Estimación insesgada, alocación de bolas en urnas y árboles aleatorios

Fernando López-Blázquez

Sanlúcar de Barrameda, Cádiz
28–30 Noviembre 2014

Índice

1. Estimación insesgada
2. Modelos de alocación aleatoria de bolas en urnas
3. Árboles aleatorios

Modelo básico

- $X_1, \ldots, X_N \overset{iid}{\sim} F(x; \theta)$
- Objetivo: Estimación puntual
  Encontrar una función de la muestra (datos) que *aproxime razonablemente* a $H(\theta)$, i.e.,

$$T_N(X_1, \ldots, X_N) \approx H(\theta)$$

Modelo básico

- $X_1, \ldots, X_N \overset{iid}{\sim} F(x;\theta)$
- Objetivo: Estimación puntual

  Encontrar una función de la muestra (datos) que *aproxime razonablemente* a $H(\theta)$, i.e.,

$$T_N(X_1, \ldots, X_N) \approx H(\theta)$$

Modelo básico

- $X_1, \ldots, X_N \overset{iid}{\sim} F(x;\theta)$
- Objetivo: Estimación puntual
  Encontrar una función de la muestra (datos) que *aproxime razonablemente* a $H(\theta)$, i.e.,

$$T_N(X_1, \ldots, X_N) \approx H(\theta)$$

Estimación insesgada

Dado $H(\theta)$ encontrar $T_N = T_N(X_1, \ldots, X_N)$ tal que:

$$\mathbb{E}_\theta T_N = H(\theta), \ \forall \theta$$

Teoría B-R-L-S

- Un estadístico $S_N = S_N(X_1, \ldots, X_N)$ es suficiente si

$$(X_1, \cdots, X_N) \mid S_N \ \text{no depende de } \theta$$

- Un estadístico $S_N = S_N(X_1, \ldots, X_N)$ es completo

$$\mathbb{E}_\theta g(S_N) = 0, \ \forall \theta \ \Rightarrow g = 0, \ \mathbb{P}_\theta - \text{c.s.}$$

Teoría B-R-L-S

- Si la familia de distribuciones $F(\cdot, \theta)$ admite un estadístico suficiente y completo, los estimadores insesgados de una función paramétrica $H(\theta)$ deben ser obtenidos, si fuera posible, como funciones del estadístico suficiente y completo:

$$\mathbb{E}_\theta T_N(S_N) = H(\theta), \ \forall \theta$$

Ejemplo

- Sea $X_1, \ldots, X_N \overset{iid}{\sim} \text{Po}(\theta)$ deseamos estimar $H(\theta) = \mathbb{P}_\theta(X = 0) = e^{-\theta}$.
- Un estadístico suficiente y completo es

$$S_N = \sum_{i=1}^{N} X_i \sim \text{Po}(N\theta)$$

- El estimador insesgado se obtiene resolviendo

$$E_\theta T_N(S_N) = e^{-\theta}$$

Ejemplo

- Sea $X_1, \ldots, X_N \stackrel{iid}{\sim} \text{Po}(\theta)$ deseamos estimar $H(\theta) = \mathbb{P}_\theta(X = 0) = e^{-\theta}$.
- Un estadístico suficiente y completo es

$$S_N = \sum_{i=1}^{N} X_i \sim \text{Po}(N\theta)$$

- El estimador insesgado se obtiene resolviendo

$$E_\theta T_N(S_N) = e^{-\theta}$$

Ejemplo

- Sea $X_1, \dots, X_N \overset{iid}{\sim} \text{Po}(\theta)$ deseamos estimar $H(\theta) = \mathbb{P}_\theta(X = 0) = e^{-\theta}$.
- Un estadístico suficiente y completo es

$$S_N = \sum_{i=1}^{N} X_i \sim \text{Po}(N\theta)$$

- El estimador insesgado se obtiene resolviendo

$$E_\theta T_N(S_N) = e^{-\theta}$$

Ejemplo - Continuación

- Es decir:

