Tesis de Licenciatura en Ciencias de la Computación

Secuencias completamente equidistribuidas basadas en secuencias de De Bruijn

Emilio Almansi

Directora: Verónica Becher Departamento de Computación Facultad de Ciencias Exactas y Naturales Universidad de Buenos Aires 4 de septiembre, 2019

¿Qué tienen en común las siguientes disciplinas?

Criptografía, seguridad informática.

- Criptografía, seguridad informática.
- Predicción del clima, medicina nuclear, simulación de proteínas.

- Criptografía, seguridad informática.
- Predicción del clima, medicina nuclear, simulación de proteínas.
- Aprendizaje automático, algoritmos probabilistas.

- Criptografía, seguridad informática.
- Predicción del clima, medicina nuclear, simulación de proteínas.
- Aprendizaje automático, algoritmos probabilistas.
- Juegos de azar, videojuegos, simulaciones físicas.

¿Qué tienen en común las siguientes disciplinas?

- Criptografía, seguridad informática.
- Predicción del clima, medicina nuclear, simulación de proteínas.
- Aprendizaje automático, algoritmos probabilistas.
- Juegos de azar, videojuegos, simulaciones físicas.

Generación de números aleatorios.

¿Qué tienen en común las siguientes disciplinas?

- Criptografía, seguridad informática.
- Predicción del clima, medicina nuclear, simulación de proteínas.
- Aprendizaje automático, algoritmos probabilistas.
- Juegos de azar, videojuegos, simulaciones físicas.

Generación de números aleatorios.

Pero, ¿qué es una secuencia de números aleatorios?

Intuición: si tiro un dado muchas veces seguidas, el resultado de cada tirada tiene que ser *impredecible* y todo número del 1 al 6 tiene que ser *equiprobable*.

Intuición: si tiro un dado muchas veces seguidas, el resultado de cada tirada tiene que ser *impredecible* y todo número del 1 al 6 tiene que ser *equiprobable*.

Respecto a la parte de impredecible:

If "random" means that the sequence satisfies no predictable rules, the title of this paper is contradictory.

Construction of a Random Sequence Donald Knuth, 1965

Intuición: si tiro un dado muchas veces seguidas, el resultado de cada tirada tiene que ser *impredecible* y todo número del 1 al 6 tiene que ser *equiprobable*.

Respecto a la parte de impredecible:

If "random" means that the sequence satisfies no predictable rules, the title of this paper is contradictory.

Construction of a Random Sequence Donald Knuth, 1965

En este trabajo, nos enfocamos en la parte de equiprobable.

Definición

Dado un entero b, una secuencia de n'umeros enteros $X=x_1,x_2,\ldots$ del conjunto $\{0,1,\ldots,b-1\}$ es **equidistribuida** si todo valor posible aparece con frecuencia asintótica igual a $\frac{1}{h}$:

Definición

Dado un entero b, una secuencia de n'umeros enteros $X=x_1,x_2,\ldots$ del conjunto $\{0,1,\ldots,b-1\}$ es **equidistribuida** si todo valor posible aparece con frecuencia asintótica igual a $\frac{1}{h}$:

$$Pr(x_i = j) = \frac{1}{h}$$
 para todo $j \in \{0, \dots, b-1\}$,

Definición

Dado un entero b, una secuencia de *números enteros* $X=x_1,x_2,\ldots$ del conjunto $\{0,1,\ldots,b-1\}$ es **equidistribuida** si todo valor posible aparece con frecuencia asintótica igual a $\frac{1}{h}$:

$$Pr(x_i = j) = \frac{1}{b}$$
 para todo $j \in \{0, \dots, b-1\}$,

donde
$$Pr(x_i = j) = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \sigma(x_i = j)$$
.

Definición

Dado un entero b, una secuencia de *números enteros* $X=x_1,x_2,\ldots$ del conjunto $\{0,1,\ldots,b-1\}$ es **equidistribuida** si todo valor posible aparece con frecuencia asintótica igual a $\frac{1}{b}$:

$$Pr(x_i=j)=rac{1}{b}$$
 para todo $j\in\{0,\ldots,b-1\}$,

donde
$$Pr(x_i = j) = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \sigma(x_i = j)$$
.

Ahora, definimos una noción equivalente para secuencias de números reales.

