Outline

- 1. Aplicación a la generación aleatoria
 - Método recursivo
 - Boltzmann samplers

Generación aleatoria

¿Por qué generación aleatoria?

- Testeo automático, pruebas de rendimiento (benchmarks),
- Analizar esatdísticamente algoritmos y estructuras de datos,
- Testear la validez de un modelo.

Muestreo uniforme

Consideremos una clase combinatoria ${\cal A}$ sin etiquetas.

Distribución uniforme

Sea $n \in \mathbb{Z}_{\geq 0}$. Queremos generar elementos de $\mathcal{A}_n = \{a \in \mathcal{A} : |a| = n\}$ según la distribución

$$\Pr_n(X=a) = \frac{1}{|\mathcal{A}_n|},$$

para cada $a \in \mathcal{A}_n$.

Muestreo uniforme

Consideremos una clase combinatoria ${\cal A}$ sin etiquetas.

Distribución uniforme

Sea $n \in \mathbb{Z}_{\geq 0}$. Queremos generar elementos de $\mathcal{A}_n = \{a \in \mathcal{A}: |a| = n\}$ según la distribución

$$\Pr_n(X=a) = \frac{1}{|\mathcal{A}_n|},$$

para cada $a \in \mathcal{A}_n$.

lacktriangle clase ${\cal A}$ dada por una especificación, consideramos **por ejemplo**

$$A = Z + Z \times A + Z \times A \times A$$

■ la talla n puede ser exacta, o aproximada $n \in [(1-\varepsilon)N, (1+\varepsilon)N]$.

Método recursivo: ejemplo

La ecuación $\mathcal{A} = \mathcal{Z} + \mathcal{Z} \times \mathcal{A} + \mathcal{Z} \times \mathcal{A} \times \mathcal{A}$ se traduce en

$$a_n = \underbrace{\mathbf{1}_{n=1}}_{\mathcal{Z}} + \underbrace{a_{n-1}}_{\mathcal{Z} \times \mathcal{A}} + \underbrace{\sum_{k=0}^{n-1} a_k a_{n-1-k}}_{\mathcal{Z} \times \mathcal{A} \times \mathcal{A}}.$$

Método recursivo: ejemplo

La ecuación $\mathcal{A} = \mathcal{Z} + \mathcal{Z} \times \mathcal{A} + \mathcal{Z} \times \mathcal{A} \times \mathcal{A}$ se traduce en

$$a_n = \underbrace{\mathbf{1}_{n=1}}_{\mathcal{Z}} + \underbrace{a_{n-1}}_{\mathcal{Z} \times \mathcal{A}} + \underbrace{\sum_{k=0}^{n-1} a_k a_{n-1-k}}_{\mathcal{Z} \times \mathcal{A} \times \mathcal{A}}.$$

Para un elemento de talla n:

lacksquare La probabilidad que provenga de $\mathcal Z$ es

$$\mathbf{1}_{n=1}/a_1 = \mathbf{1}_{n=1}$$
.

■ La probabilidad que provenga de $\mathbb{Z} \times \mathcal{A}$ es

$$\frac{a_{n-1}}{a_n} \, .$$

■ La probabilidad que provenga de $\mathcal{Z} \times \mathcal{A}_k \times \mathcal{A}_{n-k}$ es

$$\frac{1}{a_n}a_k a_{n-1-k} \, .$$

Método recursivo: ejemplo

La ecuación $\mathcal{A} = \mathcal{Z} + \mathcal{Z} \times \mathcal{A} + \mathcal{Z} \times \mathcal{A} \times \mathcal{A}$ se traduce en

$$a_n = \underbrace{\mathbf{1}_{n=1}}_{\mathcal{Z}} + \underbrace{a_{n-1}}_{\mathcal{Z} \times \mathcal{A}} + \underbrace{\sum_{k=0}^{n-1} a_k a_{n-1-k}}_{\mathcal{Z} \times \mathcal{A} \times \mathcal{A}}.$$

Para un elemento de talla n:

 \blacksquare La probabilidad que provenga de $\mathcal Z$ es

$$\mathbf{1}_{n=1}/a_1 = \mathbf{1}_{n=1}$$
.