$$\sum_{n=0}^{\infty} T_N(n) e^{-N\theta} \frac{(N\theta)^n}{n!} = e^{-\theta}$$

- luego

$$T_N(n) = \left(1 - \frac{1}{N}\right)^n$$

- y es estimador buscado es

$$T_N = \left(1 - \frac{1}{N}\right)^{\sum_{i=1}^{N} X_i}$$

Ejemplo - Continuación

- Es decir:

$$\sum_{n=0}^{\infty} T_N(n) e^{-N\theta} \frac{(N\theta)^n}{n!} = e^{-\theta}$$

- luego

$$T_N(n) = \left(1 - \frac{1}{N}\right)^n$$

- y es estimador buscado es

$$T_N = \left(1 - \frac{1}{N}\right)^{\sum_{i=1}^{N} X_i}$$

Ejemplo - Continuación

- Es decir:

$$\sum_{n=0}^{\infty} T_N(n) e^{-N\theta} \frac{(N\theta)^n}{n!} = e^{-\theta}$$

- luego

$$T_N(n) = \left(1 - \frac{1}{N}\right)^n$$

- y es estimador buscado es

$$T_N = \left(1 - \frac{1}{N}\right)^{\sum_{i=1}^{N} X_i}$$

Distribuciones de series de potencias (PSD)

Definición

Sea $G(\theta) = \sum_{k\geq 0} g_k \theta^k$ con $g_k \geq 0$, $\forall k \geq 0$. Entonces,

$$X \sim PSD\,(G(\theta)) \;\Leftrightarrow\; \mathbb{P}_\theta(X = k) = \frac{g_k \theta^k}{G(\theta)},\;\; k \geq 0$$

Ejemplos

$$\begin{aligned}
X \sim Po(\theta) &\quad\Leftrightarrow\quad X \sim PSD\left(e^{\theta}\right)\\
X \sim Ge(\theta) &\quad\Leftrightarrow\quad X \sim PSD\left((1-\theta)^{-1}\right)\\
X \sim BN(r,\theta) &\quad\Leftrightarrow\quad X \sim PSD\left((1-\theta)^{-r}\right)\\
X \sim Bi(N,\theta) &\quad\Leftrightarrow\quad X \sim PSD\left((1+\theta)^{N}\right)
\end{aligned}$$

Distribuciones de series de potencias (PSD)

Definición

Sea $G(\theta) = \sum_{k\geq 0} g_k \theta^k$ con $g_k \geq 0$, $\forall k \geq 0$. Entonces,

$$X \sim PSD\left(G(\theta)\right) \Leftrightarrow \mathbb{P}_\theta(X = k) = \frac{g_k \theta^k}{G(\theta)}, \quad k \geq 0$$

Ejemplos

$$\begin{array}{lcl}
X \sim Po(\theta) & \Leftrightarrow & X \sim PSD\left(e^\theta\right) \\
X \sim Ge(\theta) & \Leftrightarrow & X \sim PSD\left((1-\theta)^{-1}\right) \\
X \sim BN(r,\theta) & \Leftrightarrow & X \sim PSD\left((1-\theta)^{-r}\right) \\
X \sim Bi(N,\theta) & \Leftrightarrow & X \sim PSD\left((1+\theta)^N\right)
\end{array}$$

Estimación insesgada en PSD

- $$\left.\begin{array}{l} X_j \overset{iid}{\sim} PSD\,(G(\theta)) \\ independientes,\ j = 1, \ldots, N \end{array}\right\} \Rightarrow S_N = \sum_{j=1}^{N} X_j \sim PSD\left(G^N(\theta)\right)$$

  es decir:

  $$\mathbb{P}_\theta(S_N = n) = \frac{g_{N,n}\theta^n}{G^N(\theta)},\ n \geq 0,\ \theta \geq 0$$

