Universidad de Cantabria

Facultad de Ciencias

Análisis de los factores de riesgo de ahogamiento: una aproximación bayesiana.

(Analysis of Drowning Risk Factors: A Bayesian Approach)

Trabajo Fin de Grado para acceder al

GRADO EN MATEMÁTICAS

Autora: Valvanuz Saro Cobo

Director: Diego García Saiz

Codirector: Sixto Herrera Garcia

Junio 2025

Resumen / Abstract

RESUMEN

La mortalidad por ahogamiento constituye una de las principales causas de fallecimiento accidental no intencional en España. En los últimos años, estos accidentes han mostrado un claro incremento, tendencia que, en base a los datos registrados, sigue al alza en este 2025. Por lo tanto, un análisis de los factores determinantes del riesgo de mortalidad en dichos accidentes es necesario de cara a establecer medidas de prevención que permitan reducir este incremento y cambiar la tendencia observada estos últimos años.

En este Trabajo Fin de Grado se consideran los datos registrados por la Escuela Segoviana de Socorrismo, aprovechando su conocimiento experto para, mediante Redes Bayesianas, establecer las relaciones probabilísticas entre las variables de entorno del accidente y analizar los factores de riesgo más determinantes en el riesgo de mortalidad del mismo.

Palabras clave: Redes bayesianas, probabilidad condicionada, significación estadística, riesgo de ahogamiento, mortalidad.

ABSTRACT

Drowning mortality is one of the leading causes of unintentional accidental death in Spain. In recent years, these incidents have shown a clear increase—a trend that, based on recorded data, continues to rise in 2025. Therefore, analyzing the key factors that determine the risk of mortality in such accidents is essential in order to establish prevention measures that may help reverse this upward trend.

In this Bachelor's Thesis, data collected by surveillance services is used, incorporating their expert knowledge to establish probabilistic relationships between environmental variables related to the accidents through Bayesian Networks. This approach allows for the identification and analysis of the most critical risk factors contributing to mortality in drowning incidents.

Key words: Bayesian networks, conditioned probability, statistical significance, drowning risk, mortality.

Índice general

1. Introducción **1**
1.1. Motivación personal 1

2. Preliminares **3**
2.1. Introducción al Machine Learning 3
2.1.1. Clasificación según el tipo de aprendizaje 3
2.1.2. Clasificación por la naturaleza de los modelos 4
2.2. Introducción a preliminares matemáticos 5
2.2.1. Teoría de grafos 5
2.2.2. Probabilidad 8
2.2.3. Inferencia Estadística 10

3. Marco teórico **15**
3.1. Introducción a las redes bayesianas 15
3.2. Aprendizajes 22
3.3. Análisis del desempeño del modelo 25

4. Aplicación práctica **29**
4.1. Interés del estudio 29
4.1.1. Datos del estudio 30
4.2. Introducción a la metodología 32
4.3. Aplicación de la metodología 34
4.3.1. Estudio 34
4.3.2. Resultados y conclusiones 42

Capítulo 1

Introducción

1.1. Motivación personal

Hoy en día todo gira en torno a los datos. Saber recolectarlos, analizarlos y entenderlos es clave para tener éxito e innovar en cualquier área. Por eso, el aprendizaje automático basado en dichos datos se ha vuelto esencial y se ha extendido entre los procesos de toma de decisiones por parte de entidades tanto públicas como privadas.

En este ámbito, los enfoques probabilísticos juegan un papel muy importante, ya que permiten incluir en la toma de decisiones la incertidumbre existente en la vida real. Entre estos enfoques, las Redes Bayesianas permiten reflejar las dependencias probabilísticas, directas y/o condicionadas, entre los diferentes factores relevantes, analizar posibles causalidades y realizar inferencias sobre la variable de interés en base al resto de factores. Además, ofrecen una representación visual clara de las relaciones entre variables a través del grafo, lo cual facilita su interpretación. Sin embargo, también tienen algunas limitaciones, como el aumento de la complejidad computacional en modelos con muchas variables o estructuras densas, y la necesidad de contar con cierta experiencia para definir correctamente la estructura del grafo en algunos casos.

En cuanto a sus aplicaciones, las redes bayesianas se utilizan ampliamente en campos donde es importante razonar de forma probabilística y tomar decisiones informadas. En medicina, por ejemplo, se aplican en sistemas de diagnóstico para identificar enfermedades a partir de los síntomas y los antecedentes del paciente. También son comunes en la detección de fraudes financieros, en sistemas de recomendación como los de las plataformas digitales, y en biología, entre otras muchas aplicaciones [1–6].

En este trabajo de fin de grado, aplicaremos este modelo para analizar las relaciones probabilísticas entre los distintos factores de entorno de un accidente no intencional con ahogamiento en España y el riesgo de mortalidad a causa de dicho accidente. Del mismo modo, el uso de este modelo nos permitirá establecer posibles relaciones de dependencia entre las diferentes variables y el riesgo de mortalidad, y analizar el efecto de dichas variables en el riesgo de mortalidad por ahogamiento del accidente a través de la inferencia probabilística.

Esta línea de estudio se puede encontrar en varios artículos de actualidad mediante diferentes enfoques. Se pueden encontrar acercamientos con redes neuronales, comparando banderas físicas con virtuales en relación con el riesgo del entorno o centradas en categorías especificas, como en los menores [7–11, estudios de ML en el riesgo de ahogamiento]. Sin embargo, también se encuentra algún estudio en una línea similar a las que se quiere afrontar en este trabajo, como por ejemplo en el artículo [12].

La mortalidad por ahogamiento constituye una de las principales causas de fallecimiento accidental no intencional en España. En los últimos años, estos accidentes han mostrado una tendencia a ir en aumento, pero con variaciones anuales significativas. Desde 2021, no ha hecho más que aumentar el número de decesos, siendo el año 2024 el segundo peor año de la última década. Sin embargo, el contador de 2025 no ha parado, indicando una continuación de la tendencia al alza. Por ello, es necesario analizar las circunstacias concurrentes en este tipo de accidentes de cara a establecer medidas preventivas y de concienciación para revertir esta tendencia.

Este tipo de incidentes no afecta por igual a toda la población: existen variaciones significativas según la edad, la localización geográfica, la nacionalidad, las condiciones meteorológicas y muchos otros factores. Por ello, el uso de modelos probabilísticos que nos permitan analizar las relaciones entre los diferentes factores alrededer de un accidente y detectar patrones que puedan contribuir al diseño de estrategias preventivas más eficaces.

Es por ello que el presente TFG presenta los siguientes objetivos:
Profundizar en la teoría sobre la cual se fundamentan los modelos gráficos probabilísticos, en particular las Redes Bayesianas, explicando su funcionamiento, estructura y características principales. Se incluyen definiciones clave, así como propiedades fundamentales como la d-separación y el teorema de sonoridad. Además, se abordan los principales algoritmos tanto para la construcción de la estructura como para la inferencia, y se presenta un análisis detallado sobre el estudio de la calidad de las redes aprendidas. Todo esto nos permitan afrontar el segundo objetivo.

Identificar los factores que influyen en la probabilidad de fallecimiento por ahogamiento en España, a partir de un análisis probabilístico basado en Redes Bayesianas, en base a lo expuesto teóricamente. Específicamente, se buscan tres objetivos prácticos:

- Primero, modelar las relaciones entre las variables sociodemográficas y situacionales presentes en los incidentes.
- Segundo, estimar la probabilidad de muerte en función de distintos perfiles de riesgo.
- Y por último, evaluar el ajuste del modelo frente a datos reales y analizar su capacidad predictiva.

Capítulo 2

Preliminares

En este capítulo se establecerán las bases teóricas de los métodos que utilizaremos en este TFG. Por un lado, la introducción sobre machine learning nos da una visión general de este concepto y sus distintas ramas, siendo extraídas las ideas de libros como [13] y [14]. Por otro lado, en la introducción a los conceptos matématicos desarrollamos los conceptos básicos sobre grafos, probabilidades y test de inferencia, para lo cual se han considerado los apuntes proporcionados a lo largo de la carrera, como los de matemática discreta [15] o los de estadística y cálculo de probabilidades [16], así como el libro de introducción a algoritmos [17].

2.1. Introducción al Machine Learning

El Machine Learning es una rama de la inteligencia artificial que permite a los sistemas aprender a partir de datos y mejorar su rendimiento en tareas concretas, sin necesidad de ser programados explícitamente.

