Gramáticas Libre de Contexto Aleatorias.

Nicolás Rivera

20 de Marzo de 2012

Pontificia Universidad Católica de Chile

Tabla de Contenidos

Tabla de Contenidos

Tabla de Contenidos

Tabla de Contenidos

Definición (Gramática Libre de Contexto)

Una Gramática Libre de Contexto es una cuádrupla $G = (V, T, P, S)$, donde V son las variables, T el conjunto de terminales, P las producciones y $S \in V$ la variable inicial.

Definición (Gramática Libre de Contexto)

Una Gramática Libre de Contexto es una cuádrupla $G = (V, T, P, S)$, donde V son las variables, T el conjunto de terminales, P las producciones y $S \in V$ la variable inicial.

Asumiremos que V, T, P son finitos. Ejemplo:

$$G = (\{S\}, \{s\}, \{S \to s, S \to SS\}, S)$$

Recordemos que la derivación de una palabra puede ser representada por un árbol τ.

Recordemos que la derivación de una palabra puede ser representada por un árbol τ.
Sea Ω el conjunto de todos los árboles de derivación τ con raíz S. (finitos)

Recordemos que la derivación de una palabra puede ser representada por un árbol τ.
Sea Ω el conjunto de todos los árboles de derivación τ con raíz S. (finitos)
Para cada árbol τ definimos $f(A \to \alpha; \tau)$ como el número de veces que la producción $A \to \alpha$ aparece en τ y $f(A; \tau)$ el número de apariciones de la variable $A \in \tau$, es decir:

$$f(A; \tau) = \sum_{\alpha \in (N \cup T)^\star \wedge A \to \alpha} f(A \to \alpha; \tau)$$

Definimos $h(\tau)$, la altura de τ, como la cantidad de vértices no terminales (variables) en el camino más largo entre la raíz de τ y sus vértices no terminales.

Definimos $h(\tau)$, la altura de τ, como la cantidad de vértices no terminales (variables) en el camino más largo entre la raíz de τ y sus vértices no terminales.

Definimos $|\tau|$, el tamaño de τ como el número total de nodos no terminales en τ.

Definimos $h(\tau)$, la altura de τ, como la cantidad de vértices no terminales (variables) en el camino más largo entre la raíz de τ y sus vértices no terminales.
Definimos $|\tau|$, el tamaño de τ como el número total de nodos no terminales en τ.
Finalmente, denotamos por γ a las palabras en $(V \cup T)^{\star}$.

Definimos $h(\tau)$, la altura de τ, como la cantidad de vértices no terminales (variables) en el camino más largo entre la raíz de τ y sus vértices no terminales.

Definimos $|\tau|$, el tamaño de τ como el número total de nodos no terminales en τ.

Finalmente, denotamos por γ a las palabras en $(V \cup T)^{\star}$.

Por otra parte definimos $n(A; \gamma)$ como el número de ocurrencias de A en γ y denotamos por $|\gamma|$ al largo de γ.

Símbolo	Significado
τ	árbol
Ω	conjunto de árboles de derivación con raíz S
$f(A \to \alpha; \tau)$	cantidad de veces que aparece $A \to \alpha$ en τ
$f(A)$	cantidad de veces que aparece A en τ
$h(\tau)$	altura del árbol (contando solo vértices no terminales)
$\|\tau\|$	cantidad de vértices no terminales en τ
γ	palabra en $(V \cup T)^\star$
$\|\gamma\|$	largo de γ

Sistema de Producción de Probabilidades

Definición (Sistema de Producción de Probabilidades)

Un Sistema de Producción de Probabilidades de G es una función $p : P \to [0,1]$ tal que para todo $A \in V$ se tiene que

$$\sum_{\alpha : A \to \alpha \in P} p(A \to \alpha) = 1.$$

Definición (Sistema de Producción de Probabilidades)

Un Sistema de Producción de Probabilidades de G es una función $p : P \to [0,1]$ tal que para todo $A \in V$ se tiene que

$$\sum_{\alpha : A \to \alpha \in P} p(A \to \alpha) = 1.$$

Esto define una probabilidad $\mathbb{P}$ sobre los árboles de derivación (incluso infinitos!) [1].