- S_N es suficiente y completo

Estimación insesgada en PSD

- Si buscamos un estimador insesgado, T_N de $H(\theta)$ debemos resolver:

$$\sum_{n\geq 0} T_N(n)\frac{g_{N,n}\theta^n}{G^N(\theta)} = H(\theta), \ \forall\theta$$

La transformación PSD

- La relación anterior puede entenderse como un funcional:

$$T_N \;\circ\!\!-\!\!\!-\!\!\!-\!\!\!-\!\!\bullet\; H$$

  que denominaremos transformación PSD
- *Obtener un estimador insesgado consiste básicamente en invertir la transformación PSD.*

La transformación PSD

- La relación anterior puede entenderse como un funcional:

$$T_N \circ\!\!-\!\!\!-\!\!\!-\!\!\!-\!\!\bullet\, H$$

  que denominaremos transformación PSD
- *Obtener un estimador insesgado consiste básicamente en invertir la transformación PSD*.

Alocación aleatoria

- Sean N urnas distinguibles (numeradas)
- Se colocarán n bolas en las urnas.

- El conjunto de todas las posibles alocaciones es:

$$\Omega_{n,N} = \left\{ \omega = (y_1, \cdots, y_N) \in \mathbb{N}_0^N : \sum_{i=1}^{n} y_i = n \right\}$$

Alocación aleatoria

- Sean N urnas distinguibles (numeradas)
- Se colocarán n bolas en las urnas.
- El conjunto de todas las posibles alocaciones es:

$$\Omega_{n,N} = \left\{ \omega = (y_1, \cdots, y_N) \in \mathbb{N}_0^N : \sum_{i=1}^{n} y_i = n \right\}$$

Alocación aleatoria

- Sea $\{w_k\}_{k\geq 0}$ una sucesión de números reales no negativos.
- Para cada alocación $\omega = (y_1, \ldots, y_N)$ se define su peso:

$$\mathbf{w}(\omega) := \prod_{j=1}^{N} w_{y_j}$$

Alocación aleatoria

- El modelo de alocación aleatoria consiste en seleccionar una alocación en $\Omega_{n,N}$ con probabilidad proporcional al peso:

$$\mathbb{P}(Y_1 = y_1, \ldots, Y_N = y_N) \propto \mathbf{w}(\omega)$$

- es decir:

$$\mathbb{P}(Y_1 = y_1, \ldots, Y_N = y_N) = \frac{\prod_{j=1}^{N} w_{y_j}}{Z(N, n)}$$

donde $Z(N, n) = \sum_{\mathbf{y} \in \Omega_{n,N}} \prod_{j=1}^{N} w_{y_j}$ (función de partición)

Resultado fundamental

- Sea $(Y_1, \ldots, Y_N)$ es una alocación aleatoria en $\Omega_{n,N}$ (N urnas y n bolas) con pesos asociados $\{w_k\}_{k\geq 0}$.
- Sean $X_1, \ldots, X_N \overset{\text{iid}}{\sim} PSD(G(\theta))$ con $G(\theta) = \sum_k w_k \theta^k$ y $S_N = X_1 + \cdots + X_N$

Teorema

$$(Y_1, \ldots, Y_N) \overset{d}{=} (X_1, \ldots, X_N) \mid S_N = n$$

Resultado fundamental

- Sea $(Y_1, \ldots, Y_N)$ es una alocación aleatoria en $\Omega_{n,N}$ (N urnas y n bolas) con pesos asociados $\{w_k\}_{k\geq 0}$.
- Sean $X_1, \ldots, X_N \overset{\text{iid}}{\sim} PSD(G(\theta))$ con $G(\theta) = \sum_k w_k \theta^k$ y $S_N = X_1 + \cdots + X_N$

