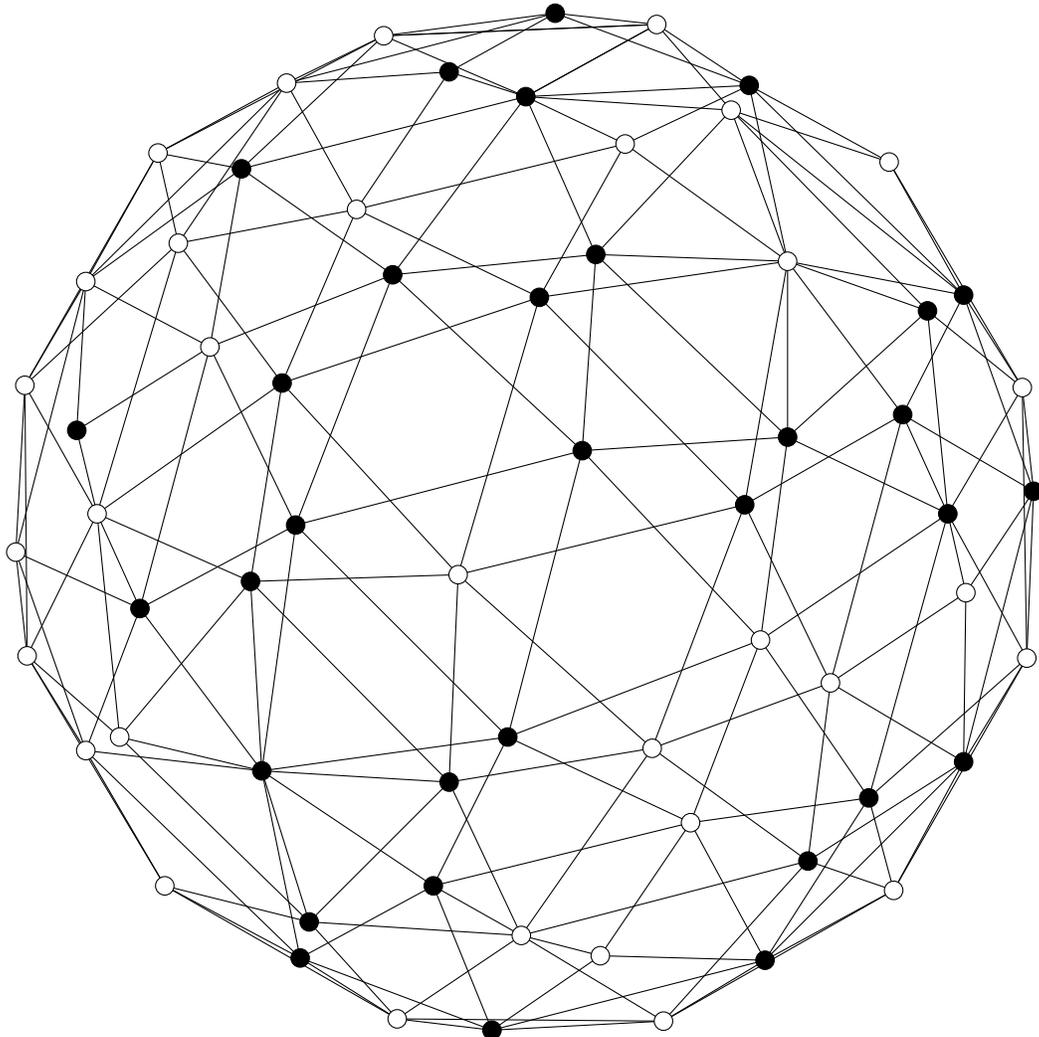


Algebrización del teorema de Hammersley-Clifford



PONTIFICIA UNIVERSIDAD JAVERIANA

FACULTAD DE CIENCIAS

DEPARTAMENTO DE MATEMÁTICAS

BOGOTÁ - COLOMBIA

2015



PONTIFICIA UNIVERSIDAD JAVERIANA

FACULTAD DE CIENCIAS
DEPARTAMENTO DE MATEMÁTICAS

Algebrización del teorema de Hammersley-Clifford

Aura Juliana Bernal Guevara.

Tesis presentada al Departamento de Matemáticas para optar al grado de Matemática.

Dirigida por:
Jesús Alonso Ochoa Arango, Ph.D.

Bogotá - Colombia
27 de julio de 2016

Nota de Aceptación _____

Jesús Alonso Ochoa Arango.
Director.

Jurado.

Bogotá - Colombia. Diciembre de 2015.

ÍNDICE GENERAL

Introducción	7
1. NOCIONES BÁSICAS DE TEORÍA DE GRAFOS	9
1.1. Grafos	9
1.1.1. Notación y terminología	9
1.1.2. Descomposición de grafos marcados	16
2. INDEPENDENCIA CONDICIONAL Y PROPIEDADES DE MARKOV	23
2.1. Independencia condicional y propiedades de Markov	23
2.1.1. Propiedades de Markov	27
2.1.2. Teorema Hammersley-Clifford	32
3. INTERPRETACIÓN ALGEBRAICA DEL TEOREMA DE HAMMERSLY-CLIFFORD	37
3.1. Modelos exponenciales, log-lineales y gráficos	37
3.1.1. Modelo algebraico	37
3.1.2. Modelo exponencial	38
3.1.3. Modelos log-lineales	39
3.2. Análisis algebraico de los modelos gráficos	41
3.2.1. Polinomios cuadráticos que representan independencia condicional.	42
3.2.2. Modelos gráficos no dirigidos y el teorema de Hammersley-Clifford	45

A mis padres

Algebrización del teorema de Hammersley-Clifford

Aura Juliana Bernal Guevara

Departamento de Matemáticas,
Facultad de Ciencias
Pontificia Universidad Javeriana

Resumen

Se planteará una reformulación del teorema de Hammersley-Clifford utilizando el lenguaje del álgebra conmutativa y las variedades afines.

Palabras Claves: Modelo gráfico, propiedades de Markov, independencia condicional, variedad afín, grafo no dirigido.

2015 Mathematics subject Classification. 14R05, 13P10, 05C20, 05C25, 05C50, 05C70, 62-09, 94C15, 97K30.

A generalization of Hammersley-Clifford theorem

Aura Juliana Bernal Guevara

Department of Mathematics,
Faculty of Sciences
Pontifical Xavierian University

Abstract

We will rewrite the Hammersley-Clifford theorem by mean of the concepts and tools from commutative algebra and algebraic varieties.

Key Words: Graphical model, Markov properties, conditional independence, undirected graph.

2015 Mathematics subject Classification. 14R05, 13P10, 05C20, 05C25, 05C50, 05C70, 62-09, 94C15, 97K30.

INTRODUCCIÓN

La teoría de modelos gráficos ha sido utilizada para modelar de forma gráfica, como su nombre lo indica, las condiciones de independencia entre las variables aleatorias utilizadas en un determinado modelo estadístico. Los mismos han sido de utilidad en aplicaciones de los métodos de probabilidad Bayesiana a problemas de visión computacional, al estudio de las cadenas de Markov, reconstrucción de imágenes, etc. Por estas razones es importante disponer de métodos computacionalmente efectivos para determinar cuando una distribución de probabilidad dada factoriza de acuerdo con un grafo específico, el teorema de Hammersley-Clifford provee un criterio *bastante simple* para dar respuesta a este interrogante.

Diaconis y Sturmfelds [DS] introducen métodos y herramientas del álgebra conmutativa para el estudio de procesos de Markov, siendo este el primer trabajo en donde se introducen métodos de naturaleza no analítica para el atacar problemas estadísticos. A partir de este trabajo, otro gran número de autores comenzaron a aplicar las herramientas del álgebra conmutativa y, como era de esperarse, de la geometría algebraica al diseño de experimentos, tablas de contingencia, biología computacional, filogenética, entre otras.

El teorema de Hammersley-Clifford (Thm. 57), estudiado en probabilidad, indica cuales distribuciones positivas factorizan de acuerdo con un grafo no dirigido dado, en este trabajo reformularemos este teorema en términos algebraicos (Thm. 87). A continuación describiremos de forma detallada el contenido de cada capítulo.

En el primer capítulo estudiaremos las nociones básicas de la teoría de grafos que serán utilizadas durante este trabajo. Los ejemplos concretos en esta sección nos permitirán visualizar de una manera sencilla los resultados. El capítulo finaliza con un resultado importante (Thm. 36) sobre la teoría de grafos marcados no dirigidos, el cual permite interpretar condiciones de descomposición en términos de propiedades de separación.

En el segundo capítulo se introduce la noción de independencia condicional, la cual toma el protagonismo en este trabajo ya que proporciona resultados admirables y se convierte en un puente para moverse entre distintas áreas de trabajo, en nuestro caso probabilidad y álgebra conmutativa ida y vuelta. Relacionamos las variables aleatorias con grafos no dirigidos, de manera que podamos representar grandes cantidades de información en términos sencillos y a su vez heredando propiedades de la teoría de grafos para ampliar las herramientas de estudio. Se definen las propiedades de Markov en grafos marcados no dirigidos y así finalmente concluir con el teorema de Hammersley-Clifford y su demostración.

En el tercer y último capítulo se definen algunos tipos de modelos estadísticos como los modelos exponenciales y log-lineales, los cuales son comunes en el estudio básico de probabilidad y estadística. Se introduce la noción de un modelo al que llamaremos algebraico, el cual será nuestro centro de atención ya que nos proporcionará una nueva notación y terminología necesaria para algebrizar el teorema de Hammersley-Clifford. Se realizan los ejemplos pertinentes para los modelos log-lineales al ser estos los más importantes del capítulo. Los modelos gráficos se definirán como un caso particular de los modelos log-lineales, donde el conjunto generador corresponderá a la colección de cliques de un grafo no dirigido. En esta última sección se introduce una familia

de polinomios cuadráticos que serán utilizados para caracterizar en términos algebraicos las *afirmaciones de independencia* y que, a su vez, nos servirán para definir la variedad algebraica que contiene como puntos todas las funciones de distribución que factorizan de acuerdo con el grafo inicial.

NOCIONES BÁSICAS DE TEORÍA DE GRAFOS

En este capítulo se intrducen algunos conceptos básicos y la notación necesaria de la teoría de grafos, requerida para comprender la teoría de modelos gráficos. Así mismo se recuerdan algunos resultados y nociones de la teoría de la probabilidad y del álgebra conmutativa. La lectura de este capítulo puede realizarse a medida que se necesiten estos conceptos para la comprensión de capítulos posteriores.

I.1 GRAFOS

I.1.1 Notación y terminología

Definición 1. Un **grafo** (finito) es un par $\mathcal{G} = (V, E)$, donde V es un conjunto finito, llamado conjunto de **vértices** y E es un subconjunto del conjunto $V \times V$, llamado conjunto de **lados**.

Definición 2. Decimos que un grafo $G = (V, E)$ es **simple** si no hay lados de la forma (x, x) , para algún $x \in V$, es decir, si no existen lazos.

Un ejemplo sencillo de grafo simple se da en la figura 1

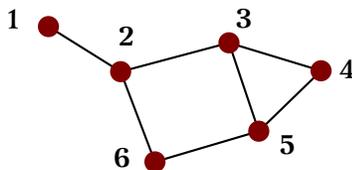


Figura 1: Grafo simple donde $V = \{1, 2, 3, 4, 5, 6\}$ es el conjunto de vértices y $E = \{(1, 2), \{2, 3\}, \{2, 6\}, \{3, 4\}, \{3, 5\}, \{4, 5\}, \{5, 6\}\}$ es el conjunto de lados.

Nota 3. De aquí en adelante los grafos utilizados siempre serán grafos simples.

Definición 4. Sea $\mathcal{G} = (V, E)$ un grafo, un lado $(\alpha, \beta) \in E$ se llama **no dirigido** cuando $(\alpha, \beta) \in E$ y $(\beta, \alpha) \in E$.

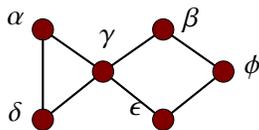


Figura 2: Grafo con todos sus lados no dirigidos.

Definición 5. Sea $\mathcal{G} = (V, E)$ un grafo, un lado $(\alpha, \beta) \in E$ se llama **dirigido** si $(\alpha, \beta) \in E$ pero $(\beta, \alpha) \notin E$.

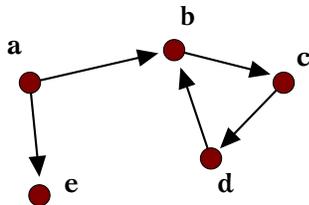


Figura 3: Grafo con todos sus lados dirigidos.

Nota 6. A lo largo de este trabajo utilizamos la siguiente notación:

$$\begin{aligned} \alpha \rightarrow \beta & \text{ si } (\alpha, \beta) \in E \wedge (\beta, \alpha) \notin E, \\ \alpha \sim \beta & \text{ si } (\alpha, \beta) \in E \wedge (\beta, \alpha) \in E, \\ \alpha \nrightarrow \beta & \text{ si } (\alpha, \beta) \notin E, \\ \alpha \nrightarrow \beta & \text{ si } (\alpha, \beta) \notin E \wedge (\beta, \alpha) \notin E. \end{aligned}$$

Nota 7. Una **línea** que une a α con β representa un lado no dirigido, mientras que una **flecha** que apunta de α a β se utiliza para los lados dirigidos $(\alpha, \beta) \in E$ para los que $(\beta, \alpha) \notin E$.

Note que si $\alpha \nrightarrow \beta$ entonces no hay ninguna línea o flecha entre los vértices α y β . No sobra aclarar que el símbolo \nrightarrow **no** está siendo utilizado como la negación de \sim .

Definición 8. Un **grafo marcado** es un grafo en el cual el conjunto de vértices está particionado en dos subconjuntos disjuntos Δ y Γ . Esto es

$$V = \Delta \cup \Gamma \quad \text{con} \quad \Delta \cap \Gamma = \emptyset.$$

Los vértices en Δ son llamados **discretos** y se representan por un **punto** o un círculo relleno, mientras que los vértices en Γ son llamados **continuos** y se representarán por un **círculo**.

Nota 9. Intuitivamente, los grafos utilizados en este trabajo serán utilizados como elemento visual para representar dependencia entre variables aleatorias en un modelo estadístico y, por lo tanto, los vértices serán indizados mediante variables aleatorias. En este caso, los círculos representarán las variables aleatorias continuas, mientras los puntos corresponderán a variables aleatorias discretas.

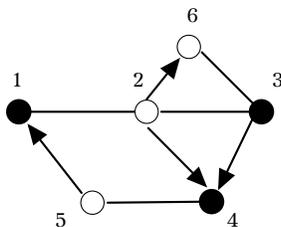


Figura 4: En este grafo encontramos $2 \rightarrow 6$ y también $5 \rightarrow 1$, pero $5 \nrightarrow 1$ ó $3 \nrightarrow 5$ mientras que por ejemplo $1 \sim 2, 3 \sim 6$ ó $4 \sim 5$.

Definición 10. Se dice que un grafo es **puro** si tiene únicamente vértices discretos o vértices continuos.



Figura 5: A la izquierda un grafo puro continuo y a la derecha un grafo puro discreto.

Definición 11. Un grafo que tiene únicamente lados no dirigidos se llama **grafo no dirigido**. Un grafo con todos sus lados dirigidos es llamado **grafo dirigido**.

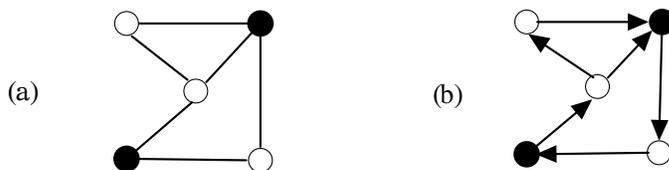


Figura 6: El grafo (a) es un grafo marcado no dirigido y el (b) es un grafo marcado dirigido.

Nota 12. Dado un grafo cualquiera, es conveniente representar los lados no dirigidos como pares no ordenados $\{\alpha, \beta\}$.

En las figuras 4 y 5 damos ejemplos de grafos donde se ilustran las nociones introducidas hasta aquí.