Definición (Sistema de Producción de Probabilidades)

Un Sistema de Producción de Probabilidades de G es una función $p : P \to [0, 1]$ tal que para todo $A \in V$ se tiene que

$$\sum_{\alpha : A \to \alpha \in P} p(A \to \alpha) = 1.$$

Esto define una probabilidad $\mathbb{P}$ sobre los árboles de derivación (incluso infinitos!) [1]. Es fácil comprobar que dado un árbol de tamaño finito τ,

Definición (Sistema de Producción de Probabilidades)

Un Sistema de Producción de Probabilidades de G es una función $p : P \to [0,1]$ tal que para todo $A \in V$ se tiene que

$$\sum_{\alpha : A \to \alpha \in P} p(A \to \alpha) = 1.$$

Esto define una probabilidad $\mathbb{P}$ sobre los árboles de derivación (incluso infinitos!) [1]. Es fácil comprobar que dado un árbol de tamaño finito τ,

$$\mathbb{P}(\tau) = \prod_{A \to \alpha} p(A \to \alpha)^{f(A \to \alpha; \tau)}$$

Definición (Sistema de Producción de Probabilidades)

Un Sistema de Producción de Probabilidades de G es una función $p : P \to [0,1]$ tal que para todo $A \in V$ se tiene que

$$\sum_{\alpha : A \to \alpha \in P} p(A \to \alpha) = 1.$$

Esto define una probabilidad $\mathbb{P}$ sobre los árboles de derivación (incluso infinitos!) [1]. Es fácil comprobar que dado un árbol de tamaño finito τ,

$$\mathbb{P}(\tau) = \prod_{A \to \alpha} p(A \to \alpha)^{f(A \to \alpha; \tau)}$$

A la construcción anterior la llamamos una *GLCA*.

Notemos que lo anterior no solo puede usarse para definir una probabilidad sobre Ω, sino que sobre Ω_A, el conjunto de todos los árboles de derivación que comienzan en A ($A \in V$). Denotamos $\mathbb{P}_A$ cuando nos referimos a la probabilidad sobre Ω_A y seguimos usando $\mathbb{P}$ para $\mathbb{P}_S$.

Notemos que lo anterior no solo puede usarse para definir una probabilidad sobre Ω, sino que sobre Ω_A, el conjunto de todos los árboles de derivación que comienzan en A ($A \in V$). Denotamos $\mathbb{P}_A$ cuando nos referimos a la probabilidad sobre Ω_A y seguimos usando $\mathbb{P}$ para $\mathbb{P}_S$.
Sea g una función $g : \Omega \to \mathbb{R}$, definimos la esperanza de g por

$$\mathbb{E}(g) = \sum_{\tau \in \Omega} g(\tau)\mathbb{P}(\tau).$$

De forma análoga se define $\mathbb{E}_A$ para funciones $g : \Omega_A \to \mathbb{R}$.

Símbolo	Significado
Ω	conjunto de árboles de derivación con raíz S
Ω_A	conjunto de árboles de derivación con raíz A ($A \in V$)
$\mathbb{P}$	probabilidad sobre Ω creada a partir de p
$\mathbb{P}_A$	probabilidad sobre Ω_A
$\mathbb{E}$	esperanza sobre Ω
$\mathbb{E}_A$	esperanza sobre Ω_A

Frecuencia Relativa de Producciones

Diremos que la medida de probabilidad $\mathbb{P}$ definida por p es propia si la probabilidad se concentra solo en árboles τ con $|\tau| < \infty$, es decir $\mathbb{P}(\Omega) = 1$.

Diremos que la medida de probabilidad $\mathbb{P}$ definida por p es propia si la probabilidad se concentra solo en árboles τ con $|\tau| < \infty$, es decir $\mathbb{P}(\Omega) = 1$.
Por ejemplo: en $G = (\{S\}, \{s\}, \{S \to s, S \to SS\}, S)$ consideremos p por $p(S \to s) = 1 - q$ y $p(S \to SS) = q$ con $q \in [0, 1]$ si τ es un árbol aleatorio en Ω entonces

