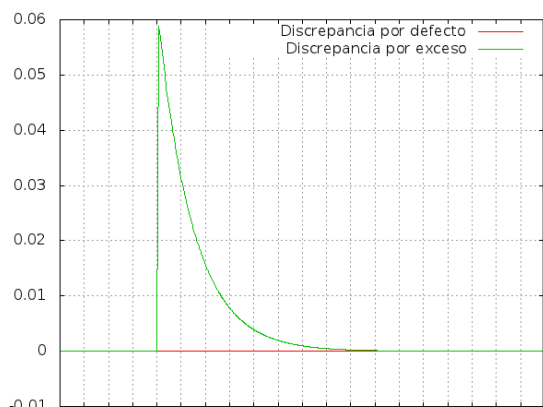
Discrepancia de Ehrenfeucht-Mycielski (Base 2)
Bloques de tamaño 14



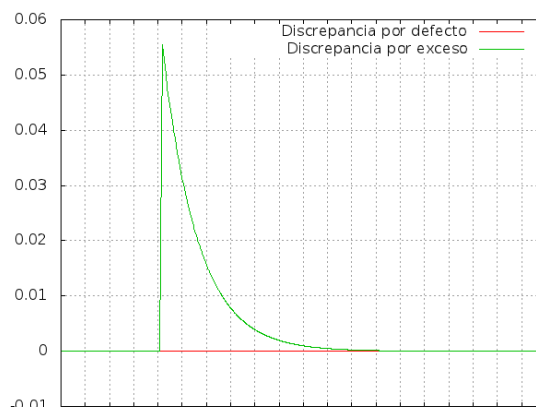
Discrepancia de Ehrenfeucht-Mycielski (Base 2)
Bloques de tamaño 15



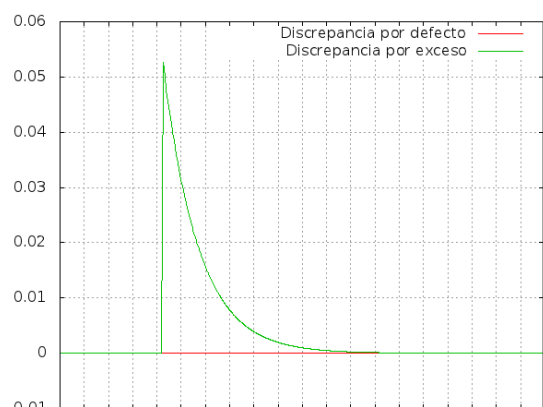
Discrepancia de Ehrenfeucht-Mycielski (Base 2)
Bloques de tamaño 16



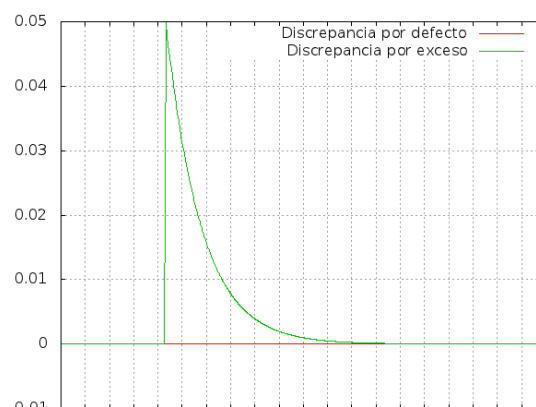
Discrepancia de Ehrenfeucht-Mycielski (Base 2)
Bloques de tamaño 17



Discrepancia de Ehrenfeucht-Mycielski (Base 2)
Bloques de tamaño 18

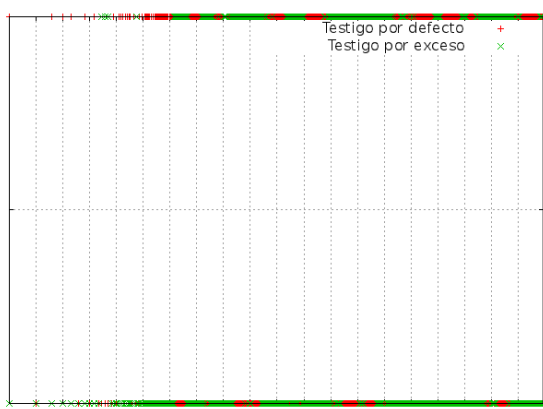


Discrepancia de Ehrenfeucht-Mycielski (Base 2)
Bloques de tamaño 19

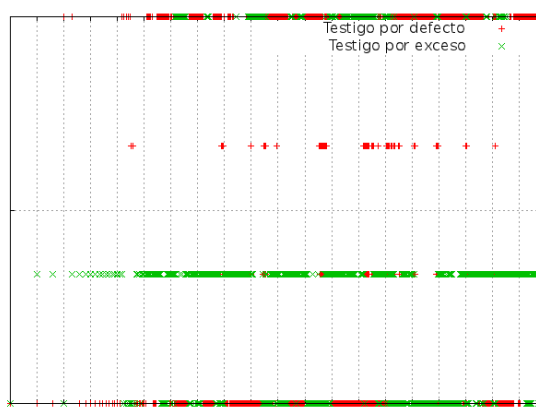


Discrepancia de Ehrenfeucht-Mycielski (Base 2)
Bloques de tamaño 20

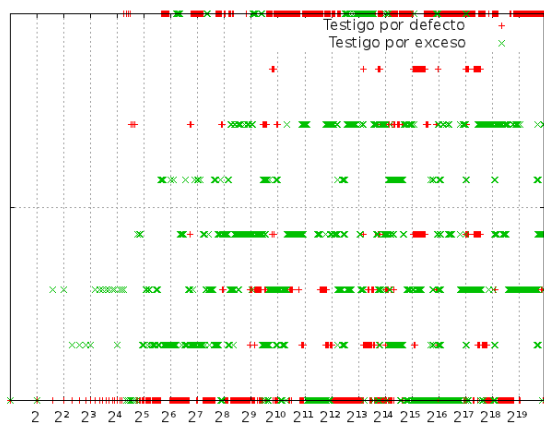
Testigos de Ehrenfeucht-Mycielski. Al igual que las variantes más aleatorias de las familias estudiadas, los testigos cambian constantemente de un paso a otro, explicando la forma de serrucho de la curva de discrepancia.



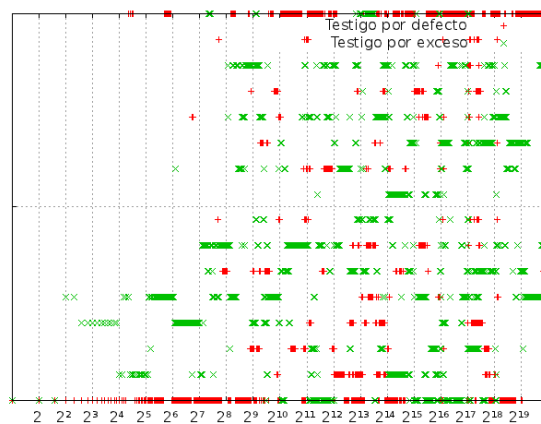
Testigos de Ehrenfeucht-Mycielski (Base 2)
Bloques de tamaño 1



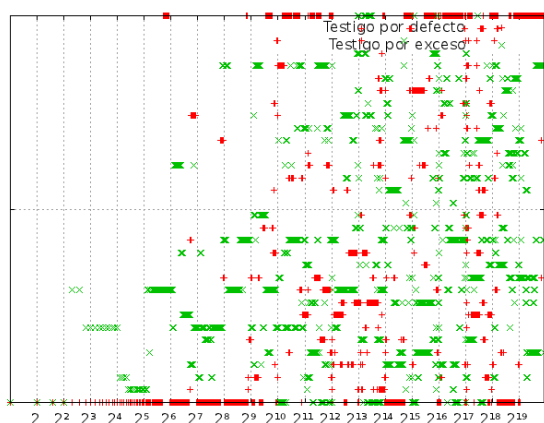
Testigos de Ehrenfeucht-Mycielski (Base 2)
Bloques de tamaño 2



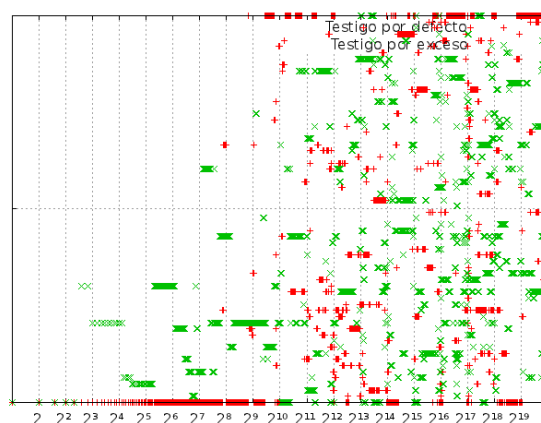
Testigos de Ehrenfeucht-Mycielski (Base 2)
Bloques de tamaño 3



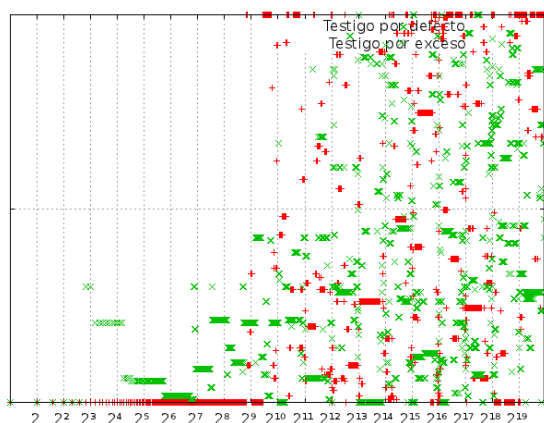
Testigos de Ehrenfeucht-Mycielski (Base 2)
Bloques de tamaño 4



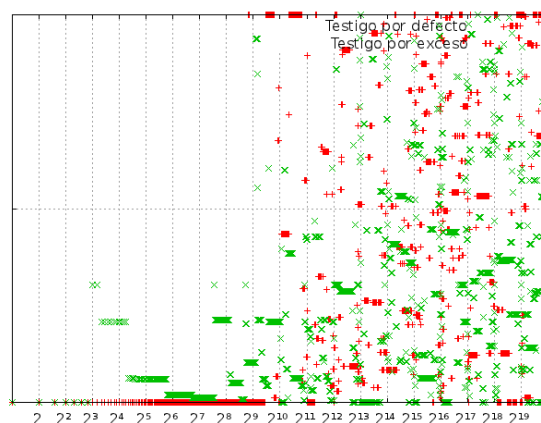
Testigos de Ehrenfeucht-Mycielski (Base 2)
Bloques de tamaño 5



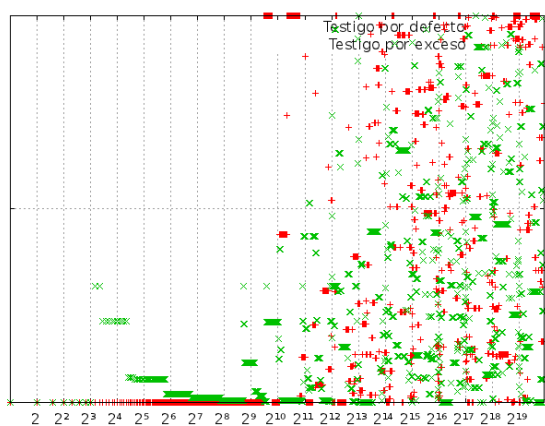
Testigos de Ehrenfeucht-Mycielski (Base 2)
Bloques de tamaño 6



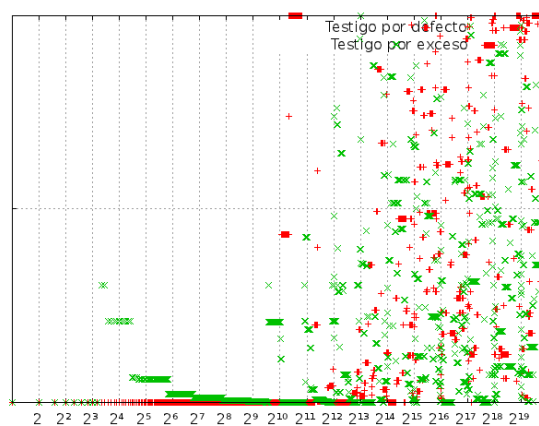
Testigos de Ehrenfeucht-Mycielski (Base 2)
Bloques de tamaño 7



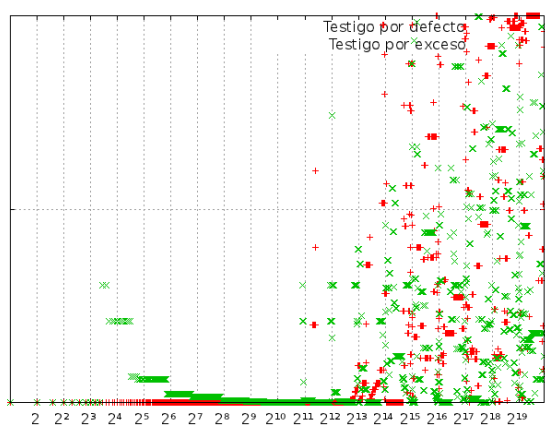
Testigos de Ehrenfeucht-Mycielski (Base 2)
Bloques de tamaño 8



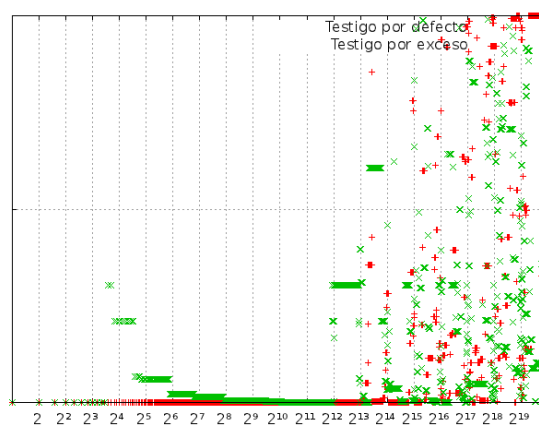
Testigos de Ehrenfeucht-Mycielski (Base 2)
Bloques de tamaño 9



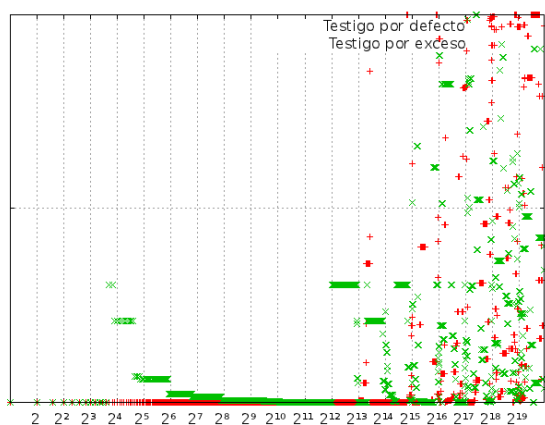
Testigos de Ehrenfeucht-Mycielski (Base 2)
Bloques de tamaño 10



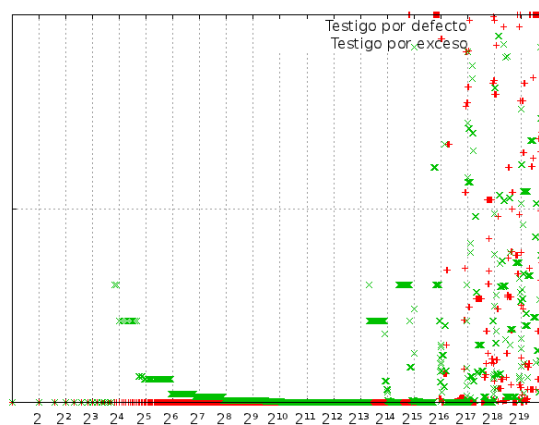
Testigos de Ehrenfeucht-Mycielski (Base 2)
Bloques de tamaño 11



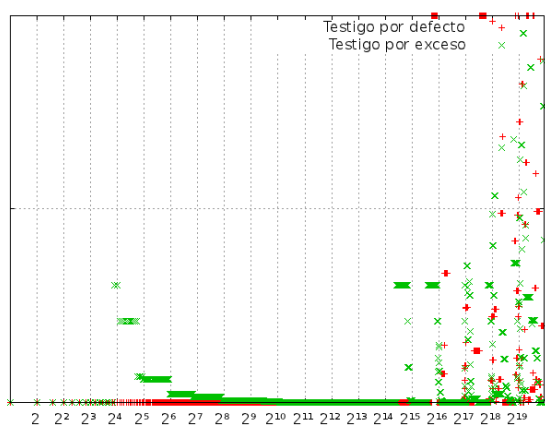
Testigos de Ehrenfeucht-Mycielski (Base 2)
Bloques de tamaño 12