Definición 13. La **versión no dirigida** \mathcal{G}^{\sim} de un grafo \mathcal{G} , es el grafo no dirigido obtenido de \mathcal{G} al sustituir flechas por líneas.

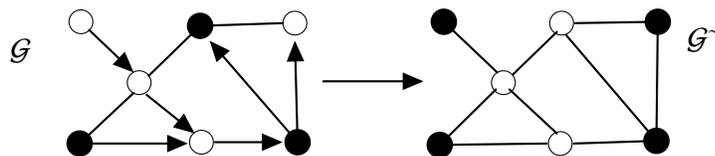


Figura 7: El grafo \mathcal{G} junto a su versión no dirigida.

Intuitivamente es claro como obtener un grafo no dirigido a partir de un grafo dirigido, pero la construcción reciproca no es inmediata si no se tiene un ingrediente adicional, una relación de orden en el conjunto de vértices.

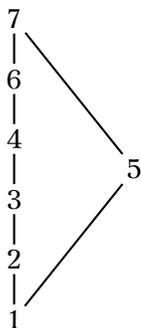
Definición 14. Sea $\mathcal{G} = (V, E)$ un grafo y supongamos que existe una relación de orden irreflexiva $<$ sobre el conjunto de vértices V . La **versión dirigida** de \mathcal{G} es el grafo obtenido al cambiar los lados no dirigidos $\{\alpha, \beta\}$ por lados dirigidos teniendo en cuenta la siguiente regla

$$\alpha \rightarrow \beta \quad \text{si y solo si} \quad \alpha < \beta.$$

El grafo resultante lo denotaremos por $\mathcal{G}^{\prec} = (V^{\prec}, E^{\prec})$, donde el conjunto de vértices V^{\prec} coincide con V , el conjunto de vértices del grafo original.

Notemos que esta construcción recuerda la construcción de una categoría a partir de un conjunto parcialmente ordenado. Solo que en este caso no estamos interesados en conservar las flechas identidad de un objeto en si mismo.

Ejemplo 15. Sea $V = \{1, 2, 3, 4, 5, 6, 7\}$ y consideremos la relación de orden \prec definida en el siguiente retículo



En la figura 8 se ilustra como pasar de un grafo a su versión no dirigida teniendo en cuenta el orden sobre el conjunto de vértices, indicado en el retículo anterior.

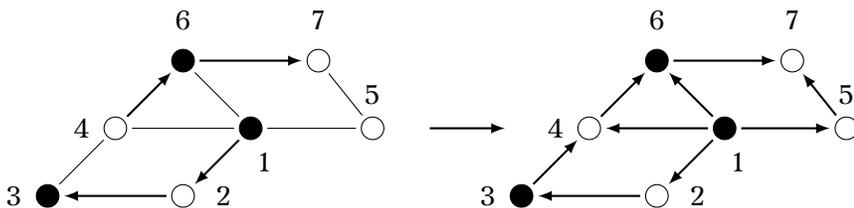


Figura 8: (a) es el Grafo \mathcal{G} y (b) su versión dirigida \mathcal{G}^{\prec} .

En ocasiones es necesario considerar fragmentos del grafo original, determinados por ciertos subconjuntos del conjunto de vértices.

Definición 16. Sea $\mathcal{G} = (V, E)$ un grafo y $A \subseteq V$. Definimos el grafo **inducido por A** y denotado por $\mathcal{G}_A = (A, E_A)$, como el grafo cuyo conjunto de vértices es A y el conjunto de lados E_A es el conjunto $E \cap (A \times A)$.

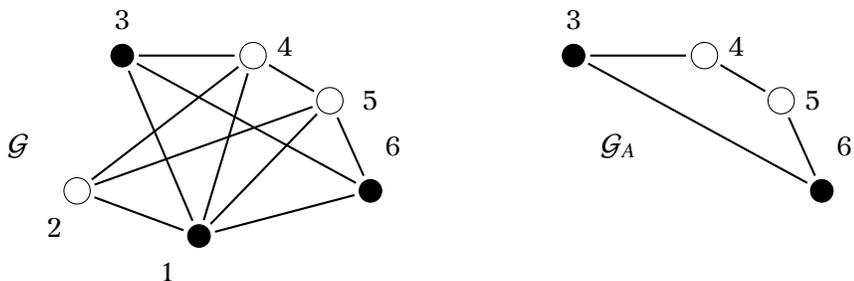


Figura 9: $V = \{1, 2, 3, 4, 5, 6\}$ y $A = \{3, 4, 5, 6\}$

Definición 17. Un grafo es **completo** si todos los vértices están unidos por una línea o una flecha.

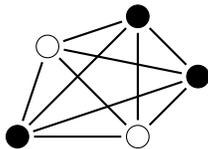


Figura 10: Grafo completo.

Definición 18. Un subconjunto es **completo** si induce un subgrafo completo. Un subconjunto completo maximal (con respecto a \subseteq) es llamado **cliqué**.

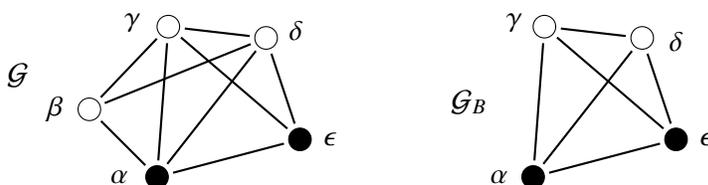


Figura 11: El conjunto $B = \{\alpha, \gamma, \delta, \epsilon\}$ es un subconjunto completo ya que induce el subgrafo completo \mathcal{G}_B , más aún es un cliqué ya que es un subconjunto completo maximal; notar que no es el único cliqué.

Definición 19. Sea $\mathcal{G} = (V, E)$ un grafo y $\alpha, \beta \in V$. Si $\alpha \rightarrow \beta$ se dice que α es **padre** de β y que β es **hijo** de α . El conjunto de padres de β se denota como $\mathbf{pa}(\beta)$ y el conjunto de hijos de α como $\mathbf{ch}(\alpha)$. Si hay una línea entre α y β , se dice que α y β son **adyacentes o vecinos**. Si no existe ninguna línea ni flecha entre α y β , esto es $\alpha \not\sim \beta$, entonces se dice que α y β son **no-adyacentes**. El conjunto de vecinos de un vértice α se denota como $\mathbf{ne}(\alpha)$.

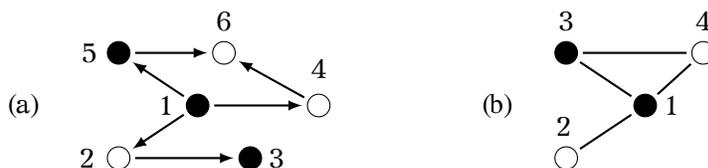


Figura 12: En el grafo (a) encontramos que $\mathbf{pa}(6) = \{5, 4\}$ y $\mathbf{ch}(1) = \{2, 4, 5\}$. Mientras que en el (b) es fácil notar que $\mathbf{ne}(4) = \{3, 1\}$, por otra parte como $2 \not\sim 3$ entonces son no-adyacentes.

Definición 20. Dado un grafo $G = (V, E)$ y un subconjunto A de V , definimos el conjunto de padres, hijos y vecinos de A , los cuales serán denotados por $\mathbf{pa}(A)$, $\mathbf{ch}(A)$ y $\mathbf{ne}(A)$ respectivamente, como

$$\mathbf{pa}(A) = \cup_{\alpha \in A} \mathbf{pa}(\alpha) \setminus A,$$

$$\mathbf{ch}(A) = \cup_{\alpha \in A} \mathbf{ch}(\alpha) \setminus A,$$

$$\mathbf{ne}(A) = \cup_{\alpha \in A} \mathbf{ne}(\alpha) \setminus A.$$

Definición 21. Dado un grafo $G = (V, E)$ y un subconjunto A de V , definimos la frontera de A , que denotaremos por $bd(A)$, como el conjunto de vértices en $V \setminus A$ que son padres o vecinos de vértices en A . La **clausura** de A se define como la unión de la frontera de A con el conjunto A ; en símbolos tenemos entonces que

$$bd(A) = pa(A) \cup ne(A),$$

$$cl(A) = A \cup bd(A).$$

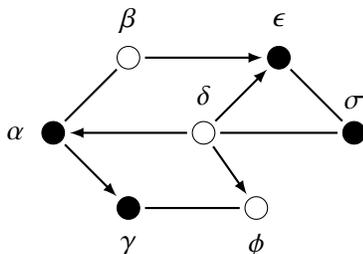


Figura 13: Sea $A = \{\beta, \delta\}$ luego $bd(A) = \{\alpha, \epsilon, \sigma, \phi\}$ y $cl(A) = \{\alpha, \epsilon, \sigma, \phi, \beta, \delta\}$.

Definición 22. Un **camino** de longitud n de α a β es una sucesión

$$\alpha = \alpha_0, \alpha_1, \dots, \alpha_{n-1}, \alpha_n = \beta$$

de vértices distintos tales que $(\alpha_{i-1}, \alpha_i) \in E$ para todo $i = 1, \dots, n$. Si existe un camino de α a β decimos entonces que α conduce a β y lo denotamos como $\alpha \mapsto \beta$. Si se tiene que $\alpha \mapsto \beta$ y $\beta \mapsto \alpha$ entonces decimos que α y β **se conectan** y se denota como $\alpha \rightleftharpoons \beta$

Nota 23. Claramente \rightleftharpoons es una relación de equivalencia y las correspondientes clases de equivalencia se denotan por $[\alpha]$. Las clases de equivalencia de \rightleftharpoons se denominan las **componentes de conectividad** de \mathcal{G} . Si $\alpha \in A \subseteq V$, el símbolo $[\alpha]_A$ denota la componente de conectividad de α en \mathcal{G}_A .

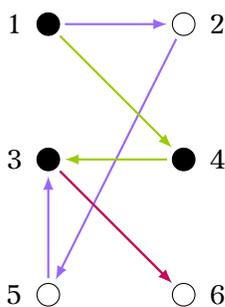


Figura 14: Camino de longitud 3 de $\alpha = 1, 4, 3, 6 = \beta$ y camino de longitud 4 de $\alpha = 1, 2, 5, 3, 6 = \beta$.

Definición 24. Sea $G = (V, E)$ un grafo y $\alpha, \beta \in V$ dos vértices cualesquiera. Un subconjunto $C \subseteq V$ es llamado (α, β) -**separador** si todos los caminos de α a β intersectan a C .

Lema 25. Sea $G = (V, E)$ un grafo no dirigido y $\alpha, \beta \in V$. Un subconjunto C del conjunto de vértices V es un (α, β) -separador si y sólo si

$$[\alpha]_{V \setminus C} \neq [\beta]_{V \setminus C}.$$

Demostración. Inmediato a partir de la definición. □

Definición 26. Sea $G = (V, E)$ un grafo y sean A, B, C subconjuntos del conjunto de vértices V . Se dice que el conjunto C separa a A de B si es un (α, β) -separador para todo $\alpha \in A$ y $\beta \in B$.

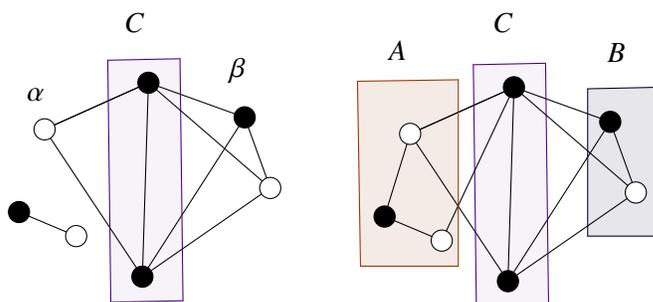


Figura 15: En el grafo de la izquierda el conjunto C es (α, β) -separador. En el grafo de la derecha C separa a A de B .

Definición 27. Una **cadena** de longitud n de α a β es una sucesión

$$\alpha = \alpha_0, \alpha_1, \dots, \alpha_{n-1}, \alpha_n = \beta$$

de vértices distintos tales que $\alpha_{i-1} \rightarrow \alpha_i$ ó $\alpha_i \rightarrow \alpha_{i-1}$ para todo $i = 1, \dots, n$.

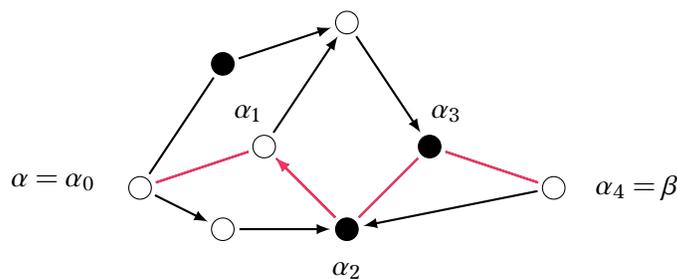


Figura 16: Cadena de α a β de longitud 4 . Sucesión $\alpha = \alpha_0, \alpha_1, \alpha_2, \alpha_3, \alpha_4 = \beta$.

Definición 28. Un n -ciclo en un grafo $G = (V, E)$ es un camino de longitud n que empieza y termina en el mismo punto vértice. Se dice que el ciclo es **dirigido** si existe una flecha entre dos vértices consecutivos que hacen parte del camino.

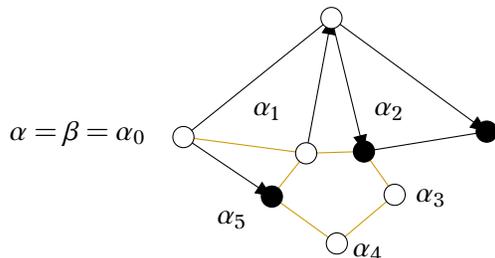


Figura 17: Un 7-ciclo no dirigido tal que $\alpha = \alpha_0, \alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_1, \alpha_0 = \beta$.

1.1.2 Descomposición de grafos marcados

Definición 29. Un triple (A, B, C) de subconjuntos disjuntos del conjunto de vértices V de un grafo marcado no dirigido, se dice que forma una **descomposición (fuerte)** de \mathcal{G} si $V = A \cup B \cup C$ y las siguientes condiciones se satisfacen:

1. C separa a A de B ;
2. C es un subconjunto completo de V ;
3. $C \subseteq \Delta \vee B \subseteq \Gamma$.

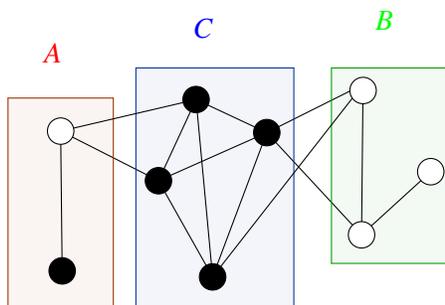


Figura 18: Descomposición fuerte de un grafo marcado no dirigido.

Cuando este es el caso decimos que (A, B, C) descompone a \mathcal{G} en componentes $\mathcal{G}_{A \cup C}$ y $\mathcal{G}_{B \cup C}$.

Para evitar confusión con el término, se utilizará la palabra *fuerte* para diferenciarla de una *descomposición débil*, la cual se define a continuación.

Definición 30. Un triple (A, B, C) de subconjuntos disjuntos del conjunto de vértices V de un grafo marcado no dirigido \mathcal{G} se dice que forma una **descomposición débil** de \mathcal{G} si $V = A \cup B \cup C$ y las siguientes dos condiciones se cumplen:

1. C separa a A de B ;
2. C es un subconjunto completo de V .

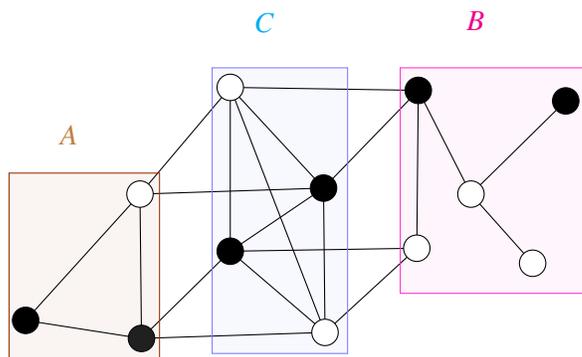


Figura 19: Descomposición débil de un grafo marcado no dirigido.