Diremos que la medida de probabilidad $\mathbb{P}$ definida por p es propia si la probabilidad se concentra solo en árboles τ con $|\tau| < \infty$, es decir $\mathbb{P}(\Omega) = 1$.
Por ejemplo: en $G = (\{S\}, \{s\}, \{S \to s, S \to SS\}, S)$ consideremos p por $p(S \to s) = 1 - q$ y $p(S \to SS) = q$ con $q \in [0, 1]$ si τ es un árbol aleatorio en Ω entonces

$$\begin{aligned}\mathbb{P}(|\tau| < \infty) &= \mathbb{P}(|\tau| < \infty)^2 p(S \to SS) + p(S \to s)\\ &= \mathbb{P}(|\tau| < \infty)^2 q + (1 - q).\end{aligned}$$

Diremos que la medida de probabilidad $\mathbb{P}$ definida por p es propia si la probabilidad se concentra solo en árboles τ con $|\tau| < \infty$, es decir $\mathbb{P}(\Omega) = 1$.
Por ejemplo: en $G = (\{S\}, \{s\}, \{S \to s, S \to SS\}, S)$ consideremos p por $p(S \to s) = 1 - q$ y $p(S \to SS) = q$ con $q \in [0, 1]$ si τ es un árbol aleatorio en Ω entonces

$$\mathbb{P}(|\tau| < \infty) = \mathbb{P}(|\tau| < \infty)^2 p(S \to SS) + p(S \to s)$$

$$= \mathbb{P}(|\tau| < \infty)^2 q + (1 - q).$$

La ecuación anterior tiene dos soluciones para $\mathbb{P}(|\tau| < \infty)$, 1 y $1/q - 1$. Se puede mostrar que la solución adecuada es

$$\mathbb{P}(|\tau| < \infty) = min(1, 1/q - 1)$$

por lo tanto si $q > 1/2$ se impone una probabilidad impropia.

Nuestro objetivo es construir buenas funciones p.

Supongamos que la gramática G es conocida, pero no la función p. Intentemos estimar p.

Supongamos que la gramática G es conocida, pero no la función p. Intentemos estimar p. Consideremos la muestra aleatoria de n árboles de derivación $\tau_1,\ldots,\tau_n$. Definimos el estimador [2]

Supongamos que la gramática G es conocida, pero no la función p. Intentemos estimar p. Consideremos la muestra aleatoria de n árboles de derivación $\tau_1, \ldots, \tau_n$. Definimos el estimador [2]

$$\hat{p}(A \to \alpha) = \frac{\sum_{i=1}^{n} f(A \to \alpha; \tau_i)}{\sum_{i=1}^{n} f(A; \tau_i)}.$$

Supongamos que la gramática G es conocida, pero no la función p. Intentemos estimar p. Consideremos la muestra aleatoria de n árboles de derivación $\tau_1,\ldots,\tau_n$. Definimos el estimador [2]

$$\hat{p}(A \to \alpha) = \frac{\sum_{i=1}^{n} f(A \to \alpha; \tau_i)}{\sum_{i=1}^{n} f(A; \tau_i)}.$$

Definimos por $\hat{\mathbb{P}}_k$ a la probabilidad sobre Ω generada por $\hat{p}$

En general no es posible observar los árboles, sino que solo podemos observar los símbolos terminales de él (es decir, la palabra). Notemos por $Y(\tau)$ a la palabra correspondiente al árbol τ. Sean $Y_1,\ldots,Y_n$ las palabras observadas entonces el estimador queda como

En general no es posible observar los árboles, sino que solo podemos observar los símbolos terminales de él (es decir, la palabra). Notemos por $Y(\tau)$ a la palabra correspondiente al árbol τ. Sean $Y_1,\ldots,Y_n$ las palabras observadas entonces el estimador queda como

$$\hat{p}(A \to \alpha) = \frac{\sum_{i=1}^{n} \mathbb{E}_{\hat{p}}(f(A \to \alpha; \tau)|\tau \in \Omega_{Y_i})}{\sum_{i=1}^{n} \mathbb{E}_{\hat{p}}(f(A; \tau)|\tau \in \Omega_{Y_i})}$$