Definición

Una secuencia de *números reales* $X=x_1,x_2,\ldots$ en el intervalo unitario [0,1) es **equidistribuida** si, dado cualquier conjunto $I\subseteq [0,1)$, la frecuencia asintótica con la que la secuencia toma valores en I es igual a su tamaño:

Definición

Una secuencia de *números reales* $X=x_1,x_2,\ldots$ en el intervalo unitario [0,1) es **equidistribuida** si, dado cualquier conjunto $I\subseteq [0,1)$, la frecuencia asintótica con la que la secuencia toma valores en I es igual a su tamaño:

$$Pr(x_i \in I) = |I|$$
 para todo $I \subseteq [0, 1)$,

Definición

Una secuencia de *números reales* $X=x_1,x_2,\ldots$ en el intervalo unitario [0,1) es **equidistribuida** si, dado cualquier conjunto $I\subseteq [0,1)$, la frecuencia asintótica con la que la secuencia toma valores en I es igual a su tamaño:

$$Pr(x_i \in I) = |I|$$
 para todo $I \subseteq [0,1)$,

donde
$$Pr(x_i \in I) = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \sigma(x_i \in I).$$

Definición

Una secuencia de *números reales* $X=x_1,x_2,\ldots$ en el intervalo unitario [0,1) es **equidistribuida** si, dado cualquier conjunto $I\subseteq [0,1)$, la frecuencia asintótica con la que la secuencia toma valores en I es igual a su tamaño:

$$Pr(x_i \in I) = |I|$$
 para todo $I \subseteq [0,1)$,

donde
$$Pr(x_i \in I) = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \sigma(x_i \in I)$$
.

Si X es equidistribuida en [0,1), entonces para cualquier b la secuencia $Y=(\lfloor bx_i\rfloor)_{i=1}^{\infty}$ es equidistribuida en $\{0,1,\ldots,b-1\}$.

Intuición: cualquier seguidilla de tiradas tiene que aparecer con igual frecuencia. Por ejemplo, (1, 1) aparece con la misma frecuencia que (2, 2) y que (6, 4).

Intuición: cualquier seguidilla de tiradas tiene que aparecer con igual frecuencia. Por ejemplo, (1, 1) aparece con la misma frecuencia que (2, 2) y que (6, 4).

Trabajamos con "ventanas" de tamaño k de la secuencia. Si $X=x_1,x_2,\ldots$ es una secuencia de *números reales*, entonces:

Intuición: cualquier seguidilla de tiradas tiene que aparecer con igual frecuencia. Por ejemplo, (1, 1) aparece con la misma frecuencia que (2, 2) y que (6, 4).

Trabajamos con "ventanas" de tamaño k de la secuencia. Si $X=x_1,x_2,\ldots$ es una secuencia de *números reales*, entonces:

$$w_1 = (x_1, x_2, \dots, x_k),$$

 $w_2 = (x_2, x_3, \dots, x_{k+1}),$
 $w_3 = (x_3, x_4, \dots, x_{k+2}),$
...

es la secuencia de ventanas de X, que llamamos $W_k(X)$.

Definición

Una secuencia de *números reales* $X=x_1,x_2,\ldots$ en el intervalo unitario [0,1) es k-distribuida si, dado cualquier conjunto $I\subseteq [0,1)^k$, la frecuencia asintótica con la que la secuencia de ventanas de X toma valores en I es igual a su tamaño:

Definición

Una secuencia de *números reales* $X=x_1,x_2,\ldots$ en el intervalo unitario [0,1) es k-distribuida si, dado cualquier conjunto $I\subseteq [0,1)^k$, la frecuencia asintótica con la que la secuencia de ventanas de X toma valores en I es igual a su tamaño:

$$Pr(w_i \in I) = |I|$$
 para todo $I \subseteq [0,1)^k$,

Definición

Una secuencia de *números reales* $X=x_1,x_2,\ldots$ en el intervalo unitario [0,1) es k-distribuida si, dado cualquier conjunto $I\subseteq [0,1)^k$, la frecuencia asintótica con la que la secuencia de ventanas de X toma valores en I es igual a su tamaño:

$$Pr(w_i \in I) = |I|$$
 para todo $I \subseteq [0,1)^k$,

donde
$$W_k(X) = (w_i)_{i=1}^{\infty}$$

= $(x_1, x_2, \dots, x_k), (x_2, x_3, \dots, x_{k+1}), \dots$

Definición

Una secuencia de *números reales* $X=x_1,x_2,\ldots$ en el intervalo unitario [0,1) es k-distribuida si, dado cualquier conjunto $I\subseteq [0,1)^k$, la frecuencia asintótica con la que la secuencia de ventanas de X toma valores en I es igual a su tamaño:

$$Pr(w_i \in I) = |I|$$
 para todo $I \subseteq [0,1)^k$,

donde
$$W_k(X) = (w_i)_{i=1}^{\infty}$$

= $(x_1, x_2, \dots, x_k), (x_2, x_3, \dots, x_{k+1}), \dots$

Si X es k-distribuida para todo k, entonces X es **completamente** equidistribuida.