Note que una descomposición débil satisface las dos primeras condiciones de la definición de descomposición fuerte 29, pero no necesariamente la tercera. En el caso de un grafo puro, esta última condición se satisface automáticamente y así todas las descomposiciones débiles son también fuertes.

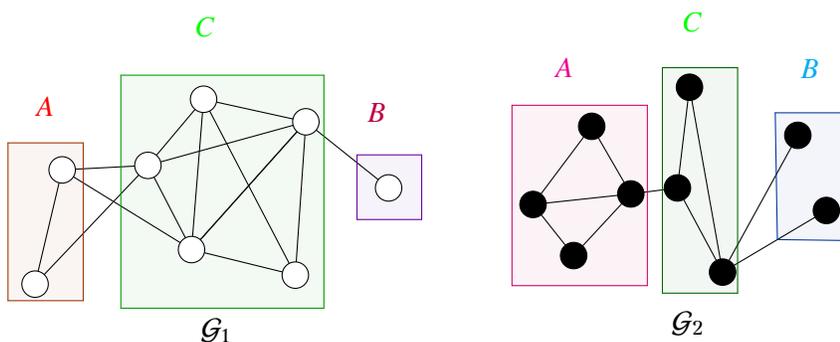


Figura 20: Descomposición de grafos puros no dirigidos.

También se admite que alguno de los conjuntos en (A, B, C) sea vacío. Ver figura 21.

Definición 31. Dado un grafo $G = (E, V)$ y una descomposición (fuerte o débil) en una terna (A, B, C) , si los conjuntos A y B son ambos no vacíos, decimos que la descomposición es **propia**.

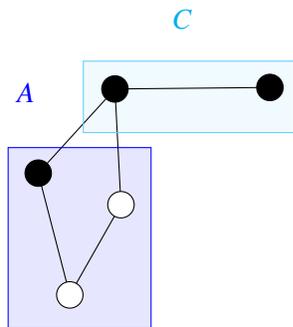


Figura 21: Descomposición fuerte de \mathcal{G} con $B = \emptyset$.

Un grafo que admite *descomposición propia* se puede descomponer sucesivamente en sus cliques.

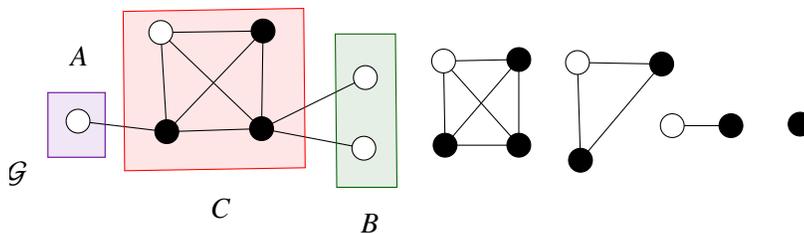


Figura 22: Descomposición en cliques del grafo \mathcal{G} .

Definición 32. Un grafo marcado no dirigido se dice que **admite descomposición** si es completo o si existe una descomposición propia (A, B, C) en subgrafos \mathcal{G}_{AUC} y \mathcal{G}_{BUC} que admiten descomposición.

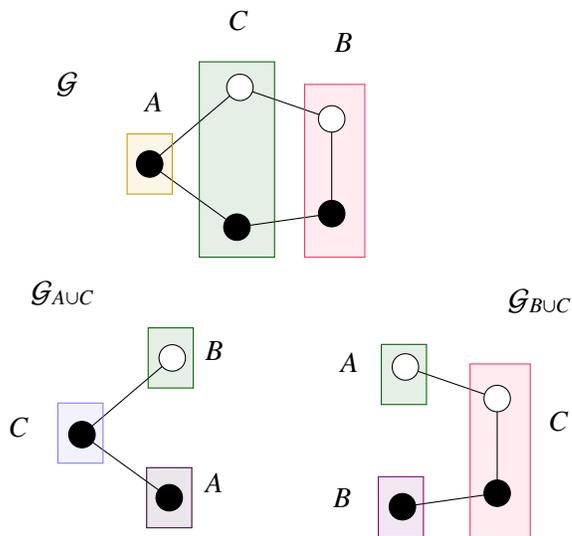


Figura 23: Descomposición propia (A, B, C) de \mathcal{G} con subgrafos \mathcal{G}_{AUC} y \mathcal{G}_{BUC} que admiten descomposición.

Note que la definición tiene sentido ya que la descomposición que se asume es propia y así los dos subgrafos $\mathcal{G}_{A \cup C}$ y $\mathcal{G}_{B \cup C}$ tienen menos vértices que el grafo original \mathcal{G} .

Nota 33. Análogamente, un grafo que admite descomposición débil se puede descomponer en sus cliques por medio de descomposición débil.

Es claro que cualquier grafo que admite descomposición también admite *descomposición débil*, pero el opuesto en general no es cierto.

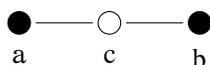


Figura 24: Este es el grafo más pequeño que no admite descomposición fuerte. En efecto, el único candidato para un conjunto C separador es el vértice c pero es continuo. No obstante, si admite la descomposición débil $(\{a\}, \{b\}, \{c\})$.

Definición 34. Un grafo **triangulado** es un grafo no dirigido con la propiedad de que cada ciclo de longitud $n \geq 4$ posee una **cuerda**, esto es, dos vértices no consecutivos que son vecinos.

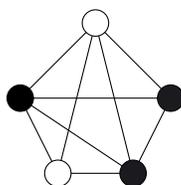


Figura 25: Grafo triangulado.

Proposición 35. Sea \mathcal{G} un grafo triangulado y $A \subset V$, entonces \mathcal{G}_A es triangulado.

Demostración. Tenemos dos casos posibles, si el número de vértices de A es menor o igual a 3, entonces no hay nada que demostrar. En otro caso, cualquier n -ciclo en \mathcal{G}_A , con $n \geq 4$ es un n -ciclo en \mathcal{G} y como este último es triangulado, entonces admite una cuerda, por lo tanto \mathcal{G}_A es triangulado. □

Teorema 36. Sea \mathcal{G} un grafo no dirigido y marcado entonces las siguientes condiciones son equivalentes:

1. \mathcal{G} admite descomposición débil;
2. \mathcal{G} es triangulado;
3. Todo (α, β) -separador minimal es completo.

Demostración. Razonaremos por inducción sobre el número de vértices $|V|$ de \mathcal{G} . Primero notemos que si el grafo no tiene más de tres vértices entonces el resultado es trivial y las tres condiciones se cumplen automáticamente.

Ahora, supongamos que el resultado se cumple para todos los grafos con $|V| \leq n$ y consideremos un grafo \mathcal{G} con $n + 1$ vértices.

- Veamos que (1) \Rightarrow (2). Supongamos que \mathcal{G} admite descomposición débil. Si el grafo es completo es evidente que es triangulado y hemos terminado. En otro caso, el grafo \mathcal{G} se puede descomponer en subgrafos \mathcal{G}_{AUC} y \mathcal{G}_{BUC} , ambos con una cantidad menor de vértices, y por hipótesis de inducción sabemos que cada uno de estos grafos son triangulados. Es decir, que todo ciclo en \mathcal{G}_{AUC} y en \mathcal{G}_{BUC} de longitud $n \geq 4$ posee una cuerda. Luego, la única opción para que exista un ciclo sin cuerda en el grafo \mathcal{G} , es uno que pase por el conjunto A y por el conjunto B . Como un ciclo es un camino de vértices distintos tales que $\alpha = \beta$ (i.e. que el camino comienza y termina en el mismo vértice) entonces ese ciclo sin cuerda debe pasar mínimo dos veces por el conjunto C , ya que este separa a A de B , pero (1) nos dice que C es completo, luego el ciclo posee una cuerda.
- Veamos (2) \Rightarrow (3). Supongamos que \mathcal{G} es triangulado. Sea C un (α, β) -separador minimal. Si C consiste de un solo vértice entonces es completo y hemos acabado. Supongamos que contiene al menos dos vértices, digamos γ_1 y γ_2 . Como C es un separador minimal, existe un camino

$$\alpha, \alpha_1, \dots, \gamma_1, \dots, \alpha_{n-1}, \alpha_n = \beta, \quad (1)$$

de α a β a través de γ_1 y otro camino,

$$\beta, \beta_1, \dots, \gamma_2, \dots, \beta_{m-1}, \beta_m = \alpha, \quad (2)$$

también de α a β , pero a través de γ_2 . Notemos que la sucesión

$$(\alpha, \alpha_1, \dots, \gamma_1, \dots, \alpha_{n-1}, \beta, \beta_{m-1}, \dots, \gamma_2, \dots, \beta_1, \alpha)$$

se diferencia de un ciclo solo en que puede tener puntos repetidos. En otras palabras, esta sucesión de vértices que comienza y termina en α puede que requiera pasar sobre un cierto vértice $\delta = \alpha_i = \beta_j$ para ir de α a β y luego para volver a α . Consideremos α_r el vértice en (1) tal que para todo $i \geq r$ se cumple que $\alpha_i \neq \beta_j$ en la componente $[\alpha]_{V \setminus C}$, es decir, el vértice más cercano a C del camino (1), en dirección α a β , que coincide con uno del camino (2). Así mismo, sea β_s el vértice en (2) tal que para todo $j \geq s$ se cumple que $\alpha_i \neq \beta_j$ en la componente $[\beta]_{V \setminus C}$, es decir, el vértice más cercano a C del camino (2), en dirección de β hacia α , que coincide con uno del camino (1).

De lo anterior, la sucesión de vértices

$$(\alpha_r, \alpha_{r+1}, \dots, \gamma_1, \dots, \beta_s, \beta_{s+1}, \dots, \gamma_2, \dots, \alpha_r), \quad (3)$$

determina un ciclo de longitud ≥ 4 , pues A , B y C son disjuntos. Luego, dado que \mathcal{G} es triangulado, podemos afirmar que posee una cuerda. Esta cuerda la podemos utilizar para encontrar un ciclo de longitud menor al anterior que comienza en un punto de A , pasa por γ_1 , luego por puntos de B y de ahí regresa a α pasando por γ_2 . Repitiendo este proceso de eliminar vértices, haciendo uso de las cuerdas encontradas, logramos obtener, finalmente, un 4-ciclo

$$(\alpha', \gamma_1, \beta', \gamma_2, \alpha'),$$

donde $\alpha' \in [\alpha]_{V \setminus C}$ y $\beta' \in [\beta]_{V \setminus C}$. Dado que C es un (α, β) separador minimal y G es triangulado, podemos afirmar que $\{\gamma_1, \gamma_2\} \in E$.

- Finalmente, veamos que (3) \Rightarrow (1). Supongamos que todo (α, β) -separador minimal es completo. Si \mathcal{G} es completo no hay nada que demostrar. En otro caso, tiene al menos dos vértices α y β que no son adyacentes. Sea C un (α, β) -separador minimal, vamos a dividir el conjunto de vértices V en 4 conjuntos disjuntos:

1. $[\alpha]_{V \setminus C}$, i.e la componente de conectividad de α en $\mathcal{G}_{V \setminus C}$,
2. $[\beta]_{V \setminus C}$, i.e la componente de conectividad de β en $\mathcal{G}_{V \setminus C}$,
3. El conjunto C ,
4. D denotará el conjunto de vértices restantes;

es claro que $V = [\alpha]_{V \setminus C} \cup [\beta]_{V \setminus C} \cup C \cup D$. Dado que C es completo, si definimos $A = [\alpha]_{V \setminus C} \cup D$ y a $B = [\beta]_{V \setminus C}$ entonces el triple (A, B, C) es una descomposición débil de \mathcal{G} .

Ahora, los subgrafos $\mathcal{G}_{A \cup C}$ y $\mathcal{G}_{B \cup C}$ también admiten descomposición. De hecho, si C_1 es un (α_1, β_1) -separador minimal en $\mathcal{G}_{A \cup C}$ entonces, es evidente que C_1 está contenido en un (α_1, β_1) -separador minimal de \mathcal{G} y por hipótesis de inducción implica que $\mathcal{G}_{A \cup C}$ admite descomposición débil y se obtiene el resultado de manera similar para $\mathcal{G}_{B \cup C}$. En conclusión, hemos encontrado una descomposición débil a partir de subgrafos que admiten descomposición débil, lo que termina la demostración.

□

Nota 37. *El grafo más pequeño que no admite descomposición débil es un 4-ciclo.*

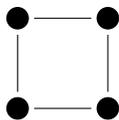


Figura 26: El grafo \mathcal{G} más pequeño que no cumple las condiciones para admitir mínimo descomposición débil.

INDEPENDENCIA CONDICIONAL Y PROPIEDADES DE MARKOV

En [D] Philip Dawid, mediante el uso de conceptos como *espacio estadístico*, *núcleos de Markov* y *conjuntos ignorables* introduce un marco teórico formal para el estudio de la *independencia condicional*, concepto fundamental para unificar áreas de la estadística que, hasta hace algún tiempo, parecían no tener una relación directa. En este capítulo estudiaremos, sin entrar en detalles de la teoría de la medida, algunas propiedades del concepto de *independencia condicional* y algunas relaciones con la teoría de grafos, finalizando con la demostración del teorema de Hammersley-Clifford.

2.1 INDEPENDENCIA CONDICIONAL Y PROPIEDADES DE MARKOV

Durante todo este trabajo, dada una variable aleatoria W usaremos la notación f_W para denotar la función de densidad de W . Así mismo, dadas variables aleatorias W_1, \dots, W_n , la función de densidad de la distribución conjunta se denotará por $f_{W_1 \dots W_n}$.

Nota 38. En este capítulo suponemos que las funciones de densidad encontradas son continuas, con el fin de evitar problemas técnicos con igualdad de funciones salvo en conjuntos de medida cero.

Definición 39. Dadas variables aleatorias $W_1, \dots, W_n, Z_1, \dots, Z_m$, definimos la función de densidad condicional $f_{W_1 \dots W_n | Z_1, \dots, Z_m}$ mediante la expresión

$$f_{W_1 \dots W_n | Z_1, \dots, Z_m}(w_1 \dots w_n | z_1, \dots, z_m) = \frac{f_{W_1 \dots W_n, Z_1, \dots, Z_m}(w_1 \dots w_n, z_1, \dots, z_m)}{f_{Z_1, \dots, Z_m}(z_1, \dots, z_m)}.$$

Definición 40. Si X, Y y Z son variables aleatorias discretas, con distribución conjunta P , decimos que X es **condicionalmente independiente** de Y dado Z bajo P , lo cual se denotará por $X \perp\!\!\!\perp Y | Z[P]$, si

$$P(X = x, Y = y | Z = z) = P(X = x | Z = z)P(Y = y | Z = z).$$

Ahora, si X, Y y Z son variables aleatorias continuas, decimos que X es **condicionalmente independiente** de Y dado Z si

$$f_{XY|Z}(x, y | z) = f_{X|Z}(x | z)f_{Y|Z}(y | z).$$

Nota 41. Note que la condición que define la independencia condicional para variables aleatorias discretas es, esencialmente, la misma que para el caso continuo. Para más detalles ver [D].

A continuación describiremos algunas propiedades del concepto de independencia condicional.

Proposición 42. Sea \mathcal{X} un espacio de probabilidad y $h : \mathcal{X} \rightarrow \mathcal{X}$ una función medible cualquiera. Dadas variables aleatorias X, Y, Z sobre \mathcal{X} la relación de independencia condicional satisface las siguientes propiedades:

(C1) si $X \perp\!\!\!\perp Y | Z$ entonces $Y \perp\!\!\!\perp X | Z$;

- (C2) si $X \perp\!\!\!\perp Y | Z$ y $U = h(X)$, entonces $U \perp\!\!\!\perp Y | Z$;
 (C3) si $X \perp\!\!\!\perp Y | Z$ y $U = h(X)$, entonces $X \perp\!\!\!\perp Y | (Z, U)$;
 (C4) si $X \perp\!\!\!\perp Y | Z$ y $X \perp\!\!\!\perp W | (Y, Z)$, entonces $X \perp\!\!\!\perp (W, Y) | Z$.