En general no es posible observar los árboles, sino que solo podemos observar los símbolos terminales de él (es decir, la palabra). Notemos por $Y(\tau)$ a la palabra correspondiente al árbol τ. Sean $Y_1,\ldots,Y_n$ las palabras observadas entonces el estimador queda como

$$\hat{p}(A \to \alpha) = \frac{\sum_{i=1}^{n} \mathbb{E}_{\hat{p}}(f(A \to \alpha; \tau)|\tau \in \Omega_{Y_i})}{\sum_{i=1}^{n} \mathbb{E}_{\hat{p}}(f(A; \tau)|\tau \in \Omega_{Y_i})}$$

Lo anterior es difícil de calcular pero hay métodos numéricos que pueden ayudar. En particular [3],

$$\hat{p}_{k+1}(A \to \alpha) = \frac{\sum_{i=1}^{n} \mathbb{E}_{\hat{p}_k}(f(A \to \alpha; \tau)|\tau \in \Omega_{Y_i})}{\sum_{i=1}^{n} \mathbb{E}_{\hat{p}_k}(f(A; \tau)|\tau \in \Omega_{Y_i})}$$

Notemos que

$$\mathbb{E}_{\hat{p}_k}(f(A \to \alpha)|\tau \in \Omega_{Y_i}) = \sum_{\tau \in \Omega_{Y_i}} f(A \to \alpha; \tau)\hat{\mathbb{P}}_k(\tau|\tau \in \Omega_{Y_i}).$$

Notemos que

$$\mathbb{E}_{\hat{p}_k}(f(A \to \alpha)|\tau \in \Omega_{Y_i}) = \sum_{\tau \in \Omega_{Y_i}} f(A \to \alpha; \tau)\hat{\mathbb{P}}_k(\tau|\tau \in \Omega_{Y_i}).$$

Sean Λ el conjunto de todos los árboles de derivación que podrían pertenecer a los datos, es decir

$$\Lambda = \{\tau : Y(\tau) \in \{Y_1, ..., Y_b\}\}$$

Notemos que

$$\mathbb{E}_{\hat{p}_k}(f(A \to \alpha)|\tau \in \Omega_{Y_i}) = \sum_{\tau \in \Omega_{Y_i}} f(A \to \alpha; \tau)\hat{\mathbb{P}}_k(\tau|\tau \in \Omega_{Y_i}).$$

Sean Λ el conjunto de todos los árboles de derivación que podrían pertenecer a los datos, es decir

$$\Lambda = \{\tau : Y(\tau) \in \{Y_1, ..., Y_b\}\}$$

Sea $\tau \in \Lambda$, sea $y = Y(\tau)$. Definamos

$$W(\tau) = \sum_{i:Y_i = y} \hat{\mathbb{P}}_k(\tau|\tau \in \Omega_y) = |i : Y_i = i|\hat{\mathbb{P}}_k(\tau).$$

Se sigue que para toda regla de producción $A \to \alpha$,

$$\sum_{i=1}^{n} \sum_{\tau \in \Omega_{Y_i}} f(A \to \alpha; \tau) \hat{\mathbb{P}}_k(\tau | \tau \in \Omega_{Y_i}) = \sum_{\tau \in \Lambda} f(A \to \alpha; \tau) W(\tau)$$

Se sigue que para toda regla de producción $A \to \alpha$,

$$\sum_{i=1}^{n} \sum_{\tau \in \Omega_{Y_i}} f(A \to \alpha; \tau) \hat{\mathbb{P}}_k(\tau | \tau \in \Omega_{Y_i}) = \sum_{\tau \in \Lambda} f(A \to \alpha; \tau) W(\tau)$$

y por lo tanto

$$\hat{p}_k(A \to \alpha) = \frac{\sum_{\tau \in \Lambda} f(A \to \alpha; \tau) W(\tau)}{\sum_{\tau \in \Lambda} f(A; \tau) W(\tau)}$$

Inspirandonos en lo anterior definimos un procedimiento simple para la construcción de un sistema de producción de probabilidades para una GLC G.

Inspirandonos en lo anterior definimos un procedimiento simple para la construcción de un sistema de producción de probabilidades para una GLC G.
Primero consideramos un conjunto finito $\Lambda \subseteq \Omega$, tal que toda regla de producción aparezca en algún árbol en Λ, luego asignamos a cada árbol un peso $W(\tau)$ tal que $\sum_{\tau \in \Lambda} W(\tau) = 1$.