¿Por qué estudiar la propiedad de equidistribución?

► Requerimiento básico de pseudo-aleatoriedad. Propiedades de equipartición y autocorrelación con retraso.

¿Por qué estudiar la propiedad de equidistribución?

- Requerimiento básico de pseudo-aleatoriedad. Propiedades de equipartición y autocorrelación con retraso.
- Calidad de un generador de números aleatorios o PNRG. Pruebas de aleatoriedad.

¿Por qué estudiar la propiedad de equidistribución?

- Requerimiento básico de pseudo-aleatoriedad. Propiedades de equipartición y autocorrelación con retraso.
- Calidad de un generador de números aleatorios o PNRG. Pruebas de aleatoriedad.
- Integración de Montecarlo, criterio de la integral de Riemann.

¿Por qué estudiar la propiedad de equidistribución?

- Requerimiento básico de pseudo-aleatoriedad. Propiedades de equipartición y autocorrelación con retraso.
- Calidad de un generador de números aleatorios o PNRG. Pruebas de aleatoriedad.
- Integración de Montecarlo, criterio de la integral de Riemann.
- Vínculo: teoría de números, computación, probabilidad y estadística.

¿Por qué estudiar la propiedad de equidistribución?

- Requerimiento básico de pseudo-aleatoriedad. Propiedades de equipartición y autocorrelación con retraso.
- Calidad de un generador de números aleatorios o PNRG. Pruebas de aleatoriedad.
- Integración de Montecarlo, criterio de la integral de Riemann.
- Vínculo: teoría de números, computación, probabilidad y estadística.

¿Qué no es la equidistribución?

Discrepancia

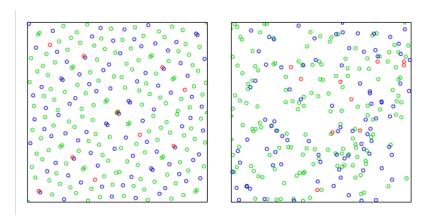


Figura 1: Izq.) Secuencia Sobol 2,3. Der.) Secuencia pseudo-aleat.

Con
$$b = 2, k = 3$$
:

Con
$$b = 2, k = 3$$
:

Con
$$b = 2, k = 3$$
:

$$0, \mathbf{0}, \mathbf{0}, \mathbf{1}, 0, 1, 1, 1$$

Con
$$b = 2, k = 3$$
:

$$0, 0, \mathbf{0}, \mathbf{1}, \mathbf{0}, 1, 1, 1$$

Con
$$b = 2, k = 3$$
:

$$0, 0, 0, \mathbf{1}, \mathbf{0}, \mathbf{1}, 1, 1$$

Con
$$b = 2, k = 3$$
:

$$0, 0, 0, 1, \mathbf{0}, \mathbf{1}, \mathbf{1}, 1$$

Con
$$b = 2, k = 3$$
:

$$0, 0, 0, 1, 0, \mathbf{1}, \mathbf{1}, \mathbf{1}$$

Con
$$b = 2, k = 3$$
:

$$\mathbf{0}, 0, 0, 1, 0, 1, \mathbf{1}, \mathbf{1}$$

Con
$$b = 2, k = 3$$
:

Con
$$b = 2, k = 3$$
:

Con
$$b = 4, k = 2$$
:

Con
$$b = 2, k = 3$$
:

Con
$$b = 4, k = 2$$
:

$$0, 0, 1, 0, 2, 0, 3, 1, 1, 2, 1, 3, 2, 2, 3, 3$$

Son secuencias muy estudiadas en combinatoria. Tienen la propiedad de "contener" a todas las posibles combinaciones de secuencias de un alfabeto y tamaño dados. Ejemplos:

Con
$$b = 2, k = 3$$
:

Con
$$b = 4, k = 2$$
:

$$0, 0, 1, 0, 2, 0, 3, 1, 1, 2, 1, 3, 2, 2, 3, 3$$

Notar que siempre tienen longitud b^k .