Esta disciplina surge a mediados del siglo XX, cuando un grupo de investigadores, entre ellos Alan Turing y Arthur Samuel, consideraron explorar el potencial de las máquinas para descubrir patrones y realizar inferencias basadas en datos históricos. Estos primeros estudios sentaron las bases teóricas y prácticas de este campo, el cual ha continuado su desarrollo de forma ininterrumpida hasta nuestros días de la mano de los desarrollos teóricos y tecnológicos experimentados. De este modo, actualmente se usan este tipo de técnicas en gran parte de los aspectos de la sociedad, tanto en el ámbito científico como en el día a día de las personas con el creciente uso y normalización de la ayuda de distintas inteligencias artificiales.

A continuación, se presenta una descripción de dos de las clasificaciones principales de este área, una según la manera de realizar el aprendizaje y otra según el tipo de modelo considerado:

2.1.1. Clasificación según el tipo de aprendizaje

Basándonos en el modo en el cual aprenden los algoritmos a partir de los datos podemos considerar tres tipos:

- **Aprendizaje supervisado:** Este tipo de aprendizaje se fundamenta en la utilización de datos *etiquetados*. Es decir, se dispone de una variable y de la cual conocemos los valores correctos observados, de modo que se pueden comparar éstos con los valores predichos por nuestro modelo y^*. De este modo, el modelo aprende a crear relaciones entre las variables de entrada (X) y las variables objetivo (Y) de salida, para poder realizar predicciones sobre un conjunto nuevo de datos. El aprendizaje de modelos como la regresión lineal (predicción) o la logística (clasificación) se enmarca dentro de este tipo.

- **Aprendizaje no supervisado:** Al contrario del aprendizaje supervisado, en esta clasificación los datos no tienen etiquetas. Por lo tanto, la finalidad es descubrir patrones y relaciones implícitas entre las variables del problema (X). El problema de agrupamiento (clustering en inglés) se enmarca en este tipo de aprendizaje, siendo algoritmos como K-means y derivados los más utilizados en este ámbito por su interpretabilidad y simplicidad.

- **Aprendizaje por refuerzo:** En esta situación, un agente autónomo aprende a actuar dentro de un entorno usando recompensas y castigos. Gracias a la exploración y a la retroalimentación que recibe, el agente modifica su comportamiento para obtener la mayor cantidad posible de recompensas a lo largo del tiempo. Algunos de los algoritmos más representativos son Q-Learning y Deep Q-Networks.

Si bien las Redes Bayesianas tratan de aprender la función de probabilidad conjunta de todas las variables $P(Y, X_1, ..., X_n)$, se clasifican, por lo general, dentro del aprendizaje supervisado, ya que se suelen utilizar para problemas de clasificación y predicción, en los cuales la red aprende las relaciones probabilísticas entre las variables a partir de los datos conocidos, para luego utilizar éstas para inferir la probabilidad de una de las variables incluidas en el modelo ($P(Y|X_1, ..., X_n)$), considerada la variable objetivo (Y).

2.1.2. Clasificación por la naturaleza de los modelos

Basándonos en la forma en la que los algoritmos procesan y aprenden a partir de los datos, los modelos de Machine Learning pueden clasificarse en cinco grandes categorías:

- **Modelos probabilísticos:** Se trata de modelos fundamentados en la teoría de la probabilidad y la estadística, diseñados para manejar la incertidumbre presente en los datos. Estos modelos describen las relaciones entre variables mediante distribuciones de probabilidad y generan predicciones junto con una medida de confianza. Resultan particularmente eficaces en situaciones donde los datos contienen ruido. Algunos de los algoritmos son las Redes Bayesianas y los modelos de Markov ocultos.

- **Modelos lineales:** Estos modelos suponen que existe una relación lineal entre las variables independientes (X) y la variable dependiente (Y). Destacan por su eficiencia y por la facilidad con la que pueden interpretarse. No obstante, su desempeño puede verse limitado en situaciones más complejas, donde las relaciones entre los datos no siguen un patrón lineal. Los algoritmos más frecuentes son la regresión lineal o los modelos lineales generalizados, en los cuales se incluye la regresión logística.

- **Modelos basados en vecindad:** Estos modelos se fundamentan en el concepto de que puntos cercanos, según alguna norma, deben presentar valores similares. De este modo, estos modelos realizan predicciones comparando nuevas observaciones con instancias previamente conocidas, utilizando métricas de similitud. Por lo tanto, son fáciles de comprender e implementar, aunque pueden ser costosos computacionalmente cuando se trabaja con grandes volúmenes de datos o con datos con un gran número de dimensiones. Descatan los algoritmos como k-Nearest Neighbors, para predicción o clasificación, o K-Means para agrupamiento.

- **Modelos basados en ensembles:** En este caso, el principio de este tipo de métodos es conocido como *"The wisdom of the crowd"* (La sabiduría de la multitud). De este modo, estos modelos combinan múltiples modelos individuales para generar la predicción final, tratando de obtener predicciones más precisas y robustas. Se suelen considerar tres familias: Bagging (Bootstrap Aggregating), Boosting y Stacking, considerando las dos primeras un mismo tipo de modelos (p.e. árboles de predicción), mientras que la tercera combina modelos diferentes (p.e. regresión lineal, K-means y Red Neuronal) a través de un meta-modelo. A través de dicha combinación se busca resolver los problemas de los modelos individuales, y así disminuir el error y aumentar la capacidad del modelo para generalizar una vez se aplique a nuevos datos. De entre esta familia de métodos, Random Forest y (Extreme) Gradient Boosting (XGBoost) son los más aplicados.

- **Redes neuronales y aprendizaje profundo:** Son modelos inspirados en el funcionamiento del cerebro humano, capaces de aprender representaciones complejas a partir de grandes volúmenes de datos. Con el desarrollo de hardware especializado, como las GPU, el aprendizaje profundo se ha consolidado como una herramienta fundamental en tareas avanzadas. Destaca su uso en los reconocimientos de imágenes o procesamiento de lenguaje natural.

En base a esta clasificación, las Redes Bayesianas se enmarcan dentro del primer grupo ya que se basan en la teoría de la probabilidad bayesiana para representar y razonar sobre la incertidumbre.

2.2. Introducción a preliminares matemáticos

Las Redes Bayesianas pertenecen a la categoría de los modelos gráficos probabilísticos [18], de modo que es necesario introducir los conceptos básicos de la teoría de grafos, de probabilidad y cómo se relacionan ambos campos en esta familia de modelos, con especial énfasis en el caso particular de las Redes Bayesianas.

2.2.1. Teoría de grafos

En primer lugar, se presentan las definiciones de los términos esenciales para el estudio de la estructura de las redes bayesianas. El punto de partida es el término más fundamental: los conceptos de grafo y grafo dirigido.

Definición 2.1. Un **grafo finito** es una terna G = (V, A , φ) , constituida por V que es un conjunto finito no vacío formado por vértices, o nodos, A que es un conjunto finito formado por aristas, y φ : A $\rightarrow$ P(V) que es una aplicación que asocia una correspondencia entre una arista y dos vértices de este modo $\varphi(a) = \{x, y\}$ siendo a una arista y x e y vértices.

Definición 2.2. Un **grafo dirigido** es un grafo finito en el cual la aplicación φ : A $\rightarrow$ P(V) asocia a cada arista un par de vértices ordenados, es decir, comienza en un nodo y acaba en el otro. De este modo, en este caso $\varphi(a) = \{x, y\}$, x representa el inicio de la arista e y el final.

Cabe recordar que los grafos son una estructura matemática que se puede representar gráficamente. Sin embargo, un mismo grafo puede tener distintas representaciones visuales que muestran las mismas conexiones y propiedades. Por eso, es importante no confundir el concepto abstracto de grafo, que es la estructura, con su forma particular de ser representado.

Como se ha definido anteriormente, en un grafo dirigido cada arista tiene una dirección que va de un nodo a otro. De este modo, se establece una relación de orden entre los nodos en la que al nodo donde comienza la arista lo llamaremos **nodo padre**, y al nodo donde termina, **nodo hijo**.

Asimismo, es natural el concepto de la ordenación topológica.

Definición 2.3. Sea G=(V,A, φ) un grafo dirigido de n variables, una **ordenación topológica** es una función

$$f : V \rightarrow \{1, 2, ..., n\} \text{ tal que } \forall a \in A \text{ si } \varphi(a) = (u, v), f(u) < f(v) \tag{2.1}$$

Es una enumeración lineal de los nodos donde para cada arista dirigida (u,v) se respeta el orden en el que aparecen los vértices.