Teorema

$$(Y_1, \ldots, Y_N) \stackrel{d}{=} (X_1, \ldots, X_N) \mid S_N = n$$

Cantidades de interés en modelos de urnas

La mayoría son de la forma:

$$E\left(g\left(Y_1,\ldots,Y_N\right)\right),\ \ g \text{ medible}$$

- **Ejemplos**
  - Función de probabilidad conjunta:

    $$P(Y_1 = y_1,\ldots,Y_N = y_N) = E\left(I\left(Y_1 = y_1,\ldots,Y_N = y_N\right)\right)$$

  - Probabilidades marginales: $P(Y_j = y_j) = E\left(I\left(Y_j = y_j\right)\right)$
  - Momentos: EY_j^k
  - Distribución de Multiplicidades

    $$\mathbb{P}\left(\sum_i \mathbf{1}\{Y_i = r\} = j\right)$$

  - Gaps, colisiones, maximas ocupaciones,...etc

Cantidades de interés en modelos de urnas

La mayoría son de la forma:

$$E\left(g\left(Y_1, \ldots, Y_N\right)\right), \quad g \text{ medible}$$

- **Ejemplos**
  - Función de probabilidad conjunta:

$$P(Y_1 = y_1, \ldots, Y_N = y_N) = E\left(I\left(Y_1 = y_1, \ldots, Y_N = y_N\right)\right)$$

  - Probabilidades marginales: $P(Y_j = y_j) = E\left(I\left(Y_j = y_j\right)\right)$
  - Momentos: EY_j^k
  - Distribución de Multiplicidades

$$\mathbb{P}\left(\sum_i \mathbf{1}\{Y_i = r\} = j\right)$$

  - Gaps, colisiones, maximas ocupaciones,...etc

Cantidades de interés en modelos de urnas

La mayoría son de la forma:

$$E\left(g\left(Y_1, \ldots, Y_N\right)\right), \ \ g \text{ medible}$$

- **Ejemplos**
  - Función de probabilidad conjunta:

    $$P(Y_1 = y_1, \ldots, Y_N = y_N) = E\left(I\left(Y_1 = y_1, \ldots, Y_N = y_N\right)\right)$$

  - Probabilidades marginales: $P(Y_j = y_j) = E\left(I\left(Y_j = y_j\right)\right)$
  - Momentos: EY_j^k
  - Distribución de Multiplicidades

    $$\mathbb{P}\Big(\sum_i \mathbf{1}\{Y_i = r\} = j\Big)$$

  - Gaps, colisiones, maximas ocupaciones,...etc

Cantidades de interés en modelos de urnas

La mayoría son de la forma:

$$E\left(g\left(Y_1,\ldots,Y_N\right)\right),\ \ g\,\text{medible}$$

- **Ejemplos**
  - Función de probabilidad conjunta:

$$P(Y_1=y_1,\ldots,Y_N=y_N)=E\left(I\left(Y_1=y_1,\ldots,Y_N=y_N\right)\right)$$

  - Probabilidades marginales: $P(Y_j=y_j)=E\left(I\left(Y_j=y_j\right)\right)$
  - Momentos: EY_j^k
  - Distribución de Multiplicidades

$$\mathbb{P}(\sum_i \mathbf{1}\{Y_i=r\}=j)$$

  - Gaps, colisiones, maximas ocupaciones,...etc

Cantidades de interés en modelos de urnas

La mayoría son de la forma:

$$E\left(g\left(Y_1,\ldots,Y_N\right)\right),\ \ g \text{ medible}$$

- **Ejemplos**
  - Función de probabilidad conjunta:

$$P(Y_1=y_1,\ldots,Y_N=y_N)=E\left(I\left(Y_1=y_1,\ldots,Y_N=y_N\right)\right)$$

  - Probabilidades marginales: $P(Y_j=y_j)=E\left(I\left(Y_j=y_j\right)\right)$
  - Momentos: EY_j^k
  - Distribución de Multiplicidades