Demostración. Sea \mathcal{X} un espacio de probabilidad y $h : \mathcal{X} \rightarrow \mathbb{R}$ una función medible cualquiera. Sean X, Y, Z sobre \mathcal{X} variables aleatorias. La primer propiedad es consecuencia inmediata de la definición de probabilidad condicional en términos de la función de densidad. En efecto, si $X \perp\!\!\!\perp Y | Z$ entonces $f_{XY|Z}(x, y | z) = f_{X|Z}(x | z)f_{Y|Z}(y | z)$ que obviamente coincide con la condición $Y \perp\!\!\!\perp X | Z$.

Para la condición 2, notemos que dada una función Borel medible $h : \mathbb{R} \rightarrow \mathbb{R}$, la variable aleatoria $h(X)$ esta definida como la composición $h \circ X$ indicada en el diagrama conmutativo

$$\begin{array}{ccc} \Omega & \xrightarrow{X} & \mathbb{R} \\ & \searrow h(X) & \downarrow h \\ & & \mathbb{R}; \end{array}$$

luego, si $X \perp\!\!\!\perp Y | Z$ entonces

$$\begin{aligned} \mathbf{P}(h(X) \in A, Y \in B | Z \in C) &= \mathbf{P}(X^{-1}(h^{-1}(A)) \cap Y^{-1}(B) | Z^{-1}(C)) \\ &= \mathbf{P}(X^{-1}(h^{-1}(A)) | Z^{-1}(C)) \mathbf{P}(Y^{-1}(B) | Z^{-1}(C)) \\ &= \mathbf{P}(h(X)^{-1}(A) | Z^{-1}(C)) \mathbf{P}(Y^{-1}(B) | Z^{-1}(C)), \end{aligned}$$

lo que indica que $h(X) \perp\!\!\!\perp Y | Z$.

En el caso de la condición 3, conservando la notación del párrafo anterior y dados $A, B, C, D \subset \mathbb{R}$ subconjuntos de Borel, tenemos

$$\begin{aligned} \mathbf{P}(X \in A, Y \in B | (Z, h(X)) \in C \times D) &= \frac{\mathbf{P}(X^{-1}(A) \cap Y^{-1}(B) \cap (Z, h(X))^{-1}(C \times D))}{\mathbf{P}((Z, h(X))^{-1}(C \times D))} \\ &= \frac{\mathbf{P}(X^{-1}(A) \cap Y^{-1}(B) \cap Z^{-1}(C) \cap h(X)^{-1}(D))}{\mathbf{P}(Z^{-1}(C) \cap h(X)^{-1}(D))} \\ &= \frac{\mathbf{P}(X^{-1}(A \cap h^{-1}(D)) \cap Y^{-1}(B) \cap Z^{-1}(C))}{\mathbf{P}(Z^{-1}(C) \cap h(X)^{-1}(D))} \\ &= \frac{\mathbf{P}(X^{-1}(A \cap h^{-1}(D)) \cap Y^{-1}(B) \cap Z^{-1}(C))}{\mathbf{P}(Z^{-1}(C))} \frac{\mathbf{P}(Z^{-1}(C))}{\mathbf{P}(Z^{-1}(C) \cap h(X)^{-1}(D))} \\ &= \frac{\mathbf{P}(X^{-1}(A \cap h^{-1}(D)) \cap Z^{-1}(C))}{\mathbf{P}(Z^{-1}(C))} \frac{\mathbf{P}(Y^{-1}(B) \cap Z^{-1}(C))}{\mathbf{P}(Z^{-1}(C))} \\ &\quad \frac{\mathbf{P}(Z^{-1}(C))}{\mathbf{P}(Z^{-1}(C) \cap h(X)^{-1}(D))} \\ &= \frac{\mathbf{P}(X^{-1}(A) \cap h(X)^{-1}(D) \cap Z^{-1}(C))}{\mathbf{P}(Z^{-1}(C) \cap h(X)^{-1}(D))} \frac{\mathbf{P}(Y^{-1}(B) \cap Z^{-1}(C) \cap h(X)^{-1}(D))}{\mathbf{P}(Z^{-1}(C) \cap h(X)^{-1}(D))} \\ &\quad \frac{\mathbf{P}(Z^{-1}(C) \cap h(X)^{-1}(D))}{\mathbf{P}(Y^{-1}(B) \cap Z^{-1}(C) \cap h(X)^{-1}(D))} \frac{\mathbf{P}(Y^{-1}(B) \cap Z^{-1}(C))}{\mathbf{P}(Z^{-1}(C))} \\ &= \mathbf{P}(X^{-1}(A) | (Z, h(X))^{-1}(C \times D)) \mathbf{P}(Y^{-1}(B) | (Z, h(X))^{-1}(C \times D)); \end{aligned}$$

esto último dado que

$$\frac{\mathbf{P}(Y^{-1}(B) \cap Z^{-1}(C) \cap h(X)^{-1}(D))}{\mathbf{P}(Z^{-1}(C))} = \frac{\mathbf{P}(Y^{-1}(B) \cap Z^{-1}(C))}{\mathbf{P}(Z^{-1}(C))} \frac{\mathbf{P}(h(X)^{-1}(D) \cap Z^{-1}(C))}{\mathbf{P}(Z^{-1}(C))}.$$

La condición 4 se demuestra de manera análoga. \square

Algunas propiedades del concepto de independencia condicional se pueden expresar fácilmente si utilizamos las funciones de densidad, tal como se ilustra en la siguiente proposición.

Proposición 43. *Si se utiliza f como un símbolo genérico para la función de densidad de las variables aleatorias correspondientes a sus argumentos, las siguientes afirmaciones son ciertas:*

$$X \perp\!\!\!\perp Y | Z \iff f(x, y, z) = f(x, z)f(y, z) / f(z); \quad (a)$$

$$X \perp\!\!\!\perp Y | Z \iff f(x | y, z) = f(x | z); \quad (b)$$

$$X \perp\!\!\!\perp Y | Z \iff f(x, z | y) = f(x | z)f(z | y); \quad (c)$$

$$X \perp\!\!\!\perp Y | Z \iff f(x, y, z) = h(x, z)k(y, z), \quad \text{para algún par de funciones } h \text{ y } k; \quad (d)$$

$$X \perp\!\!\!\perp Y | Z \iff f(x, y, z) = f(x | z)f(y, z). \quad (e)$$

Demostración. Según la definición tenemos $X \perp\!\!\!\perp Y | Z$ si y solo si $f_{XY|Z}(x, y | z) = f_{X|Z}(x | z)f_{Y,Z}(y | z)$, de aquí que $X \perp\!\!\!\perp Y | Z$ si y solo si $\frac{f(x, y, z)}{f(z)} = \frac{f(x, z)}{f(z)} \frac{f(y, z)}{f(z)}$, lo cual se cumple si y solo si

$$f(x, y, z) = \frac{f(x, z)f(y, z)}{f(z)}. \quad (4)$$

con lo cual obtenemos la propiedad (a).

Para (b) notemos, de lo anterior, que $X \perp\!\!\!\perp Y | Z$ si y solo si $f(x, y, z) = (f(x, z)f(y, z)) / f(z)$. Así, suponiendo que ninguna de las cantidades en cuestión se anulan, se sigue que $X \perp\!\!\!\perp Y | Z$ si y solo si $f(x, y, z) / f(y, z) = f(x, z) / f(z)$, es decir, si y solo si $f(x | y, z) = f(x | z)$.

La propiedad (c) se demuestra de manera análoga a la anterior, basta dividir a ambos lados de la igualdad (4) por $f(y)$.

Para la propiedad (d) basta demostrar que $X \perp\!\!\!\perp Y | Z$ si $f(x, y, z) = h(x, z)k(y, z)$. En efecto, *marginalizando* sobre y tenemos

$$\begin{aligned} \int_{\Omega} f(x, y, z) \, d\mathbf{P}(y) &= \int_{\Omega} h(x, z)k(y, z) \, d\mathbf{P}(y) \\ &= h(x, z) \int_{\Omega} k(y, z) \, d\mathbf{P}(y); \end{aligned}$$

esto es $f(x, z) = h(x, z)\tilde{k}(z)$, donde $\tilde{k}(z) = \int_{\Omega} k(y, z) \, d\mathbf{P}(y)$. Si *marginalizamos* de nuevo, pero esta vez con respecto a x , tenemos,

$$\begin{aligned} \int_{\Omega} \int_{\Omega} f(x, y, z) \, d\mathbf{P}(y) \, d\mathbf{P}(x) &= \int_{\Omega} h(x, z)\tilde{k}(z) \, d\mathbf{P}(x) \\ &= \tilde{k}(z) \int_{\Omega} h(y, z) \, d\mathbf{P}(y) \\ &= \tilde{k}(z)\tilde{h}(z); \end{aligned}$$

donde $\tilde{h}(z) = \int_{\Omega} h(y, z) \, d\mathbf{P}(y)$ y por lo tanto $f(z) = \tilde{k}(z)\tilde{h}(z)$. Repitiendo este proceso, pero integrando esta vez con respecto a x , obtenemos $f(y, z) = \tilde{h}(z)k(y, z)$.

De lo anterior se sigue que

$$\begin{aligned} f(x,y,z) &= h(x,z)k(y,z) \\ &= \frac{\tilde{k}(z)h(x,z)\tilde{h}(z)k(y,z)}{\tilde{k}(z)\tilde{h}(z)} \\ &= \frac{f(x,z)f(y,z)}{f(z)}, \end{aligned}$$

y así concluimos que $X \perp\!\!\!\perp Y | Z$.

La propiedad (e) es inmediata a partir de la ecuación (4). \square

Nota 44. Existe una propiedad adicional, intuitivamente plausible, para la relación de independencia condicional,

$$(C5) \text{ si } X \perp\!\!\!\perp Y | Z \text{ y } X \perp\!\!\!\perp Z | Y, \text{ entonces } X \perp\!\!\!\perp (Y,Z),$$

sin embargo esta no se cumple en general.

Ejemplo 45. Un contraejemplo se tiene cuando $X = Y = Z$ son variables aleatorias binarias con probabilidad $P(X = 1) = P(X = 0) = \frac{1}{2}$. Veamos que las hipótesis de la condición (C5) se cumplen. Notemos que $X \perp\!\!\!\perp Y | Z$, se cumple si y solo si

$$P(X = x, Y = y, Z = z) = \frac{P(X = x, Z = z)P(Y = y, Z = z)}{P(Z = z)}.$$

Los valores posibles para las ternas (x,y,z) están en el conjunto

$$\mathcal{X} = \{(0,0,0), (0,0,1), (0,1,0), (1,0,0), (1,1,0), (1,0,1), (0,1,1), (1,1,1)\},$$

de lo cuales $P(0,0,0) = P(1,1,1) = \frac{1}{2}$ y en las demás ternas P se anula, esto dado que $X = Y = Z$.

Por otra parte, $P(X = x) = \frac{1}{2}$, $P(Z = z) = \frac{1}{2}$ y $P(Y = y, Z = z) = \frac{1}{2}$ cuando $y = z$. Por lo tanto,

$$\frac{P(X = x)P(Y = y, Z = z)}{P(Z = z)} = \frac{1}{2}.$$

Este mismo resultado se obtiene con $X \perp\!\!\!\perp Z | Y$. Ahora veamos que $X \perp\!\!\!\perp (Y,Z)$ no se cumple. Es decir,

$$P(X = x, Y = y, Z = z) \neq P(X = x)P((Y,Z) = (y,z)).$$

Si tenemos que $x \neq y, x \neq z$ o $y \neq z$ entonces $P(x,y,z) = 0$, como es el caso de $(x,y,z) = (0,1,1)$. Ahora

$$P(X = 0)P(Y = 1, Z = 1) = \left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = \frac{1}{4}.$$

Proposición 46. Si la densidad conjunta de todas las variables con respecto a la medida producto es positiva y continua entonces la condición (C5) se cumple.

Demostración. Supongamos que la función de densidad conjunta es continua y positiva, esto es $f(x,y,z) > 0$, y supongamos también que $X \perp\!\!\!\perp Y | Z$ y $X \perp\!\!\!\perp Z | Y$. Entonces, por la proposición 43 tenemos que para todos los valores (x,y,z) :

$$f(x,y,z) = k(x,z)l(y,z) = g(x,y)h(y,z)$$

para funciones estrictamente positivas g, h, k, l . Entonces, como la función de densidad se asume continua (ver nota 38) para todo z tenemos que

$$g(x, y) = \frac{k(x, z)l(y, z)}{h(y, z)}.$$

Ahora fijamos $z = z_0$ y se hace $\pi(x) = k(x, z_0)$ y $\gamma(y) = l(y, z_0)/h(y, z_0)$ entonces $g(x, y) = \pi(x)\gamma(y)$. Entonces $f(x, y, z) = \pi(x)\gamma(y)h(y, z)$ y por lo tanto $X \perp\!\!\!\perp (Y, Z)$. \square

2.1.1 Propiedades de Markov

En esta sección estudiaremos algunas relaciones que existen entre el concepto de independencia condicional de variables aleatorias y grafos. Sea $\mathcal{G} = (V, E)$ un grafo y $\{X_\alpha : \Omega \rightarrow \mathcal{X}_\alpha\}_{\alpha \in V}$ una colección de variables aleatorias, donde cada \mathcal{X}_α es un espacio medible y a su vez, espacios vectoriales de dimensión finita o conjuntos discretos (finitos). Dado $A \subseteq V$ denotaremos por X_A a la colección de variables aleatorias $\{X_\alpha\}_{\alpha \in A}$ y $\mathcal{X}_A = \prod_{\alpha \in A} \mathcal{X}_\alpha$, los elementos de \mathcal{X}_A se denotan por $x_A = (x_\alpha)_{\alpha \in A}$ y, además, denotaremos por X al producto cartesiano \mathcal{X}_V de todos los espacios medibles indizados por el conjunto de vértices de grafo en cuestión.

Por último, dados $A, B, C \subseteq V$ usaremos la notación $A \perp\!\!\!\perp B | C$ para escribir la propiedad de independencia condicional de los vectores aleatorios $X_A \perp\!\!\!\perp X_B | X_C$.