Para poder definir correctamente el término de grafo dirigido acíclico es necesario introducir los siguientes conceptos.

Definición 2.4. Sea G=(V,A φ) un grafo.
Un **recorrido** en G es una sucesión $x_0a_1x_1a_2x_2..x_{k-1}a_kx_k$ donde $x_i \in$ V y a_i es una arista incidente a x_{i-1} y a x_i.

Definición 2.5. Un recorrido se dice **camino** si todas las aristas son distintas.

Definición 2.6. Un **camino cerrado** empieza y acaba en el mismo nodo.

Definición 2.7. Un **camino simple** es aquel en el que no se repiden los arcos, es decir, todos los nodos, excepto el primero y el último son distintos dos a dos, en el caso de que sea un camino cerrado.

Definición 2.8. Un **lazo** es un camino que conecta a un nodo consigo mismo, es decir, si se da que $\varphi(a) = \{x, x\}$ para un vértice x.

Definición 2.9. Un **ciclo** es un camino simple, cerrado, no trivial (no es un nodo aislado o un lazo).

Definición 2.10. Un **grafo dirigido acíclico** (DAG, por sus siglas en inglés) es un grafo G=(V,A, φ) que cumple los requisitos expuestos de grafo dirigido y se le añade la característica de no tener ciclos.

Ahora vamos a exponer un teorema de identicación de grafos dirigidos acíclicos.

Teorema 2.11. ***Teorema de caracterización de DAGs***
Un grafo dirigido tiene una ordenación topológica $\leftrightarrow$ *no tiene ciclos dirigidos.*

Demostración. Vamos a demotrar primero este sentido de la afirmación $\leftarrow$

Procederemos por inducción.

Supongamos que tenemos G = (V,E) un grafo dirigido sin ciclos. Queremos probar que existe una ordenación topológica. Aplicamos inducción sobre el número de vértices $|V| = n$.

·Primero, caso base, en donde $n = 0$
Como G no tiene vértices, la ordenación topológica vacía es válida.
Esto implica que la propiedad se cumple.

·Aplicamos el paso inductivo:
Supongamos que todo grafo dirigido acíclico con $n - 1$ vértices admite una ordenación topológica.

Sea G un grafo dirigido acíclico con n vértices. Como G no contiene ciclos, debe existir al menos un vértice $v \in V$ sin predecesores, es decir, con grado de entrada 0 en dicho nodo. Si no existiera tal vértice, entonces podríamos seguir una cadena infinita de predecesores, lo cual implicaría un ciclo, contradiciendo la hipótesis.

Eliminamos dicho vértice v y todas sus aristas salientes, obteniendo un nuevo grafo $G' = (V', E')$, donde $|V'| = n - 1$.

Por hipótesis de inducción, G' admite una ordenación topológica.
Ahora insertamos el vértice v al inicio de la ordenación, como no tiene predecesores, implica que se conserva la ordenación para G.

Ahora veamos el otro sentido $\rightarrow$.

Supongamos que un grafo dirigido $G = (V, E)$ tiene una ordenación topológica, pero que G contiene un ciclo dirigido.
Sea $v_1, v_2, \ldots, v_k, v_1$ un ciclo en G, con $k \geq 2$ vértices distintos. Como se trata de un ciclo, todas las aristas (v_i, v_{i+1}) para $1 \leq i < k$ y (v_k, v_1) están en E.

En una ordenación topológica, si $(u, v) \in E$, entonces u debe aparecer antes que v.

Entonces, aplicando esto al ciclo, tendríamos:

$$v_1 < v_2 < \ldots < v_k < v_1$$

lo cual es una contradicción, ya que implica que $v_1 < v_1$, lo cual es imposible.

Por tanto, nuestra suposición es falsa. □

2.2.2. Probabilidad

En esta subsección vamos a exponer los pilares matemáticos probabilísticos sobre los que se sustentan las redes bayesianas.

Definición 2.12. Dado un fenómeno, tenemos que un **espacio muestral** Ω es el conjunto de todos los posibles resultados, cada elemento se corresponde univocamente con uno de estos resultados.

Una vez definido el espacio de todos los resultados posibles, podemos centrarnos en los subconjuntos de interés dentro de ese espacio.

Definición 2.13. Un **suceso** es cualquier subconjunto de resultados del espacio muestral.

Para poder trabajar rigurosamente con estos subconjuntos, se requiere una estructura algebraica que garantice ciertas propiedades cerradas. Ahí entra en juego la σ-álgebra:

Definición 2.14. Una **σ-álgebra** F es un conjunto de subconjuntos del espacio muestral Ω que cumple las siguientes propiedades:

1. $\Omega \in F$
2. Si $A \in F$ entonces $A^C \in F$
3. Si $\{A_n\}_{n\in\mathbb{N}} \subseteq F$ entonces $\cup_{n=1}^{\infty} A_n \in F$

Una vez establecidos estos conceptos previos fundamentales, ya podemos definir el marco completo que utilizaremos para las probabilidades.

Definición 2.15. Un **espacio probabilístico** es una terna (Ω, F, P), donde:

- Ω es el espacio muestral,
- F es una σ-álgebra sobre Ω,
- $P : F \to [0, 1]$ es una función de probabilidad que cumple:
  1. P(Ω)=1
  2. P(A)$\geq$0 $\forall A \in F$
  3. P($\cup_{n=1}^{\infty} A_n$) = $\sum_{n=1}^{\infty} P(A_n)$ si $A_i \cap A_j = \emptyset$ para i$\neq$j.

Definición 2.16. Una **variable aleatoria** es una función $X : \Omega \to \mathbb{R}$ para cualquier $x \in \mathbb{R}$, el conjunto $\{\omega \in \Omega : X(\omega) \leq x\} \in F$.

Una variable aleatoria puede ser discreta, cuando toma valores en un conjunto finito o numerable; o continua que es cuando toma valores en un intervalo de la recta real, una unión de varios intervalos, o la totalidad de la recta real. En el caso de las redes bayesianas, cada nodo del grafo representa una variable aleatoria. En concreto, dada la naturaleza de las variables consideradas en este TFG, solo vamos a considerar variables discretas, de modo que sólo vamos a profundizar en los conceptos asociados a éstas.

Cuando trabajamos con dos o más variables aleatorias discretas, nos interesa describir cómo se relacionan entre sí. Para ello, utilizamos la función de probabilidad conjunta, que nos permite conocer la probabilidad de que todas las variables tomen simultáneamente determinados valores.

Definición 2.17. La **función de probabilidad** de una variable discreta es P(X= x_i) ≥ 0, tal que cumple que $\sum_i P(X = x_i) = 1 \ \forall x_i \in \Omega$.

Si bien la expresión anterior nos permite conocer la probabilidad de ocurrencia de uno o varios valores de X, en el presente TFG se requieren métricas más complejas que permitan ver la relación entre las variables, así como saber las probabilidades de un suceso sabiendo que se ha dado otro cierto suceso. Esto nos lleva a los conceptos de función de probabilidad conjunta y probabilidad condicionada, que definimos a continuación.

Definición 2.18. Sean X e Y dos variables aleatorias discretas definidas sobre un mismo espacio probabilístico, la **función de probabilidad conjunta** de X e Y es la función:

$$p : X \times Y \to [0, 1] \text{ tal que } p_{X,Y}(x, y) = P(X = x, Y = y) \tag{2.2}$$

que cumple que

- $p_{X,Y}(x, y) \geq 0 \ \forall (x, y)$
- $\sum_{x \in X} \sum_{y \in Y} p_{X,Y}(x, y) = 1$

A partir de una distribución conjunta es posible obtener la distribución de una sola variable.

Definición 2.19. Dada una distribución conjunta de dos variables aleatorias discretas X y Y, la **probabilidad marginal** de $X = x$ se obtiene sumando sobre todos los posibles valores de Y.

$$P(X = x) = \sum_y P(X = x, Y = y) \tag{2.3}$$

Definición 2.20. Dado el espacio probabilístico (Ω, F, P) y los sucesos X, Y con P(Y=y)>0 para los sucesos de cualquera de las dos variables, llamamos **probabilidad de X a que se dé el suceso x condicionada a que haya ocurrido el suceso Y=y**, al valor

$$P(X = x|Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)} \tag{2.4}$$

Dichas definiciones pueden ampliarse a un mayor número de variables y sucesos. Por ejemplo:

$$P(X = x|Y = y, Z = z) = \frac{P(X = x, Y = y, Z = z)}{P(Y = y, Z = z)} \tag{2.5}$$

$$P(X = x, Y = y|Z = z) = \frac{P(X = x, Y = y, Z = z)}{P(Z = z)} \tag{2.6}$$

Otra noción clave es la de independencia probabilística.