$$\mathbb{P}\left(\sum_i \mathbf{1}\{Y_i=r\}=j\right)$$

  - Gaps, colisiones, maximas ocupaciones,...etc

Cantidades de interés en modelos de urnas

La mayoría son de la forma:

$$E\left(g\left(Y_1,\ldots,Y_N\right)\right), \ g \text{ medible}$$

- **Ejemplos**
  - Función de probabilidad conjunta:

    $$P(Y_1=y_1,\ldots,Y_N=y_N)=E\left(I\left(Y_1=y_1,\ldots,Y_N=y_N\right)\right)$$

  - Probabilidades marginales: $P(Y_j=y_j)=E\left(I\left(Y_j=y_j\right)\right)$
  - Momentos: EY_j^k
  - Distribución de Multiplicidades

    $$\mathbb{P}(\sum_i \mathbf{1}\{Y_i=r\}=j)$$

  - Gaps, colisiones, maximas ocupaciones,...etc

Transformaciones en modelos de urnas

Se verifica que $E_\theta\left(g\left(X_1,\ldots,X_N\right)\right)$ es la $PSD(G^N(\theta))$-transformada de $E\left(g\left(Y_1,\ldots,Y_N\right)\right)$, es decir

$$E\left(g\left(Y_1,\ldots,Y_N\right)\right) \circ\!\!-\!\!\!-\!\!\!-\!\!\bullet E_\theta\left(g\left(X_1,\ldots,X_N\right)\right)$$

Maxwell-Boltzmann

- n bolas numeradas en N urnas distinguibles
- Sucesión de pesos: $w_k = \frac{1}{k!}$
- Función de probabilidad conjunta de las alocaciones

$$P\left(Y_1 = y_1 \ldots, Y_N = y_N\right) = \frac{n!}{\prod_j y_j!}\frac{1}{N^n}$$

para $y_1 + \cdots, y_N = n$

- PSD asociada: $X_j \overset{iid}{\sim} PSD(\exp(\theta)) \equiv Po(\theta)$
- $S_N = \sum_{j=1}^{N} X_j(\theta) \sim Po(N\theta)$
- La transformación asociada es

$$H(\theta) = \sum_{n=0}^{\infty} T_N(n) \exp(-N\theta) \frac{(N\theta)^n}{n!}$$

Poissonización

Maxwell-Boltzmann

- n bolas numeradas en N urnas distinguibles
- Sucesión de pesos: $w_k = \frac{1}{k!}$
- Función de probabilidad conjunta de las alocaciones

$$P\left(Y_1 = y_1 \ldots, Y_N = y_N\right) = \frac{n!}{\prod_j y_j!}\frac{1}{N^n}$$

para $y_1 + \cdots , y_N = n$

- PSD asociada: $X_j \overset{iid}{\sim} PSD(\exp(\theta)) \equiv Po(\theta)$
- $S_N = \sum_{j=1}^{N} X_j(\theta) \sim Po(N\theta)$
- La transformación asociada es

$$H(\theta) = \sum_{n=0}^{\infty} T_N(n) \exp(-N\theta) \frac{(N\theta)^n}{n!}$$

Poissonización

Maxwell-Boltzmann

- n bolas numeradas en N urnas distinguibles
- Sucesión de pesos: $w_k = \frac{1}{k!}$
- Función de probabilidad conjunta de las alocaciones

$$P\left(Y_1 = y_1 \ldots, Y_N = y_N\right) = \frac{n!}{\prod_j y_j!} \frac{1}{N^n}$$

para $y_1 + \cdots, y_N = n$

- PSD asociada: $X_j \overset{iid}{\sim} PSD(\exp(\theta)) \equiv Po(\theta)$
- $S_N = \sum_{j=1}^{N} X_j(\theta) \sim Po(N\theta)$
- La transformación asociada es