■ La probabilidad que provenga de $\mathbb{Z} \times \mathcal{A}$ es

$$\frac{a_{n-1}}{a_n}$$

■ La probabilidad que provenga de $\mathcal{Z} \times \mathcal{A}_k \times \mathcal{A}_{n-k}$ es

$$\frac{1}{a_n}a_ka_{n-1-k}.$$

Luego se continua recursivamente (e independientemente) con cada factor.

Para clases \mathcal{A} , \mathcal{B} y \mathcal{C} , denotamos sus generadores uniformes $\Gamma A[i]$, $\Gamma B[i]$, $\Gamma C[i]$.

Método recursivo: suma

Si $C = A + B \operatorname{con} A \cap B = \emptyset$:

$$\Gamma C[n] = \begin{cases} \Gamma A[n] & \text{con proba. } \frac{a_n}{c_n} = \frac{a_n}{a_n + b_n}, \\ \Gamma B[n] & \text{con proba. } \frac{b_n}{c_n} = \frac{b_n}{a_n + b_n}. \end{cases}$$

Método recursivo: producto

Si $C = A \times B$:

$$\Gamma C[n] = (\Gamma A[k], \Gamma B[n-k])$$
 con proba. $\frac{a_k b_{n-k}}{c_n}$.

Primero elegimos k con $Pr(k=j) = \frac{a_j b_{n-j}}{c_n}$.

Teorema

El método recursivo produce elementos de tamaño exacto dado n con la distribución uniforme.

Teorema

El método recursivo produce elementos de tamaño exacto dado n con la distribución uniforme.

Necesario precalcular $a_0, a_1, a_2, \ldots, a_n$!

Problema

¿Cuál es el costo esperado para extraer un elemento de talla n de $\mathcal{A} = \mathcal{Z} + \mathcal{Z} \times \mathcal{A} + \mathcal{Z} \times \mathcal{A} \times \mathcal{A}$, suponiendo dados los coeficientes a_0, \dots, a_n ?

Teorema

El método recursivo produce elementos de tamaño exacto dado n con la distribución uniforme.

Necesario precalcular $a_0, a_1, a_2, \ldots, a_n$!

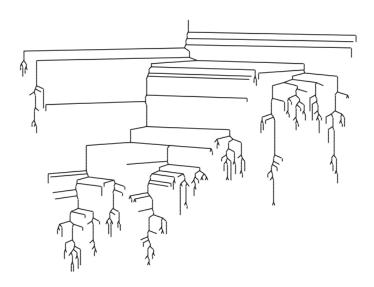
Problema

¿Cuál es el costo esperado para extraer un elemento de talla n de $\mathcal{A} = \mathcal{Z} + \mathcal{Z} \times \mathcal{A} + \mathcal{Z} \times \mathcal{A} \times \mathcal{A}$, suponiendo dados los coeficientes a_0, \dots, a_n ?^a

Se puede modelar con funciones generatrices!

^aRespuesta $\Theta(n^{3/2})$.

Ejemplo de árbol uniforme de \mathcal{A} de talla n = 1000



Desventaja principal del método recursivo: necesario conocer la secuencia de conteo (costoso calcular, números grandes en memoria).

Desventaja principal del método recursivo: necesario conocer la secuencia de conteo (costoso calcular, números grandes en memoria).

El método de **Boltzmann** considera

- ullet Sólo las evaluaciones de las funciones generatrices en un punto heta.
- El tamaño del objeto producido es una variable aleatoria, pero se puede controlar con el parámetro θ .
- Permite muestreo exacto y muestreo aproximado (más rápido!).
- Muy general: una especificación se traduce en un generador.

Desventaja principal del método recursivo: necesario conocer la secuencia de conteo (costoso calcular, números grandes en memoria).

El método de **Boltzmann** considera

- ullet Sólo las evaluaciones de las funciones generatrices en un punto heta.
- El tamaño del objeto producido es una variable aleatoria, pero se puede controlar con el parámetro θ .
- Permite muestreo exacto y muestreo aproximado (más rápido!).
- Muy general: una especificación se traduce en un generador.

Fijamos θ tal que la suma $A(\theta)$ converge y definimos:

$$\Pr(\Gamma A(\theta) = a) = \frac{\theta^{|a|}}{A(\theta)},$$

para todo $a \in \mathcal{A}$.