Definición 2.21. Un conjunto de variables aleatorias $X_1, ..., X_n$ son **independientes** si y solo si se cumple que $P(X_1 = x_1, ..., X_k = x_k) = \prod_{i=1}^{k} P(X_i = x_i)$. Si esta igualdad no se cumple, son variables dependientes.

Definición 2.22. Dos variables X e Y son **condicionamente independientes** $X \perp Y|Z$ dado un conjunto Z si $P(X, Y|Z) = P(X|Z)P(Y|Z)$.

Las definiciones anteriores nos introducen el Teorema de Bayes, el cual es esencial para el desarollo de las Redes Bayesianas:

Teorema 2.23. ***Teorema de Bayes.***
Consideremos el espacio probabilístico (Ω,F, P). Sean $A_1, ...A_k, B \in \Omega$ tales que

- *Los eventos $A_1, \ldots, A_k$ forman una partición de Ω*
- *$P(A_i) > 0 \ \forall i$*
- *$P(B) > 0$*

Entonces:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{2.7}$$

$$P(B) = P(B|A)P(A) \tag{2.8}$$

o ampliado:

$$P(A_i|B) = \frac{P(B|A_i)P(A)}{P(B)} \tag{2.9}$$

$$P(B) = \sum_i P(B|A_i)P(A_i) \tag{2.10}$$

2.2.3. Inferencia Estadística

En este apartado se abordan los test de hipótesis que nos permitan contrastar la dependencia e independencia de las variables incluidas en el modelo (*test de independencia*). Para ello, comenzaremos por establecer los fundamentos teóricos de este tipo de pruebas y posteriormente la metodología de aplicación.

Definición 2.24. La **hipótesis nula** H_0 es la proposión que se somete a prueba.
En el caso de la prueba de independencia χ^2, H_0 implica suponer que dos variables son independientes, mientras que en la prueba de proporciones de dos muestras, H_0 implica suponer que ambas proporciones son iguales, es decir, $p_1 = p_2$

Definición 2.25. La **hipótesis alternativa** H_1 es la afirmación complementaria a la hipótesis nula. Representa lo que se desea demostrar si ésta se rechaza. Es decir, en las pruebas anteriores, que las variables son dependientes y que las proporciones son diferentes, respectivamente.

Definición 2.26. El **error de tipo I** ocurre cuando se rechaza la hipótesis nula siendo verdadera.

Definición 2.27. El **nivel de significancia** α es la probabilidad de cometer un error de tipo I.

$$\alpha = P(\text{Rechazar } H_0 | H_0 \text{ es cierta}) \tag{2.11}$$

Definición 2.28. El **p-valor** es la probabilidad, bajo la hipótesis nula, de obtener una estadística de prueba al menos tan extrema como la observada, donde un valor se define como extremos según se aleje de la hipótesis nula H_0.

Definición 2.29. Sea X una variable con r valores e Y otra con c valores, la **tabla de contingencia** es una matriz de tamaño $r \times c$ que en cada posición $[i, j]$ contiene el número de observaciones de la intersección de los valores del suceso $X = x_i$ y el suceso $Y = y_j$. Es decir, organiza las frecuencias absolutas de las combinaciones entre los valores de ambas variables. Esto permite ver su distribución conjunta y sirve como pilar para realizar el estudio de la independencia.

Primero, vamos a exponer el funcionamiento de los test de independencia. Como comienzo, se plantea la relación entre dos variables cuya dependencia se quiere analizar estableciendo tanto la hipótesis nula como la alternativa. La hipótesis nula debe ser conservadora, con el propósito de reducir el error de tipo I. Se escoge un valor para el nivel de signficancia α, siendo $\alpha = 0{,}05$ un valor comúnmente utilizado.

En este punto, es el momento de seleccionar un estadístico. Para evaluar las dependencias se aplica el test χ^2, en el cual se comparan las frecuencias reales registradas en la tabla de contingencia con las esperadas si, en efecto, estas variables fueran independientes. Es decir, si se cumpliese la hipótesis nula H_0

Como hemos expuesto, es necesario comparar las frecuencias, para lo cual se considera la tabla de contingencia. Para calcular las frecuencias esperadas bajo la hipótesis de independencia, multiplicamos el total de su fila por el total de su columna, y dividimos entre el número de observaciones, es decir, aplicando

$$E_{ij} = \frac{(\text{total fila}_i \times \text{total columna}_j)}{N}$$

donde E_{ij} es el valor para la celda (i, j), y N es el total de observaciones.

A partir de la distribución de frecuencias observadas (O) y esperadas bajo la hipótesis nula (E), estimamos el estadístico asociado a la prueba (χ^2):

$$T = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}},$$

donde O_{ij} representa la frecuencia observada y E_{ij} la frecuencia teórica estimada bajo H_0 para esos valores (i, j).

El estadístico anterior se distribuye como una χ^2 de $(r-1)(c-1)$ grados de libertad, siendo r el número de filas y c el número de columnas de la tabla de contingencia. Por lo tanto, dado el nivel de significancia, a través de la tabla asociada a esta distribución podemos establecer el umbral para el valor del estadístico T a partir del cual se puede rechazar la hipótesis nula a dicho nivel de significancia. En caso contrario, se acepta H_0.

De forma alternativa a este último paso, en vez de comparar estadísticos, se podría calcular el p-valor, que representa la probabilidad de obtener un estadístico igual o más extremo que el observado, asumiendo que la hipótesis nula es verdadera. Para determinar el p-valor, se utiliza la función de distribución acumulada de la χ^2 con los grados de libertad correspondientes. Dicha aproximación es ideal para ser usada en software estadístico y es la habitualmente utilizada en este tipo de análisis.

Si el p-valor obtenido es:

$$p < \alpha \rightarrow \text{rechazar } H_0 \tag{2.12}$$

$$p \geq \alpha \rightarrow \text{no rechazar } H_0 \tag{2.13}$$

La traducción de esto es que un p-valor pequeño proporciona evidencia estadística suficiente para rechazar la hipótesis nula, lo cual sugiere que las variables no son independientes. En cambio, lo contrario indica que las diferencias observadas podrían deberse al azar, y por lo tanto no se rechaza la independencia entre las variables.

Cabe resaltar que para que el test sea válido, se requiere que las frecuencias esperadas en cada celda sean suficientemente grandes. En caso contrario, pueden utilizarse métodos alternativos como el test exacto de Fisher.

De modo similar, es posible aplicar otros tipo de pruebas en el contexto del experimento, como los z-test, con el objeto de analizar si las proporciones observadas de un determinado suceso difieren significativamente entre distintas categorías.

Este tipo de prueba tiene dos vertientes, la primera mediante la comparación entre una proporción observada y una proporción teórica esperada, llamada prueba de una proporción, o la segunda a través de la comparación entre dos proporciones poblacionales independientes, la prueba de dos proporciones.

Esta segunda prueba se fundamenta en la suposición de que, bajo ciertas condiciones, la distribución muestral de las proporciones puede aproximarse a una distribución normal. Entre estas condiciones, destaca la necesidad de contar con un tamaño muestral n lo suficientemente grande como para garantizar la validez de dicha aproximación (para que se cumpla $n * p \geq 5$ y $n(1-p) \geq 5$ siendo p la proporción de éxito).

Para proceder con esta prueba se siguen los siguientes pasos:
Primero se realiza la formulación de las hipótesis. En este caso, se toma como hipótesis nula que las proporciones son iguales $H_0 : p_1 = p_2$, lo que implica que la hipótesis alter-

nativa sería que las proporciones son distintas, $H_1 : p_1 \neq p_2$.

Después, se procede al cálculo del estadístico, que para una proporción viene dado por la expresión:

$$Z = \frac{p_1 - p_2}{\sqrt{p(1-p)(\frac{1}{n_1} + \frac{1}{n_2})}} \tag{2.14}$$

donde:

- p_1, p_2 son las proporciones muestrales
- $p = \frac{x_1+x_2}{n_1+n_2}$ proporción combinada
- x_1, x_2 éxitos en cada muestra
- n_1, n_2 los tamaños de muestra

El estadístico Z se distribuye como una $N(0, 1)$ de modo que, dado un nivel de significación α, empleamos la siguiente lógica para evaluar si se acepta o rechaza la hipótesis nula.

$$\text{Si } Z > N^{-1}_{(0,1)}(\alpha) \rightarrow p_1 \neq p_2 \tag{2.15}$$

También es posible usar la tabla de la distribución normal para encontrar el p-valor y compararlo con el nivel de significancia escogido, de modo que podamos rechazar o no la hipótesis nula con el mismo criterio considerado para el test χ^2.