$$H(\theta) = \sum_{n=0}^{\infty} T_N(n) \exp(-N\theta) \frac{(N\theta)^n}{n!}$$

Poissonización

Maxwell-Boltzmann

- n bolas numeradas en N urnas distinguibles
- Sucesión de pesos: $w_k = \frac{1}{k!}$
- Función de probabilidad conjunta de las alocaciones

$$P\left(Y_1 = y_1 \ldots, Y_N = y_N\right) = \frac{n!}{\prod_j y_j!} \frac{1}{N^n}$$

para $y_1 + \cdots, y_N = n$

- PSD asociada: $X_j \overset{iid}{\sim} PSD(\exp(\theta)) \equiv Po(\theta)$
- $S_N = \sum_{j=1}^{N} X_j(\theta) \sim Po(N\theta)$
- La transformación asociada es

$$H(\theta) = \sum_{n=0}^{\infty} T_N(n) \exp(-N\theta) \frac{(N\theta)^n}{n!}$$

Poissonización

Maxwell-Boltzmann

- n bolas numeradas en N urnas distinguibles
- Sucesión de pesos: $w_k = \frac{1}{k!}$
- Función de probabilidad conjunta de las alocaciones

$$P(Y_1 = y_1 \ldots, Y_N = y_N) = \frac{n!}{\prod_j y_j!} \frac{1}{N^n}$$

para $y_1 + \cdots, y_N = n$
- PSD asociada: $X_j \overset{iid}{\sim} PSD(\exp(\theta)) \equiv Po(\theta)$
- $S_N = \sum_{j=1}^{N} X_j(\theta) \sim Po(N\theta)$
- La transformación asociada es

$$H(\theta) = \sum_{n=0}^{\infty} T_N(n) \exp(-N\theta) \frac{(N\theta)^n}{n!}$$

Poissonización

Maxwell-Boltzmann

- n bolas numeradas en N urnas distinguibles
- Sucesión de pesos: $w_k = \frac{1}{k!}$
- Función de probabilidad conjunta de las alocaciones

$$P\left(Y_1 = y_1 \ldots, Y_N = y_N\right) = \frac{n!}{\prod_j y_j!} \frac{1}{N^n}$$

para $y_1 + \cdots, y_N = n$

- PSD asociada: $X_j \overset{iid}{\sim} PSD(\exp(\theta)) \equiv Po(\theta)$
- $S_N = \sum_{j=1}^{N} X_j(\theta) \sim Po(N\theta)$
- La transformación asociada es

$$H(\theta) = \sum_{n=0}^{\infty} T_N(n) \exp(-N\theta) \frac{(N\theta)^n}{n!}$$

Poissonización

Ejercicio...

Calcular la probabilidad de que haya k urnas vacías

Solución

- El número de urnas vacías es

$$M_0 := \sum_{i=1}^{N} \mathbf{1}\{Y_i = 0\}$$

- Se tiene:

$$\mathbb{P}\left(\sum_{i=1}^{N} \mathbf{1}\{Y_i = 0\} = k\right) \circ\!\!-\!\!\!-\!\!\!-\!\!\bullet\, \mathbb{P}\left(\sum_{i=1}^{N} \mathbf{1}\{X_i = 0\} = k\right)$$

- Pero $\sum_{i=1}^{N} \mathbf{1}\{X_i = 0\} \sim \mathrm{Bi}(N, e^{-\theta})$

Solución

- El número de urnas vacías es

$$M_0 := \sum_{i=1}^{N} \mathbf{1}\{Y_i = 0\}$$

- Se tiene:

$$\mathbb{P}\left(\sum_{i=1}^{N} \mathbf{1}\{Y_i = 0\} = k\right) \circ\!\!\longrightarrow\!\!\bullet\, \mathbb{P}\left(\sum_{i=1}^{N} \mathbf{1}\{X_i = 0\} = k\right)$$