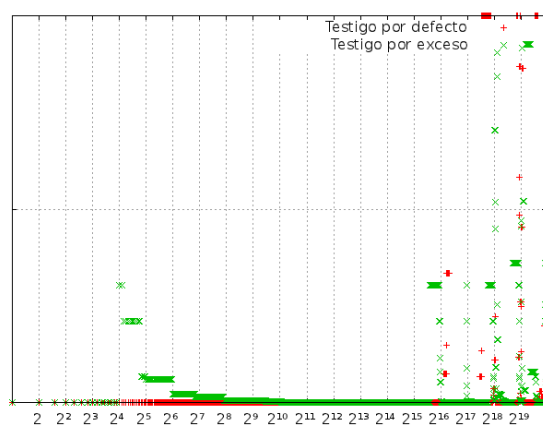
Testigos de Ehrenfeucht-Mycielski (Base 2)
Bloques de tamaño 13



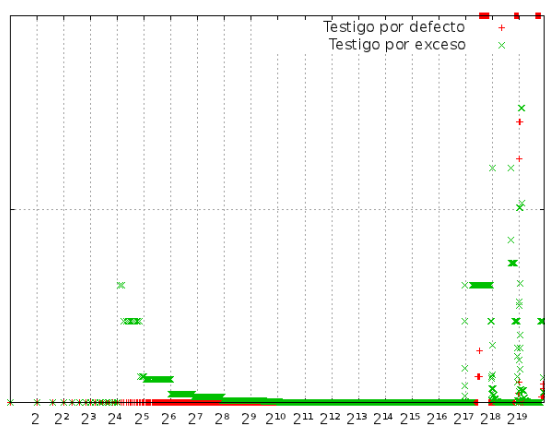
Testigos de Ehrenfeucht-Mycielski (Base 2)
Bloques de tamaño 14



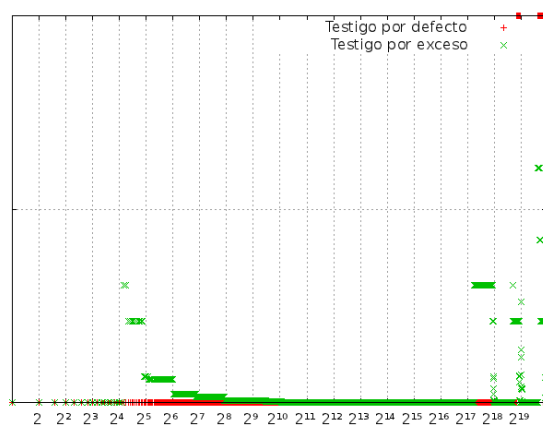
Testigos de Ehrenfeucht-Mycielski (Base 2)
Bloques de tamaño 15



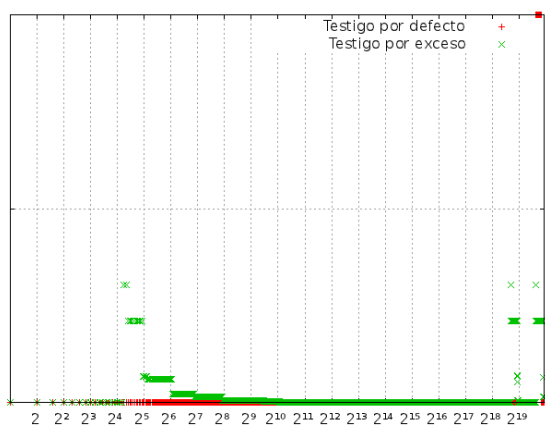
Testigos de Ehrenfeucht-Mycielski (Base 2)
Bloques de tamaño 16



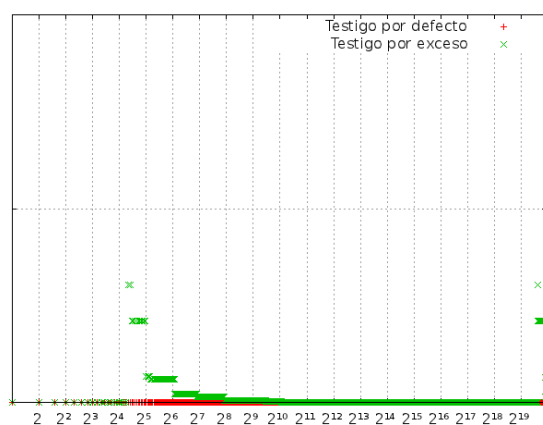
Testigos de Ehrenfeucht-Mycielski (Base 2)
Bloques de tamaño 17



Testigos de Ehrenfeucht-Mycielski (Base 2)
Bloques de tamaño 18



Testigos de Ehrenfeucht-Mycielski (Base 2)
Bloques de tamaño 19



Testigos de Ehrenfeucht-Mycielski (Base 2)
Bloques de tamaño 20

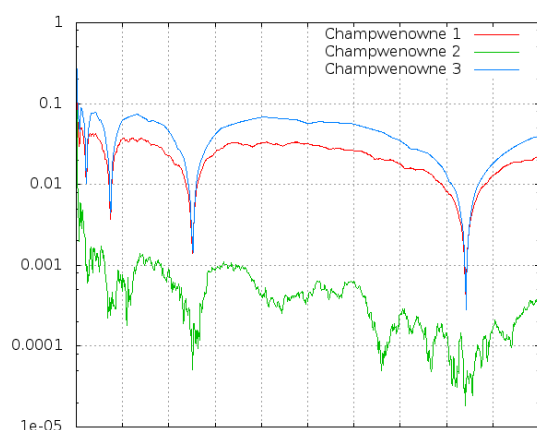
4.3. Resultados de comparación

Ahora veremos cómo se comparan entre sí las discrepancias de las diferentes secuencias estudiadas. Los siguientes gráficos muestran la discrepancia de normalidad para cada tamaño de bloque. No se discrimina por exceso y defecto, sino que se toma la mayor de ambas como en la definición tradicional. El eje x representa las posiciones en la secuencia, en escala lineal. El eje y indica el valor de la discrepancia para la posición correspondiente en escala logarítmica para resaltar las diferencias para los valores cercanos a 0 que se alcanzan en las colas de las funciones.

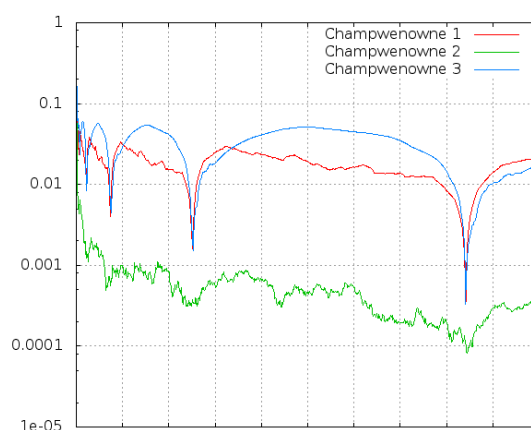
Con el objetivo de suavizar las curvas, se graficaron sólo mil puntos en cada una. Los puntos elegidos fueron aquellos que maximizan la discrepancia para cada intervalo de 1000 posiciones. Para los bloques más grandes, es posible que dos líneas se superpongan y una de ellas quede parcial o totalmente oculta.

4.3.1. Todas las variantes de Champernowne tienen la misma discrepancia?

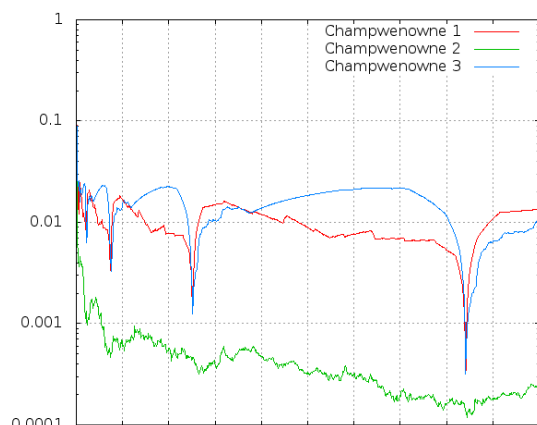
La variante aleatoria de Champernowne tiene una discrepancia un orden de magnitud menor que las otras dos variantes, que tienen una magnitud comparable. La variante 3, que fue construida con el objetivo de maximizar la discrepancia, efectivamente resultó tener una discrepancia ligeramente mayor a la variante ordenada lexicográficamente.



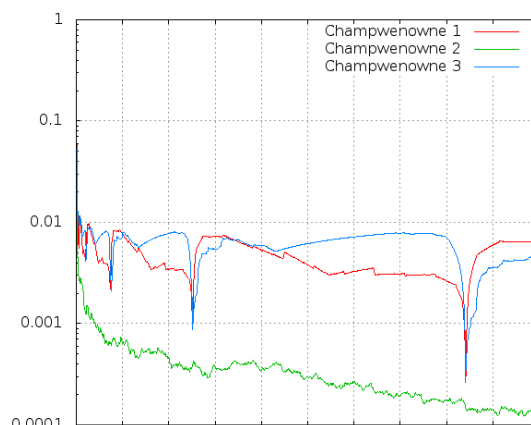
Comparación entre variantes de Champernowne (Base 3)
Bloques de tamaño 1



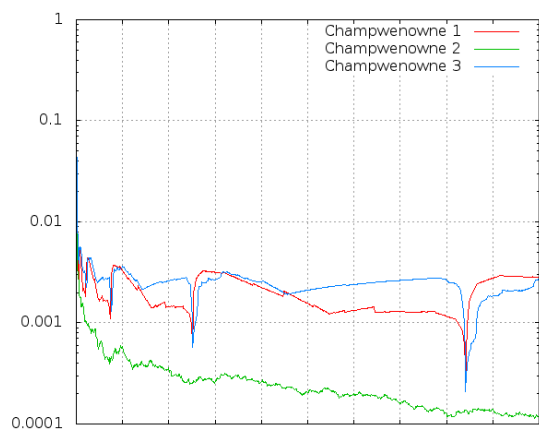
Comparación entre variantes de Champernowne (Base 3)
Bloques de tamaño 2



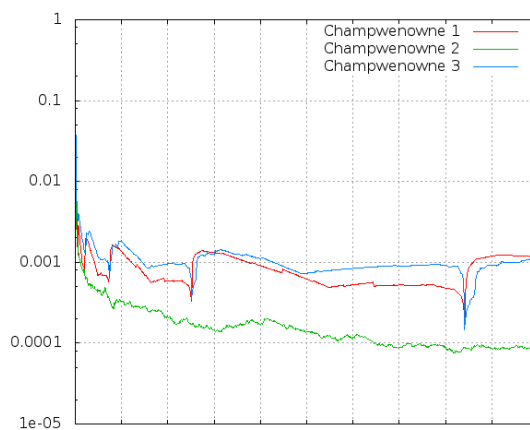
Comparación entre variantes de Champernowne (Base 3)
Bloques de tamaño 3



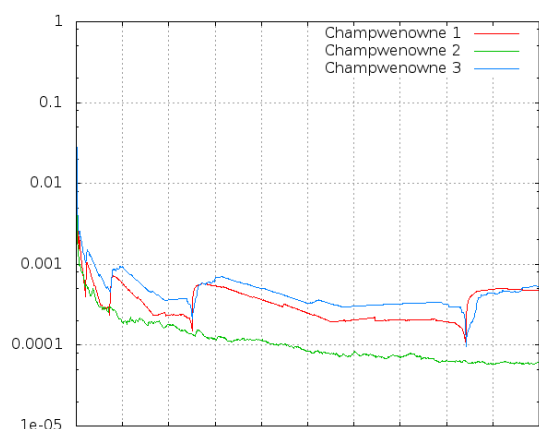
Comparación entre variantes de Champernowne (Base 3)
Bloques de tamaño 4



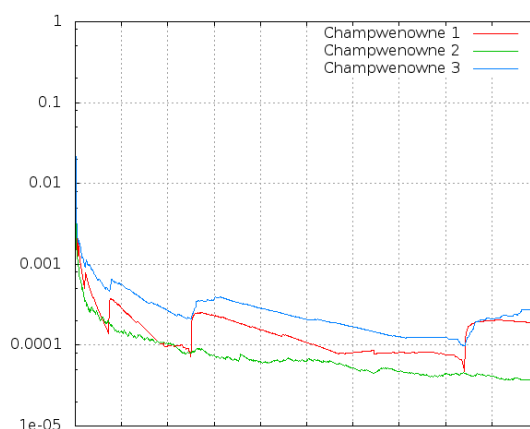
Comparación entre variantes de Champwenowne (Base 3)
Bloques de tamaño 5



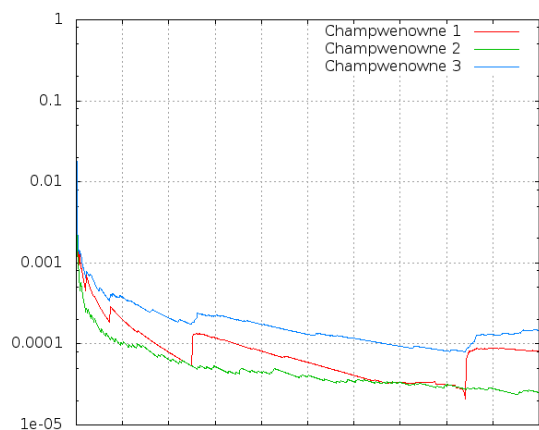
Comparación entre variantes de Champwenowne (Base 3)
Bloques de tamaño 6



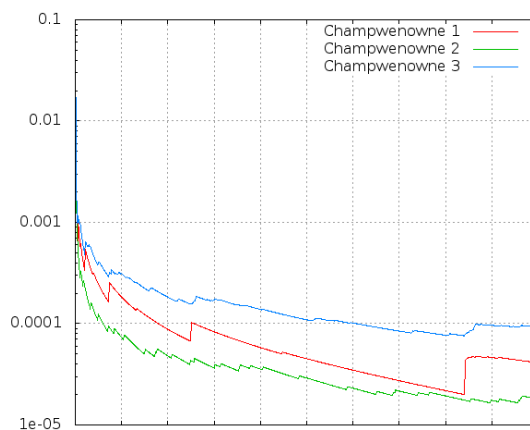
Comparación entre variantes de Champwenowne (Base 3)
Bloques de tamaño 7



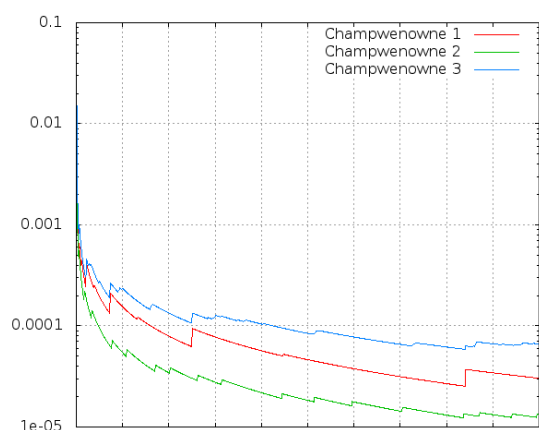
Comparación entre variantes de Champwenowne (Base 3)
Bloques de tamaño 8



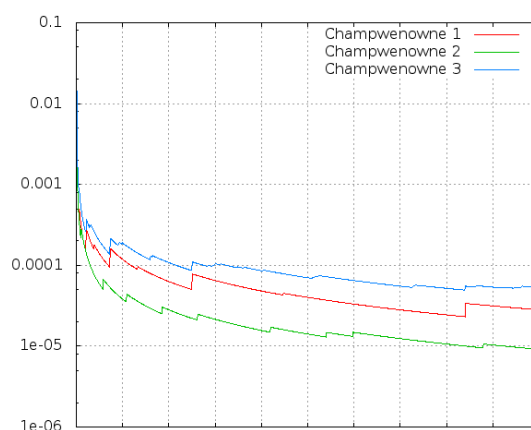
Comparación entre variantes de Champwenowne (Base 3)
Bloques de tamaño 9



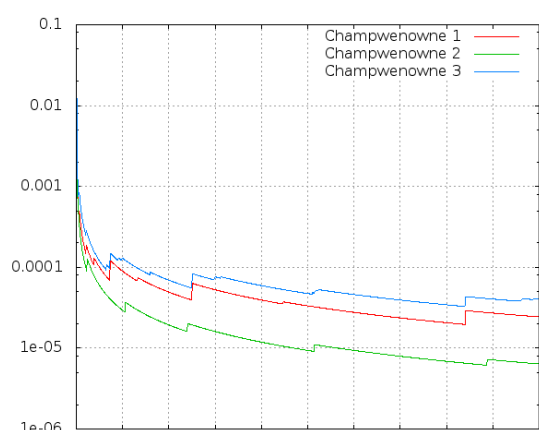
Comparación entre variantes de Champwenowne (Base 3)
Bloques de tamaño 10



Comparación entre variantes de Champwenowne (Base 3)
Bloques de tamaño 11



Comparación entre variantes de Champwenowne (Base 3)
Bloques de tamaño 12

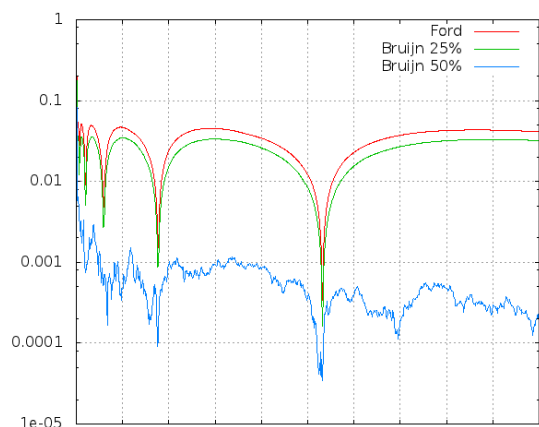


Comparación entre variantes de Champwenowne (Base 3)
Bloques de tamaño 13

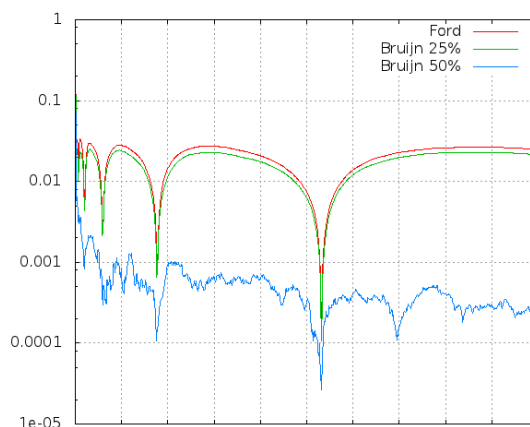
4.3.2. Todas las de Bruijn extendidas tienen la misma discrepancia?