Definición 47. Una medida de probabilidad P sobre X se dice que cumple

(P) **La propiedad de Markov por parejas**, en relación con \mathcal{G} , si para cada par (α, β) de vértices no adyacentes

$$\alpha \perp\!\!\!\perp \beta | V \setminus \{\alpha, \beta\};$$

(L) **La propiedad local de Markov**, relativo a \mathcal{G} , si para cualquier vértice $\alpha \in V$

$$\alpha \perp\!\!\!\perp V \setminus cl(\alpha) | ne(\alpha);$$

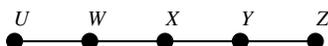
(G) **La propiedad global de Markov**, relativo a \mathcal{G} , si para cualquier terna (A, B, S) de subconjuntos disjuntos de V tal que S separa a A de B en \mathcal{G}

$$A \perp\!\!\!\perp B | S.$$

Ejemplo 48. Consideremos una colección de cinco variables aleatorias binarias U, W, X, Y, Z que satisfacen las siguientes condiciones:

- U y Z son independientes,
- $W = U, Y = Z$ y $X = WY$,
- $P(U = 1) = P(Z = 1) = P(U = 0) = P(Z = 0) = 1/2$.

y consideremos el grafo



Veamos que la propiedad global de Markov no se cumple. Consideremos $A = \{U, W\}$, $B = \{Y, Z\}$ y $S = \{X\}$. Luego

$$P(U = u, W = w, Y = y, Z = z, X = x) \neq \frac{P(U = u, W = w, X = x)P(Y = y, Z = z, X = x)}{P(X = x)}.$$

Veamos los valores para los cuales la distribución conjunta de cada uno de los factores involucrados en la relación anterior no se anula:

$$\begin{aligned}\Omega_{u,w,y,z,x} &= \{(0,0,0,0,0), (1,1,0,0,0), (0,0,1,1,0), (1,1,1,1,1)\} \\ \Omega_{u,w,x} &= \{(0,0,0), (1,1,1)\} \\ \Omega_{y,z,x} &= \{(0,0,0), (1,1,1)\} \\ \Omega_x &= \{0,1\},\end{aligned}$$

donde la probabilidad conjunta del primero es de $1/4$ y la de los demás $1/2$. Por lo tanto, en particular si elegimos la terna $(u, w, x) = (1,1,1)$ y $(y, z, w) = (1,1,1)$ tenemos

$$\begin{aligned}P(U = 1, W = 1, Y = 1, Z = 1, X = 1) &\neq \frac{P(U=1, W=1, X=1)P(Y=1, Z=1, X=1)}{P(X=1)} \\ \frac{1}{4} &\neq \frac{1}{2}.\end{aligned}$$

Por otra parte, la propiedad local de Markov sí se cumple para este grafo. En un primer caso, consideremos $\alpha = U$ luego $V \setminus cl(U) = \{X, Y, Z\}$ y $ne(U) = \{W\}$, por lo que debemos verificar que se cummple

$$P(U = u, X = x, Y = y, Z = z, W = w) = \frac{P(U = u, W = w)P(X = x, Y = y, Z = z, W = w)}{P(W = w)}$$

De nuevo veamos los valores para los cuales la distribución conjunta de cada factor involucrado en la igualdad anterior no se anula:

$$\begin{aligned}\Omega_{u,x,y,z,w} &= \{(0,0,0,0,0), (0,0,1,1,0), (1,0,0,0,1), (1,1,1,1,1)\} \\ \Omega_{x,y,z,w} &= \{(0,0,0,0), (0,0,0,1), (0,1,1,0), (1,1,1,1)\} \\ \Omega_{u,w} &= \{(0,0), (1,1)\} \\ \Omega_w &= \{0,1\},\end{aligned}$$

donde la probabilidad conjunta de (U, X, Y, Z) y (X, Y, Z, W) es de $1/4$ y la de (U, W) y W es de $1/2$ por lo tanto la igualdad se cumple. Veamos ahora el caso en que $\alpha = X$, luego $V \setminus cl(X) = \{U, Z\}$ y $ne(X) = \{W, Y\}$. Verifiquemos la igualdad

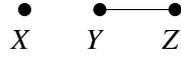
$$P(X = x, U = u, Z = z, W = w, Y = y) = \frac{P(X = x, W = w, Y = y)P(U = u, Z = z, W = w, Y = y)}{P(W = w, Y = y)},$$

siguiendo el mismo razonamiento anterior tenemos los conjuntos

$$\begin{aligned}\Omega_{x,u,z,w,y} &= \{(0,0,0,0,0), (0,0,1,0,1), (0,1,0,1,0), (1,1,1,1,1)\} \\ \Omega_{u,z,w,y} &= \{(0,0,0,0), (0,1,0,1), (1,0,1,0), (1,1,1,1)\} \\ \Omega_{x,w,y} &= \{(0,0,0), (0,0,1), (0,1,0), (1,1,1)\} \\ \Omega_{w,y} &= \{(0,0), (0,1), (1,0), (1,1)\},\end{aligned}$$

donde la probabilidad conjunta de todos es de $1/4$. Y así vemos que se cumple la propiedad local de Markov.

Ejemplo 49. Veamos un ejemplo sencillo que satisface la propiedad por parejas de Markov con respecto al siguiente grafo



Para tres variables aleatorias binarias $X = Y = Z$ con $P(X = 1) = P(X = 0) = 1/2$. Sean $\alpha, \beta \in V$ no adyacentes, en particular tomemos $\alpha = X$ y $\beta = Z$, veamos que se cumple la siguiente igualdad

$$P(X = x, Y = y, Z = z) = \frac{P(X = x, Z = z)P(Y = y, Z = z)}{P(Z = z)},$$

donde los valores para los cuales la distribución conjunta de cada factor involucrado en la igualdad anterior no se anulan son los siguientes

$$\begin{aligned}\Omega_{x,y,z} &= \{(0,0,0), (1,1,1)\} \\ \Omega_{x,z} &= \{(0,0), (1,1)\} \\ \Omega_{y,z} &= \{(0,0), (1,1)\} \\ \Omega_w &= \{0,1\},\end{aligned}$$

y como cada probabilidad conjunta es de $1/2$ se cumple la propiedad por parejas. Por otra parte, con este mismo grafo podemos ver que la propiedad local de Markov no se cumple. Si tomamos $\alpha = X$ tenemos que $V \setminus cl(X) = \{Y, Z\}$ y $ne(X) = \{\emptyset\}$ luego tenemos la relación de independencia $X \perp\!\!\!\perp (Y, Z)$. Veamos que no se cumple la igualdad

$$P(X = x, Y = y, Z = z) = P(X = x)P(Y = y, Z = z)$$

Como los valores de la distribución conjunta donde no se anulan los factores involucrado en la igualdad anterior son

$$\begin{aligned}\Omega_{x,y,z} &= \{(0,0,0), (1,1,1)\} \\ \Omega_{y,z} &= \{(0,0), (1,1)\} \\ \Omega_x &= \{0,1\},\end{aligned}$$

y su probabilidad conjunta es de $1/2$ para todos, si tomamos en particular el valor $(x,y,z) = (0,1,1)$ tenemos que

$$\begin{aligned}P(X = 0, Y = 1, Z = 1) &= P(X = 0)P(Y = 1, Z = 1) \\ 0 &= \frac{1}{4}.\end{aligned}$$

Estas propiedades de Markov están relacionadas de la siguiente manera:

Proposición 50. Para cualquier grafo no dirigido \mathcal{G} y cualquier distribución de probabilidad sobre X se cumple que

$$(G) \implies (L) \implies (P).$$

Demostración. $(G) \implies (L)$ Sea $\alpha \in V$ utilizando $S = ne(\alpha)$, $V \setminus cl(\alpha)$ y $\{\alpha\} = A$, vemos que S separa a α de $V \setminus cl(\alpha)$, por lo tanto (G) implica (L) .

$(L) \implies (P)$. Supongamos que para cualquier vértice $\alpha \in V$ se cumple

$$\alpha \perp\!\!\!\perp V \setminus cl(\alpha) \mid ne(\alpha).$$

Si α y β no son adyacentes entonces $\beta \in V \setminus cl(\alpha)$, por lo tanto

$$V \setminus \{\alpha, \beta\} = ne(\alpha) \cup ((V \setminus cl(\alpha)) \setminus \{\beta\}),$$

de donde por (L) y (C3), donde

$$V \setminus cl(\alpha) = \{\beta, \gamma_1, \gamma_2, \dots, \gamma_s\},$$

y las funciones $h_1 = \pi_{\gamma_1}$, proyección al espacio \mathcal{X}_{γ_1} , y $U_1 = h_1(X)$ dadas por

$$\begin{array}{ccc} \Omega & \xrightarrow{X} & \mathcal{X}_{V \setminus cl(\alpha)} = (\mathcal{X}_\beta \times \mathcal{X}_{\gamma_1} \times \dots \times \mathcal{X}_{\gamma_s}) \\ & \searrow^{U_1 = h_1(X)} & \downarrow h_1 = \pi_{\gamma_1} \\ & & \mathcal{X}_{\gamma_1} \end{array}$$

podemos afirmar que

$$\alpha \perp\!\!\!\perp V \setminus cl(\alpha) \mid ne(\alpha) \cup \{\gamma_1\}.$$

De nuevo por (C3), si definimos las funciones $h_2 = \pi_{\gamma_2}$ y $U_2 = h_2(X)$ dadas por

$$\begin{array}{ccc} \Omega & \xrightarrow{X} & \mathcal{X}_{V \setminus cl(\alpha)} = (\mathcal{X}_\beta \times \mathcal{X}_{\gamma_1} \times \mathcal{X}_{\gamma_2} \times \dots \times \mathcal{X}_{\gamma_s}) \\ & \searrow^{U_2 = h_2(X)} & \downarrow h_2 = \pi_{\gamma_2} \\ & & \mathcal{X}_{\gamma_2} \end{array}$$

se sigue que

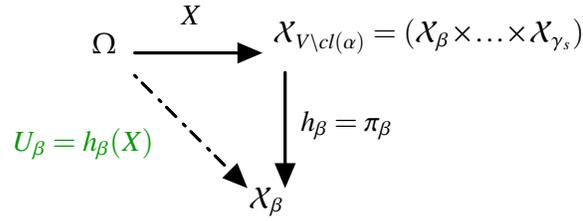
$$\alpha \perp\!\!\!\perp V \setminus cl(\alpha) \mid ne(\alpha) \cup \{\gamma_1\} \cup \{\gamma_2\}.$$

Continuando de esta forma, utilizando todas las funciones proyección, hasta la última $h_s = \pi_{\gamma_s}$ y $U_s = h_s(X)$ dadas por

$$\begin{array}{ccc} \Omega & \xrightarrow{X} & \mathcal{X}_{V \setminus cl(\alpha)} = (\mathcal{X}_\beta \times \dots \times \mathcal{X}_{\gamma_s}) \\ & \searrow^{U_s = h_s(X)} & \downarrow h_s = \pi_{\gamma_s} \\ & & \mathcal{X}_{\gamma_s} \end{array}$$

obtenemos finalmente que

$$\alpha \perp\!\!\!\perp V \setminus cl(\alpha) \mid ne(\alpha) \cup \{\gamma_1\} \cup \dots \cup \{\gamma_s\},$$



es decir,

$$\alpha \perp\!\!\!\perp V \setminus cl(\alpha) \mid V \setminus \{\alpha, \beta\}.$$

Con el último resultado y la condición (C2), si definimos $h_\beta = \pi_\beta$ y a $U_\beta = h_\beta(X)$, dadas por obtenemos el resultado deseado $\alpha \perp\!\!\!\perp \beta \mid V \setminus \{\alpha, \beta\}$. \square

Si se cumple para todos los subconjuntos A, B, C , y D que

$$\text{si } A \perp\!\!\!\perp B \mid (C \cup D) \text{ y } A \perp\!\!\!\perp C \mid (B \cup D) \text{ entonces } A \perp\!\!\!\perp (B \cup C) \mid D \quad (\star)$$

entonces todas las propiedades de Markov son equivalentes. El resultado se enuncia en el siguiente teorema.

Teorema 51 (Pearl-Paz). *Si una distribución de probabilidad sobre X es tal que se cumple (\star) para subconjuntos disyuntos A, B, C, D entonces*

$$(G) \iff (L) \iff (P)$$

Demostración. Solo nos falta demostrar que $(P) \implies (G)$. Supongamos que la propiedad por parejas se cumple y que S separa a A de B en \mathcal{G} , además (\star) se cumple. Sin pérdida de generalidad podemos suponer que A y B son no vacíos.

Demostramos por *inducción inversa* sobre el número de vértices $n = |S|$ en S .

- Si $n = |V| - 2$ entonces A y B consisten en un solo punto cada uno, digamos $\{\alpha\} = A$ y $\{\beta\} = B$ y la condición de independencia se tiene por (P)

$$\alpha \perp\!\!\!\perp \beta \mid S \text{ es exactamente } A \perp\!\!\!\perp B \mid S.$$

- Supongamos que $n < |V| - 2$ y esta separación implica independencia condicional para cualquier conjunto S que separa a A de B que tengan más de n vértices. Primero suponemos que $V = A \cup B \cup S$, lo que implica que al menos A o B tiene más de un elemento, digamos que es A . Si $\alpha \in A$ entonces $S \cup \{\alpha\}$ separa a $A \setminus \{\alpha\}$ de B y también que $S \cup A \setminus \{\alpha\}$ separa a α de B . Entonces por hipótesis de inducción se tiene que

$$A \setminus \{\alpha\} \perp\!\!\!\perp B \mid S \cup \{\alpha\} \text{ y } \alpha \perp\!\!\!\perp B \mid S \cup A \setminus \{\alpha\}.$$

Ahora, por (\star) si $A = B$, $B = A \setminus \{\alpha\}$, $C = \alpha$, y $D = S$ entonces se tiene que

$$A \perp\!\!\!\perp B \mid S$$

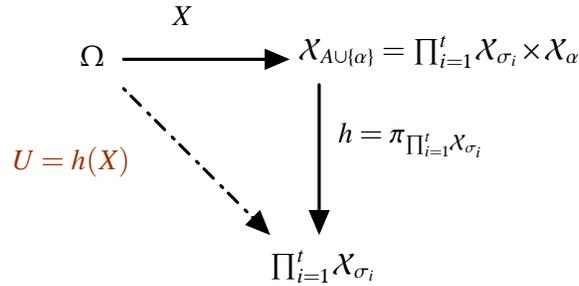
Si $A \cup B \cup S \subset V$, existe al menos un $\alpha \in V \setminus (A \cup B \cup S)$. Entonces $S \cup \{\alpha\}$ separa a A de B , esto es

$$A \perp\!\!\!\perp B \mid S \cup \{\alpha\}.$$

Más aún, $A \cup S$ separa a B de $\{\alpha\}$ o $B \cup S$ separa a A de $\{\alpha\}$. Supongamos que pasa lo primero, es decir que $\alpha \perp\!\!\!\perp B \mid A \cup S$. Entonces utilizando una vez más (\star) si $A = B$, $B = A$, $C = \{\alpha\}$, y $D = S$. se tiene que

$$B \perp\!\!\!\perp A \cup \{\alpha\} \mid S.$$

Sea entonces $A \cup \{\alpha\} = \{\sigma_1, \sigma_2, \dots, \sigma_t, \alpha\}$ y utilizando la condición (C2), definimos funciones $h = \pi_{\prod_{i=1}^t X_{\sigma_i}}$ y $U = h(X)$ dadas por



De lo cual tenemos que $A \perp\!\!\!\perp B \mid S$ y Y así completando la prueba.

□

La propiedad global de Markov es importante ya que proporciona un criterio general para decidir cuando dos grupos de variables A y B son condicionalmente independientes dado un tercer grupo de variables S .

La condición de independencia está íntimamente relacionada con la factorización, así como las propiedades de Markov.

2.1.2 Teorema Hammersley-Clifford

Definición 52. Una medida de probabilidad P sobre X se dice que **factoriza** con respecto a \mathcal{G} si para todos los subconjuntos completos $a \subseteq V$ existen funciones no negativas ψ_a que depende de x a través de x_a únicamente y existe una medida producto $\mu = \otimes_{a \in V} \mu_a$ sobre X , tal que P tiene densidad f con respecto a μ donde f tiene la forma

$$f = \prod_{a \text{ completo}} \psi_a(x).$$

Si P factoriza de acuerdo con \mathcal{G} , decimos que P tiene la propiedad (F) para \mathcal{G} y el conjunto de tales medidas de probabilidad se denota por $M_F(\mathcal{G})$.

Ejemplo 53. Sea \mathcal{G} el siguiente grafo



La medida de probabilidad P sobre el espacio $X = X_{\alpha_1} \times X_{\alpha_2}$ factoriza con respecto al grafo anterior, si y solo si

$$f = \psi_1(x_1)\psi_2(x_2),$$

ya que α_1 y α_2 son los únicos subconjuntos completos de \mathcal{G} .

Nota 54. Notemos que si P se factoriza de acuerdo con \mathcal{G} entonces

- Pueden existir varias medidas producto μ ,
- Las funciones ψ_a no se determinan de forma única,
- Sin pérdida de generalidad se puede suponer que solo los cliques aparecen como los subconjuntos a , es decir que

$$f = \prod_{c \in C} \psi_c(x),$$

donde C es el conjunto de cliques de \mathcal{G} .

Proposición 55. Para cualquier grafo no dirigido \mathcal{G} y cualquier distribución de probabilidad X se cumple que

$$(F) \implies (G) \implies (L) \implies (P)$$

Demostración. Solo debemos demostrar que $(F) \implies (G)$ ya que las demás implicaciones se demostraron en la proposición anterior al Teorema de Pearl & Paz.