Un punto clave del muestreo de Boltzmann es el siguiente:

Proposición

La distribución condicionada a un tamaño dado n es uniforme:

$$\Pr\left(\Gamma A(\theta) = a \mid |\Gamma A(\theta)| = n\right) = \frac{1}{a_n},$$

para todo $a \in \mathcal{A}_n$.

Un punto clave del muestreo de Boltzmann es el siguiente:

Proposición

La distribución condicionada a un tamaño dado n es uniforme:

$$\Pr\left(\Gamma A(\theta) = a \mid |\Gamma A(\theta)| = n\right) = \frac{1}{a_n},$$

para todo $a \in \mathcal{A}_n$.

¿Cómo tomar una muestra de la distribución de Boltzmann?

Muestreo de Boltzmann: ejemplo

La ecuación $\mathcal{A} = \mathcal{Z} + \mathcal{Z} \times \mathcal{A} + \mathcal{Z} \times \mathcal{A} \times \mathcal{A}$ se traduce en

$$A(\theta) = \underbrace{\theta}_{\mathcal{Z}} + \underbrace{\theta A(\theta)}_{\mathcal{Z} \times \mathcal{A}} + \underbrace{\theta \cdot A(\theta) \cdot A(\theta)}_{\mathcal{Z} \times \mathcal{A} \times \mathcal{A}}.$$

Muestreo de Boltzmann: ejemplo

La ecuación $\mathcal{A} = \mathcal{Z} + \mathcal{Z} \times \mathcal{A} + \mathcal{Z} \times \mathcal{A} \times \mathcal{A}$ se traduce en

$$A(\theta) = \underbrace{\theta}_{\mathcal{Z}} + \underbrace{\theta A(\theta)}_{\mathcal{Z} \times \mathcal{A}} + \underbrace{\theta \cdot A(\theta) \cdot A(\theta)}_{\mathcal{Z} \times \mathcal{A} \times \mathcal{A}}.$$

Para un elemento según $\Gamma A(\theta)$

lacksquare Será igual a $\mathcal Z$ con proba.

$$\frac{\theta}{A(\theta)}$$
.

■ Proviene de $\mathcal{Z} \times \mathcal{A}$ con proba.

$$\frac{\theta A(\theta)}{A(\theta)}$$

lacksquare Proviene de $\mathcal{Z} \times \mathcal{A} \times \mathcal{A}$ es

$$\frac{\theta A(\theta)^2}{A(\theta)}.$$

Muestreo de Boltzmann: ejemplo

La ecuación $\mathcal{A} = \mathcal{Z} + \mathcal{Z} \times \mathcal{A} + \mathcal{Z} \times \mathcal{A} \times \mathcal{A}$ se traduce en

$$A(\theta) = \underbrace{\theta}_{\mathcal{Z}} + \underbrace{\theta A(\theta)}_{\mathcal{Z} \times \mathcal{A}} + \underbrace{\theta \cdot A(\theta) \cdot A(\theta)}_{\mathcal{Z} \times \mathcal{A} \times \mathcal{A}}.$$

Para un elemento según $\Gamma A(\theta)$

• Será igual a \mathcal{Z} con proba.

$$\frac{\theta}{A(\theta)}$$
.

■ Proviene de $\mathcal{Z} \times \mathcal{A}$ con proba.

$$\frac{\theta A(\theta)}{A(\theta)}$$
,

■ Proviene de $\mathcal{Z} \times \mathcal{A} \times \mathcal{A}$ es

$$\frac{\theta A(\theta)^2}{A(\theta)}$$

Luego se continua recursivamente (e independientemente).

Muestreo de Boltzmann: construcciones

Muestreo de Boltzmann: suma

Si
$$C = A + B \operatorname{con} A \cap B = \emptyset$$
:

$$\Gamma C(\theta) = \begin{cases} \Gamma A(\theta) & \text{con proba. } \frac{A(\theta)}{C(\theta)} = \frac{A(\theta)}{A(\theta) + B(\theta)}, \\ \Gamma B(\theta) & \text{con proba. } \frac{B(\theta)}{C(\theta)} = \frac{B(\theta)}{A(\theta) + B(\theta)}. \end{cases}$$