De Bruijn extendidas en base 3

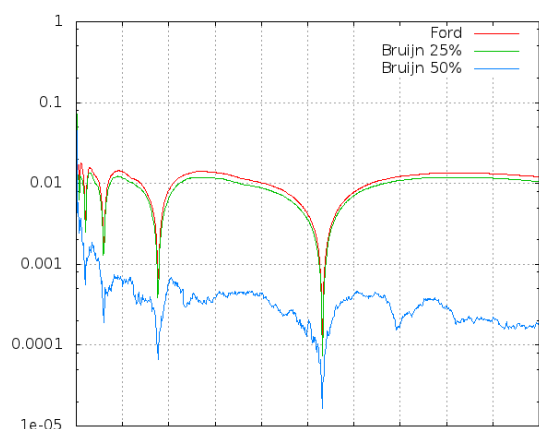
A diferencia de las variantes de Champwenowne, en este caso es la versión más aleatoria la que alcanza un orden de magnitud menor en su discrepancia. La variante ligeramente sesgada y la variante lexicográficamente menor (la Ford Extendida) tienen discrepancias casi iguales, aunque la oportunidad de no siempre elegir el camino lexicográficamente menor de la variante 25% reduce mínimamente la discrepancia.



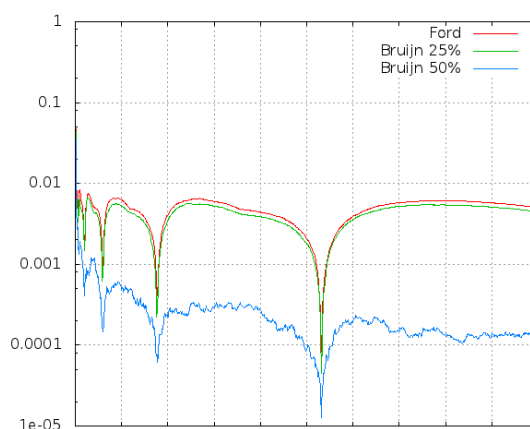
Comparación entre secuencias de Bruijn (Base 3)
Bloques de tamaño 1



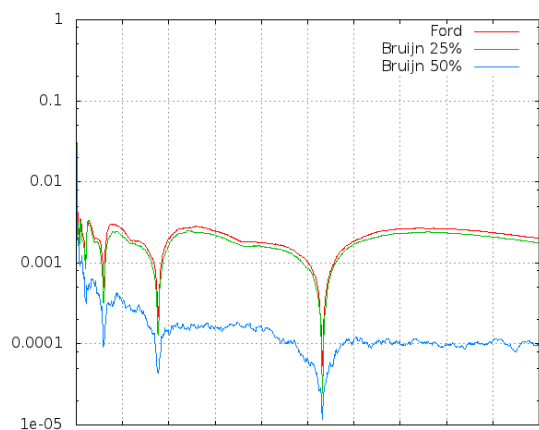
Comparación entre secuencias de Bruijn (Base 3)
Bloques de tamaño 2



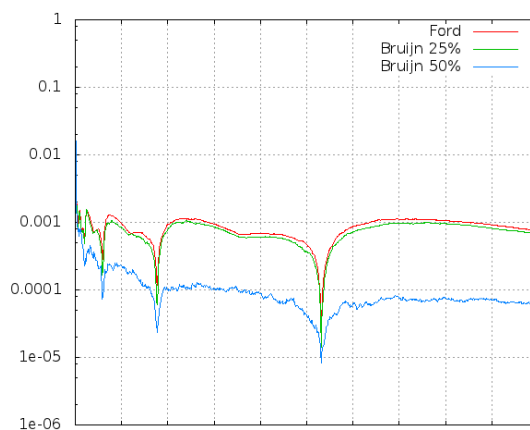
Comparación entre secuencias de Bruijn (Base 3)
Bloques de tamaño 3



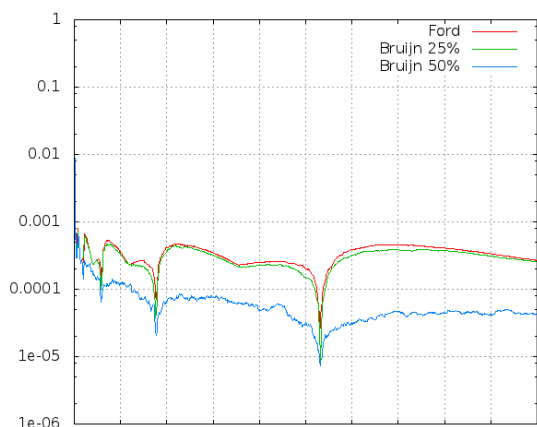
Comparación entre secuencias de Bruijn (Base 3)
Bloques de tamaño 4



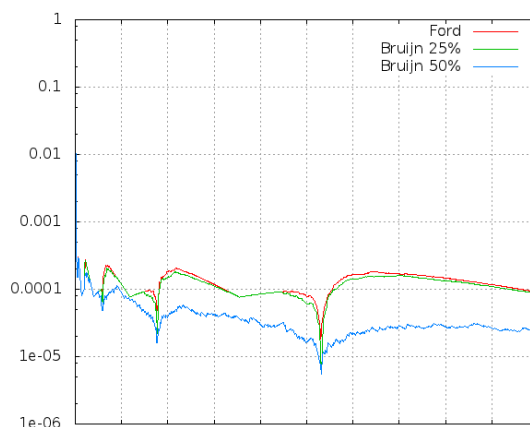
Comparación entre secuencias de Bruijn (Base 3)
Bloques de tamaño 5



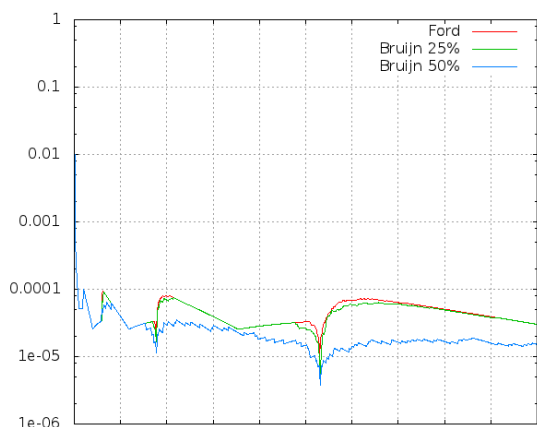
Comparación entre secuencias de Bruijn (Base 3)
Bloques de tamaño 6



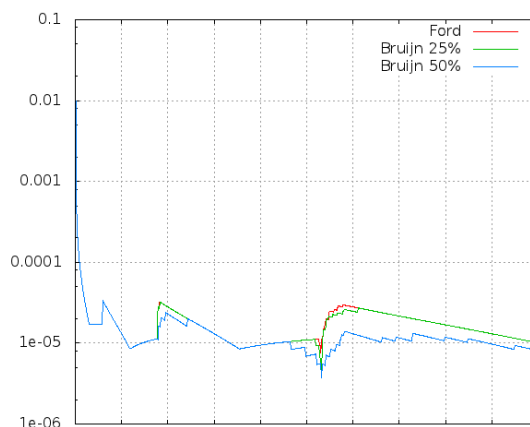
Comparación entre secuencias de Bruijn (Base 3)
Bloques de tamaño 7



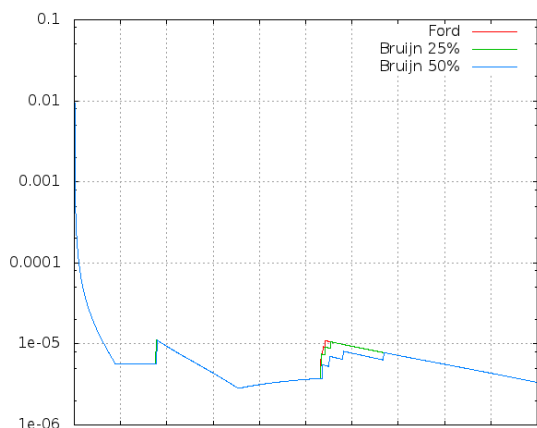
Comparación entre secuencias de Bruijn (Base 3)
Bloques de tamaño 8



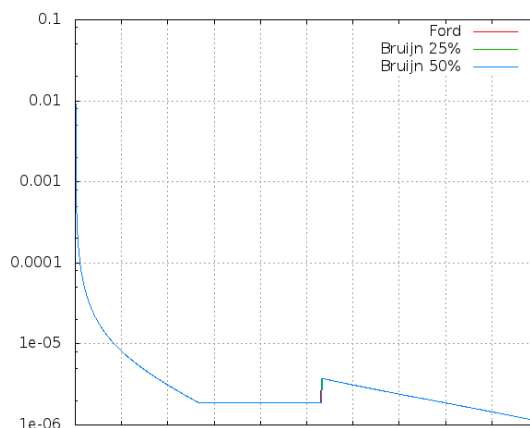
Comparación entre secuencias de Bruijn (Base 3)
Bloques de tamaño 9



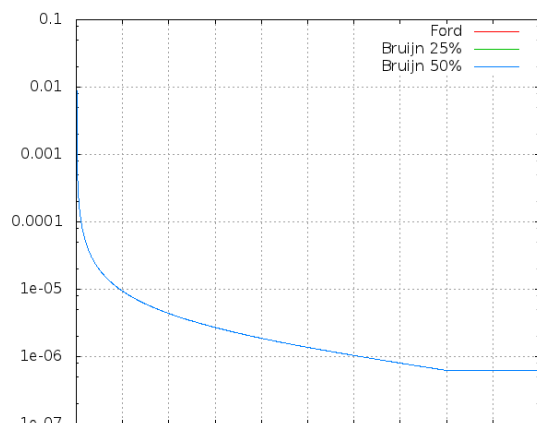
Comparación entre secuencias de Bruijn (Base 3)
Bloques de tamaño 10



Comparación entre secuencias de Bruijn (Base 3)
Bloques de tamaño 11



Comparación entre secuencias de Bruijn (Base 3)
Bloques de tamaño 12

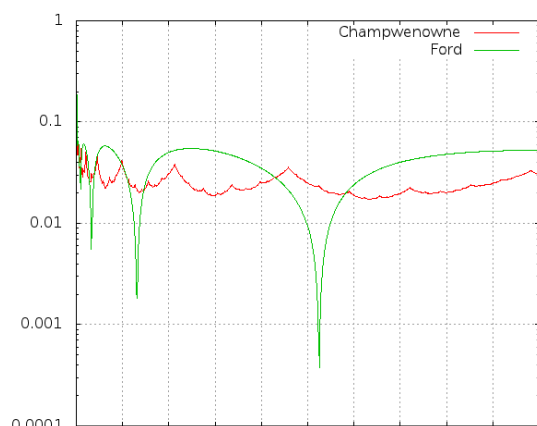


Comparación entre secuencias de Bruijn (Base 3)
Bloques de tamaño 13

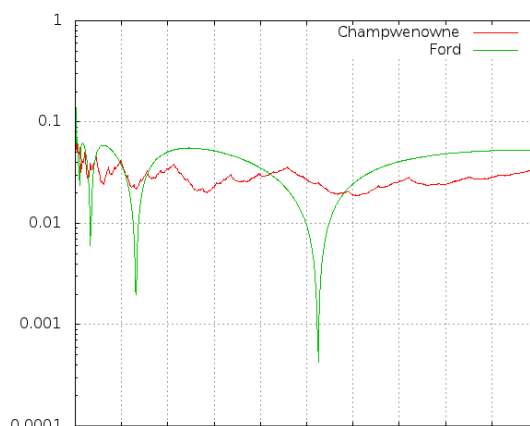
4.3.3. Las de Bruijn extendidas y Champernowne tienen la misma discrepancia?

De Bruijn extendidas en base 2

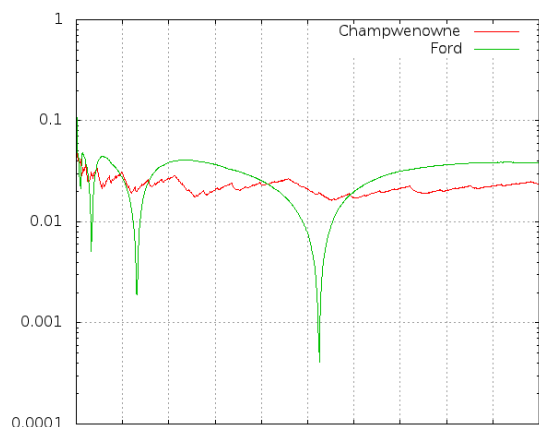
Esta secuencia de gráficos es muy curiosa. Para bloques de tamaño chico la secuencia de Champernowne tiene menor discrepancia, salvo en posiciones cercanos a potencias impares de dos donde sabemos que la discrepancia alcanza al valor 0. Recordemos que estos gráficos toman una muestra cada 1000 puntos y por ende el gráfico no llega a tocar al eje de las abscisas. No obstante, a medida que crece el tamaño del bloque, la diferencia se achica hasta que, finalmente, la discrepancia de Champernowne excede a la de Ford Extendida para los bloques más grandes.