Inspirandonos en lo anterior definimos un procedimiento simple para la construcción de un sistema de producción de probabilidades para una GLC G.
Primero consideramos un conjunto finito $\Lambda \subseteq \Omega$, tal que toda regla de producción aparezca en algún árbol en Λ, luego asignamos a cada árbol un peso $W(\tau)$ tal que $\sum_{\tau\in\Lambda} W(\tau) = 1$.
y finalmente definimos p como

$$p(A \to \alpha) = \frac{\sum_{\tau\in\Lambda} f(A \to \alpha; \tau) W(\tau)}{\sum_{\tau\in\Lambda} f(A; \tau) W(\tau)}.$$

Inspirandonos en lo anterior definimos un procedimiento simple para la construcción de un sistema de producción de probabilidades para una GLC G.
Primero consideramos un conjunto finito $\Lambda \subseteq \Omega$, tal que toda regla de producción aparezca en algún árbol en Λ, luego asignamos a cada árbol un peso $W(\tau)$ tal que $\sum_{\tau \in \Lambda} W(\tau) = 1$.
y finalmente definimos p como

$$p(A \to \alpha) = \frac{\sum_{\tau \in \Lambda} f(A \to \alpha; \tau) W(\tau)}{\sum_{\tau \in \Lambda} f(A; \tau) W(\tau)}.$$

El método anterior es llamado "Método de Frecuencia Relativa"

Teorema

Si todos los elementos de T aparecen en algún árbol de Λ y todas las producciones tienen probabilidades positivas, entonces el método de frecuencia relativa impone una distribución propia sobre Ω.

Corolario

Bajo las condiciones anteriores, para todo $A \in V$ el método de frecuencia relativa impone una distribución propia sobre Ω_A.

Demostración: Pizarra.

Un poco de notación. Dados $\Lambda = \{\tau_1, ..., \tau_n\}$, definimos

$$F(A \to \alpha) = \sum_{\tau \in \Lambda} f(A \to \lambda; \tau) W(\tau),$$

Un poco de notación. Dados $\Lambda = \{\tau_1, ..., \tau_n\}$, definimos

$$F(A \to \alpha) = \sum_{\tau \in \Lambda} f(A \to \lambda; \tau) W(\tau),$$

y

$$F(A) = \sum_{\alpha : A \to \alpha} F(A \to \alpha) = \sum_{\tau \in \Lambda} f(A; \tau) W(\tau),$$

Un poco de notación. Dados $\Lambda = \{\tau_1, ..., \tau_n\}$, definimos

$$F(A \to \alpha) = \sum_{\tau \in \Lambda} f(A \to \lambda; \tau) W(\tau),$$

y

$$F(A) = \sum_{\alpha : A \to \alpha} F(A \to \alpha) = \sum_{\tau \in \Lambda} f(A; \tau) W(\tau),$$

por lo tanto

$$p(A \to \alpha) = \frac{F(A \to \alpha)}{F(A)},$$

Teorema

Supongamos que todos las producciones de P de la gramática G aparecen en algún árbol de τ y supongamos que $W(\tau) > 0$ para todo $\tau \in \Lambda$, entonces y supongamos además que p fue asignada con el método de frecuencia relativa, entonces si Υ es un árbol aleatorio entonces

$$\mathbb{E}(f(A \to \alpha; \Upsilon)) = \sum_{\tau \in \Lambda} f(A \to \alpha; \tau) W(\tau)$$

Demostración: Pizarra.

Sea $\mathbb{P}$ la probabilidad inducida por el método de frecuencia relativa. La entropía de $\mathbb{P}$ es

$$H(\mathbb{P}) = \sum_{\tau \in \Omega} \frac{1}{\mathbb{P}(\tau)} \log(\mathbb{P}(\tau))$$

Sea $\mathbb{P}$ la probabilidad inducida por el método de frecuencia relativa. La entropía de $\mathbb{P}$ es

$$H(\mathbb{P}) = \sum_{\tau \in \Omega} \frac{1}{\mathbb{P}(\tau)} \log(\mathbb{P}(\tau))$$

Teorema

$$H(\mathbb{P}) = \sum_{A \to \alpha \in P} F(A \to \alpha) \log(\frac{1}{F(A \to \alpha)}) - \sum_{A \in V} F(A) \log(\frac{1}{F(A)})$$

Demostración: Pizarra.