Supongamos que (A, B, C) es cualquier triple de subconjuntos disyuntos tales que S separa a A de B . Sea \hat{A} el conjunto de las componentes de conectividad en $\mathcal{G}_{V \setminus S}$ que continen a A y sea $\hat{B} = V \setminus (\hat{A} \cup S)$. Como A y B están separados por S sus elementos están en distintas componentes de conectividad de $\mathcal{G}_{V \setminus S}$ y cualquier clique de \mathcal{G} es subconjunto de $\hat{A} \cup S$ o de $\hat{B} \cup S$. Si C_A denota los cliques contenidos en $\hat{A} \cup S$, obtenemos que

$$f = \prod_{c \in C} \psi_c(x) = \prod_{c \in C_A} \psi_c(x) \prod_{c \in C \setminus C_A} C_A \psi_c(x) = h(x_{\hat{A} \cup S}) k(x_{\hat{B} \cup S}).$$

Luego, por la propiedad

$$X \perp\!\!\!\perp Y \mid Z \iff f(x, y, z) = h(x, z) k(y, z) \text{ para algún } h, k$$

tenemos que $\hat{A} \perp\!\!\!\perp \hat{B} \mid S$. Finalmente, por propiedad (C2) donde $A = \{\gamma_1, \dots, \gamma_r\}$ y $\hat{A} = \{\gamma_1, \dots, \gamma_r, \sigma_1, \dots, \sigma_s\}$ y definimos las funciones $h_1 = \pi_{\prod_{i=1}^r \gamma_i}$ y $U_1 = h_1(X)$ dadas por

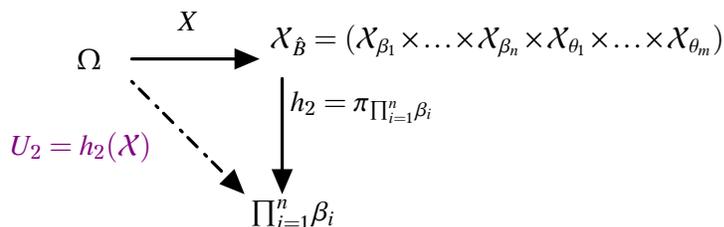
$$\begin{array}{ccc} \Omega & \xrightarrow{X} & \mathcal{X}_{\hat{A}} = (\mathcal{X}_{\gamma_1} \times \dots \times \mathcal{X}_{\gamma_r} \times \mathcal{X}_{\sigma_1} \times \dots \times \mathcal{X}_{\sigma_s}) \\ & \searrow^{U_1 = h_1(X)} & \downarrow h_1 = \pi_{\prod_{i=1}^r \gamma_i} \\ & & \prod_{i=1}^r \gamma_i \end{array}$$

De lo cual obtenemos

$$A \perp\!\!\!\perp \hat{B} \mid S.$$

Aplicando una vez más la propiedad (C2) con $B = \{\beta_1, \dots, \beta_n\}$ y $\hat{B} = \{\beta_1, \dots, \beta_n, \theta_1, \dots, \theta_m\}$ y definiendo funciones $h_2 = \pi_{\prod_{i=1}^n \beta_i}$ y $U = h_2(X)$ dadas por

ya así obtenemos $A \perp\!\!\!\perp B \mid S$ que es la propiedad independencia deseada. \square



En el caso en que P tenga densidad positiva y continua podemos utilizar el Lema de inversión de Möbius para demostrar que (P) implica (F) y de esta manera todas las propiedades de Markov son equivalentes.

Lema 56 (Inversión de Möbius.). Sean Ψ y Φ funciones definidas sobre el conjunto de todos los subconjuntos de un conjunto finito V , que toman valores en un grupo abeliano. Las siguientes dos condiciones son equivalentes:

1. Para todo $a \subseteq V$: $\Psi(a) = \sum_{b: b \subseteq a} \Phi(b)$;
2. Para todo $a \subseteq V$: $\Phi(a) = \sum_{b: b \subseteq a} (-1)^{|a \setminus b|} \Psi(b)$.

Demostración. Veamos primero que $(2) \Rightarrow (1)$. Supongamos que (2) se cumple para dos funciones Ψ y Φ que toman valores en un grupo abeliano. Como vale para todo $a \subseteq V$, sea $b \subseteq V$ tal que:

$$\begin{aligned}
 \sum_{b: b \subseteq a} \Phi(b) &= \sum_{b: b \subseteq a} \sum_{c: c \subseteq b} (-1)^{|b \setminus c|} \Psi(c) \\
 &= \sum_{c: c \subseteq b} \Psi(c) \left\{ \sum_{b: c \subseteq b \subseteq a} (-1)^{|b \setminus c|} \right\} \\
 &= \sum_{c: c \subseteq b} \Psi(c) \left\{ \sum_{h: h \subseteq a \setminus c} (-1)^{|h|} \right\}.
 \end{aligned}$$

La última suma es cero excepto cuando $a \setminus c = \emptyset$, i.e si $c = a$, ya que todo conjunto finito no vacío tiene la misma cantidad de subconjuntos con cardinalidad par que impar. La demostración de $(1) \Rightarrow (2)$ se hace de manera análoga. \square

Teorema 57 (Hammersley-Clifford.). Una distribución de probabilidad P con densidad f continua y positiva con respecto a la medida producto μ , satisface la propiedad por parejas de Markov con respecto a un grafo no dirigido \mathcal{G} si y solo si se factoriza de acuerdo a \mathcal{G} .

Demostración. Por proposición anterior sabemos que si P factoriza de acuerdo a \mathcal{G} entonces cumple la propiedad por parejas de Markov. Únicamente debemos ver que $(P) \Rightarrow (F)$.

Como la densidad es positiva podemos tomar logaritmos a ambos lados de la definición de factorización

$$f = \prod_{a \text{ completo}} \psi_a(x)$$

Entonces, si $\phi_a(x) = \log \psi_a(x)$ podemos escribir la ecuación como

$$\log f(x) = \sum_{a: a \subseteq V} \phi_a(x)$$

donde, $\phi_a(x) \equiv 0$ excepto cuando a es subconjunto completo de V . Así que para la prueba debemos escribir la propiedad (F) de esa manera.

Supongamos que P cumple la propiedad por parejas de Markov y elegimos un elemento fijo arbitrario $x^* \in \mathcal{X}$. Para todo $a \subseteq V$ definimos la función

$$H_a(x) = \log f(x_a, x_{a^c}^*),$$

donde el elemento $y = (x_a, x_{a^c}^*)$ con $y_\gamma = x_\gamma$ para $\gamma \in a$ y $y_\gamma = x_\gamma^*$ para $\gamma \notin a$. Note que $H_V(x) = \log f(x)$. Como x^* es fijo, $H_a(x)$ depende únicamente de x a través de los valores de x_a más no de los de x_{a^c} . Ahora definimos, para todo $a \subseteq V$ la función

$$\phi_a(x) = \sum_{b:b \subseteq a} (-1)^{|a \setminus b|} H_b(x).$$

De esto, es claro que $\phi_a(x)$ también depende de x únicamente por $H_b(x)$ y como b se elige como subconjunto de a , $\phi_a(x)$ depende x únicamente a través de x_a . Ahora, ya tenemos dos funciones definidas sobre todos los subconjuntos de V , así que podemos aplicar el Lema de Möbius y obtenemos

$$\log f(x) = H_V(x) = \sum_{a:a \subseteq V} \phi_a(x).$$

Solo falta ver que $\phi_a \equiv 0$ cuando $a \subseteq V$ no es completo. Para esto, supongamos que a no es completo, entonces existen $\alpha, \beta \in a$ tal que $\alpha \neq \beta$. Sea $c = a \setminus \{\alpha, \beta\}$. Para simplificar la notación escribimos H_a por $H_c(x)$. Ahora expandimos $\phi_a(x)$ y obtenemos

$$\begin{aligned} \phi_a(x) &= \sum_{b:b \subseteq a} (-1)^{|a \setminus b|} H_b(x) \\ &= \sum_{b:b \subseteq (c \cup \{\alpha, \beta\})} (-1)^{|a \setminus b|} H_b(x) \\ &= \sum_{\substack{b \subseteq a \\ b \subseteq c}} (-1)^{|a \setminus b|} H_b + \sum_{\substack{b \cup \{\alpha\} \subseteq a \\ b \subseteq c}} (-1)^{|a \setminus b \cup \{\alpha\}|} H_{b \cup \{\alpha\}} + \sum_{\substack{b \cup \{\beta\} \subseteq a \\ b \subseteq c}} (-1)^{|a \setminus b \cup \{\beta\}|} H_{b \cup \{\beta\}} \\ &\quad + \sum_{\substack{b \cup \{\alpha, \beta\} \subseteq a \\ b \subseteq c}} (-1)^{|a \setminus b \cup \{\alpha, \beta\}|} H_{b \cup \{\alpha, \beta\}} \\ &= \sum_{b:b \subseteq c} (-1)^{|a \setminus b|} \{H_b - H_{b \cup \{\alpha\}} - H_{b \cup \{\beta\}} + H_{b \cup \{\alpha, \beta\}}\} \\ &= \sum_{b:b \subseteq c} (-1)^{|c \setminus b|} \{H_b - H_{b \cup \{\alpha\}} - H_{b \cup \{\beta\}} + H_{b \cup \{\alpha, \beta\}}\} \end{aligned}$$

Para terminar, debemos demostrar que

$$H_b - H_{b \cup \{\alpha\}} - H_{b \cup \{\beta\}} + H_{b \cup \{\alpha, \beta\}} = 0.$$

Para facilitar la notación, sea $d = V \setminus \{\alpha, \beta\}$. Entonces, por la propiedad por parejas de Markov sabemos que $\alpha \perp\!\!\!\perp \beta \mid V \setminus \{\alpha, \beta\}$ y además por proposición

$$X \perp\!\!\!\perp Y \mid Z \iff f(x, y, z) = f(x \mid z) f(y, z)$$

tenemos que

$$\begin{aligned}
H_{b \cup \{\alpha, \beta\}}(x) - H_{b \cup \{\alpha\}} &= \log \frac{f(x_b, x_\alpha, x_\beta, x_{d \setminus b}^*)}{f(x_b, x_\alpha, x_\beta^*, x_{d \setminus b}^*)} \\
&= \log \frac{f(x_\alpha | x_b, x_{d \setminus b}^*) f(x_\beta, x_b, x_{d \setminus b}^*)}{f(x_\alpha | x_b, x_{d \setminus b}^*) f(x_\beta^*, x_b, x_{d \setminus b}^*)} \\
&= \log \frac{f(x_\alpha^* | x_b, x_{d \setminus b}^*) f(x_\beta, x_b, x_{d \setminus b}^*)}{f(x_\alpha^* | x_b, x_{d \setminus b}^*) f(x_\beta^*, x_b, x_{d \setminus b}^*)} \\
&= \log \frac{f(x_b, x_\beta, x_\alpha^*, x_{d \setminus b}^*)}{f(x_b, x_\alpha^*, x_\beta^*, x_{d \setminus b}^*)} \\
&= H_{b \cup \{\beta\}}(x) - H_b(x).
\end{aligned}$$

En conclusión

$$H_b(x) - H_{b \cup \{\alpha\}} - H_{b \cup \{\beta\}}(x) + H_{b \cup \{\alpha, \beta\}}(x) = 0$$

Esto es cuando a no es completo, por lo tanto la suma completa es igual a cero lo cual completa la demostración.

□

INTERPRETACIÓN ALGEBRAICA DEL TEOREMA DE HAMMERSLY-CLIFFORD

3.1 MODELOS EXPONENCIALES, LOG-LINEALES Y GRÁFICOS

En este capítulo consideraremos espacios de probabilidad finitos, donde la σ -álgebra correspondiente es la colección de todos sus subconjuntos.

Definición 58. Dado un conjunto Θ , un modelo estadístico parametrizado por Θ es una colección $(\mathcal{X}, \mathcal{F}, \{P_\theta : \theta \in \Theta\})$ donde $(\mathcal{X}, \mathcal{F})$ es un espacio medible y $P_\theta : \mathcal{F} \rightarrow \mathbb{R}$ es una distribución de probabilidad sobre \mathcal{X} , para todo $\theta \in \Theta$.

Nota 59. En un modelo estadístico, las medidas de probabilidad P_θ están determinadas por las funciones

$$\begin{aligned} p_\theta : \mathcal{X} &\rightarrow \mathbb{R} \\ x &\mapsto p_\theta(x) = P_\theta(\{x\}). \end{aligned}$$

Entonces, especificar el modelo estadístico $(\mathcal{X}, \mathcal{F}, \{P_\theta : \theta \in \Theta\})$ es equivalente a especificar una familia $(\mathcal{X}, \mathcal{F}, \{p_\theta : \theta \in \Theta\})$ donde p_θ son las funciones de densidad de P_θ , para todo $\theta \in \Theta$. Cuando $\mathcal{X} = \{x_1, \dots, x_m\}$ y $\Theta = \{\theta_1, \dots, \theta_d\}$, tenemos d -medidas de probabilidad $p_{\theta_1}, \dots, p_{\theta_d}$, donde cada p_i está determinada por su valor en \mathcal{X}_i

$$\begin{aligned} p_1 : \mathcal{X} \rightarrow \mathbb{R} &\quad \therefore (p_1(x_1), \dots, p_1(x_m)) \in \mathbb{R}_{\geq 0}^m, \\ &\quad \vdots \\ p_d : \mathcal{X} \rightarrow \mathbb{R} &\quad \therefore (p_d(x_1), \dots, p_d(x_m)) \in \mathbb{R}_{\geq 0}^m, \end{aligned}$$

por lo tanto, tener d -funciones de densidad p_1, \dots, p_d es equivalente a tener una matriz

$$A = \begin{bmatrix} p_1(x_1) & p_1(x_2) & \dots & p_1(x_m) \\ p_2(x_1) & p_2(x_2) & \dots & p_2(x_m) \\ \vdots & \vdots & & \vdots \\ p_d(x_1) & p_d(x_2) & \dots & p_d(x_m) \end{bmatrix}.$$

A continuación, siguiendo esta idea, mostraremos como se pueden representar algunas familias de modelos estadísticos discretos con una notación matricial conveniente, $(\mathcal{X}, \mathcal{P}(\mathcal{X}), A, \Theta)$.

3.1.1 Modelo algebraico

La clase de modelos que vamos a considerar en esta sección son llamados **modelos algebraicos**. Nuestro objetivo en este capítulo es mostrar que algunos modelos conocidos, como la familia de *modelos exponenciales* y de *modelos log-lineales*, están incluidos en la familia de los modelos algebraicos.

Definición 60. Un modelo estadístico algebraico consiste en una par (\mathcal{X}, A) , donde \mathcal{X} es un conjunto finito y $A = (a_{ij}) \in \mathbb{M}_{d \times m}(\mathbb{Z}_{\geq 0})$ es una matriz de enteros no negativos tal que

$$\sum_{i=1}^d a_{i1} = \sum_{i=1}^d a_{i2} = \cdots = \sum_{i=1}^d a_{im}.$$

Definición 61. Sea (\mathcal{X}, A) un modelo estadístico algebraico. Consideremos la aplicación ϕ_A que toma d -vectores reales no negativos y los envía a m -vectores reales no negativos definida como

$$\phi_A : \mathbb{R}_{\geq 0}^d \rightarrow \mathbb{R}_{\geq 0}^m, \quad (t_1, \dots, t_d) \mapsto \left(\prod_i t_i^{a_{i1}}, \prod_i t_i^{a_{i2}}, \dots, \prod_i t_i^{a_{im}} \right).$$

Decimos que una distribución p pertenece al modelo (\mathcal{X}, A) si y solo si p esta en la imagen de la aplicación ϕ_A .

Observación 62. Cuando p pertenece al modelo estadístico algebraico (\mathcal{X}, A) también diremos que p factoriza de acuerdo con A .

3.1.2 Modelo exponencial

Definición 63. Consideremos $\Theta = [-\infty, \infty)^d \subset \bar{\mathbb{R}}^d$, el conjunto de reales extendidos, un **modelo exponencial**, parametrizado por Θ , es una colección $(\mathcal{X}, T, \{p_\theta : \theta \in \Theta\})$ donde \mathcal{X} es un conjunto finito y p_θ es la aplicación definida sobre \mathcal{X} , definida por

$$p_\theta(x) = Z(\theta) e^{\langle \theta, T(x) \rangle},$$

para todo $\theta \in \Theta$. Donde $x \in \mathcal{X}$, $Z(\theta) : \Theta \rightarrow \mathbb{R}$ es la constante de normalización y $T : \mathcal{X} \rightarrow \mathbb{Z}^d \setminus \{0\}$ es llamado un estadístico.

Proposición 64. La clase de los modelos exponenciales está incluida en la clase de los modelos algebraicos.