Muestreo de Boltzmann: producto

Si
$$C = A \times B$$
:

$$\Gamma C(\theta) = (\Gamma A(\theta), \Gamma B(\theta)).$$

Muestreo de Boltzmann: construcciones

Muestreo de Boltzmann: suma

Si
$$C = A + B \operatorname{con} A \cap B = \emptyset$$
:

$$\Gamma C(\theta) = \begin{cases} \Gamma A(\theta) & \text{con proba. } \frac{A(\theta)}{C(\theta)} = \frac{A(\theta)}{A(\theta) + B(\theta)}, \\ \Gamma B(\theta) & \text{con proba. } \frac{B(\theta)}{C(\theta)} = \frac{B(\theta)}{A(\theta) + B(\theta)}. \end{cases}$$

Muestreo de Boltzmann: producto

Si
$$C = A \times B$$
:

$$\Gamma C(\theta) = (\Gamma A(\theta), \Gamma B(\theta)).$$

Proposición

Las construcciones producen elementos con la distribución de Boltzmann.

Muestreo de Boltzmann: construcciones II

Escribimos $X \sim \text{Geom}(\lambda)$ sii $\Pr(X = k) = \lambda^k (1 - \lambda)$ para $k \ge 0$.

Secuencia

Si
$$C = Seq(A)$$
: tomar $k \leftarrow Geom(A(\theta))$ y

$$\Gamma C(\theta) \leftarrow (\underbrace{\Gamma A(\theta), \dots, \Gamma A(\theta)}_{k \text{ independientes}}).$$

Muestreo de Boltzmann: construcciones II

Escribimos $X \sim \text{Geom}(\lambda)$ sii $\Pr(X = k) = \lambda^k (1 - \lambda)$ para $k \ge 0$.

Secuencia

Si C = Seq(A): tomar $k \leftarrow Geom(A(\theta))$ y

$$\Gamma C(\theta) \leftarrow (\underbrace{\Gamma A(\theta), \dots, \Gamma A(\theta)}_{k \text{ independientes}}).$$

Ejemplo

Las particiones enteras con partes $\leq K$ están especificadas por

$$Seq(1) \times Seq(2) \times ... \times Seq(K)$$
, $|i| = i$.

Muestreo de Boltzmann: construcciones II

Escribimos $X \sim \text{Geom}(\lambda)$ sii $\Pr(X = k) = \lambda^k (1 - \lambda)$ para $k \ge 0$.

Secuencia

Si C = Seq(A): tomar $k \leftarrow Geom(A(\theta))$ y

$$\Gamma C(\theta) \leftarrow (\underbrace{\Gamma A(\theta), \dots, \Gamma A(\theta)}_{k \text{ independientes}}).$$

Ejemplo

Las particiones enteras con partes $\leq K$ están especificadas por

$$Seq(1) \times Seq(2) \times ... \times Seq(K)$$
, $|i| = i$.

Para muestrear con Boltzmann (param. θ) consideramos

$$k_1 \leftarrow \text{Geom}\left(\theta^1\right), k_2 \leftarrow \text{Geom}\left(\theta^2\right), \dots k_K \leftarrow \text{Geom}\left(\theta^K\right).$$

Muestro de Boltzmann: complejidad y tamaño

Teorema

Sea \mathcal{A} una clase especificada en términos de $\{+, \times, \operatorname{Seq}\}$ y $\{\mathcal{E}, \mathcal{Z}\}$. La generación de un objeto $a \in \mathcal{A}$ requiere tiempo O(|a|).

^aSuponemos que podemos calcular perfectamente las series.

Muestro de Boltzmann: complejidad y tamaño

Teorema

Sea \mathcal{A} una clase especificada en términos de $\{+, \times, \operatorname{Seq}\}$ y $\{\mathcal{E}, \mathcal{Z}\}$. La generación de un objeto $a \in \mathcal{A}$ requiere tiempo^a O(|a|).

^aSuponemos que podemos calcular perfectamente las series.

Sea N = |a| para a tomado con la distribución de Boltzmann de A,

$$\mathbb{E}_{\theta}[N] = \sum_{a \in \mathcal{A}} |a| \frac{\theta^a}{A(\theta)} = \frac{\theta A'(\theta)}{A(\theta)} ,$$

$$\mathbb{E}_{\theta}[N^2] = \sum_{a \in \mathcal{A}} |a|^2 \frac{\theta^a}{A(\theta)} = \frac{\theta A'(\theta) + \theta^2 A''(\theta)}{A(\theta)}.$$

Problema. Encontrar $a \in \mathcal{A}$ con $|a| \in I_n(\varepsilon) = [(1 - \varepsilon)n, (1 + \varepsilon)n]$ para $\varepsilon > 0$ fijo:

- el muestreo aproximado es más rápido que el exacto,
- útil si propiedades de objetos de tamaños cercanos son similares.