Capítulo 3

Marco teórico

En este capítulo se presenta una introducción a las redes bayesianas, partiendo ya de los preliminares matemáticos necesarios expuestos en el capítulo anterior. Se abordarán los conceptos fundamentales, para posteriormente en el siguiente capítulo realizar una aplicación práctica. El capítulo también incluye propiedades, tipos de aprendizaje y métricas para poder evaluar con precisión los resultados obtenidos con consultas a fuentes como los siguientes apuntes del Máster Interuniversitario de Ciencia de Datos [19] o los libros [14], [13] y [20].

3.1. Introducción a las redes bayesianas

Las redes bayesianas se enmarcan dentro de los modelos gráficos probabilísticos. De este modo, constan de un grafo dirigido acíclico (DAG, por sus siglas en inglés) que representa dependencias probabilísticas entre el conjunto de variables consideradas en el modelo. En estos modelos, cada nodo del grafo corresponde a una variable, y las aristas dirigidas indican las dependencias, directas o condicionadas, entre ellas. Por otro lado, el grafo definido anteriormente permite definir la función de probabilidad conjunta ($P(X_1, ..., X_n)$) de las variables de forma eficiente, así como realizar inferencias con la red.

A continuación, se presenta la base teórica de este concepto.

Definición 3.1. Una **red bayesiana** consta de un grafo dirigido y acíclico G=(V,A, φ), en el que

- A cada nodo de V le corresponde una única variable de un conjunto de variables X, sobre el cual existe una distribución de probabilidad P.
- G es un I-mapa minimal de dicha probabilidad P.

  Los **I-mapas** son grafos dirigidos acíclicos que representan las independencias condicionales presentes en una distribución de probabilidad P. Se profundizará en el concepto de las independencias más adelante en esta sección.

  Un I-mapa se considera **minimal** cuando ninguna de sus aristas puede eliminarse sin perder alguna independencia representada correctamente.

- Cada variable X_i tiene asociada una distribución de probabilidad condicionada a las variables $X_{\pi(i)}$, denotada $\mathrm{P}(X_i|X_{\pi(i)})$, siendo $\pi(i)$ los padres de X_i en el grafo. Por lo tanto cada nodo tiene una Tabla de la Probabilidad Condicional (CPT) asociada.

  Una **CPT** es una estructura tabular propia de los modelos probabilísticos en la cual se almacenan las probabilidades de cada nodo para cada posible combinación de valores de sus padres. Si no tiene padres, contiene su probabilidad marginal.

Resumiendo, las redes bayesianas se componen de un DAG para representar las relaciones entre variables y CPTs asociados a cada nodo, que proporcionan las probabilidades condicionadas.

La siguiente propiedad probabilística es fundamental porque permite descomponer una distribución de probabilidad conjunta compleja en una serie de probabilidades condicionales más simples, teniendo en cuenta sus relaciones de dependencia, ayudándonos de la estructura del grafo de la red.

Proposición 3.1.1. (Regla de la cadena)
Dada una red bayesiana sobre un conjunto X de n variables, la distribución de probabilidad conjunta de X es el producto de las distribuciones de probabilidad condicionada especificas en la red, es decir, para cada variable X_i su probabilidad condicional depende de todas sus variables padre $X_{\pi(i)}$,

$$P(X_1, .., X_n) = \prod_{i=1}^{n} P(X_i|X_{\pi(i)}) \tag{3.1}$$

En caso de no usarse, esta distribución conjunta sobre las n variables requeriría especificar $2^n - 1$ probabilidades independientes, en el caso más simple de considerar variables binarias. Sin embargo, aplicando la regla de la cadena se reduce únicamente a considerar los padres de dichos nodos. Si supusieramos que los nodos tienen valores binarios y a lo sumo k padres, siendo k menor que n, entonces la complejidad de la red habría disminuido hasta ser de complejidad $O(n * 2^k)$.

Otro término importante con respecto al tamaño de las redes, es su dimensión, que mide cuántos parámetros son necesarios almacenar para representar la distribución conjunta de sus variables. Depende de la cantidad de variables, sus dependencias y su número de estados posibles.

Proposición 3.1.2. (Dimensión de una RB)
La dimensión de una red bayesiana, es el número total de parámetros de la red,

$$P = \sum_{i=1}^{n}((k_i - 1) \prod_{X_i \in X_{\pi(i)}} k_j) \tag{3.2}$$

siendo n el número de variables, k_i el número de posibles valores para la variable X_i, significando el productorio el número de combinaciones posibles de los padres del nodo i-ésimo ($\pi(i)$). Cuantas más variables y más relaciones entre ellas, mayor será la dimensión de la red.

Recordemos que las redes bayesianas representan las relaciones de dependencia gráficamente, a través de las aristas. Para poder determinar formalmente este criterio surge el concepto de d-separación.

Definición 3.2. Sea G=(V,A, φ) un grafo dirigido acíclico formado por un conjunto de nodos V que representan las variables y un conjunto de aristas A que representan sus relaciones, y sean X,Y,Z $\subseteq$ V subconjuntos disjuntos de nodos, decimos que Z **d-separa** a X de Y en G si todo camino no dirigido entre un nodo en X y un nodo en Y está bloqueado por Z según unos criterios sobre su dirección de enlace siguiendo unas reglas gráficas. En concreto hay tres estructuras clave:

- Cadena:
$$A \rightarrow B \rightarrow C \tag{3.3}$$
  - Sin evidencia: A y C son dependientes, como si se aplicase la propiedad transitiva.
  - Daba evidencia: Sabemos B, entonces A y C están d-separadas.
- Divergencia:
$$A \leftarrow B \rightarrow C \tag{3.4}$$
  - Sin evidencia: A y C son dependientes. (B causa A y C)
  - Daba evidencia: Sabemos B, entonces A y C están d-separadas.
- Convergencia/estructura en v:
$$A \rightarrow B \ \leftarrow C \tag{3.5}$$
  - Sin evidencia: A y C son independientes y causan B.
  - Daba evidencia: Sabemos B, entonces A y C están d-separadas.

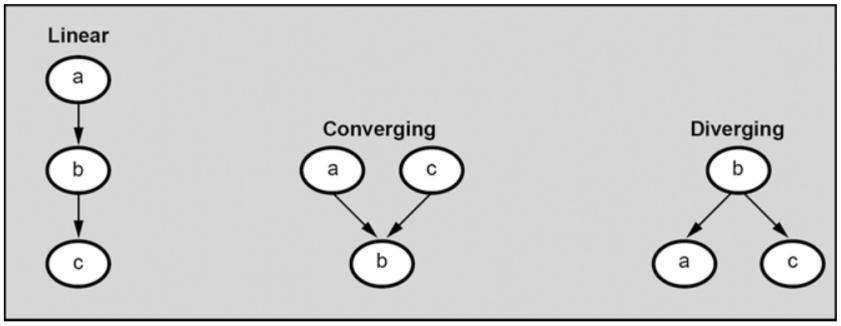


Figura 3.1: Esquema de los tres tipos de relación.

A cualquier red bayesiana se le puede aplicar un análisis de independencia por d-separación empleando combinaciones de estos tres tipos de estructuras.

Dando un paso más allá, y trás un riguroso proceso de demostración se puede establecer que la d-separación implica independencia condicional. Esta propiedad se conoce como el **Teorema de Sonoridad**.

Teorema 3.3. ***Teorema de sonoridad de la d-separación***
Dada una red bayesiana G=(V,A, φ), si dos variables X e Y están d-separadas por Z, entonces son condicionalmente independendientes.

Demostración. Esta demostración se va a realizar teniendo en cuenta dos casos:

1. Primero el caso en el cual se da que X∪Y∪Z=V, es decir, la unión es todo el conjunto de variables.

2. Después el caso en el cual se da que X∪Y∪Z⊂V, es decir, la unión está contenida en todo el conjunto de variables pero no es igual.

Veamos cada caso:

1. El conjunto Z se puede descomponer disjutamente como Z=$Z_1 \cup Z_2$, siendo los nodos de Z_1 con padres en X y siendo los nodos de Z_2 con padres en Y.