Demostración. Supongamos que tenemos un modelo exponencial determinado por la familia exponencial

$$p_\theta(x) = Z(\theta) e^{\langle \theta, T(x) \rangle}.$$

En primer lugar, notemos que $T : \mathcal{X} \rightarrow \mathbb{Z}_{\geq 0}^d \setminus \{0\}$ no depende de $\theta \in \Theta$. Ahora, como \mathcal{X} es finito podemos evaluar los elementos x_1, \dots, x_m en T :

$$T(x_1) = \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{d1} \end{bmatrix}, T(x_2) = \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{d2} \end{bmatrix}, \dots, T(x_m) = \begin{bmatrix} a_{1m} \\ a_{2m} \\ \vdots \\ a_{dm} \end{bmatrix}.$$

donde $\sum_{i=1}^d a_{ij} = k$ para $j = 1, \dots, m$. Consideremos la matriz A como

$$A = \begin{bmatrix} T(x_1) & T(x_2) & \cdots & T(x_m) \end{bmatrix} = \begin{bmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & & \vdots \\ a_{d1} & \cdots & a_{dm} \end{bmatrix} \in \mathbb{M}_{d \times m}(\mathbb{Z}_{\geq 0}).$$

Por otra parte, como $Z(\theta) \neq 0 \forall \theta \in \Theta$ y $\theta = \{\theta_1, \dots, \theta_d\} \in \bar{\mathbb{R}}^d$, para cada $x_j \in \mathcal{X}$ tenemos

$$\begin{aligned} p_\theta(x_j) &= Z(\theta) e^{\langle \theta, T(x_j) \rangle} \\ \frac{p_\theta(x_j)}{Z(\theta)} &= e^{\sum_{i=1}^d \theta_i a_{ij}} \\ \frac{p_\theta(x_j)}{Z(\theta)} &= \prod_{i=1}^d e^{\theta_i a_{ij}} \\ \frac{p_\theta(x_j)}{Z(\theta)} &= (e^{\theta_1})^{a_{1j}} (e^{\theta_2})^{a_{2j}} \dots (e^{\theta_d})^{a_{dj}} \\ \frac{p_\theta(x_j)}{Z(\theta)} &= t_1^{a_{1j}} \dots t_d^{a_{dj}}, \end{aligned}$$

donde $(e^{\theta_i}) = t_i$. Sea $\tilde{p}_\theta(x_j) \equiv \frac{p_\theta(x_j)}{Z(\theta)}$, por lo tanto $\tilde{p}_\theta(x_j) \in \phi_A(\mathbb{R}_{\geq 0}^d)$.

□

3.1.3 Modelos log-lineales

Definición 65. Un modelo log-lineal está definido por el siguiente conjunto de datos:

- $\mathcal{X} = \prod_{i=1}^n I_{X_i}$, donde $I_{X_j} = \text{Im}(\mathcal{X}_j)$ es el conjunto de todos los estados posibles para \mathcal{X}_j , los cuales llamaremos niveles.
- $\mathbb{X} = \{X_1, X_2, \dots, X_r\}$, es la familia de variables aleatorias que llamaremos factores.
- $\mathbb{G} = \{\mathbb{G}_1, \dots, \mathbb{G}_m\} \subseteq \mathcal{P}(\mathbb{X})$ es decir, cada \mathbb{G}_i es un subconjunto de \mathbb{X} , i.e $\mathbb{G}_i = \{X_{i1}, \dots, X_{im_i}\} \subseteq \mathbb{X}$. Esta colección será llamada la familia de generadores del modelo.

Definimos el **modelo log-lineal** con conjunto generador \mathbb{G} como

$$p(x) \propto \prod_{i=1}^m \psi_{\mathbb{G}_i}(x),$$

donde $x \in \mathcal{X}$ y $\psi_{\mathbb{G}_i}(x)$ depende únicamente de $I_{X_{i1}}, \dots, I_{X_{im_i}}$.

Proposición 66. La clase de los modelos log-lineales está incluida en la clase de los modelos algebraicos.

Demostración. Supongamos que tenemos un modelo log-lineal.

$$p(x) \propto \prod_{\mathbb{G}_i \in \mathbb{G}} \psi_{\mathbb{G}_i}(x).$$

Ahora consideremos una matriz A tal que:

- Las columnas de A estarán indizadas por los elementos de $\mathcal{X} = I_{X_1} \times I_{X_2} \times \dots \times I_{X_n}$,
- Las filas estarán indizadas por los pares $(\psi_{\mathbb{G}_i}, \bar{x})$, donde $\bar{x} \in \mathcal{X}_{\mathbb{G}_i}$,

- Todas la entradas de la matriz son 1 o 0, dependiendo de la siguiente regla. La entrada correspondiente a la fila (ψ_{G_i}, \bar{x}) y a la columna \bar{y} es 1 si y solo si $\pi_{G_i}(\bar{y}) = \bar{x}$, en otro caso la entrada es cero.

Se puede demostrar que la distribución P que define el model log-lineal con generador G pertenece a la imagen de la aplicación ϕ_A . □

A continuación ilustraremos el método de la demostración mediante un ejemplo concreto.

Ejemplo 67 (Sin tercera opción). Sea $X = \{X_1, X_2, X_3\}$, un conjunto de variables binarias y sea $G = \{G_1, G_2, G_3\}$ la familia de generadores, donde $G_1 = \{X_1, X_2\}$, $G_2 = \{X_2, X_3\}$ y $G_3 = \{X_1, X_3\}$. En este caso, por ser variables aleatorias binarias tenemos que $I_{X_1} = I_{X_2} = I_{X_3} = \{0, 1\}$ luego

$$X = \prod_{i=1}^3 I_{X_i} = \{000, 001, 010, 011, 100, 110, 101, 111\}.$$

Es decir que la matriz consiste en 8 columnas. Ahora, encontremos la cantidad de filas. Recordemos que estarán indizadas por los pares (ψ_{G_i}, \bar{x}) , donde

$$\begin{aligned} X_{G_1} &= I_{X_1} \times I_{X_2} = \{00, 01, 10, 11\}, \\ X_{G_2} &= I_{X_2} \times I_{X_3} = \{00, 01, 10, 11\}, \\ X_{G_3} &= I_{X_1} \times I_{X_3} = \{00, 01, 10, 11\}. \end{aligned}$$

Es decir que la matriz tiene 12 filas. Veamos los valores de las entradas en la siguiente tabla

	000	001	010	011	100	101	110	111
$t_1(\psi_{G_1}, 00)$	1	1	0	0	0	0	0	0
$t_2(\psi_{G_1}, 01)$	0	0	1	1	0	0	0	0
$t_3(\psi_{G_1}, 10)$	0	0	0	0	1	1	0	0
$t_4(\psi_{G_1}, 11)$	0	0	0	0	0	0	1	1
$t_5(\psi_{G_2}, 00)$	1	0	0	0	1	0	0	0
$t_6(\psi_{G_2}, 01)$	0	1	0	0	0	1	0	0
$t_7(\psi_{G_2}, 10)$	0	0	1	0	0	0	1	0
$t_8(\psi_{G_2}, 11)$	0	0	0	1	0	0	0	1
$t_9(\psi_{G_3}, 00)$	1	0	1	0	0	0	0	0
$t_{10}(\psi_{G_3}, 01)$	0	1	0	1	0	0	0	0
$t_{11}(\psi_{G_3}, 10)$	0	0	0	0	1	0	1	0
$t_{12}(\psi_{G_3}, 11)$	0	0	0	0	0	1	0	1

Por lo tanto la matriz $A \in \mathbb{M}_{12 \times 8}(\mathbb{Z}_{\geq 0})$ es

$$A = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \end{bmatrix}.$$

Entonces, la distribución de probabilidad $p = (p_{000}, p_{001}, p_{010}, p_{011}, p_{100}, p_{101}, p_{110}, p_{111})$ pertenece al modelo sin tercera opción si está en la imagen de la aplicación $\phi_A : \mathbb{R}_{\geq 0}^{12} \rightarrow \mathbb{R}_{\geq 0}^8$, asociada a la matriz A , la cual esta definida por

$$(t_1, \dots, t_{12}) \mapsto (t_1 t_5 t_9, t_1 t_6 t_{10}, t_2 t_7 t_9, t_2 t_8 t_{10}, t_3 t_5 t_{11}, t_3 t_6 t_{12}, t_4 t_7 t_{11}, t_4 t_8 t_{12}).$$

3.2 ANÁLISIS ALGEBRAICO DE LOS MODELOS GRÁFICOS

Existen ciertos modelos log-lineales importantes en nuestro objetivo de algebrizar el teorema de Hammersley-Clifford, estos son los modelos gráficos no dirigidos. Este modelo se especifica por un grafo no dirigido \mathcal{G} , tal que su conjunto de vértices es \mathbb{X} y su conjunto de lados es \mathbb{E} , i.e. $\mathcal{G} = (\mathbb{X}, \mathbb{E})$.

Definición 68. El **modelo gráfico no dirigido** del grafo \mathcal{G} es el modelo log-lineal cuyo conjunto generador son los cliques del grafo no dirigido.

La matriz A asociada con el modelo algebraico, la cual es función del grafo \mathcal{G} , la denotaremos por $A(\mathcal{G})$.

Definición 69. Un modelo gráfico no dirigido se dice que **es un modelo gráfico que admite descomposición** si y solo si el grafo \mathcal{G} tiene una cuerda.

Para escribir el teorema de Hammersley-Clifford en términos geométrico-algebraicos, necesitamos la noción de variedad algebraica afín.

Definición 70. Sea k un campo y sean f_1, \dots, f_s polinomios en $k[x_1, \dots, x_n]$. Llamamos **variedad afín** definida por f_1, \dots, f_s al conjunto

$$\mathbb{V}(f_1, \dots, f_s) = \left\{ (a_1, \dots, a_n) \in k^n : f_i(a_1, \dots, a_n) = 0, \quad \forall 1 \leq i \leq s \right\}.$$

Definición 71. Definimos el **espacio proyectivo complejo n -dimensional**, denotado por $\mathbb{P}^n(\mathbb{C})$, como el conjunto de las clases de equivalencia en $\mathbb{C}^{n+1} \setminus \{0\}$, determinadas por la relación de equivalencia \sim definida por

$$(x_0, \dots, x_n) \sim (y_0, \dots, y_n)$$

si y solo si $\exists \alpha \in \mathbb{C}^*$ tal que $(y_0, \dots, y_n) = \alpha(x_0, \dots, x_n)$.

Definición 72. Sean f_1, \dots, f_s polinomios homogéneos en $\mathbb{C}[x_0, \dots, x_n]$, el conjunto

$$\mathbb{V}_P(f_1, \dots, f_s) = \left\{ [x] \in \mathbb{P}^n(\mathbb{C}) \mid f_i(x) = 0, \forall i = 1, \dots, s \right\},$$

es llamado la **variedad proyectiva** determinada por los polinomios f_1, \dots, f_s .

3.2.1 Polinomios cuadráticos que representan independencia condicional.

Un conjunto de distribuciones de probabilidad que satisfacen una condición de independencia se puede representar como un punto en una cierta variedad algebraica. Estas variedades estarán definidas a partir de ideales que consisten de ciertos polinomios cuadráticos llamados *diferencia de productos cruzados* (CPDs).

Definición 73. Dadas tres variables aleatorias discretas X, Y, Z , definimos la **diferencia de productos cruzados** como

$$\text{cpd}(X = \{x, \tilde{x}\}, Y = \{y, \tilde{y}\} \mid Z = z) \equiv p(x, y, z)p(\tilde{x}, \tilde{y}, z) - p(\tilde{x}, y, z)p(x, \tilde{y}, z),$$

y definimos el **cociente de productos cruzados** por

$$\text{cpr}(X = \{x, \tilde{x}\}, Y = \{y, \tilde{y}\} \mid Z = z) \equiv \frac{p(x, y, z)p(\tilde{x}, \tilde{y}, z)}{p(\tilde{x}, y, z)p(x, \tilde{y}, z)},$$

donde p es la función de densidad conjunta para X, Y, Z , y x, \tilde{x} son niveles de X , y, \tilde{y} son niveles de Y y z es nivel de Z .

Observación 74. Sea $\mathbb{X} = \{X_1, \dots, X_n\}$ una familia de variables aleatorias, definidas sobre un conjunto finito \mathcal{X} , y sean I_{X_1}, \dots, I_{X_n} los niveles de las variables X_1, \dots, X_n , respectivamente.

Supongamos que $\mathcal{X} = I_{X_1} \times \dots \times I_{X_n}$ y consideremos $\mathbb{R}[\mathcal{X}]$ el anillo de polinomios en las variables

$$\{p_{a_1 a_2 \dots a_n} \mid (a_1, \dots, a_n) \in \mathcal{X}\}.$$

Por ejemplo, si X_1, X_2 son variables aleatorias binarias entonces $\mathcal{X} = \{00, 01, 10, 11\}$, y por lo tanto tenemos

$$\mathbb{R}[\mathcal{X}] = \mathbb{R}[p_{00}, p_{01}, p_{10}, p_{11}].$$

Definición 75. Dado un conjunto de variables aleatorias \mathbb{X} , una afirmación de independencia es una **afirmación del estilo**

$$X \text{ es independiente de } Y \text{ dado } Z \quad \diamond$$

donde $X, Y, Z \subseteq \mathbb{X}$ son disjuntos de \mathbb{X} .

Nota 76. La condición de independencia de la definición 75 se puede escribir de la siguiente forma. Supongamos que nuestro conjunto de variables aleatorias \mathbb{W} está dado por $\mathbb{W} = \{W_1, W_2, \dots, W_n\}$, donde cada W_i es una variable aleatoria y

$$\left. \begin{array}{l} X = \{X_1, \dots, X_l\} \\ Y = \{Y_1, \dots, Y_k\} \\ Z = \{Z_1, \dots, Z_t\} \end{array} \right\} \subseteq \mathbb{W},$$

son los subconjuntos de variables aleatorias indicados en la definición de afirmación de independencia. Si P es la distribución condicional de X, Y dado Z , denotada por

$$P((X_1, \dots, X_n) = (x_1, \dots, x_n), (Y_1, \dots, Y_k) = (y_1, \dots, y_k) \mid (Z_1, \dots, Z_t) = (z_1, \dots, z_t)),$$

entonces, la afirmación de independencia indica que se puede escribir como el producto de las distribuciones condicionales

$$P((X_1, \dots, X_n) = (x_1, \dots, x_n) \mid (Z_1, \dots, Z_t) = (z_1, \dots, z_t))$$

con

$$P((Y_1, \dots, Y_k) = (y_1, \dots, y_k) \mid (Z_1, \dots, Z_t) = (z_1, \dots, z_t)).$$

Es claro que con cada afirmación de independencia podemos asociar una diferencia de productos cruzados; lo interesante es que estos polinomios, son polinomios cuadráticos en $\mathbb{R}[X]$.

Proposición 77. Si X, Y y Z variables aleatorias definidas sobre un conjunto finito X entonces

$$X \perp\!\!\!\perp Y \mid Z \iff \text{cpd}(X = \{x, \tilde{x}\}, Y = \{y, \tilde{y}\} \mid Z = z) = 0.$$

Demostración. (\Rightarrow) Supongamos que X es independiente de Y dado Z . Sabemos que

$$p(X = x, Y = y \mid Z = z) = \frac{p(X = x, Y = y, Z = z)}{p(Z = z)},$$

lo cual, con la notación abreviada para funciones de densidad, se puede escribir como

$$p(x, y \mid z) = \frac{p(x, y, z)}{p(z)}. \quad (5)$$

De otro lado, dado que X es independiente de Y dado Z se sigue que

$$p(X = x, Y = y \mid Z = z) = p(X = x \mid Z = z)p(Y = y \mid Z = z);$$

lo cual, utilizando de nuevo la notación abreviada se escribe como

$$p(x, y \mid z) = p(x \mid z)p(y \mid z). \quad (6)$$

Igualando las expresiones obtenidas en (5) y (6) resulta

$$p(x, y, z) = \frac{p(x, z)p(y, z)}{p(z)}.$$

Ahora, la diferencia de productos cruzados

$$\begin{aligned} \text{cpd}(X = \{x, \tilde{x}\}, Y = \{y, \tilde{y}\} \mid Z = z) &= p(x, y, z)p(\tilde{x}, \tilde{y}, z) - p(\tilde{x}, y, z)p(x, \tilde{y}, z) \\ &= \frac{p(x, z)p(y, z)}{p(z)} \cdot \frac{p(\tilde{x}, z)p(\tilde{y}, z)}{p(z)} - \frac{p(x, z)p(\tilde{y}, z)}{p(z)} \cdot \frac{p(\tilde{x}, z)p(y, z)}{p(z)} \\ &= 0. \end{aligned}$$

(\Leftarrow) Supongamos ahora que

$$\text{cpd}(X = \{x, \tilde{x}\}, Y = \{y, \tilde{y}\} \mid Z = z) = 0,$$

veamos que $X \perp\!\!\!\perp Y | Z$.