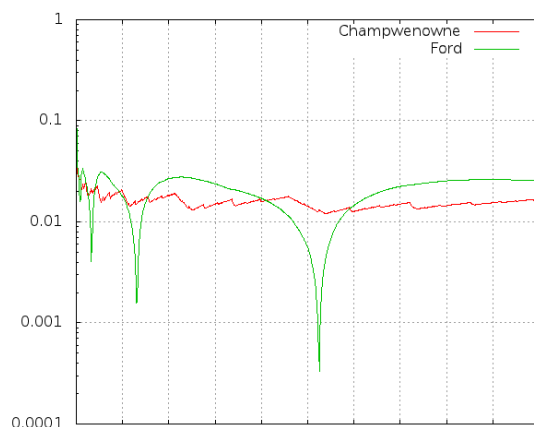
Comparación entre Ford Extendida y Champernowne (Base 2)
Bloques de tamaño 1



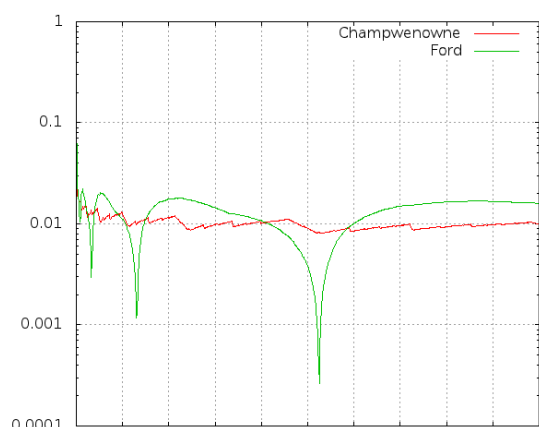
Comparación entre Ford Extendida y Champernowne (Base 2)
Bloques de tamaño 2



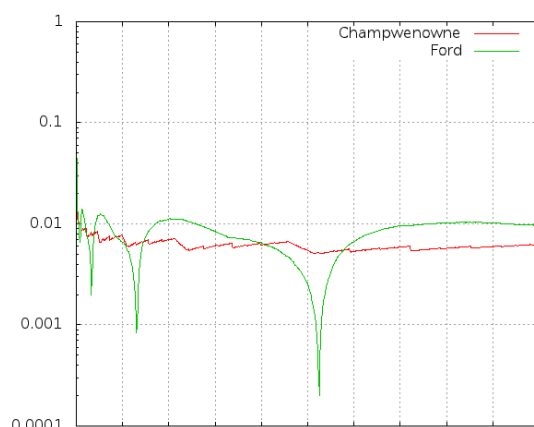
Comparación entre Ford Extendida y Champwenowne (Base 2)
Bloques de tamaño 3



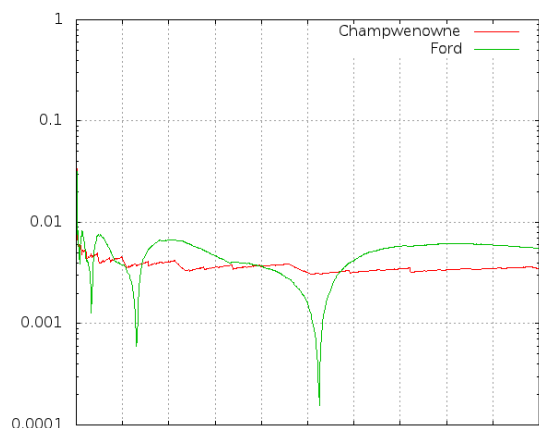
Comparación entre Ford Extendida y Champwenowne (Base 2)
Bloques de tamaño 4



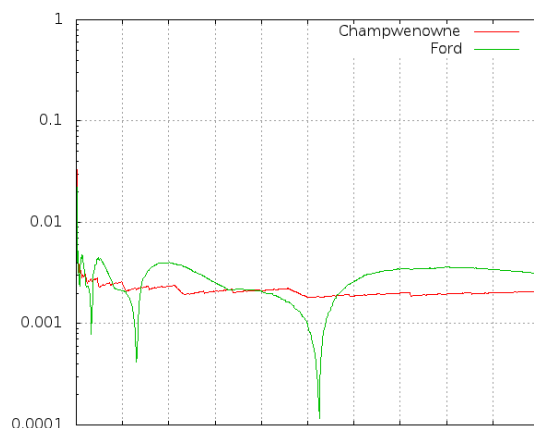
Comparación entre Ford Extendida y Champwenowne (Base 2)
Bloques de tamaño 5



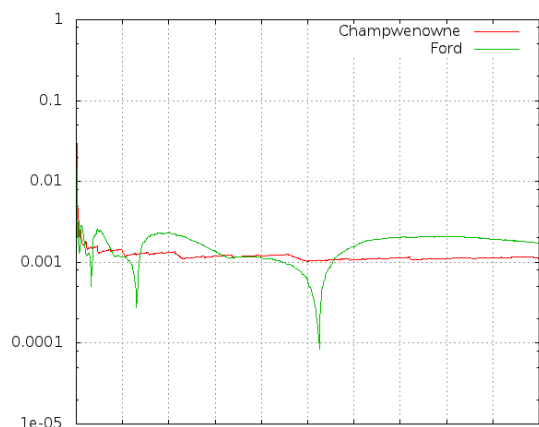
Comparación entre Ford Extendida y Champwenowne (Base 2)
Bloques de tamaño 6



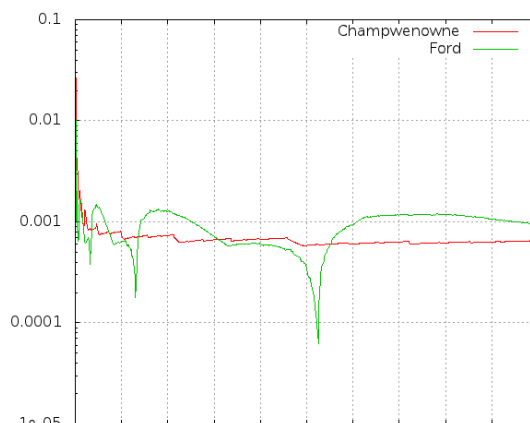
Comparación entre Ford Extendida y Champwenowne (Base 2)
Bloques de tamaño 7



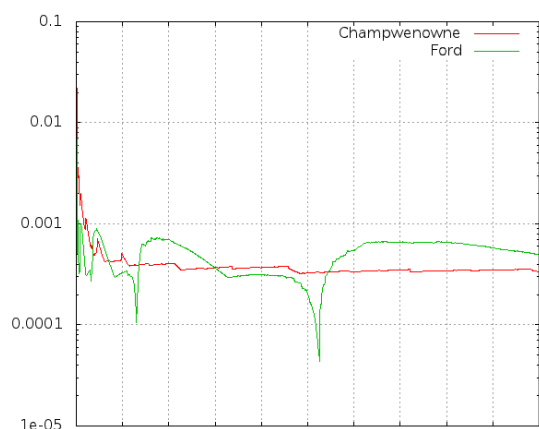
Comparación entre Ford Extendida y Champwenowne (Base 2)
Bloques de tamaño 8



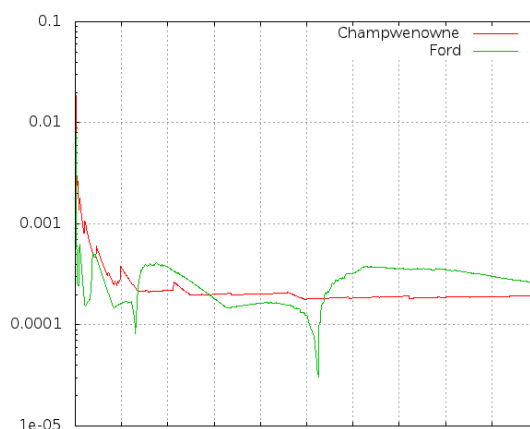
Comparación entre Ford Extendida y Champernowne (Base 2)
Bloques de tamaño 9



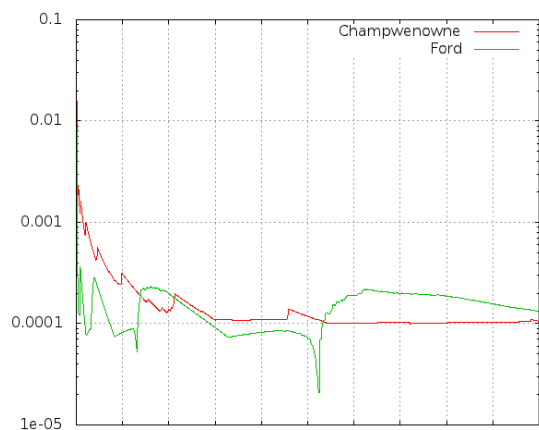
Comparación entre Ford Extendida y Champernowne (Base 2)
Bloques de tamaño 10



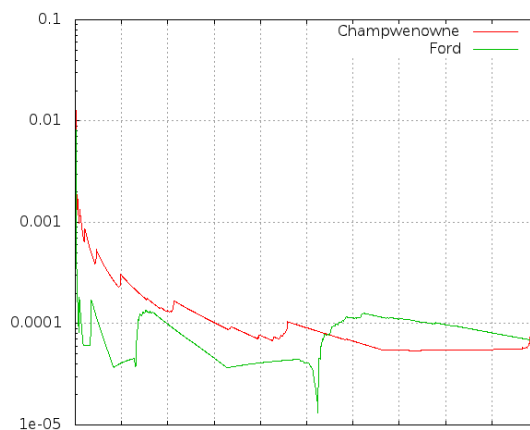
Comparación entre Ford Extendida y Champernowne (Base 2)
Bloques de tamaño 11



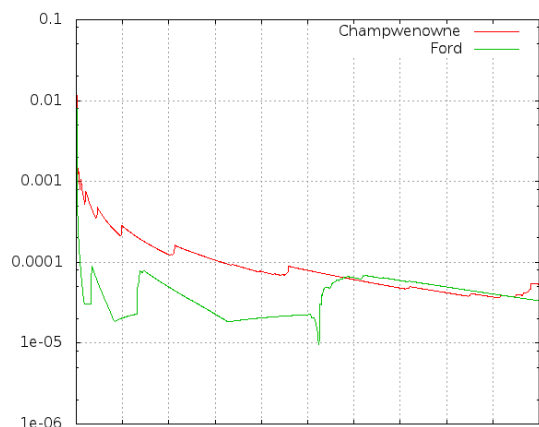
Comparación entre Ford Extendida y Champernowne (Base 2)
Bloques de tamaño 12



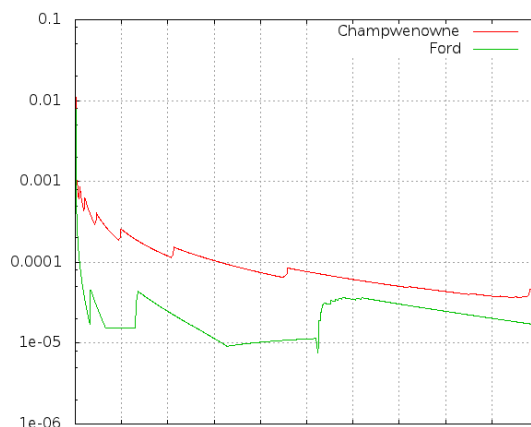
Comparación entre Ford Extendida y Champernowne (Base 2)
Bloques de tamaño 13



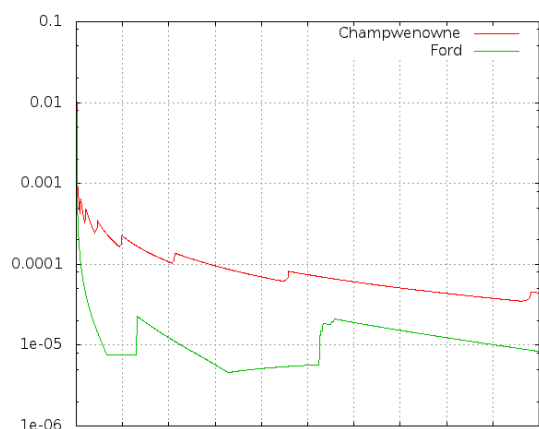
Comparación entre Ford Extendida y Champernowne (Base 2)
Bloques de tamaño 14



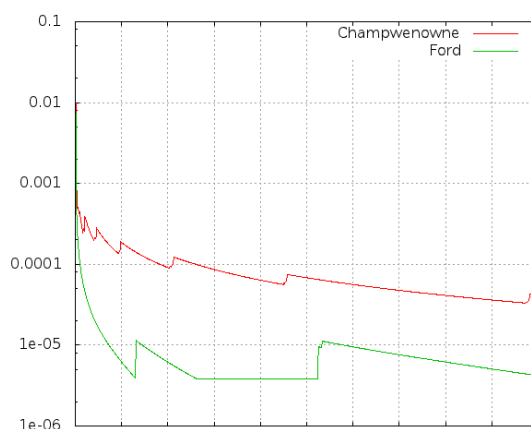
Comparación entre Ford Extendida y Champwenowne (Base 2)
Bloques de tamaño 15



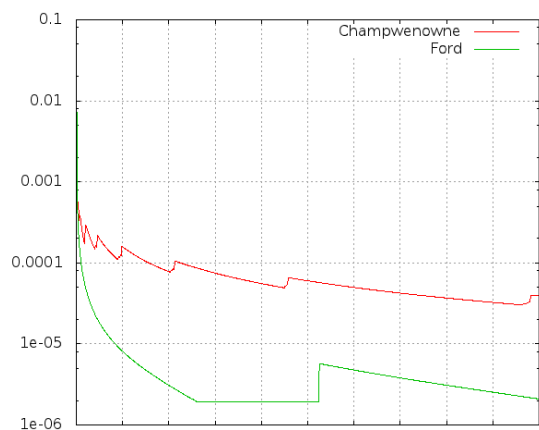
Comparación entre Ford Extendida y Champwenowne (Base 2)
Bloques de tamaño 16



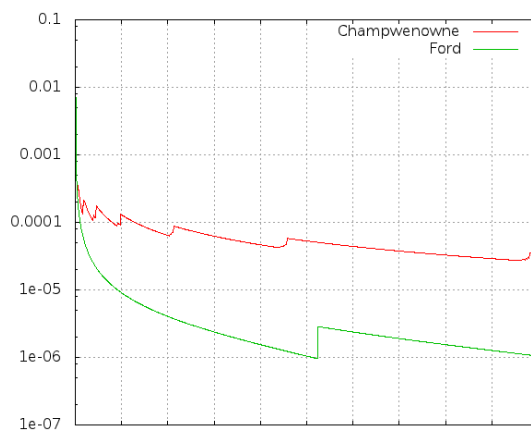
Comparación entre Ford Extendida y Champwenowne (Base 2)
Bloques de tamaño 17



Comparación entre Ford Extendida y Champwenowne (Base 2)
Bloques de tamaño 18



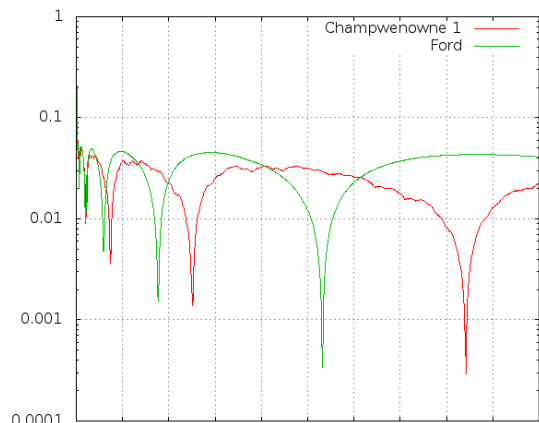
Comparación entre Ford Extendida y Champwenowne (Base 2)
Bloques de tamaño 19



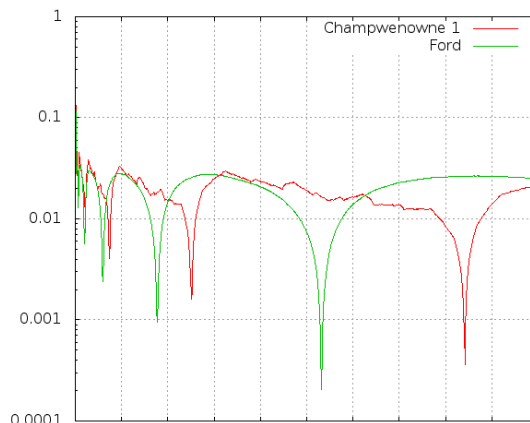
Comparación entre Ford Extendida y Champwenowne (Base 2)
Bloques de tamaño 20

De Bruijn extendidas en base 3

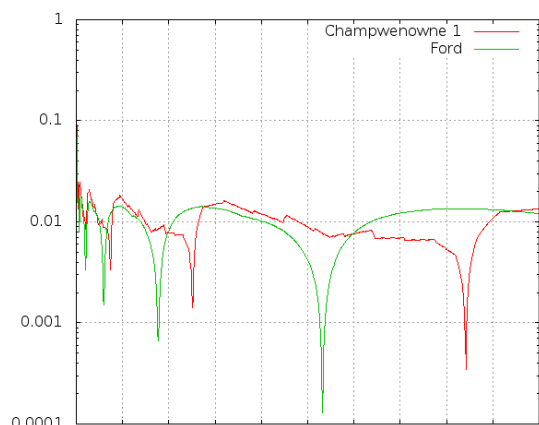
Para base 3 sucede lo mismo que para base 2. Ford Extendida comienza con una discrepancia mayor a Champewowne, aunque es más difícil de apreciar porque ambas secuencias siguen un patrón de arcos con distinto ritmo. La diferencia entre las discrepancias se achica a medida que crece el tamaño del bloque hasta que eventualmente la discrepancia de Ford está por debajo de la de Champewowne.