- Pero $\sum_{i=1}^{N} \mathbf{1}\{X_i = 0\} \sim \mathrm{Bi}(N, e^{-\theta})$

Solución

- El número de urnas vacías es

$$M_0 := \sum_{i=1}^{N} \mathbf{1}\{Y_i = 0\}$$

- Se tiene:

$$\mathbb{P}\left(\sum_{i=1}^{N} \mathbf{1}\{Y_i = 0\} = k\right) \multimap \mathbb{P}\left(\sum_{i=1}^{N} \mathbf{1}\{X_i = 0\} = k\right)$$

- Pero $\sum_{i=1}^{N} \mathbf{1}\{X_i = 0\} \sim \mathrm{Bi}(N, e^{-\theta})$

Solución

- Entonces:

$$\mathbb{P}\left(\sum_{i=1}^{N} \mathbf{1}\{Y_i = 0\} = k\right) \circ\!\!-\!\!\!-\!\!\!-\!\!\bullet\, \mathbb{P}\left(\sum_{i=1}^{N} \mathbf{1}\{X_i = 0\} = k\right) =$$

$$= \binom{N}{k} e^{-k\theta}(1 - e^{-\theta})^{N-k} \bullet\!\!-\!\!\!-\!\!\!-\!\!\circ$$

$$\binom{N}{k} \sum_{j=0}^{N-k-1} \binom{N-k}{j} (-1)^j \left(1 - \frac{k+j}{N}\right)^n$$

Ejercicio....

¿Cuál es el valor esperado del número de urnas vacías?

Solución

- Sean

$$M_0 := \sum_{i=1}^{N} \mathbf{1}\{Y_i = 0\}$$
$$M_0^* := \sum_{i=1}^{N} \mathbf{1}\{X_i = 0\} \sim \mathrm{Bi}(N, e^{-\theta})$$

- Entonces:

$$\mathbb{E}M_0 \qquad \mathbb{E}M_0^* = Ne^{-\theta} \qquad \left(1 - \frac{1}{N}\right)^n$$

Solución

- Sean

$$M_0 := \sum_{i=1}^{N} \mathbf{1}\{Y_i = 0\}$$

$$M_0^* := \sum_{i=1}^{N} \mathbf{1}\{X_i = 0\} \sim \mathrm{Bi}(N, e^{-\theta})$$

- Entonces:

$$\mathbb{E}M_0 \circ\!\!-\!\!\!-\!\!\!-\!\!\bullet\, \mathbb{E}M_0^* = Ne^{-\theta} \bullet\!\!-\!\!\!-\!\!\!-\!\!\circ \left(1 - \frac{1}{N}\right)^n$$

Árboles aleatorios finitamente generados

- Consideremos el conjunto de árboles con n nodos, $\mathfrak{T}_n$.
- Sea $d^+(v)$ el número de nodos que cuelgan del vértice v (outdegree)
- Sea la sucesión de pesos (no negativos) $\{w_k\}_{k\geq 0}$
- El peso asociado a un árbol $T \in \mathfrak{T}_n$ es

$$w(T) := \prod_{v \in T} w_{d^+(v)}$$

Árboles aleatorios finitamente generados

- Seleccionar un árbol en $\mathfrak{T}_n$ con probabilidad proporcional al peso:

$$\mathbb{P}(\mathcal{T} = T) \propto \prod_{v \in T} w_{d^+(v)}, \quad T \in \mathfrak{T}_n.$$

- es decir:

$$\mathbb{P}(\mathcal{T} = T) = \frac{\prod_{v \in T} w_{d^+(v)}}{Z_n}, \quad T \in \mathfrak{T}_n,$$

donde

$$Z_n = \sum_{T \in \mathfrak{T}_n} \prod_{v \in T} w_{d^+(v)}$$

Árboles aleatorios finitamente generados

- Seleccionar un árbol en $\mathfrak{T}_n$ con probabilidad proporcional al peso:

$$\mathbb{P}(\mathcal{T} = T) \propto \prod_{v \in T} w_{d^+(v)}, \quad T \in \mathfrak{T}_n.$$