Problema. Encontrar $a \in \mathcal{A}$ con $|a| \in I_n(\varepsilon) = [(1 - \varepsilon)n, (1 + \varepsilon)n]$ para $\varepsilon > 0$ fijo:

- el muestreo aproximado es más rápido que el exacto,
- útil si propiedades de objetos de tamaños cercanos son similares.

Procedimiento de Boltzmann con rechazo. $\mu A(\theta; I)$

- 1. Rechazo temprano: si el objeto que estamos produciendo con $\Gamma A(\theta)$ supera $M = \max I$ durante la construcción, $\Gamma^{\leq M} A(\theta) := \mathbf{nil}$.
- 2. Producir elementos usando $\Gamma^{\leq M}A(\theta)$, rechazando hasta producir un elemento que tenga talla en nuestro target set I y aceptar.

```
[importante: si falla (nil) \Rightarrow recomenzar de 0]
```

Problema. Encontrar $a \in \mathcal{A}$ con $|a| \in I_n(\varepsilon) = [(1 - \varepsilon)n, (1 + \varepsilon)n]$ para $\varepsilon > 0$ fijo:

- el muestreo aproximado es más rápido que el exacto,
- útil si propiedades de objetos de tamaños cercanos son similares.

Procedimiento de Boltzmann con rechazo. $\mu A(\theta; I)$

- 1. Rechazo temprano: si el objeto que estamos produciendo con $\Gamma A(\theta)$ supera $M = \max I$ durante la construcción, $\Gamma^{\leq M} A(\theta) := \mathbf{nil}$.
- 2. Producir elementos usando $\Gamma^{\leq M}A(\theta)$, rechazando hasta producir un elemento que tenga talla en nuestro target set I y aceptar. [importante: si falla (nil) \Rightarrow recomenzar de 0]

Tuning paramétro. Encontrar θ_n tal que $\mathbb{E}_{\theta_n}[N] = n$, i.e., $\frac{\theta_n A'(\theta_n)}{A(\theta_n)} = n$.

Complejidad de $\mu A(\theta_n; I_n(\varepsilon))$ depende de la distribución de N:

- $\Theta(n)$ cuando N es concentrada cuando $n \to \infty$ [Chebyshev],
- $-\Theta(n)$ si $[z^n]A(z) \sim c n^{\alpha-1}R^{-n}$ con $\alpha > 0$,
- pero $\Theta(n^2)$ para $\mathcal{A} = \mathcal{Z} + \mathcal{Z} \times \mathcal{A} + \mathcal{Z} \times \mathcal{A} \times \mathcal{A}$.

Complejidad de $\mu A(\theta_n; I_n(\varepsilon))$ depende de la distribución de N:

- $-\Theta(n)$ cuando N es concentrada cuando $n \to \infty$ [Chebyshev],
- $-\Theta(n)$ si $[z^n]A(z) \sim c n^{\alpha-1}R^{-n}$ con $\alpha > 0$,
- pero $\Theta(n^2)$ para $\mathcal{A}=\mathcal{Z}+\mathcal{Z}\times\mathcal{A}+\mathcal{Z}\times\mathcal{A}\times\mathcal{A}$. ¿ mejor complejidad ?

Singular samplers

La ecuación $\mathcal{A} = \mathcal{Z} + \mathcal{Z} \times \mathcal{A} + \mathcal{Z} \times \mathcal{A} \times \mathcal{A}$ se traduce en

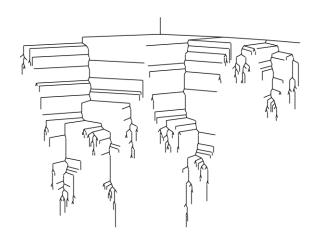
$$A(z) = \frac{1 - z - \sqrt{(z+1) \cdot (1-3z)}}{2z}.$$

- Singularidad dominante $\rho = 1/3$, donde $A(\rho) = 1 < \infty$.
- Se puede aplicar Boltzmann con $\theta = \rho$. Notar $\mathbb{E}[N] = \infty$.
- Considerar $\theta = \rho$ en lugar de $\theta = \theta_n$, $\mu A(\rho; I_n)$ es un singular sampler.