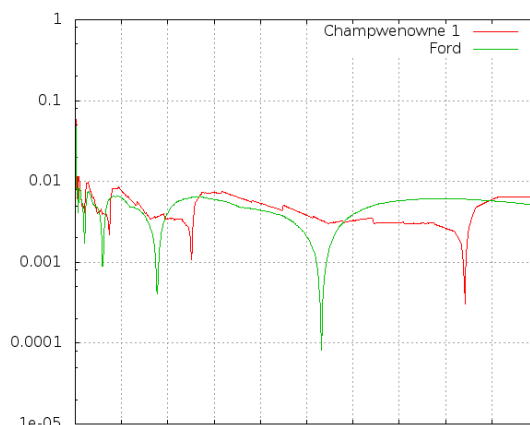
Comparación entre Ford Extendida y Champewowne (Base 3)
Bloques de tamaño 1



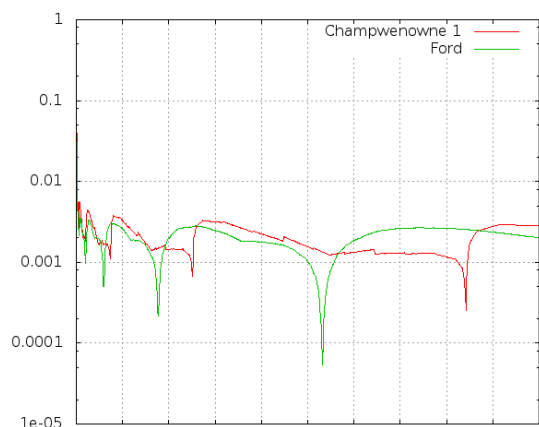
Comparación entre Ford Extendida y Champewowne (Base 3)
Bloques de tamaño 2



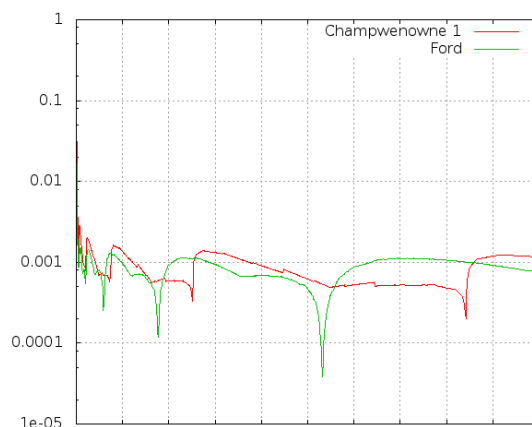
Comparación entre Ford Extendida y Champewowne (Base 3)
Bloques de tamaño 3



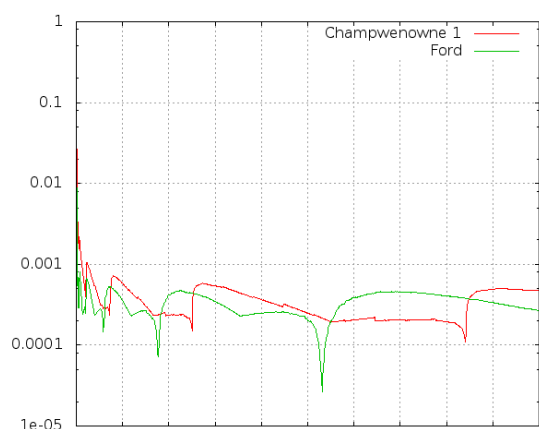
Comparación entre Ford Extendida y Champewowne (Base 3)
Bloques de tamaño 4



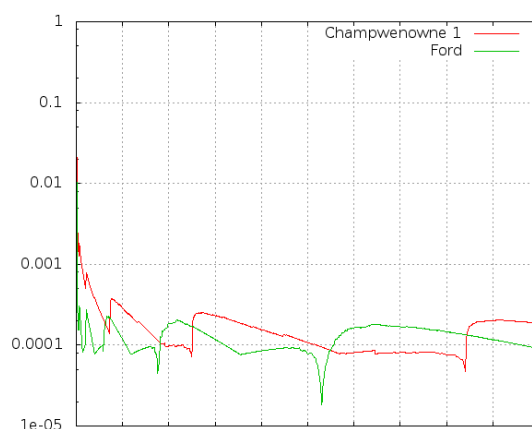
Comparación entre Ford Extendida y Champwenowne (Base 3)
Bloques de tamaño 5



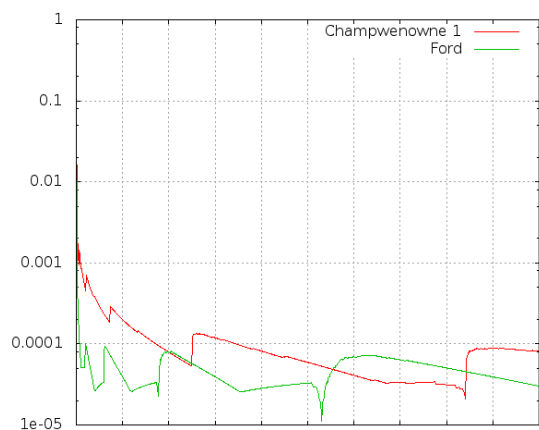
Comparación entre Ford Extendida y Champwenowne (Base 3)
Bloques de tamaño 6



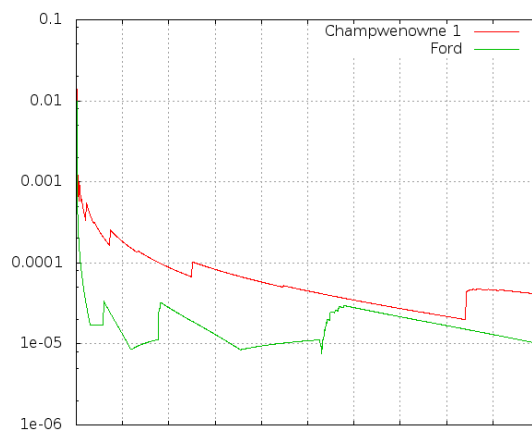
Comparación entre Ford Extendida y Champwenowne (Base 3)
Bloques de tamaño 7



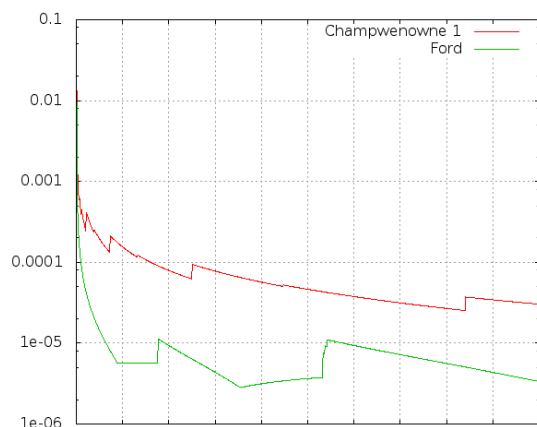
Comparación entre Ford Extendida y Champwenowne (Base 3)
Bloques de tamaño 8



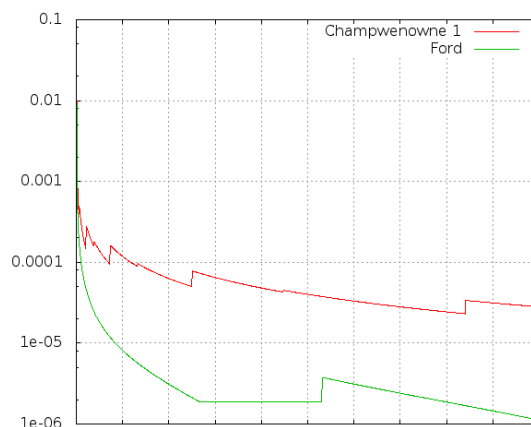
Comparación entre Ford Extendida y Champwenowne (Base 3)
Bloques de tamaño 9



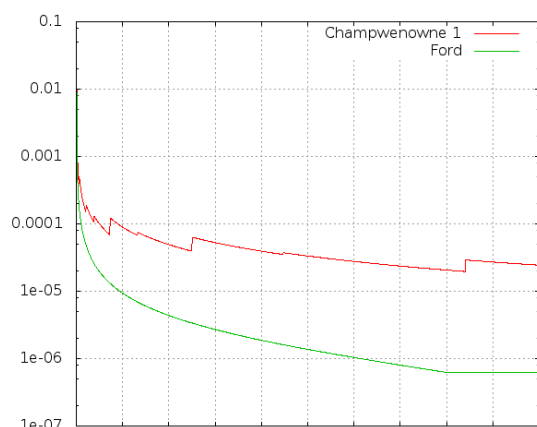
Comparación entre Ford Extendida y Champwenowne (Base 3)
Bloques de tamaño 10



Comparación entre Ford Extendida y Champernowne (Base 3)
Bloques de tamaño 11



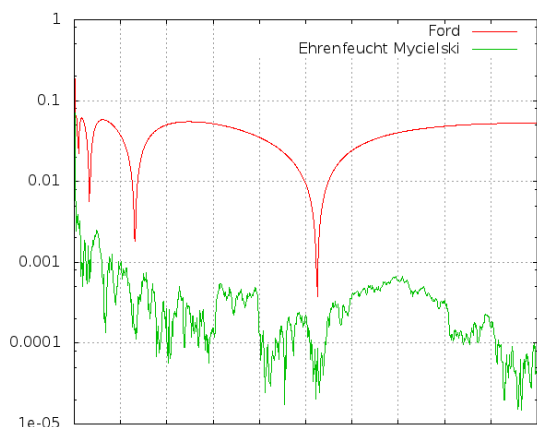
Comparación entre Ford Extendida y Champernowne (Base 3)
Bloques de tamaño 12



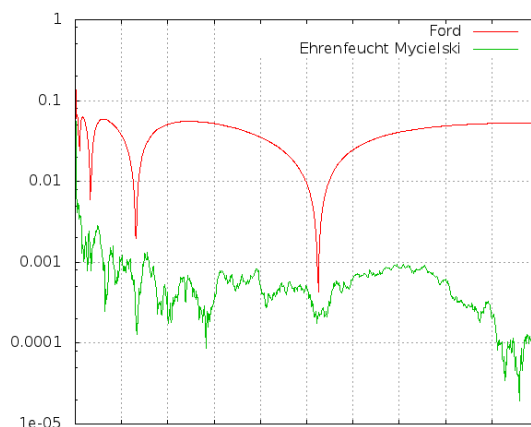
Comparación entre Ford Extendida y Champernowne (Base 3)
Bloques de tamaño 13

4.3.4. Cuánta discrepancia tiene la secuencia de Ehrenfeucht-Mycielski?

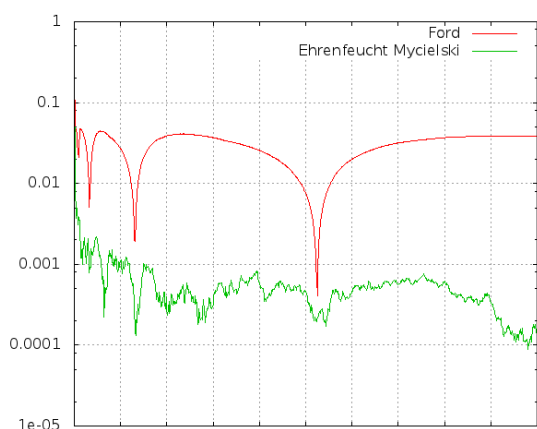
La discrepancia de Ehrenfeucht-Mycielski se mantiene consistentemente debajo de la de Ford Extendida. Como sabemos que la secuencia de Ford Extendida es normal, este gráfico parecería indicar que Ehrenfeucht-Mycielski también lo es.



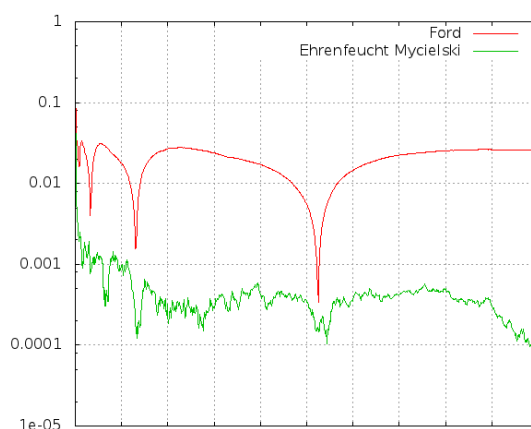
Comparación entre Ford Extendida y Ehrenfeucht-Mycielski (Base 2)
Bloques de tamaño 1



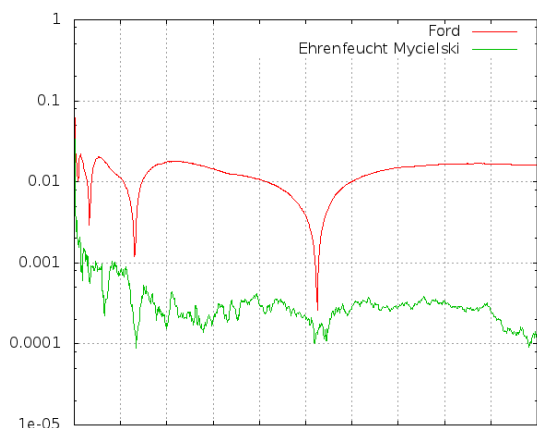
Comparación entre Ford Extendida y Ehrenfeucht-Mycielski (Base 2)
Bloques de tamaño 2



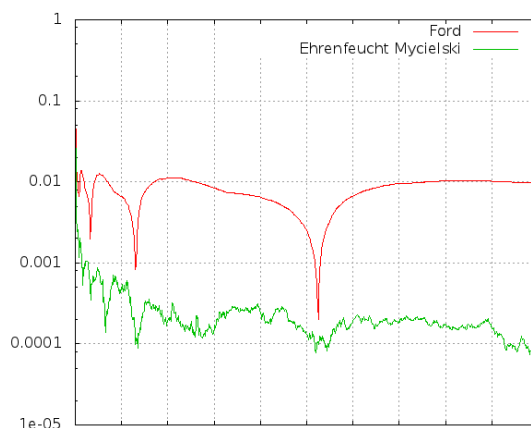
Comparación entre Ford Extendida y Ehrenfeucht-Mycielski (Base 2)
Bloques de tamaño 3



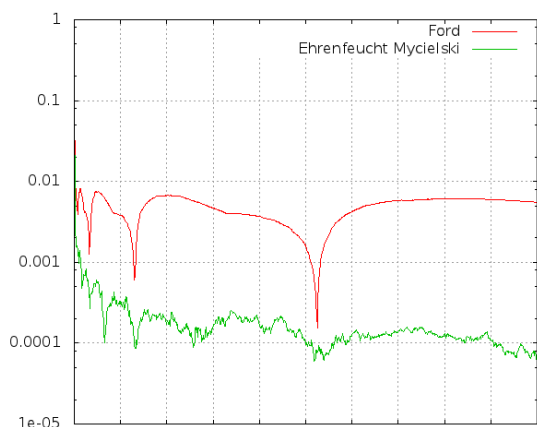
Comparación entre Ford Extendida y Ehrenfeucht-Mycielski (Base 2)
Bloques de tamaño 4



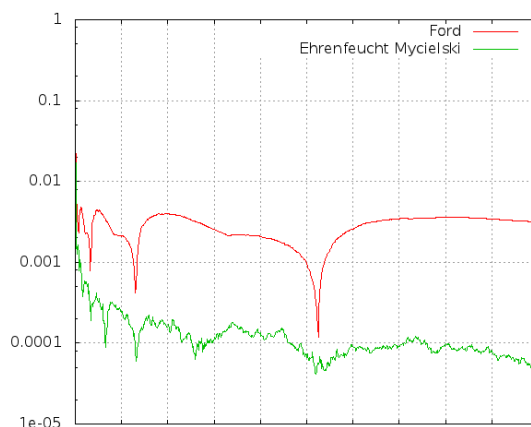
Comparación entre Ford Extendida y Ehrenfeucht-Mycielski (Base 2)
Bloques de tamaño 5



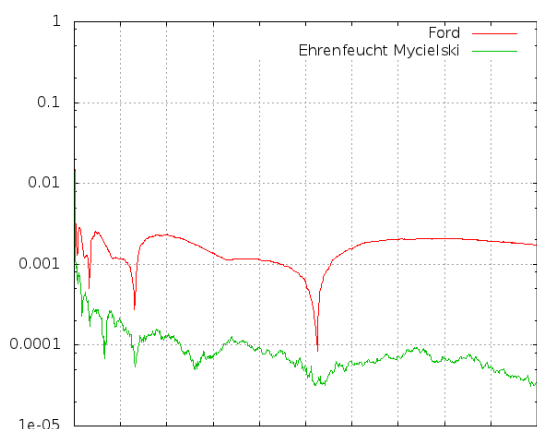
Comparación entre Ford Extendida y Ehrenfeucht-Mycielski (Base 2)
Bloques de tamaño 6