- es decir:

$$\mathbb{P}(\mathcal{T} = T) = \frac{\prod_{v \in T} w_{d^+(v)}}{Z_n}, \quad T \in \mathfrak{T}_n,$$

donde

$$Z_n = \sum_{T \in \mathfrak{T}_n} \prod_{v \in T} w_{d^+(v)}$$

Relación entre árboles aleatorios y modelos de alocación

- Un árbol puede asociarse a un modelo de alocación:

$$\begin{array}{lcl} \text{Nodo} & \longrightarrow & \text{Urna} \\ \text{Outdegree} & \longrightarrow & \text{Bolas} \end{array}$$

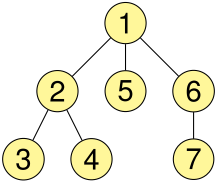
Relación entre árboles aleatorios y modelos de alocación

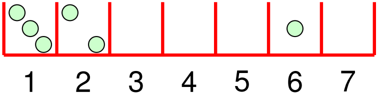
- Un árbol puede asociarse a un modelo de alocación:

$$\begin{array}{lcl} \text{Nodo} & \longrightarrow & \text{Urna} \\ \text{Outdegree} & \longrightarrow & \text{Bolas} \end{array}$$

Relación entre árboles aleatorios y modelos de alocación

- Consideramos orden depth-first en los árboles
- Construimos un modelo de alocación con n urnas (vértices) y $n-1$ bolas (outdegree)





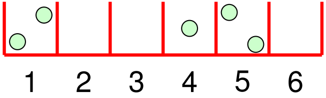
Relación entre árboles aleatorios y modelos de alocación

- Desafortunadamente la relación entre árboles en $\mathfrak{T}_n$ y alocaciones en $\Omega_{n-1,n}$ no es biyectiva
- Por ejemplo la alocación $(2, 0, 0, 1, 2, 0) \in \Omega_{n-1,n}$

no se corresponde con ningún árbol en $\mathfrak{T}_6$.

Relación entre árboles aleatorios y modelos de alocación

- Desafortunadamente la relación entre árboles en $\mathfrak{T}_n$ y alocaciones en $\Omega_{n-1,n}$ no es biyectiva
- Por ejemplo la alocación $(2,0,0,1,2,0) \in \Omega_{n-1,n}$



no se corresponde con ningún árbol en $\mathfrak{T}_6$.

Relación entre árboles aleatorios y modelos de alocación

- Sin embargo una y solo una permutación circular de una alocación en $\Omega_{n-1,n}$ se corresponde con un árbol en $\mathfrak{T}_n$.

Relación entre árboles aleatorios y modelos de alocación

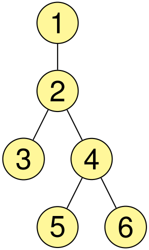
- En la alocación anterior

$(2,0,0,1,2,0)$ NO
$(0,2,0,0,1,2)$ NO
$(2,0,2,0,0,1)$ NO
$(1,2,0,2,0,0)$ SI
$(0,1,2,0,2,0)$ NO
$(0,0,1,2,0,2)$ NO

Relación entre árboles aleatorios y modelos de alocación

- En la alocación anterior

$$
\begin{array}{ll}
(2,0,0,1,2,0) & \text{NO} \\
(0,2,0,0,1,2) & \text{NO} \\
(2,0,2,0,0,1) & \text{NO} \\
(1,2,0,2,0,0) & \text{SI} \\
(0,1,2,0,2,0) & \text{NO} \\
(0,0,1,2,0,2) & \text{NO}
\end{array}
$$



Conclusión

La estimación insesgada en *PSD*, el estudio de las distribuciones asociadas a modelos de alocación aleatoria de bolas en urnas y las distribuciones de árboles aleatorios finitamente generados son problemas distintos que pueden ser abordados con una herramienta común:

la transformada PSD!