Por hipótesis tenemos que

$$p(x, y, z)p(\tilde{x}, \tilde{y}, z) - p(\tilde{x}, y, z)p(x, \tilde{y}, z) = 0,$$

es decir,

$$p(x, y, z)p(\tilde{x}, \tilde{y}, z) = p(\tilde{x}, y, z)p(x, \tilde{y}, z).$$

Si *marginalizamos* con respecto a \tilde{x} obtenemos

$$\sum_{\tilde{x}} p(x, y, z)p(\tilde{x}, \tilde{y}, z) = \sum_{\tilde{x}} p(\tilde{x}, y, z)p(x, \tilde{y}, z).$$

Teniendo en cuenta que al realizar la suma indicada, sobre todos los valores posibles de X , de la función de densidad de la distribución conjunta, obtenemos la función de densidad de la distribución conjunta de las demás variables, obtenemos

$$p(x, y, z)p(\tilde{y}, z) = p(y, z)p(x, \tilde{y}, z).$$

Repitiendo este proceso con Y , obtenemos

$$\sum_{\tilde{y}} p(x, y, z)p(\tilde{y}, z) = \sum_{\tilde{y}} p(y, z)p(x, \tilde{y}, z),$$

es decir,

$$p(x, y, z)p(z) = p(y, z)p(x, z),$$

de donde

$$p(x, y, z) = \frac{p(y, z)p(x, z)}{p(z)},$$

lo que significa que $X \perp\!\!\!\perp Y | Z$. □

Observación 78. *El procedimiento anterior se puede generalizar para cualquier condición de independencia, en un número arbitrario de variables aleatorias.*

Teorema 79. *Dada una familia de variables aleatorias $\mathbb{W} = \{W_1, W_2, \dots, W_n\}$ y subcolecciones*

$$\left. \begin{array}{l} X = \{X_1, \dots, X_l\} \\ Y = \{Y_1, \dots, Y_k\} \\ Z = \{Z_1, \dots, Z_t\} \end{array} \right\} \subseteq \mathbb{W},$$

la afirmación de independencia

X es independiente de Y dado Z,

es equivalente a la ecuación

$$\text{cpd}(X = \{x, \tilde{x}\}, Y = \{y, \tilde{y}\} | Z = z) = 0,$$

donde X, Y y Z , en la última ecuación, denotan las distribuciones conjuntas de las familias X, Y y Z .

Demostración. Basta repetir el razonamiento de la prueba del teorema 77, teniendo en cuenta que, si marginalizamos la función de densidad de la distribución conjunta $p(w)$, con respecto a las variables que no están en la familia X ,

$$p(x) = \sum_{w-X} p(w)$$

obtenemos las funciones de densidad de la distribución conjunta de las variables aleatorias que están en X . \square

Definición 80. La condición de independencia \diamond se llama **saturada** si $X \cup Y \cup Z = \{X_1, \dots, X_n\}$.

Nota 81. El polinomio cpd para una condición de independencia saturada es un binomio cuadrático.

3.2.2 Modelos gráficos no dirigidos y el teorema de Hammersley-Clifford

En esta sección vamos a describir conjuntos de condiciones de independencia derivados de las condiciones de separación en grafos no dirigidos y su conexión con la factorización de distribuciones. Para finalizar nuestro objetivo, vamos a describir el teorema de Hammersley-Clifford en lenguaje de ideales y variedades.

Definición 82. Sea \mathcal{G} un modelo gráfico no dirigido con variables $\{X_1, \dots, X_n\}$. Definimos el **ideal por parejas** de Markov, denotado por $I_{(p)\mathcal{G}}$, como el ideal en $\mathbb{R}[X]$ generado por los binomios cuadráticos correspondientes a todas las condiciones de independencia saturadas

$$X_i \text{ es independiente de } X_j \text{ dado } \{X_1, \dots, X_n\} \setminus \{X_i, X_j\},$$

para todo par no adjacente (X_i, X_j) del grafo \mathcal{G} .

Observación 83. Note que los polinomios cpd asociados con las condiciones por parejas de Markov

$$X_i \perp\!\!\!\perp X_j \mid \{X_1, \dots, X_n\} \setminus \{X_i, X_j\},$$

las cuales son saturadas, corresponden a los binomios cuadráticos

$$p(x_i, y_j, z)p(\tilde{x}_i, \tilde{y}_j, z) - p(\tilde{x}_i, y_j, z)p(x_i, \tilde{y}_j, z) = 0;$$

donde $Z = \{X_1, \dots, X_n\} \setminus \{X_i, X_j\}$ y $z = (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_{j-1}, x_{j+1}, \dots, x_n)$.

Definición 84. Definimos la **variedad por parejas de Markov**, como la variedad definida por el ideal $I_{(p)\mathcal{G}}$, es decir, como el conjunto

$$X_{(p)\mathcal{G}}^k = \{(a_1, \dots, a_n) \in \mathbb{R}_{\geq 0}^n \mid f = 0 \text{ para todo } f \in I_{(p)\mathcal{G}}\},$$

donde $k = \mathbb{R}_{\geq 0}$.

Ejemplo 85. Consideremos el grafo \mathcal{G} del ejemplo 53, donde X_1 y X_2 son variables aleatorias binarias. Entonces, tenemos dos generadores $G_1 = \{X_1\}$ y $G_2 = \{X_2\}$, de manera que las distribuciones de probabilidad para el modelo gráfico tienen la forma

$$P(x_1, x_2) \propto \psi_{(1)}(x_1)\psi_{(2)}(x_2).$$

Por ser variables binarias, tenemos el espacio $\mathcal{X} = \{00, 01, 10, 11\}$. Hallemos los polinomios del ideal $I_{(p)\mathcal{G}}$ los cuales cumplen la condición de independencia

$$X_i \perp\!\!\!\perp X_j \mid \{X_1, \dots, X_n\} \setminus \{X_i, X_j\} \iff \text{cpd}(X_i = \{x_i, \bar{x}_i\}, Y_j = \{y_j, \bar{y}_j\} \mid Z = z) = 0.$$

En este caso, como el grafo tiene únicamente dos vértices y además son no adyacentes, la condición de independencia se convierte en

$$X_1 \perp\!\!\!\perp X_2$$

lo que es equivalente a afirmar

$$\text{cpd}(X_1 = \{x_1, \bar{x}_1\}, X_2 = \{x_2, \bar{x}_2\}) = 0,$$

es decir,

$$p(x_1, x_2)p(\bar{x}_1\bar{x}_2) - p(\bar{x}_1x_2)p(x_1, \bar{x}_2) = 0.$$

Recordemos que x_1, \bar{x}_1 son niveles de X_1 y x_2, \bar{x}_2 son niveles de X_2 entonces x_i tendrá nivel 1 ó 0. Al calcular todos los polinomios cpds algunos resultan triviales y otros se repiten. En particular tenemos

$$\begin{aligned} \text{cpd}(X_1 = \{0, 1\}, X_2 = \{0, 1\}) &= 0 \\ p(0, 0)p(1, 1) - p(1, 0)p(0, 1) &= 0 \\ p_{00}p_{11} - p_{10}p_{01} &= 0. \end{aligned}$$

Entonces, como este polinomio no se anula trivialmente, pertenece al ideal $I_{(p)}$. También tenemos que

$$\begin{aligned} \text{cpd}(X_1 = \{1, 0\}, X_2 = \{0, 1\}) &= 0 \\ p(1, 0)p(0, 1) - p(0, 0)p(1, 1) &= 0 \\ p_{10}p_{01} - p_{00}p_{11} &= 0. \end{aligned}$$

Por lo tanto, obtenemos el ideal por parejas generado por los siguientes binomios

$$I_{(p)\mathcal{G}} = \langle p_{00}p_{11} - p_{10}p_{01}, p_{10}p_{01} - p_{00}p_{11} \rangle.$$

Entonces, este es el ideal binomial en un anillo de polinomios de 4 indeterminantes

$$I_{(p)\mathcal{G}} \subset \mathbb{R}[\mathcal{X}] = \mathbb{R}[p_{00}, p_{01}, p_{10}, p_{11}].$$

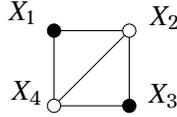
Ahora veamos la matriz $A(\mathcal{G})$ asociada al modelo gráfico, con generadores $\mathbb{G}_1 = \{X_1\}$ y $\mathbb{G}_2 = \{X_2\}$. Tenemos que

$$\begin{aligned} \mathcal{X}_{\mathbb{G}_1} &= I_{X_1} = \{0, 1\}, \\ \mathcal{X}_{\mathbb{G}_2} &= I_{X_2} = \{0, 1\}, \end{aligned}$$

entonces $A(\mathcal{G}) \in M_{4 \times 4}(\mathbb{Z}_{\geq 0})$ es la matriz

$$A(\mathcal{G}) = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}.$$

Ejemplo 86. Consideremos el modelo gráfico no dirigido de variables aleatorias binarias con grafo \mathcal{G}



Este grafo tiene dos cliques por lo que tiene dos generadores $\mathbb{G} = \{\{X_1, X_2, X_4\}, \{X_2, X_3, X_4\}\}$, luego las distribuciones de probabilidad para este modelo tienen la forma

$$P(x_1, x_2, x_3, x_4) \propto \psi_{(1,2,4)}(x_1, x_2, x_4) \psi_{(2,3,4)}(x_2, x_3, x_4).$$

Como las variables son binarias tenemos que $\mathcal{X} = \{0000, 0001, 0010, 0100, \dots, 1110, 1111\}$ y queremos ver los polinomios del ideal $I_{(p)\mathcal{G}}$, los cuales cumplen la condición de independencia

$$X_i \perp\!\!\!\perp X_j \mid \{X_1, \dots, X_n\} \setminus \{X_i, X_j\} \iff \text{cpd}(X_i = \{x_i, \tilde{x}_i\}, Y_j = \{y_j, \tilde{y}_j\} \mid Z = z) = 0.$$

Ya que los únicos vértices no adjacentes son X_1 y X_3 existe una única condición de independencia

$$X_1 \perp\!\!\!\perp X_3 \mid \{X_2, X_4\},$$

la cual es equivalente a afirmar que

$$\text{cpd}(X_1 = \{x_1, \tilde{x}_1\}, X_3 = \{x_3, \tilde{x}_3\} \mid (X_2, X_4) = \{x_2, x_4\}) = 0,$$

esto es,

$$p(x_1, x_3, x_2, x_4) p(\tilde{x}_1, \tilde{x}_3, x_2, x_4) - p(\tilde{x}_1, x_3, x_2, x_4) p(x_1, \tilde{x}_3, x_2, x_4) = 0.$$

Recordemos que cada x_i es nivel de X_i , luego las únicas opciones que tenemos para x_i son 0 ó 1. Note que algunos de los polinomios serán triviales y otros estarán repetidos. Veamos unos casos particulares,

$$\begin{aligned} \text{cpd}(X_1 = \{1, 1\}, X_3 = \{0, 1\} \mid (X_2, X_4) = \{1, 0\}) &= 0 \\ p(1, 0, 1, 0) p(1, 1, 1, 0) - p(1, 0, 1, 0) p(1, 1, 1, 0) &= 0 \\ p(1, 1, 0, 0) p(1, 1, 1, 0) - p(1, 1, 0, 0) p(1, 1, 1, 0) &= 0 \\ p_{1100} p_{1110} - p_{1100} p_{1110} &= 0, \end{aligned}$$

donde las dos últimas igualdades se obtiene al organizar los niveles y obtener el orden $p(x_1, x_2, x_3, x_4)$; este es un ejemplo de un resultado trivial.

Ahora, sea

$$\begin{aligned} \text{cpd}(X_1 = \{0, 1\}, X_3 = \{0, 1\} \mid (X_2, X_4) = \{1, 1\}) &= 0 \\ p(0, 0, 1, 1) p(1, 1, 1, 1) - p(1, 0, 1, 1) p(0, 1, 1, 1) &= 0 \\ p(0, 1, 0, 1) p(1, 1, 1, 1) - p(1, 1, 0, 1) p(0, 1, 1, 1) &= 0 \\ p_{0101} p_{1111} - p_{1101} p_{0111} &= 0, \end{aligned}$$

por lo tanto este binomio pertenece a $I_{(p)\mathcal{G}}$. Si continuamos cambiando todos los niveles posibles para $x_1, \tilde{x}_1, x_2, x_3, \tilde{x}_3$ y x_4 , obtenemos el ideal por parejas generado por los siguiente binomios

$$I_{(p)\mathcal{G}} = \langle p_{0111} p_{1101} - p_{0101} p_{1111}, p_{0110} p_{1100} - p_{0100} p_{1110}, p_{0011} p_{0110} - p_{0010} p_{0111}, p_{0010} p_{1000} - p_{0000} p_{1010} \rangle.$$

Luego, este es un ideal binomial en un anillo de polinomios de 16 indeterminantes:

$$I_{(p)\mathcal{G}} \subset \mathbb{R}[\mathcal{X}] = \mathbb{R}[p_{0000}, p_{0001}, p_{0010}, \dots, p_{1111}].$$

La matriz $A(\mathcal{G})$ asociada al modelo gráfico, con generadores $\mathbb{G}_1 = \{X_1, X_2, X_4\}$ y $\mathbb{G}_2 = \{X_2, X_3, X_4\}$ se construye de la siguiente manera

$$\begin{aligned} \mathcal{X}_{\mathbb{G}_1} &= I_{X_1} \times I_{X_2} \times I_{X_4} = \{000, 001, 010, 100, 011, 101, 110, 111\}, \\ \mathcal{X}_{\mathbb{G}_2} &= I_{X_2} \times I_{X_3} \times I_{X_4} = \{000, 001, 010, 100, 011, 101, 110, 111\}. \end{aligned}$$

Entonces $A(\mathcal{G}) \in M_{16 \times 16}(\mathbb{Z}_{\geq 0})$ es la matriz cuadrada que sigue

$$A(\mathcal{G}) = \begin{pmatrix} 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \end{pmatrix}.$$

Finalmente, tenemos todas las herramientas y la notación necesaria para describir el Teorema 54 de Hammersley-Clifford en términos de modelos gráficos, el cual relaciona el ideal de las afirmaciones de independencia condicional por parejas con la factorización.

Teorema 87 (Hammersley-Clifford). *Sea \mathcal{G} un modelo gráfico no dirigido. Una distribución de probabilidad P estrictamente positiva factoriza de acuerdo a $A(\mathcal{G})$ si y solo si P pertenece a la variedad $X_{(p)}^{>0}$.*

La demostración de este teorema es equivalente a la demostración del Teorema 57 realizada en el capítulo 2.

BIBLIOGRAFÍA

- [CD] COX, DAVID A., JOHN B. LITTLE, AND DONAL O'SHEA., *Ideals, Varieties, and Algorithms: An Introduction to Computational Algebraic Geometry and Commutative Algebra*, 3rd edition. Springer-Verlag, New York, 1992.
- [DS] DIACONIS P. AND STURMFELS, B., *Algebraic algorithms for sampling from conditional distributions*, Ann. Statist. 26 , no. 1, 363–397 (1998).
- [GD] GEIGER, DAN, CHRISTOPHER MEEK, AND BERND STURMFELS., "*On the Toric Algebra of Graphical Models.*", Ann. Statist. 34.3 (2006): 1463-492. Web.
- [D] DAWID, P., *Conditional Independence for Statistical operations*. Ann. Statist. Vol 8, No. 3, 598-617, 1980.
- [LS] LAURITZEN, S., *Graphical Models*, Clarendon Press, Oxford, 1996.