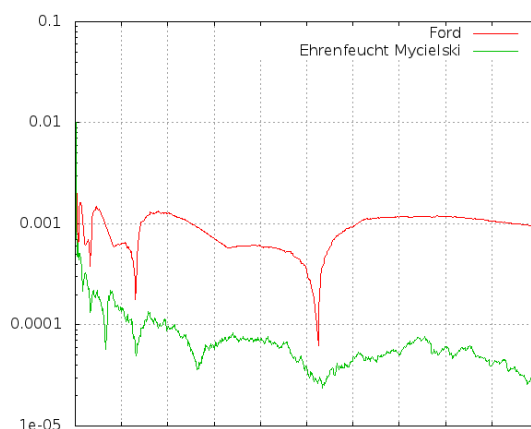
Comparación entre Ford Extendida y Ehrenfeucht-Mycielski (Base 2)
Bloques de tamaño 7



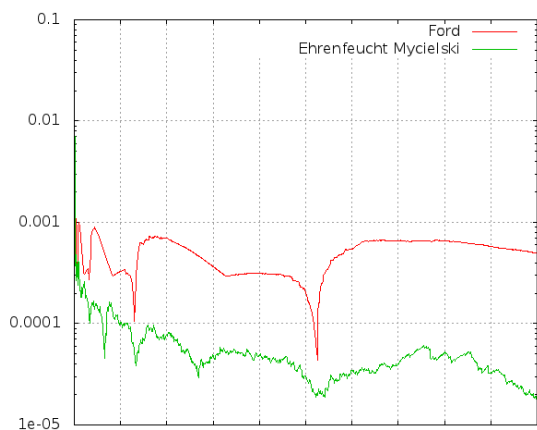
Comparación entre Ford Extendida y Ehrenfeucht-Mycielski (Base 2)
Bloques de tamaño 8



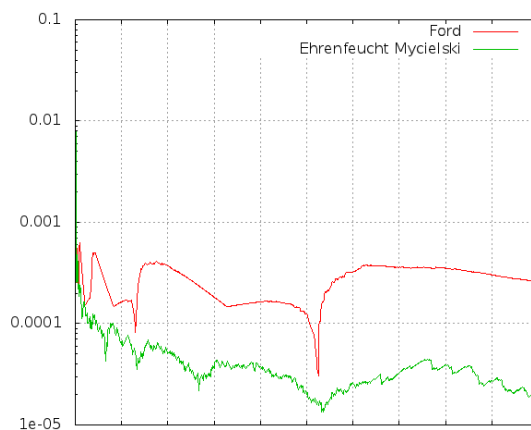
Comparación entre Ford Extendida y Ehrenfeucht-Mycielski (Base 2)
Bloques de tamaño 9



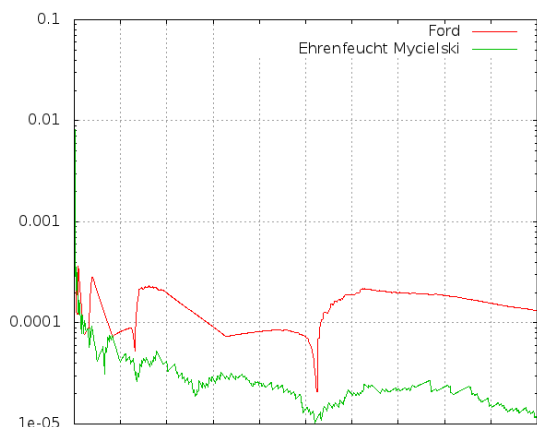
Comparación entre Ford Extendida y Ehrenfeucht-Mycielski (Base 2)
Bloques de tamaño 10



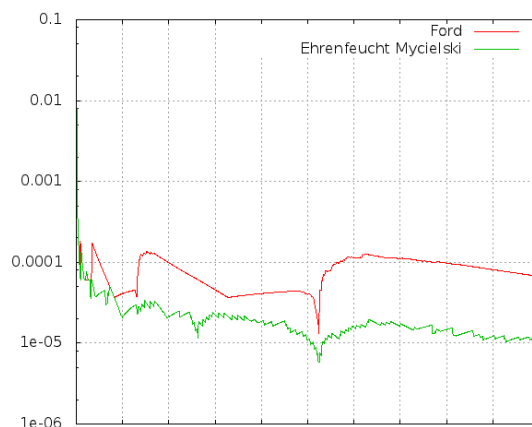
Comparación entre Ford Extendida y Ehrenfeucht-Mycielski (Base 2)
Bloques de tamaño 11



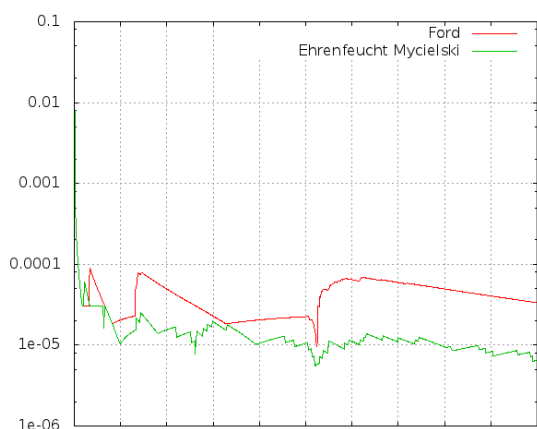
Comparación entre Ford Extendida y Ehrenfeucht-Mycielski (Base 2)
Bloques de tamaño 12



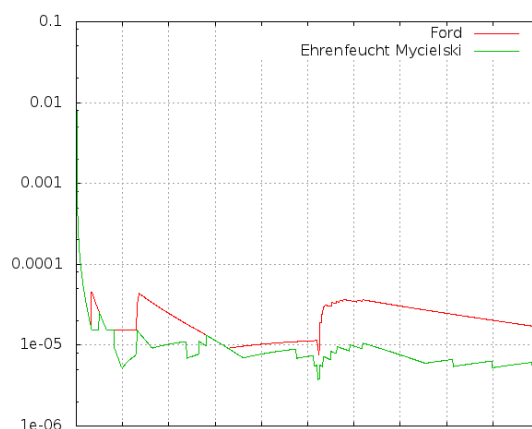
Comparación entre Ford Extendida y Ehrenfeucht-Mycielski (Base 2)
Bloques de tamaño 13



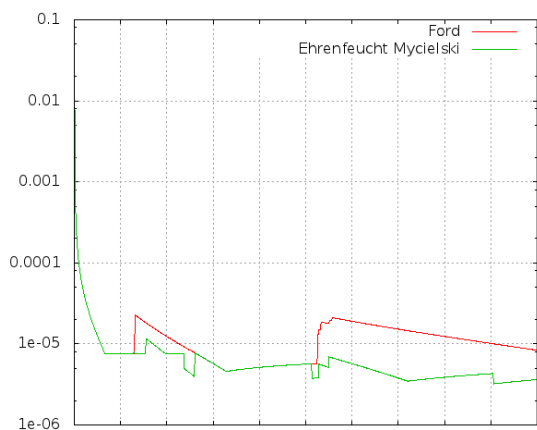
Comparación entre Ford Extendida y Ehrenfeucht-Mycielski (Base 2)
Bloques de tamaño 14



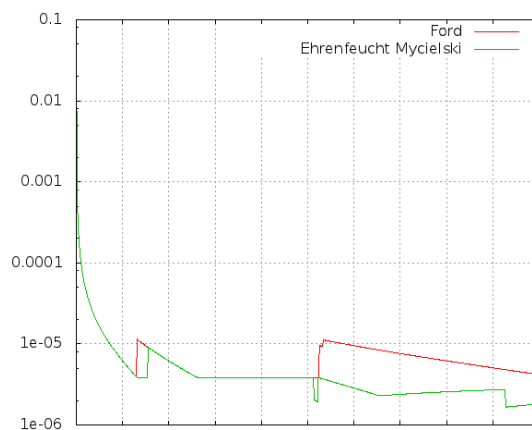
Comparación entre Ford Extendida y Ehrenfeucht-Mycielski (Base 2)
Bloques de tamaño 15



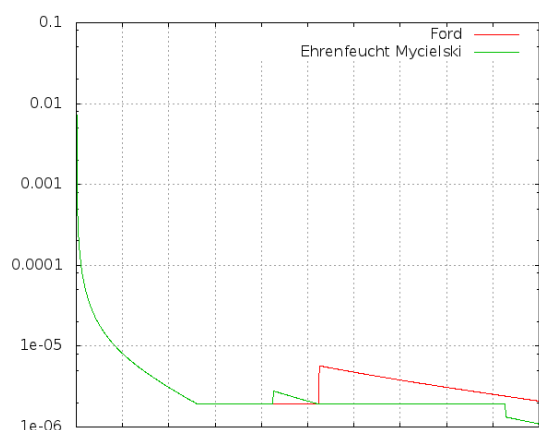
Comparación entre Ford Extendida y Ehrenfeucht-Mycielski (Base 2)
Bloques de tamaño 16



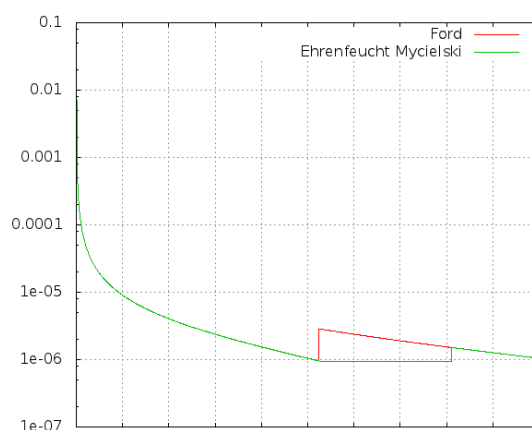
Comparación entre Ford Extendida y Ehrenfeucht-Mycielski (Base 2)
Bloques de tamaño 17



Comparación entre Ford Extendida y Ehrenfeucht-Mycielski (Base 2)
Bloques de tamaño 18



Comparación entre Ford Extendida y Ehrenfeucht-Mycielski (Base 2)
Bloques de tamaño 19

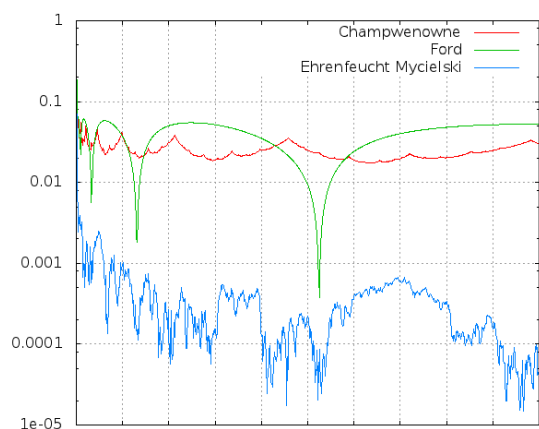


Comparación entre Ford Extendida y Ehrenfeucht-Mycielski (Base 2)
Bloques de tamaño 20

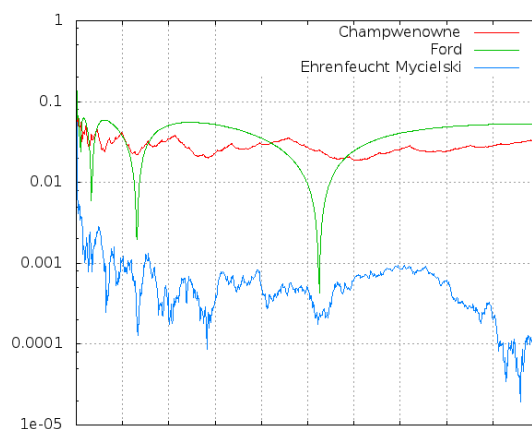
4.3.5. Todas en el mismo gráfico

Base 2

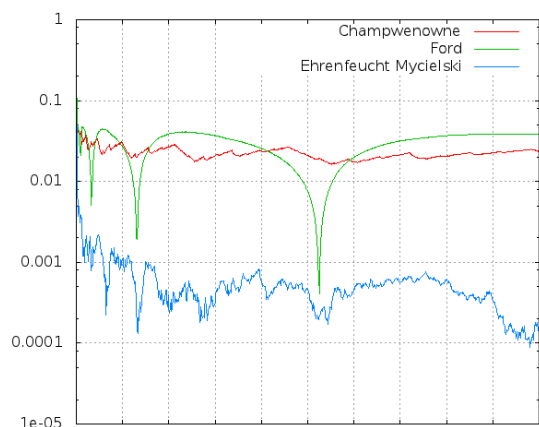
Aquí podemos ver las dos cosas que observábamos antes. Por un lado tenemos a Champwenowne y a la secuencia de Ford Extendida que van invirtiéndose y por el otro tenemos a la secuencia de Ehrenfeucht-Mycielski que se mantiene por debajo de ambas.



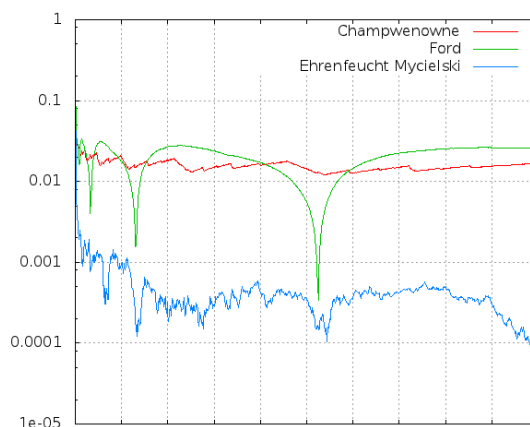
Comparación entre todas las secuencias (Base 2)
Bloques de tamaño 1



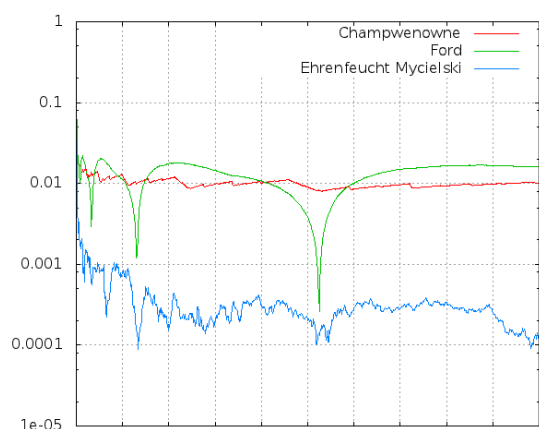
Comparación entre todas las secuencias (Base 2)
Bloques de tamaño 2



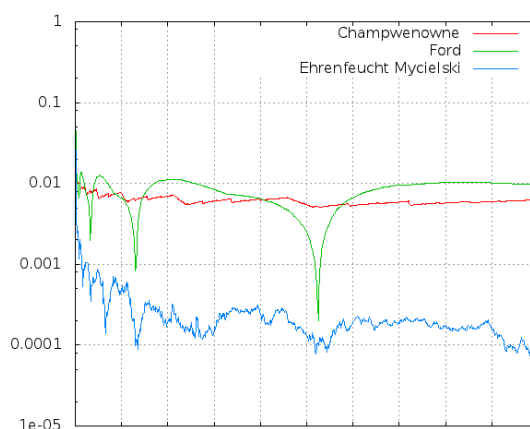
Comparación entre todas las secuencias (Base 2)
Bloques de tamaño 3



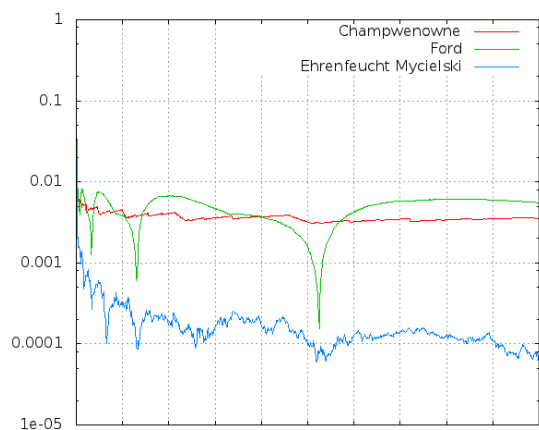
Comparación entre todas las secuencias (Base 2)
Bloques de tamaño 4