Diremos que M es la matríz de media de una $GMLA$ si M es una matriz de $|V|x|V|$ donde $M(A, B)$ es el promedio de variables B resultantes de expandir A, es decir

Diremos que M es la matríz de media de una $GMLA$ si M es una matriz de $|V|x|V|$ donde $M(A,B)$ es el promedio de variables B resultantes de expandir A, es decir

$$M(A,B) = \sum_{\alpha:\mathcal{A}\to\alpha} p(A\to\alpha)n(B;\alpha),$$

Diremos que M es la matríz de media de una $GMLA$ si M es una matriz de $|V|x|V|$ donde $M(A, B)$ es el promedio de variables B resultantes de expandir A, es decir

$$M(A, B) = \sum_{\alpha:\mathcal{A}\to\alpha} p(A \to \alpha)n(B; \alpha),$$

donde $n(B; \alpha)$ es la cantidad de veces que aparece B en α.

Una herramienta útil:

Una herramienta útil:

Teorema (Perron-Frobenius)

Sea M una matriz cuadrada no negativa. Entonces

1. *Existe un valor propio no negativo $\rho = \rho(M)$ tal que ningún valor propio de A tiene valor absoluto más grande que ρ.*
2. *El vector propio asociado a ρ tiene componentes no negativas.*
3. *Si M es irreducible, es decir, para todo i, j existe n tal que $M^{(n)}(i,j) > 0$, entonces ρ es único y su vector propio es estrictamente positivo.*

Estudiaremos ρ y su relación con las *GLCA*.

La norma de una matriz está dada po $\|M\| = \sup_{|\vec{v}|=1} |M\vec{v}|$. Otro útil resultado es

Lema

$$\lim_{n\to\infty} \|M^{(n)}\|^{1/n} = \rho(M).$$

Teorema

Sea M la matriz de media de una GLCA, si $\rho(M) < 1$ entonces para cada $m \in \mathbb{N}$ se tiene que

$$\mathbb{E}(|\tau|^m) < \infty$$

Demostración: Pizarra.

Teorema

Si p es asignada por el método de frecuencia relativa entonces la matriz de media M de la GLCA cumple con $\rho(M) < 1$.

Demostración: Pizarra.

Una distribución de Gibbs sobre los árboles de derivación tiene la siguiente forma:

$$\mathbb{P}(\tau) = \frac{e^{\lambda \cdot U(\tau)}}{Z_\lambda},$$

Una distribución de Gibbs sobre los árboles de derivación tiene la siguiente forma:

$$\mathbb{P}(\tau) = \frac{e^{\lambda \cdot U(\tau)}}{Z_\lambda},$$

donde λ y $U(\tau)$ son vectores de la misma dimensión y $\lambda \cdot U(\tau)$ es su producto interno. Por otra parte $Z_\lambda = \sum_{\tau \in \Omega} e^{\lambda \cdot U(\tau)}$.

Mostraremos que una distribución de Gibbs es más general que una $GLCA$.

Mostraremos que una distribución de Gibbs es más general que una *GLCA*.
Consideremos

1. $U(\tau) = f(\tau) = \{f(A \to \alpha; \tau)\}_{(A \to \alpha) \in P}$
2. $\lambda_{A \to \alpha} = \log(p(A \to \alpha))$

Entonces si p forma una distribución propia $\mathbb{P}$, se tiene que

Mostraremos que una distribución de Gibbs es más general que una *GLCA*.
Consideremos

1. $U(\tau) = f(\tau) = \{f(A \to \alpha; \tau)\}_{(A\to\alpha)\in P}$
2. $\lambda_{A\to\alpha} = \log(p(A \to \alpha))$

Entonces si p forma una distribución propia $\mathbb{P}$, se tiene que

$$Z_\lambda = \sum_{\tau\in\Omega} \exp\{\lambda \cdot f(\tau)\} = \sum_{\tau\in\Omega} \mathbb{P}(\tau) = 1$$

Mostraremos que una distribución de Gibbs es más general que una *GLCA*.