Son secuencias muy estudiadas en combinatoria. Tienen la propiedad de "contener" a todas las posibles combinaciones de secuencias de un alfabeto y tamaño dados. Ejemplos:

Con
$$b = 2, k = 3$$
:

Con
$$b = 4, k = 2$$
:

$$0, 0, 1, 0, 2, 0, 3, 1, 1, 2, 1, 3, 2, 2, 3, 3$$

Notar que siempre tienen longitud b^k . ¿Qué relación tienen con las secuencias equidistribuidas?

Definición

Una **secuencia** A **de órden** n es la secuencia que se obtiene al dividir cada elemento de una secuencia de De Bruijn de base 2^n y de órden n por su base:

Definición

Una **secuencia** A **de órden** n es la secuencia que se obtiene al dividir cada elemento de una secuencia de De Bruijn de base 2^n y de órden n por su base:

$$A^{(n)} = \frac{f_1}{2^n}, \frac{f_2}{2^n}, \dots, \frac{f_{2^{n^2}}}{2^n}$$

Definición

Una **secuencia** A **de órden** n es la secuencia que se obtiene al dividir cada elemento de una secuencia de De Bruijn de base 2^n y de órden n por su base:

$$A^{(n)} = \frac{f_1}{2^n}, \frac{f_2}{2^n}, \dots, \frac{f_{2^{n^2}}}{2^n}$$

donde $F^{(2^n,n)}=f_1,\ldots,f_{2^{n^2}}$ denota una secuencia de De Bruijn de base 2^n y de órden n.

Definición

Una **secuencia** A **de órden** n es la secuencia que se obtiene al dividir cada elemento de una secuencia de De Bruijn de base 2^n y de órden n por su base:

$$A^{(n)} = \frac{f_1}{2^n}, \frac{f_2}{2^n}, \dots, \frac{f_{2^{n^2}}}{2^n}$$

donde $F^{(2^n,n)}=f_1,\dots,f_{2^{n^2}}$ denota una secuencia de De Bruijn de base 2^n y de órden n.

Una **secuencia** B **de órden** n es la secuencia que se obtiene de concatenar $n2^{2n}$ copias de una secuencia A de órden n:

Definición

Una **secuencia** A **de órden** n es la secuencia que se obtiene al dividir cada elemento de una secuencia de De Bruijn de base 2^n y de órden n por su base:

$$A^{(n)} = \frac{f_1}{2^n}, \frac{f_2}{2^n}, \dots, \frac{f_{2^{n^2}}}{2^n}$$

donde $F^{(2^n,n)}=f_1,\dots,f_{2^{n^2}}$ denota una secuencia de De Bruijn de base 2^n y de órden n.

Una **secuencia** B **de órden** n es la secuencia que se obtiene de concatenar $n2^{2n}$ copias de una secuencia A de órden n:

$$B^{(n)} = \left\langle \underbrace{A^{(n)}; A^{(n)}; \dots; A^{(n)}}_{n2^{2n} \text{ veces}} \right\rangle.$$

$$F^{(4,2)} = 0, 0, 1, 0, 2, 0, 3, 1, 1, 2, 1, 3, 2, 2, 3, 3$$

$$F^{(4,2)} = 0, 0, 1, 0, 2, 0, 3, 1, 1, 2, 1, 3, 2, 2, 3, 3$$

$$A^{(2)} = \frac{0}{4}, \frac{0}{4}, \frac{1}{4}, \frac{0}{4}, \frac{2}{4}, \frac{0}{4}, \frac{3}{4}, \frac{1}{4}, \frac{1}{4}, \frac{2}{4}, \frac{1}{4}, \frac{3}{4}, \frac{2}{4}, \frac{2}{4}, \frac{3}{4}, \frac{3}{4}$$

$$\begin{split} F^{(4,2)} &= 0, 0, 1, 0, 2, 0, 3, 1, 1, 2, 1, 3, 2, 2, 3, 3 \\ A^{(2)} &= \frac{0}{4}, \frac{0}{4}, \frac{1}{4}, \frac{0}{4}, \frac{2}{4}, \frac{0}{4}, \frac{3}{4}, \frac{1}{4}, \frac{1}{4}, \frac{2}{4}, \frac{1}{4}, \frac{3}{4}, \frac{2}{4}, \frac{2}{4}, \frac{3}{4}, \frac{3}{4} \end{split}$$