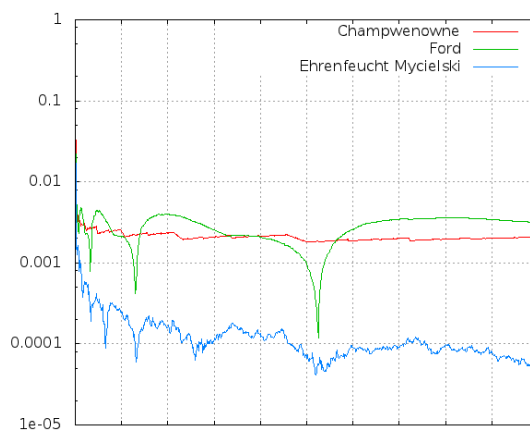
Comparación entre todas las secuencias (Base 2)
Bloques de tamaño 5



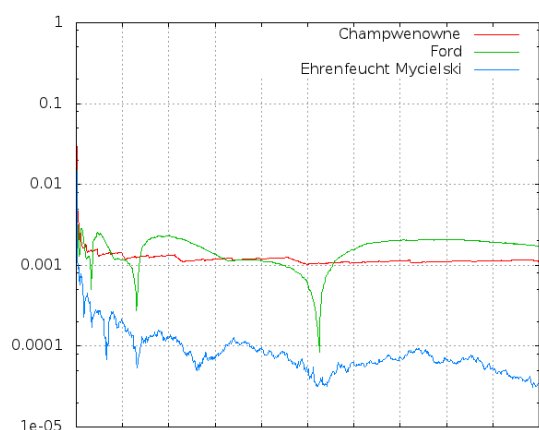
Comparación entre todas las secuencias (Base 2)
Bloques de tamaño 6



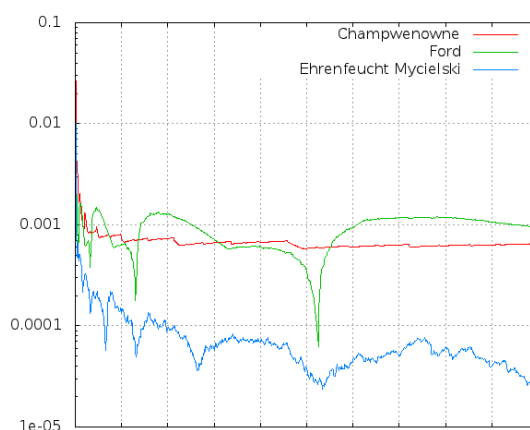
Comparación entre todas las secuencias (Base 2)
Bloques de tamaño 7



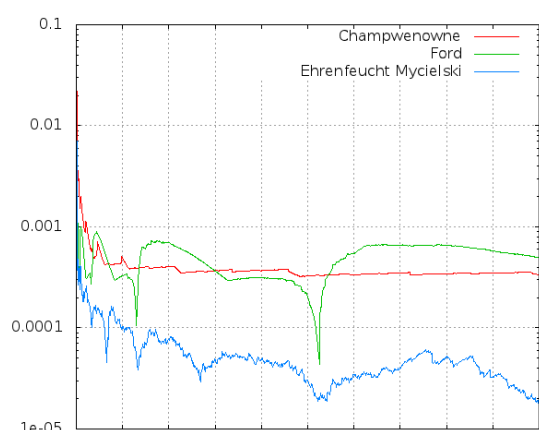
Comparación entre todas las secuencias (Base 2)
Bloques de tamaño 8



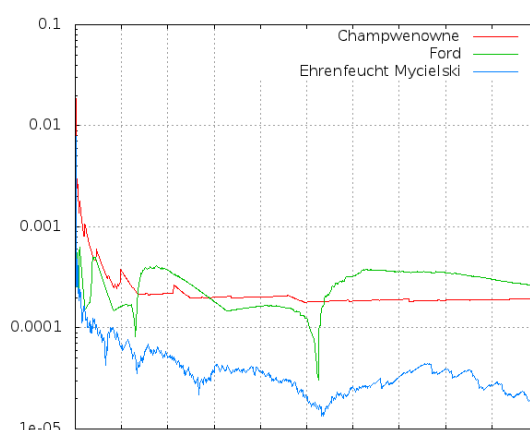
Comparación entre todas las secuencias (Base 2)
Bloques de tamaño 9



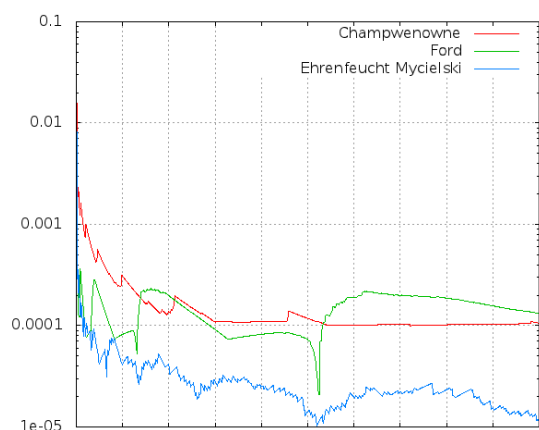
Comparación entre todas las secuencias (Base 2)
Bloques de tamaño 10



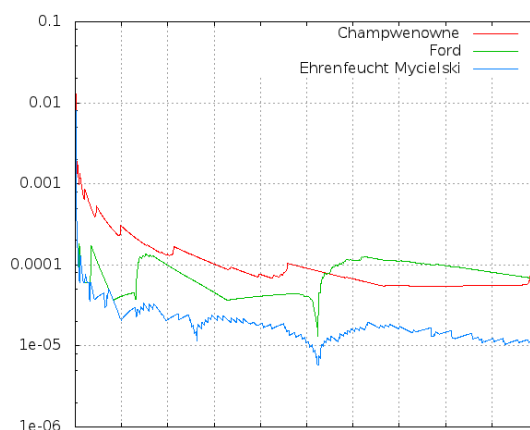
Comparación entre todas las secuencias (Base 2)
Bloques de tamaño 11



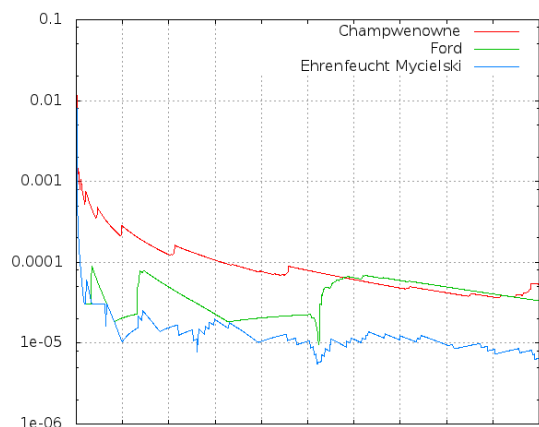
Comparación entre todas las secuencias (Base 2)
Bloques de tamaño 12



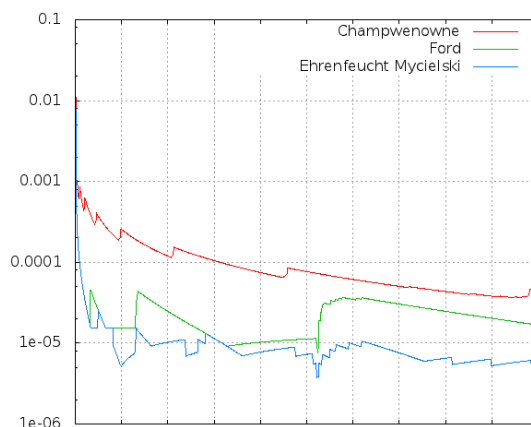
Comparación entre todas las secuencias (Base 2)
Bloques de tamaño 13



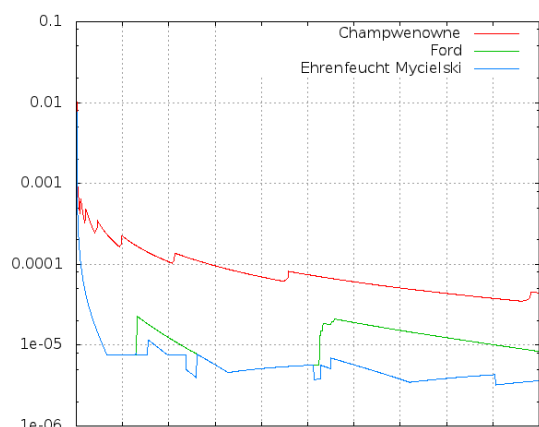
Comparación entre todas las secuencias (Base 2)
Bloques de tamaño 14



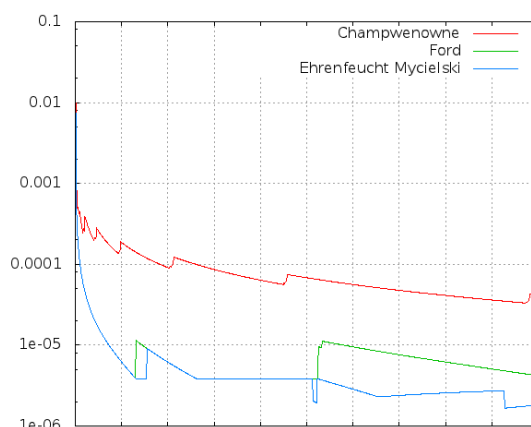
Comparación entre todas las secuencias (Base 2)
Bloques de tamaño 15



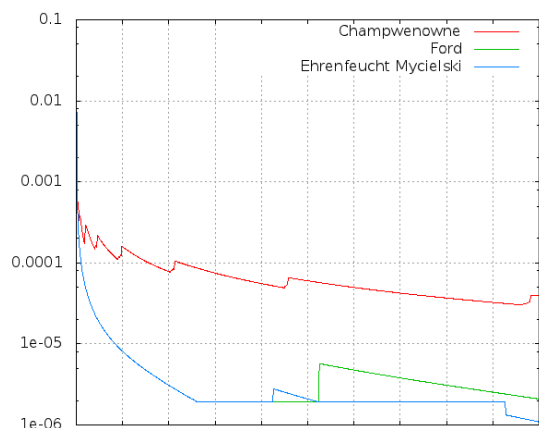
Comparación entre todas las secuencias (Base 2)
Bloques de tamaño 16



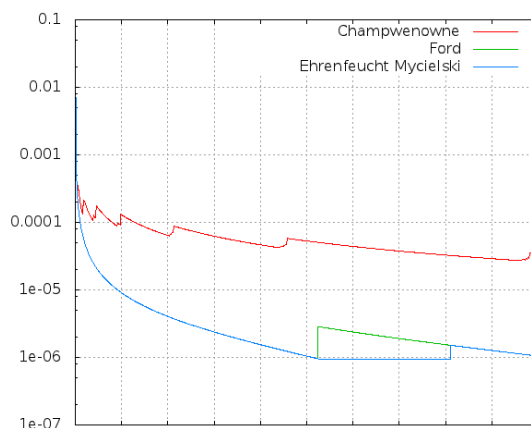
Comparación entre todas las secuencias (Base 2)
Bloques de tamaño 17



Comparación entre todas las secuencias (Base 2)
Bloques de tamaño 18



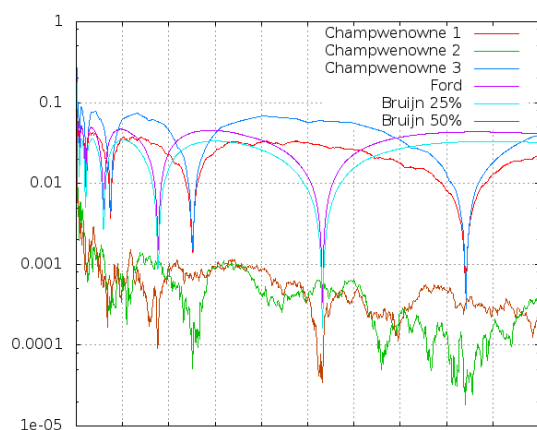
Comparación entre todas las secuencias (Base 2)
Bloques de tamaño 19



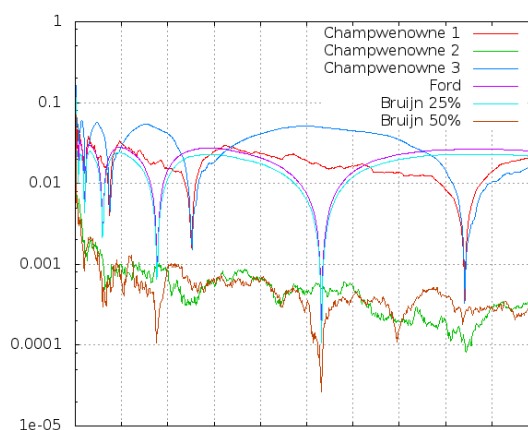
Comparación entre todas las secuencias (Base 2)
Bloques de tamaño 20

Base 3

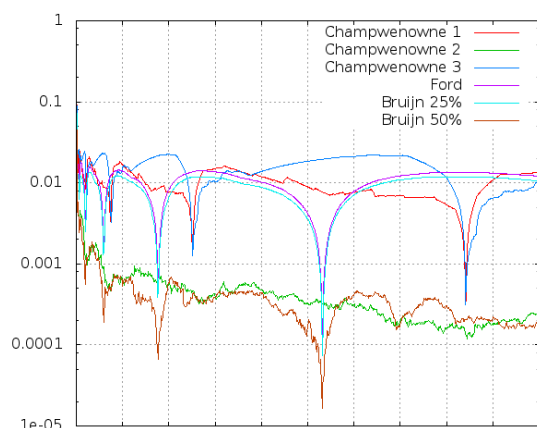
En esta serie de gráficos podemos ver las comparaciones dentro de ambas clases, pero también entre las dos clases. Para los bloques de tamaño chico, tenemos 3 pares de secuencias con discrepancias similares. Por debajo de todo, tenemos a las variantes más aleatorias de ambas familias de secuencias. Notablemente por encima, tenemos a las otras dos variantes de la secuencia de Bruijn y, con magnitud comparable pero con arcos que se repiten a distinto ritmo, las otras dos variantes de Champernowne. Como ya sabemos, a medida que crece el tamaño del bloque, la discrepancia de las secuencias de Bruin decrece más rápidamente que la de las variantes de Champernowne hasta que las 3 variantes de la secuencia de Bruin quedan por debajo de la menos discrepante de las variantes de Champernowne.