   Como Z d-separa X e Y, se cumple que:

   $$\forall w \in X \cup Z_1 \rightarrow \pi(w) \in X \cup Z \tag{3.6}$$

   $$\forall v \in Y \cup Z_2 \rightarrow \pi(v) \in Y \cup Z \tag{3.7}$$

   Esto significa que todos los nodos en el primer conjunto, tienen sus padres en el segundo conjunto, v,w son los nodos y $\pi(v), \pi(w)$ sus padres. Esto permite descomponer la distribución conjunta P(X,Y,Z) en dos partes independientes:

   $$P(X,Y,Z) = \prod_{v \in V} P(v|\pi(v)) = \prod_{v \in X \cup Z_1} P(v|\pi(v)) \cdot \prod_{v \in Y \cup Z_2} P(v|\pi(v)) \tag{3.8}$$

   donde los factores del producto ϕ=$\prod_{v \in X \cup Z_1} P(v|\pi(v))$ y ψ= $\prod_{v \in Y \cup Z_2} P(v|\pi(v))$ son funciones que dependen exclusivamente de las variables de los subconjuntos correspondientes, pues solo se ven afectados por los valores que se tomen dentro.

   Entonces la probabilidad condicional la podemos expresar como:

   $$P(X,Y|Z) = \frac{P(X,Y,Z)}{P(Z)} = \frac{P(X,Y,Z)}{\sum_{X,Y} P(X,Y,Z)} = \frac{\phi(X,Z)}{\sum_X \phi(X,Z)} \cdot \frac{\psi(Y,Z)}{\sum_Y \psi(Y,Z)} \tag{3.9}$$

   Habiendo aplicado las definiciones de probabilidad condicionada, probabilidad marginal, la descomposición de la probabillidad conjunta.
   Como $P(X|Z) = \frac{\phi(X,Z)}{\sum_X \phi(X,Z)}$ y$P(Y|Z) = \frac{\psi(Y,Z)}{\sum_Y \psi(Y,Z)}$, hemos demostrado la independencia condicional para este caso.

2. Consideramos el caso en el cual la unión es un subconjunto de nodos del grafo. Por ello, vamos a definir los siguientes conjuntos U=$X \cup Y \cup Z$ y $U^* = U \cup \pi(U)$, y denotamos a G^* al subgrafo de dichos nodos.

Vamos a emplear la siguiente propiedad, añadiendo su demostración para poder aplicar su uso:
Sea Y un nodo sin hijos de G, G^* el subgrafo obtenido tras eliminar dicho nodo y X un subconjunto de nodos de G, se cumple que $P_G(X) = P_{G^*}(X)$

Demostración. Veamos que la distribución marginal del conjunto de nodos X no cambia, aplicando la regla de la cadenay sabiendo que dado que Y no tiene hijos se cumple que $\sum_Y P_G(Y|\pi(Y)) = 1$:

$P_G(X) = \sum_Y P_G(X,Y) = \sum_Y P_G(Y|\pi(Y)) \prod_{v\in X} P_G(v|\pi(v))$
$= \prod_{v\in X} P_G(v|\pi(v)) = P_{G^*}(X)$ □

Entonces podemos aplicar que $P_G(X,Y,Z) = P_{G^*}(X,Y,Z)$, y tenemos que $U \subset V$ y $U^* = V^*$. Podemos dividir U^* en dos conjuntos X^* e Y^* tal que estén d-separados por Z de modo que $Y^* = U^* \setminus (X^* \cup Z)$. De este modo por el primer caso podemos tener la siguiente descomposición dado la d-separación: $P_{G^*}(X^*, Y^*, Z) = \phi(X^*, Z)\psi(Y^*, Z)$, y no solo eso, si no que también podemos tomar $X' = X^* \setminus X$ y $Y' = Y^* \setminus Y$.

Vamos a marginalizar estas variables más lejanas añadidas,
$P_{G^*}(X,Y,Z) = \sum_{X',Y'} \phi(X,X',Z)\psi(Y,Y',Z) = \sum_{X'} \phi(X,X',Z) \sum_{Y'} \psi(Y,Y',Z) = \phi'(X,Z)\psi'(Y,Z)$.

De este modo, podemos usar el mismo argumento que en el primer caso para probar que $P_{G*}(X|Z) = P_{G^*}(Y|Z)$ y añadiendo la propiedad antes expuesta de $P_G(X,Y,Z) = P_{G^*}(X,Y,Z)$, quedaría la demostración completada.

□

Se va a introducir ahora el recíproco del teorema anterior. En conjunto, ambos teoremas aseguran que la d-separación constituye un criterio completo para representar las relaciones de independencia condicional en redes bayesianas, proporcionando así una base formal para razonar gráficamente sobre la estructura de dependencias en modelos probabilísticos.

Teorema 3.4. ***Teorema de la completitud de la d-separación***
Sea G un grafo dirigido acíclico, y sea P el conjunto de todas las distribuciones de probabilidad que factorizan sobre G, entonces en cualquier subconjunto de variables X,Y,Z:

$$\forall P \in P(G), X \perp\!\!\!\perp Y|Z \text{ en } P \rightarrow \text{d-sep}_G(X,Y|Z) \tag{3.10}$$

Este teorema afirma que si en toda distribución que factoriza sobre un grafo G se cumple que X,Y son condionalmente independientes dado Z, entonces X e Y están d-separados por Z, es decir, estará representado estructuralmente en el grafo.

Demostración. Para demostrarlo, usamos el contrarrecíproco:

Si X e Y no están d-separados por Z, entonces existe al menos una distribución que factoriza sobre G tal que $X \not\perp\!\!\!\perp Y|Z$.

Supongamos que existe un camino no bloqueado entre X e Y dado Z:
$X = U_1 - U_2 - ... - -U_k = Y$.

Construimos una distribución de probabilidad tal que las variables del camino estén perfectamente correlacionadas. Esto se puede hacer definiendo:
$P(U_1 = u_1, U_2 = u_2, ..., U_k = u_k) = \delta(u_1 = u_2 = ... = u_k)$, siendo δ la función delta de Kronecker, que caracteriza con el valor 1 si todos los valors son iguales y 0 en caso contrario. Ignoramos el resto de nodos fuera del camino y consideramos sus distribuciones como independientes con respecto a las variables del camino construido.

Como hemos visto, hay un camino activo entre ambas variables X e Y entonces no son condicionalmente independientes dado Z. Por lo tanto, la d-separación no se cumple. □

No obstante, aunque el teorema de completitud garantiza que toda independencia condicional válida se refleja en la estructura del grafo mediante d-separación, en la práctica pueden existir independencias que no se identifiquen directamente a partir de una estructura aprendida. Por esta razón, resulta relevante complementar el análisis estructural con pruebas estadísticas como el test χ^2, que permiten detectar relaciones de independencia que podrían no ser evidentes a simple vista.

En este contexto, contamos con un caso concreto de independencia condicional, reduciendo significativamente las variables involucradas en el modelo, el concepto de cobertura de Markov.

Definición 3.5. La **cobertura de Markov** de un nodo X es el conjunto mínimo de nodos que, una vez conocidos, hacen que X sea condicionalmente independiente de todos los demás nodos en la red. Este conjunto está formado por sus padres, sus hijos y los otros padres de sus hijos.

Este concepto es un caso particular de la d-separación. Esto quiere decir que, si conocemos la cobertura de Markov de un nodo, entonces ese nodo está d-separado del resto de la red que no pertenece a dicho subconjunto. Por lo tanto, la probabilidad condicional de una variable X dada toda la red es igual a la probabilidad condicional de X dada su cobertura de Markov:

$$P(X|\text{cobertura de Markov}) = P(X|\text{nodos de toda la red}) \tag{3.11}$$

La demostración de esta afirmación es sencilla, dado que es el conjunto mínimo de variables d-separadas del resto de la red por este conjunto, entonces son independientes condicionalmente dada la cobertura, lo que impacta sobre la probabilidad condicionada.

Así, se reduce el espacio de búsqueda para inferencia y aprendizaje, mejorando la eficiencia y reduciendo el coste computacional de ambas, porque solo necesitamos los nodos más cercanos.

En este punto, es donde se muestra la verdadera importancia del teorema de Bayes, empleado localmente, ya que nos permite actualizar nuestras creencias sobre el valor de

un nodo a partir de la información disponible en dicho campo de influencia (la cobertura).