$$B^{(2)} &= \left\langle \underbrace{A^{(2)}; \dots; A^{(2)}}_{2 \times 2^{2 \times 2} = 32 \text{ veces}} \right\rangle$$

$$F^{(4,2)} = 0, 0, 1, 0, 2, 0, 3, 1, 1, 2, 1, 3, 2, 2, 3, 3$$

$$A^{(2)} = \frac{0}{4}, \frac{0}{4}, \frac{1}{4}, \frac{0}{4}, \frac{2}{4}, \frac{0}{4}, \frac{3}{4}, \frac{1}{4}, \frac{1}{4}, \frac{2}{4}, \frac{1}{4}, \frac{3}{4}, \frac{2}{4}, \frac{2}{4}, \frac{3}{4}, \frac{3}{4}$$

$$B^{(2)} = \left\langle \underbrace{A^{(2)}; \dots; A^{(2)}}_{2 \times 2^{2 \times 2} = 32 \text{ veces}} \right\rangle$$

$$= \underbrace{\frac{0}{4}, \frac{0}{4}, \dots, \frac{3}{4}, \frac{3}{4}}_{A^{(2)}}, \dots, \underbrace{\frac{0}{4}, \frac{0}{4}, \dots, \frac{3}{4}, \frac{3}{4}}_{A^{(2)}}.$$

Ahora sí, ya estamos en condiciones de definir la secuencia de Knuth.

Ahora sí, ya estamos en condiciones de definir la secuencia de Knuth.

Definición

La secuencia de Knuth, que denominamos K, se define como la concatenación de todas las posibles secuencias B en órden creciente:

Ahora sí, ya estamos en condiciones de definir la secuencia de Knuth.

Definición

La secuencia de Knuth, que denominamos K, se define como la concatenación de todas las posibles secuencias B en órden creciente:

$$K = \langle B^{(1)}; B^{(2)}; B^{(3)}; \dots \rangle.$$

Ahora sí, ya estamos en condiciones de definir la secuencia de Knuth.

Definición

La secuencia de Knuth, que denominamos K, se define como la concatenación de todas las posibles secuencias B en órden creciente:

$$K = \langle B^{(1)}; B^{(2)}; B^{(3)}; \dots \rangle.$$

Teorema (Knuth, 1965)

La secuencia K es completamente equidistribuida.

Knuth primero formula el lema 2 y reduce el problema a probar la k-distribución de una familia de secuencias de números enteros.

Knuth primero formula el lema 2 y reduce el problema a probar la k-distribución de una familia de secuencias de números enteros.

Para cada entero m, prueba que la secuencia $\lfloor 2^m K \rfloor$ de números enteros es k-distribuida para todo k. Esto implica que K también lo es.

Knuth primero formula el lema 2 y reduce el problema a probar la *k*-distribución de una familia de secuencias de números enteros.

Para cada entero m, prueba que la secuencia $\lfloor 2^m K \rfloor$ de números enteros es k-distribuida para todo k. Esto implica que K también lo es.

Definición

Fusce sit amet lacus viverra, viverra massa sit amet, placerat neque. Integer ipsum sapien, efficitur quis dui vitae, facilisis tempus dolor.

Knuth primero formula el lema 2 y reduce el problema a probar la *k*-distribución de una familia de secuencias de números enteros.

Para cada entero m, prueba que la secuencia $\lfloor 2^m K \rfloor$ de números enteros es k-distribuida para todo k. Esto implica que K también lo es.

Definición

Fusce sit amet lacus viverra, viverra massa sit amet, placerat neque. Integer ipsum sapien, efficitur quis dui vitae, facilisis tempus dolor.

Duis ornare volutpat libero, at sodales dolor porttitor at.

Knuth primero formula el lema 2 y reduce el problema a probar la *k*-distribución de una familia de secuencias de números enteros.

Para cada entero m, prueba que la secuencia $\lfloor 2^m K \rfloor$ de números enteros es k-distribuida para todo k. Esto implica que K también lo es.

Definición

Fusce sit amet lacus viverra, viverra massa sit amet, placerat neque. Integer ipsum sapien, efficitur quis dui vitae, facilisis tempus dolor.

Duis ornare volutpat libero, at sodales dolor porttitor at.

In rutrum dapibus justo, at mattis lacus ultrices sed. Suspendisse suscipit luctus fermentum.