Ejemplo: generador $A = Z + Z \times A + Z \times A \times A$

```
def randA(tS) :
    if tS <= 0 :
    return False, None, None
   tS -= 1
    x = rng.random()
    if x < theta :
       return True,(), 1
    found,1,sl = randA(tS)
    if not found :
        return False . None . None
    if x < theta + theta * eA :
        return True,((),1),sl+1
    found, r, sr = randA(tS-sl)
    if not found :
        return False, None, None
    return True, ((),1,r),1+s1+sr
while True :
    found, tree, size = randA(MAX)
    if found and size >= MIN :
        print(size)
        break
```

Ejemplo de un árbol aproximado (1023 nodos)



Singular sampler II

Para nuestros árboles unarios-binarios ${\cal A}$ tenemos

Proposición

Existe c > 0 tal que $\Pr(|\Gamma A(\rho)| = k) \sim c \times k^{-3/2}$.

Singular sampler II

Para nuestros árboles unarios-binarios ${\cal A}$ tenemos

Proposición

Existe c > 0 tal que $\Pr(|\Gamma A(\rho)| = k) \sim c \times k^{-3/2}$.

Demostración.

Aplicar el Teorema de Transferencia para estimar a_k .

Singular sampler II

Para nuestros árboles unarios-binarios ${\cal A}$ tenemos

Proposición

Existe c > 0 tal que $\Pr(|\Gamma A(\rho)| = k) \sim c \times k^{-3/2}$.

Demostración.

Aplicar el Teorema de Transferencia para estimar a_k .

Sumando deducimos

$$\Pr(|\Gamma A(\rho)| \in [(1-\varepsilon)n, (1+\varepsilon)n]) = \Theta(n^{-1/2}).$$

■ En media $\Theta(\sqrt{n})$ iteraciones para obtener

$$|\Gamma A(\rho)| \in [(1-\varepsilon)n, (1+\varepsilon)n].$$

lacksquare A priori la complejidad global es $O(n^{3/2})$.

Singular samplers

Un análisis más preciso, que considera los tamaños de los objetos rechazados produce:

Teorema (Duchon, Flajolet, Louchard, Schaeffer, '04 – Theorem 7.3)

Para una clase definida recursivamente por una ecuación^a, el singular sampler aproximado $\mu A(\rho;I_n(\varepsilon))$ tiene una complejidad acumulada, sumando la complejidad de todos los rechazos, de $\Theta(n)$.

^aTienen el mismo tipo de singularidad. Se puede extender a un sistema.

Muestreo con tamaño exacto

Si queremos producir un elemento uniforme de ${\cal A}$ con tamaño exacto n, este es un problema más complicado. La probabilidad es menor:

$$\Pr(|\Gamma A(\theta)| = n) = \frac{a_n \theta^n}{A(\theta)}$$

Muestreo con tamaño exacto

Si queremos producir un elemento uniforme de $\mathcal A$ con tamaño exacto n, este es un problema más complicado. La probabilidad es menor:

$$\Pr(|\Gamma A(\theta)| = n) = \frac{a_n \theta^n}{A(\theta)}$$

Por ejemplo, para los singular samplers tenemos:

Teorema (Duchon, Flajolet, Louchard, Schaeffer, '04 – Theorem 7.3)

Para una clase definida recursivamente por una ecuación^a, el singular sampler para muestreo exacto $\mu A(\rho;\{n\})$ tiene una complejidad acumulada, sumando la complejidad de todos los rechazos, de $\Theta(n^2)$.

^aTienen el mismo tipo de singularidad. Se puede extender a un sistema.

Para aprender más

- Philippe Duchon. Random generation of combinatorial structures: Boltzmann samplers and beyond. https://arxiv.org/abs/1112.5071
- Philippe Duchon, Philippe Flajolet, Guy Louchard y Gilles Schaeffer Boltzmann Samplers for the Random Generation of Combinatorial Structures. https://doi.org/10.1017/S0963548304006315
- Philippe Flajolet, Éric Fusy y Carine Pivoteau Boltzmann Sampling of Unlabelled Structures. https://doi.org/10.1137/1.9781611972979.5
- Albert Nijenhuis y Herbert S. Wilf Combinatorial Algorithms.