Realizando un resumen, la figura 4.2 muestra de forma esquemática el proceso de aprendizaje de un modelo de machine learning, en este caso considerando como familia de modelos a las redes bayesianas. Partiendo de una muestra de datos observados $(X_1, ..., X_n)$ y, a través de un algoritmo de aprendizaje (p.e. Hill-Climbing), se aprenden tanto el grafo, en este caso dirigido y acíclico, como las tablas de probabilidad de la red bayesiana, dando lugar a una distribución de probabilidad conjunta $(P(X_1, ..., X_n))$ con la cual poder hacer inferencia considerando nuevos casos $(P(X_i = x_i|e))$.

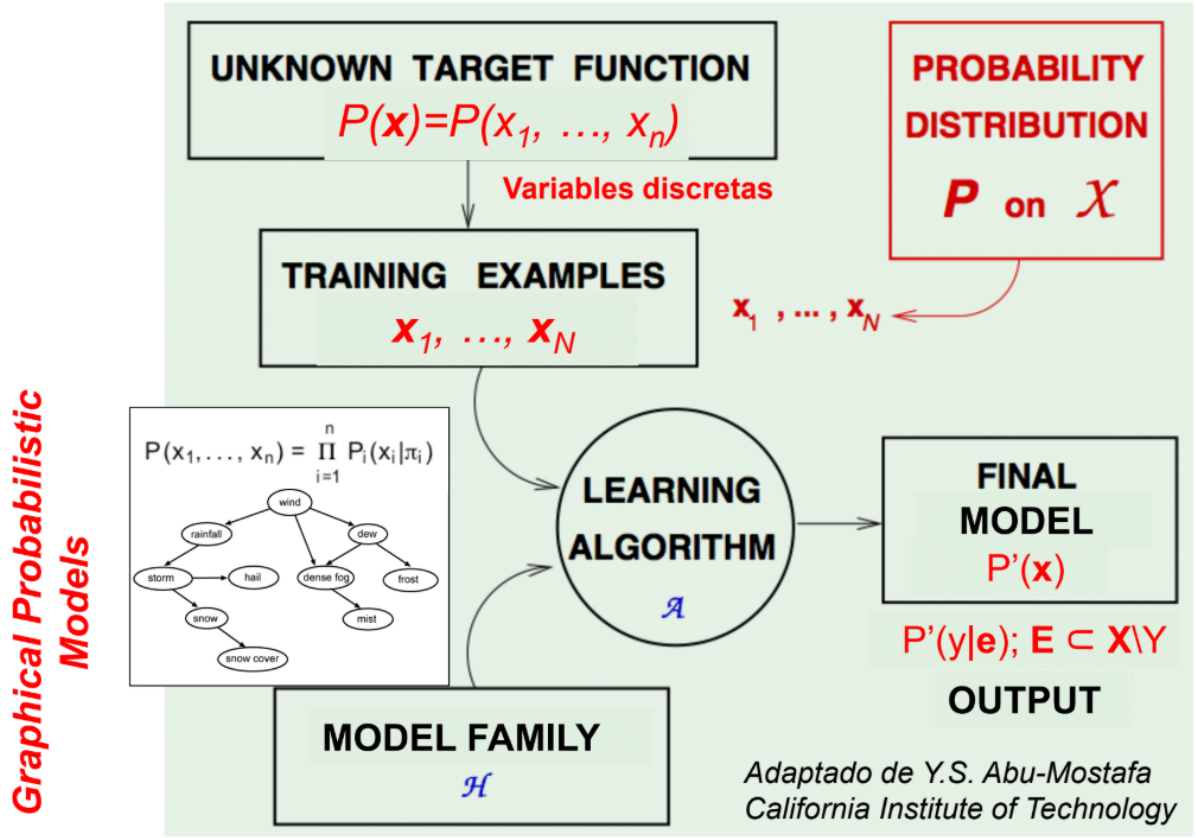


Figura 3.2: Esquema del proceso de aprendizaje considerando como modelo las redes bayesianas.

Una descripción en detalle del proceso esquematizado en la figura 4.2 para el aprendizaje de redes bayesianas, de arriba a abajo:

- Función Objetivo Desconocida.
  Representa la distribución de probabilidad conjunta sobre un conjunto de variables discretas. Esta es la distribución real de la población que queremos aprender o aproximar a partir de un conjunto finito de muestras.

- Muestra de Entrenamiento.
  Son las observaciones que se utilizan para aprender el modelo. Este conjunto finito de datos conforman una muestra tomada de la distribución real.

- Algoritmo de Aprendizaje.
  Es el mecanismo que, a partir de la muestra de entrenamiento y considerando una familia de modelos, estima los parámetros del modelo para que se aproxime a la función objetivo desconocida.

- Familia de Modelos.
  Corresponde a los posibles modelos que el algoritmo puede considerar. En redes bayesianas, cada modelo está definido por una estructura de grafo dirigida acíclica y un conjunto de distribuciones de probabilidad condicional.

- Modelo Probabilístico Gráfico Final.
  Es el modelo aprendido por el algoritmo. Idealmente, este modelo se aproxima bien a la distribución verdadera $P(X_1, ..., X_n)$. Una vez aprendido, éste puede ser usado para realizar inferencias ($P(X_i = x_i|e)$) ante nuevas evidencias (e).

3.2. Aprendizajes

Una vez explicados los conceptos básicos en los que se fundamentan las Redes Bayesianas, como se ha mostrado en la figura 4.2 necesitamos conocer cómo aprender esa estructura gráfica y las tablas de probabilidad condicionada, para poder realizar la inferencia ante eventos predefinidos. Para ello deberemos considerar diferentes algoritmos:

1. En primer lugar, deberemos considerar un algoritmo de **aprendizaje de la estructura gráfica**, que representa las relaciones a partir de los datos. Hay dos tipos principales, cada uno con su propia manera de proceder.

   - **Algoritmos basados en búsqueda y puntuación (Score-based)**
     Este tipo de algoritmos se centran en buscar la estructura que maximiza una función objetivo, llamada score, que mide lo bien que una red explica los datos.

     Dentro de esta categoría se encuentra el algoritmo Hill Climbing. Se trata de un algoritmo heurístico que parte de una red inicial (puede ser vacía) y que realiza modificaciones iterativas en las que va añadiendo, eliminando o revirtiendo aristas. En cada una de estas acciones evalúa el score a optimizar, ya sea, por ejemplo, BIC, BDeu u otros.

     La resolución de este problema se modela de la misma forma que un problema de optimización:

     $$max_{G \in \mathcal{G}} \text{Score}(\mathcal{G}, D)$$

     siendo $\mathcal{G}$ el conjunto de los posibles subgrafos y D el conjunto de datos.

     Este algoritmo es ventajoso por su simplicidad y eficiencia para datasets no demasiado grandes. Sin embargo, podría encontrarse un óptimo local provocando que no encuentre la mejor estructura global. Esto se ve explicado porque la estructura del espacio de búsqueda en el aprendizaje de estructuras de redes bayesianas no es convexa y, por lo tanto, no se puede garantizar que cualquier mínimo local sea global.

- **Algoritmos basados en independencia condicional (Constraint-based)** Buscan aprender la estructura basándose en pruebas estadísticas de independencia condicional entre variables.

  Dentro de este tipo está el algoritmo PC, que se basa en realizar pruebas de independencia condicional entre pares de variables, condicionando en diferentes subconjuntos de otras variables. Realiza los siguientes pasos:

  a) Primero parte de una red completa, es decir hay conexión entre todas las variables. Se trata del esqueleto completo del grafo, que es el grafo no dirigido subyacente.
  b) Segundo, se realizan pruebas de independencia condicional entre pares de variables, condicionando sobre diferentes subconjuntos del resto de variables. Si dos variables resultan ser condicionalmente independientes dado algún conjunto, se elimina la arista que las une. El resultado de este paso es el esqueleto reducido, que contiene solo las relaciones directas necesarias.
  c) Tercero, una vez obtenido el esqueleto, se busca identificar las estructuras en v del el grafo. Posteriormente, se aplican reglas de orientación adicionales para definir la dirección de las aristas, evitando la creación de ciclos dirigidos y asegurando que las relaciones de dependencia se mantengan consistentes con los datos.

  Este método aplica evidencia estadística y puede identificar la estructura causal bajo ciertas condiciones. Sin embargo, es muy sensible a la calidad de las pruebas de independencia y a la cantidad de datos provocando ser computacionalmente costoso.

2. Una vez que se ha aprendido la estructura del grafo de una red bayesiana, el siguiente paso es el aprendizaje paramétrico. Este proceso consiste en estimar los parámetros de las distribuciones de probabilidad condicional asociadas a cada nodo dado su conjunto de padres. A continuación, se presentan dos de los métodos más utilizados en contextos con variables discretas y datos completos.