Consideremos

1. $U(\tau) = f(\tau) = \{f(A \to \alpha; \tau)\}_{(A\to\alpha)\in P}$
2. $\lambda_{A\to\alpha} = \log(p(A \to \alpha))$

Entonces si p forma una distribución propia $\mathbb{P}$, se tiene que

$$Z_\lambda = \sum_{\tau\in\Omega} \exp\{\lambda \cdot f(\tau)\} = \sum_{\tau\in\Omega} \mathbb{P}(\tau) = 1$$

y luego

$$\mathbb{P}(\tau) = \prod_{A\to\alpha} p(A \to \alpha)^{f(A\to\alpha;\tau)} =$$

$$= \exp \sum_{(A\to\alpha)\in P} \log(p(A \to \alpha))f(A \to \alpha; \tau) = \exp\{\lambda \cdot f(\tau)\}.$$

Lo anterior sirve para normalizar gramáticas, es decir, convertir $GLCA$ impropias en propias.

Lo anterior sirve para normalizar gramáticas, es decir, convertir *GLCA* impropias en propias. Si $\mathbb{P}$ es una probabilidad impropia, es decir $\mathbb{P}(\Omega) < 1$ entonces definimos

Lo anterior sirve para normalizar gramáticas, es decir, convertir *GLCA* impropias en propias. Si $\mathbb{P}$ es una probabilidad impropia, es decir $\mathbb{P}(\Omega) < 1$ entonces definimos

$$P\tilde{(\tau)} = \frac{\mathbb{P}(\tau)}{\mathbb{P}(\Omega)}.$$

Con lo anterior vimos como construir una distribución de Gibbs dada la $GLCA$. De la misma forma, sí la distribución de Gibbs tiene la forma vista en la parte anterior entonces forma una $GLCA$.

Con lo anterior vimos como construir una distribución de Gibbs dada la *GLCA*. De la misma forma, sí la distribución de Gibbs tiene la forma vista en la parte anterior entonces forma una *GLCA*.
Lo anterior tiene una aplicación. Sea una *GLCA* con $\mathbb{P}$ impropia, es decir $\mathbb{P}(\Omega) < 1$. Luego

$$\tilde{\mathbb{P}}(\tau) = \frac{\mathbb{P}(\tau)}{\mathbb{P}(\Omega)},$$

es una normalización de $\mathbb{P}$. En general se puede normalizar de la misma forma $\mathbb{P}_A$.

Teorema (Normalización)

Supuponga que las probabilidades de producción impuestas por p impropia son positivas para toda regla de producción. Entonces el sistema de producción de probabilidades normalizado p̃ se induce por

$$\tilde{p}(A \to \alpha) = \frac{1}{\mathbb{P}(\Omega_A)} p(A \to \alpha) \prod_{B \in V} \mathbb{P}(\Omega_B)^{n(B;\alpha)}.$$

Entonces p̃ induce una GLCA propia.

Resumen.

- Vimos que es una *GLCA*.

Resumen.

- Vimos que es una $GLCA$.
- Construimos una clase particular de $GLCA$ basadas en la intuición estadística.

Resumen.

- Vimos que es una *GLCA*.
- Construimos una clase particular de *GLCA* basadas en la intuición estadística.
- Estudiamos dicha gramática desde el punto de vista analítico.

Resumen.

- Vimos que es una *GLCA*.
- Construimos una clase particular de *GLCA* basadas en la intuición estadística.
- Estudiamos dicha gramática desde el punto de vista analítico.
- Vimos un tipo de distribución más general que las *GLCA*, las distribuciones de Gibbs.

¿Qué hay más adelante?

- Estudiar las *GLCA* de forma general.

¿Qué hay más adelante?

- Estudiar las *GLCA* de forma general.
- Resultados límites.

¿Qué hay más adelante?

- Estudiar las *GLCA* de forma general.
- Resultados límites.
- Tener más tipo de estadísticas para las *GLCA*

¿Qué hay más adelante?

- Estudiar las *GLCA* de forma general.
- Resultados límites.
- Tener más tipo de estadísticas para las *GLCA*
- Estudio algorítmico de las gramáticas.

¿Qué hay más adelante?

- Estudiar las *GLCA* de forma general.
- Resultados límites.
- Tener más tipo de estadísticas para las *GLCA*
- Estudio algorítmico de las gramáticas.
- Aplicaciones.