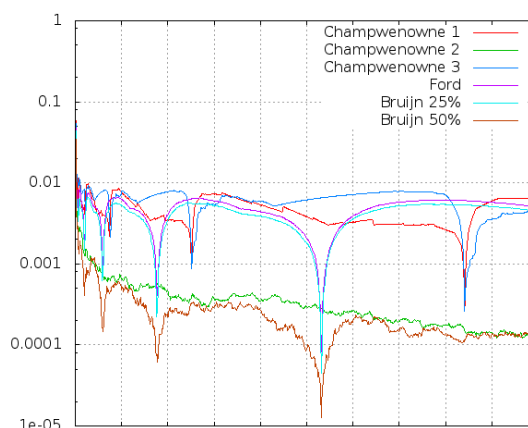
Comparación entre todas las secuencias (Base 3)
Bloques de tamaño 1



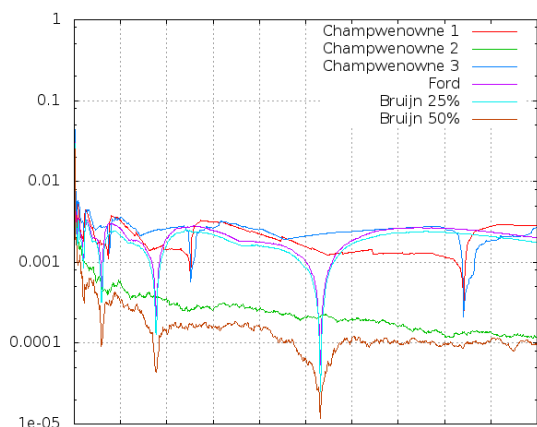
Comparación entre todas las secuencias (Base 3)
Bloques de tamaño 2



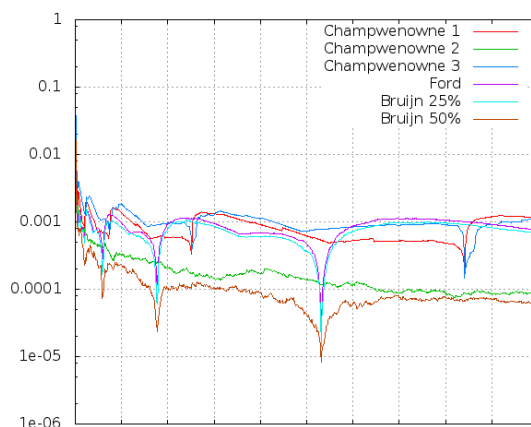
Comparación entre todas las secuencias (Base 3)
Bloques de tamaño 3



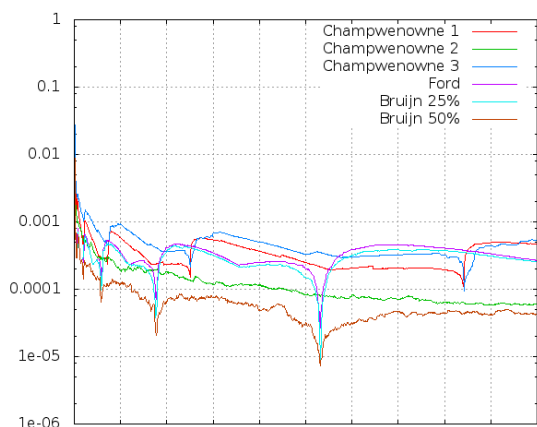
Comparación entre todas las secuencias (Base 3)
Bloques de tamaño 4



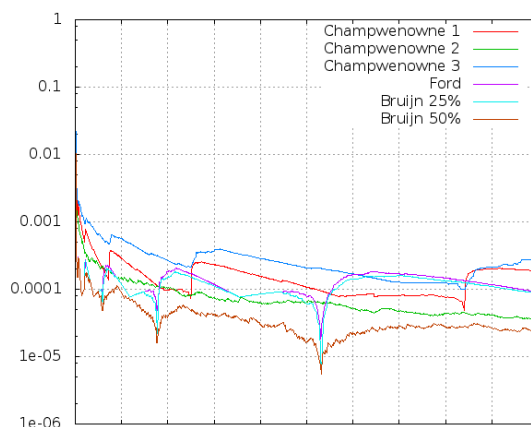
Comparación entre todas las secuencias (Base 3)
Bloques de tamaño 5



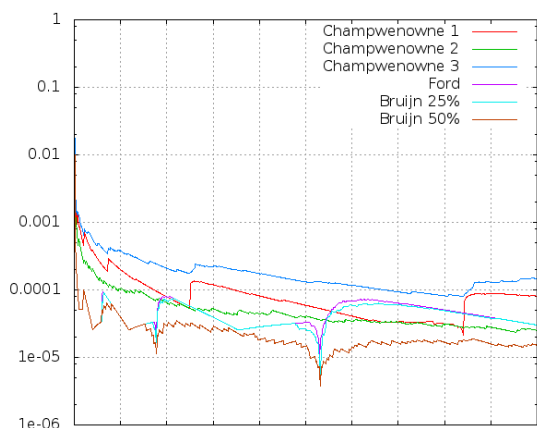
Comparación entre todas las secuencias (Base 3)
Bloques de tamaño 6



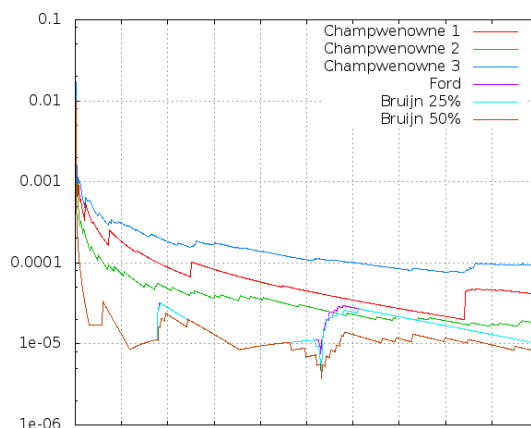
Comparación entre todas las secuencias (Base 3)
Bloques de tamaño 7



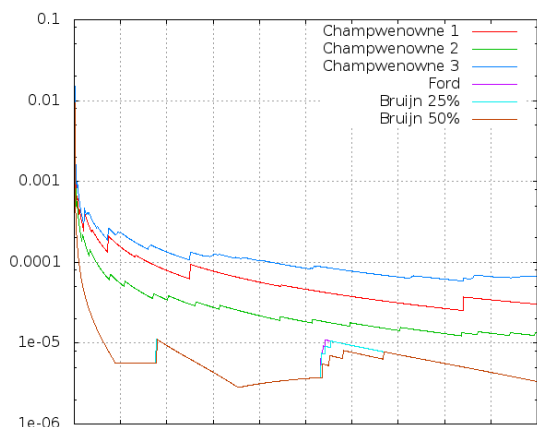
Comparación entre todas las secuencias (Base 3)
Bloques de tamaño 8



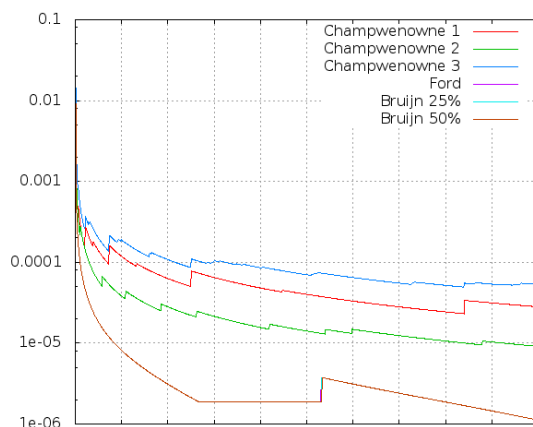
Comparación entre todas las secuencias (Base 3)
Bloques de tamaño 9



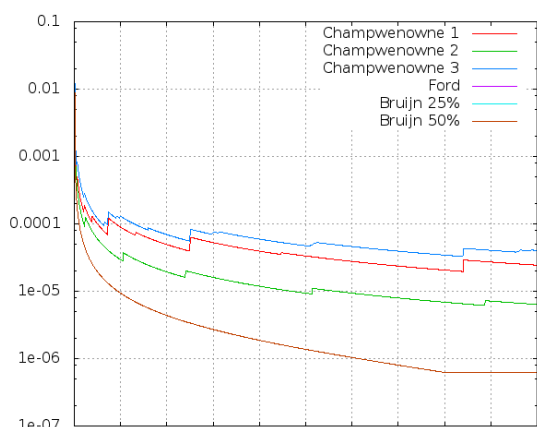
Comparación entre todas las secuencias (Base 3)
Bloques de tamaño 10



Comparación entre todas las secuencias (Base 3)
Bloques de tamaño 11



Comparación entre todas las secuencias (Base 3)
Bloques de tamaño 12



Comparación entre todas las secuencias (Base 3)
Bloques de tamaño 13

Capítulo 5

Conclusiones

5.1. Variantes de secuencias de Bruijn

Por definición, una secuencia de Bruijn extendida tiene discrepancia de normalidad 0 en las posiciones de la forma $|A|^i$. Esto se ve reflejado en el gráfico de discrepancia, que presenta una sucesión de arcos que se repiten regularmente, si usamos una escala logarítmica para el eje x . Esta regularidad es más evidente en la variante Ford Extendida que sobre las otras variantes que no siempre favorecen la minimización lexicográfica. A medida que disminuye la probabilidad de tomar la ruta lexicográficamente menor, la regularidad se mantiene pero la curva se hace menos suave. Esto se debe a que el bloque testigo que maximiza la discrepancia cambia bruscamente de un valor a otro para las variantes más aleatorias. Por contraste, la variante Ford Extendida tiene como testigo al bloque 0^i más consistentemente.

Respecto a cómo cambia la discrepancia entre las variantes, resulta que la secuencia de Ford Extendida es la que maximiza la discrepancia, como esperábamos. Además, la discrepancia aumenta rápidamente cuanto más aleatoria es la construcción: La variante con 25 % de probabilidad de tomar la ruta lexicográficamente menor tiene una discrepancia comparable a la de Ford Extendida, pero la que tiene un 50 % de probabilidad muestra una discrepancia mucho menor, alrededor de un orden de magnitud.

5.2. Variantes de Champernowne

Recordemos que incluso la variante ordenada lexicográficamente (variante 1), es ligeramente distinta a la secuencia de Champernowne original. La variante estudiada incluye todos los bloques de tamaño 1, luego los de tamaño 2 y así sucesivamente (0 1 2 00 01 02 10 11 12...), mientras que la secuencia original se “saltea” los bloques con *leading zeros* (0 1 2 10 11 12...). A causa de esto, incluso esta variante ordenada lexicográficamente se comporta de manera distinta a la secuencia de Champernowne original usada en base 2.

En particular, estas variantes presentan una regularidad similar a las secuencias de Bruijn mencionada previamente, mostrando una sucesión de arcos, aunque no tan regulares y con otro ritmo. En las variantes de Champernowne, la discrepancia se acerca a 0 (valles entre los arcos) no en las posiciones 3^i , sino en los puntos donde ya aparecieron todos los bloques de tamaño menor a i . Es decir, luego del paso marcado como (*) en el algoritmo.

De las tres variantes, la del orden aleatorio (la segunda) tiene una discrepancia significativamente menor a las otras dos, con una diferencia cercana a un orden de magnitud. Como esperábamos, la variante que posterga un dígito lo máximo posible (tercera variante) fue la que maximizó la discrepancia ubicándose por encima de la variante ordenada lexicográficamente.

5.3. Relación entre los diferentes tipos de secuencia

Las secuencias de Champernowne y de Bruijn Extendidas tienen discrepancias comparables. Sin embargo, la discrepancia de las secuencias de Bruijn Extendidas decrece más rápidamente a medida que aumenta el tamaño del bloque. Incluso la secuencia de Ford Extendida, que aparentemente maximiza la discrepancia de su clase de secuencias eventualmente alcanza la discrepancia de la secuencia de Champernowne y luego toma valores aún menores. Por otro lado, la secuencia de Ehrenfeucht-Mycielski se mantiene por debajo de las otras dos curvas, lo cual nos permite sostener la conjetura de que la secuencia de Ehrenfeucht-Mycielski es normal en base dos.

5.4. Conjeturas sobre las secuencias estudiadas

Recordemos que sólo hemos estudiado el comportamiento de la discrepancia de las secuencias estudiadas para el primer millón de dígitos. Debido a que la discrepancia es una propiedad asintótica, el comportamiento en un segmento inicial tiene escaso valor. No obstante, como en cada uno de los casos estudiados la tendencia es muy marcada creemos que los resultados de estos experimentos sí exhiben propiedades de la discrepancia de normalidad.

Viendo los gráficos donde se comparan las secuencias de Bruijn Extendidas y de Champernowne, creemos que para tamaños de bloque mayores a los estudiados la discrepancia de la secuencia de Ford Extendida y de las otras de Bruijn Extendidas es acotada superiormente por la de Champernowne. Conjeturamos que este comportamiento se extiende a alfabetos de mayor tamaño, dado que para tamaño 2 y 3 se observó el mismo comportamiento. Nuestra conjetura sugiere una cota más ajustada para el resultado teórico de Ugalde [23] de que la discrepancia de las secuencias de Bruijn Extendidas es $O(\sqrt{(\log \log n)/\log n})$.

Conjetura 1. *Las secuencias de Bruijn extendidas tienen una discrepancia de orden $O(1/\log n)$.*

También conjeturamos que la discrepancia de las secuencias de Champernowne (original) y Ford Extendida puede calcularse observando sólo los bloques que son rachas de ceros, es decir, bloques de la forma 0^ℓ , para cada $\ell \geq 1$.

Conjetura 2. *Si $\alpha = \text{Champernowne}$ o Ford Extendida , para la posición n y tamaño de bloque ℓ :*

$$D(\alpha, n, \ell) \leq \left| \frac{\text{occ}(0^\ell, \alpha[1..n])}{n} - \frac{1}{|A|^\ell} \right|$$

Donde $\text{occ}(x, s) = |\{i : x = s[i..i + |x| - 1]\}|$ es la cantidad de ocurrencias de x en la cadena s .

A partir de los experimentos conjeturamos que la secuencia Ford extendida tiene una discrepancia máxima entre las secuencias de Bruijn extendidas.

Conjetura 3. *Fijemos un alfabeto de al menos dos símbolos, y sea α la secuencia Ford extendida. Para toda secuencia de Bruijn extendida β en ese alfabeto, para toda posición n y para todo tamaño bloque ℓ ,*

$$D(\alpha, n, \ell) \geq D(\beta, n, \ell).$$

Recordemos también que el algoritmo para generar la secuencia de Ford Extendida involucra extender un camino Hamiltoniano a un ciclo Euleriano. Una variante de este algoritmo, tal vez un poco más elegante, es en el grafo de línea cerrar el ciclo Hamiltoniano antes de extenderlo. Conjeturamos que la secuencia que se obtiene de esta manera se comporta de manera similar a la secuencia Ford extendida que estudiamos.

Bibliografía

- [1] David H. Bailey and Richard E. Crandall. On the random character of fundamental constant expansions. *Experimental Mathematics*, 10(2):175–190, 2001.
- [2] Verónica Becher, Santiago Figueira, and Rafael Picchi. Turing’s unpublished algorithm for normal numbers. *Theoretical Computer Science*, 377:126–138, 2007.
- [3] Verónica Becher and Pablo A. Heiber. On extending de Bruijn sequences. *Information Processing Letters*, 111(930–932), 2011.
- [4] Jean Berstel and Dominique Perrin. The origins of combinatorics on words. *European Journal on Combinatorics*, 28(3):996–1022, 2007.
- [5] Émile Borel. Les probabilités dénombrables et leurs applications arithmétiques. *Rendiconti del Circolo Matematico di Palermo*, 27:247–271, 1909.
- [6] Jonathan M. Borwein and David H. Bailey. *Mathematics by Experiment: Plausible Reasoning in the 21st Century*. AK Peters, second edition, 2008.
- [7] N.G. Bruijn. A combinatorial problem. *Koninklijke Nederlandse Akademie v. Wetenschappen*, 49:758–764, 1946.
- [8] David G. Champernowne. The construction of decimals in the scale of ten. *Journal of the London Mathematical Society*, 8:254–260, 1933.
- [9] Joshua Cooper and Christine Heitsch. The discrepancy of the lex-least de bruijn sequence. *Discrete Mathematics*, 310:1152–1159, 2010.
- [10] Arthur H. Copeland and Paul Erdős. Note on normal numbers. *Bulletin American Mathematical Society*, 52:857–860, 1946.
- [11] Andrzej Ehrenfeucht and Jan Mycielski. A pseudorandom sequence: how random is it? In Richard K. Guy, Unsolved problems’. *American Mathematical Monthly*, 99:373–375, 1992.
- [12] I. Gál and L.Gál. The discrepancy of the sequence $(2^n x)$. *Indagationes Mathematicae*, 26:129–143, 1964.
- [13] I.J. Good. Normal recurring decimals. *Journal of the London Mathematical Society*, 21(3):167–169, 1946.
- [14] Lauwerens Kuipers and Harald Niederreiter. *Uniform distribution of sequences*. Dover Books on Mathematics, New York, 2006.
- [15] Henri Lebesgue. Sur certaines démonstrations d’existence. *Bulletin de la Société Mathématique de France*, 45:132–144, 1917.
- [16] M.B. Levin. On absolutely normal numbers. *Moscow University Mathematics Bulletin*, 1(34):32–39, 1979.
- [17] M.B. Levin. On the discrepancy estimate of normal numbers. *Acta Arithmetica*, 88:99–111, 1999.

- [18] Alan M. Turing. A note on normal numbers. In J.L. Britton, editor, *Collected Works of A.M. Turing: Pure Mathematics*, pages 117–119. North Holland, Amsterdam, 1992. With notes of the editor in pages 263–265.
- [19] Camille Flye Sainte-Marie. Question 48. *L'intermédiaire des mathématiciens*, 1:107–110, 1894.
- [20] J. Schiffer. Discrepancy of normal numbers. *Acta Arithmetica*, 47(2):175– 186, 1986.
- [21] W. Schmidt. On normal numbers. *Pacific Journal of Mathematics*, 10:661–672, 1960.
- [22] Waclaw Sierpiński. Démonstration élémentaire du théorème de M. Borel sur les nombres absolument normaux et détermination effective d'un tel nombre. *Bulletin de la Société Mathématique de France*, 45:127–132, 1917.
- [23] Edgardo Ugalde. An alternative construction of normal numbers. *Journal de Théorie des nombres de Bordeaux*, 12:165–177, 2000.
- [24] Y. Bugeaud. *Distribution Modulo One and Diophantine Approximation*. Cambridge Tracts in Mathematics 193. Cambridge University Press, 2012.