   - **Estimación por máxima verosimilitud (Maximum Likelihood Estimation)**

     La estimación por máxima verosimilitud consiste en identificar los valores de los parámetros que maximizan la función de verosimilitud. En el contexto de redes bayesianas con variables discretas, esta técnica se basa en el conteo de frecuencias relativas:

     $$P(X = x|\pi(X) = u) = \frac{N(x, u)}{N(u)} \tag{3.12}$$

     siendo el numerador el número de veces de esa combinación de valores, mientras que el denominador es el número total de ocurrencias del valor u de la variable padre.

- **Estimación bayesiana (Bayesian Parameter Estimation)**
  La estimación bayesiana introduce una perspectiva probabilística sobre los parámetros, al tratarlos como variables aleatorias con una distribución a priori que se actualiza mediante la observación de los datos. En redes bayesianas con variables discretas, es habitual utilizar la distribución de Dirichlet:

$$P(X = x|\pi(X) = u) = \frac{N(x,u) + \alpha_{x,u}}{N(u) + \sum_{x'} \alpha_{x',u}} \tag{3.13}$$

  siendo $\alpha_{x,u}$ los hiperparámetros del prior Dirichlet, que pueden interpretarse como conteos ficticios o pseudo-observaciones. Esta técnica tiene la ventaja de evitar asignaciones de probabilidad cero y permite incorporar conocimiento previo en forma de distribuciones a priori.

3. Una vez construida la red bayesiana y halladas las probabilidades condicionadas de cada variable, necesitamos algún modo de evaluar los cambios en la probabilidad de un evento asociado a nuevas evidencias sobre una o varias variables de la red. Para ello, se considerarán algoritmos de **inferencia probabilística**. En este campo encontramos tres aproximaciones:

   - **Inferencia exacta**.
     Se utiliza cuando es posible calcular de manera exacta las probabilidades de interés a partir del modelo probabilístico dado por la red bayesiana.

     Un algoritmo representativo es la eliminación de variables, que consiste en reordenar y marginalizar variables no relevantes para obtener la distribución deseada. Esto se hace mediante operaciones algebraicas sobre las tablas de probabilidad condicional, reduciendo progresivamente el problema hasta obtener la respuesta. Proporciona resultados exactos pero su complejidad crece exponencialmente con el número de variables y relaciones.

   - **Inferencia aproximada**.
     Se utiliza cuando la inferencia exacta es computacionalmente inviable debido a la gran complejidad de la red.

     Un método ampliamente usado es el muestreo de Monte Carlo, en particular el algoritmo Gibbs Sampling [21]. Estos métodos se basan en generar muestras aleatorias de las variables, respetando las dependencias, para posteriormente estimar las probabilidades deseadas a partir de la frecuencia relativa de los eventos observados en las muestras. Proporciona resultados aproximados pero permite operar con redes más grandes y/o más densas.

   - **Inferencia con evidencia parcial**.
     Las redes bayesianas permiten hacer inferencia conociendo solo los valores de un subconjunto de variables. Tanto los métodos exactos como los aproximados pueden aplicarse en estos casos, adaptando el cálculo a las variables observadas y no observadas.

3.3. Análisis del desempeño del modelo

La validación de las redes bayesianas implica revisar cómo está construida su estructura. En particular, es importante asegurarse de que las relaciones de dependencia entre las variables que la red ha aprendido reflejen correctamente la realidad del problema que se está modelando.

Además de revisar la estructura del modelo, es fundamental evaluar lo bien que se ajusta la red bayesiana a los datos reales con los que se está trabajando. Esto significa comprobar si las probabilidades que el modelo asigna a los distintos eventos de variables coinciden con lo que realmente se observa en los datos. Un buen ajuste implica que la red es capaz de representar correctamente las distribuciones probabilísticas presentes en el conjunto de entrenamiento. Si el modelo se ajusta mal, es probable que sus predicciones sean poco confiables, incluso si la estructura es razonable.

Sin embargo, cuando se utiliza o se puede utilizar una red bayesiana para hacer predicciones sobre nuevos datos, es esencial evaluar lo precisas que son esas predicciones. Esto permite saber si el modelo puede generalizar bien más allá de los datos con los que fue entrenado.

Por ello, vamos a realizar un análisis de cada una de las tres categorías:

Evaluación de la estructura
Existen distintas estrategias para validar esta estructura, tanto desde el punto de vista gráfico como estadístico. A continuación, veremos tres enfoques comunes que se pueden utilizar para este propósito:

- Análisis gráfico de la propiedad de la d-separación:
  Es una forma de verificar si dos variables son independientes dado un conjunto de nodos en una red bayesiana, simplemente observando la estructura. Si los datos contradicen esto, la estructura está mal.

- Pruebas sobre la independencia condicional entre las variables:
  Para comprobar si dos variables dependen entre si o no, podemos aplicar unos tests estadísticos de independencia para validar si las relaciones entre nodos se sostienen en los datos. Un ejemplo de estos tests son los test de chi-cuadrado.

- Reentrenamiento y análisis de sensibilidad.
  Este método es más rudimentario, se basa en modificar o eliminar relaciones de la red y ver como afecta al rendimiento, si mejora tanto la verosimilitud como la complejidad por los criterios AIC y BIC al quitar alguna, esto implica que la relación estaba probablemente mal, si al igual mejora al añadir, faltaba una dependencia.

Evaluación del ajuste al modelo

Antes de utilizar la red es esencial evaluar si representa adecuadamente los datos observados. Esta evaluación permite medir el grado de correspondencia entre el modelo y la realidad que intenta capturar. Para ello existen diversas métricas que determinan esta

necesidad, son los scores de estructura antes mencionados, entre ellos se encuentran:

- Log-Likelihood: mide la probabilidad de que el modelo haya generado los datos observados.
$$log(R) = \sum_{i=1}^{n} logP(x_i|x_{\pi(i)}) \tag{3.14}$$
Un valor más alto indica mejor ajuste, es decir, es más probable que los datos provengan del modelo.

- Criterio de información Akaine AIC, se trata de un criterio que penaliza la log-verosimilitud por la complejidad del modelo t el sobreajuste.
$$AIC = -2 \cdot log(R) + 2k \tag{3.15}$$
siendo k el número de parámetros independientes. Cuando se obtiene un menor AIC implica un mejor modelo con un buen balance entre ajuste y simplicidad.

- Criterio de información bayesiano BIC tiene un uso similar al AIC, pero con una penalización más fuerte por la complejidad. También se busca el valor más bajo. BIC favorece modelos más simples que el AIC cuando el tamaño de muestra es grande. Esto se debe a que BIC está en función del número de observaciones n y AIC del número de parámetros k.
$$BIC = -2log(R) + k \cdot log(n) \tag{3.16}$$

- Divergencia de Kullback-Leibler mide la diferencias entre dos distribuciones de probabilidad, por ejemplo, entre la distribución Q estimada por la red bayesiana y la distribución verdadera P. Si las distribuciones coinciden, la divergencia KL será cero, si no, mide esta pérdida de información. En este caso para casos con variables disscretas se expresa:
$$D_{KL}(P||Q) = \sum_{x} P(x)log\frac{P(x)}{Q(x)} \tag{3.17}$$

Evaluación de la calidad predictiva
Si la red bayesiana se usa para hacer predicciones sobre nuevas observaciones, el error de predicción es una medida importante. Éste se puede evaluar de distintas formas.

- Matriz de confusión: Esta matriz proporciona una representación que permite visualizar el desempeño del algoritmo en términos de predicciones correctas e incorrectas, segmentadas por clases específicas. En concreto, se trata de una tabla de contingencia de dimensión mxm siendo m el número de clases posibles. La matriz contiene cuatro elementos:

  1. Verdaderos Positivos (VP): Casos en los que el modelo predice correctamente la clase positiva.
  2. Falsos Positivos (FP): Casos en los que el modelo predice la clase positiva incorrectamente.

3. Verdaderos Negativos (VN): Casos en los que el modelo predice correctamente la clase negativa.
4. Falsos Negativos (FN): Casos en los que el modelo no logra identificar la clase positiva.

A partir de estos elementos, se pueden derivar múltiples métricas de evaluación, entre ellas la precisión, el recall, la exactitud, la medida F1, y también se emplean algunos de sus valores en la curva ROC.

- Curva ROC y AUC: Si el modelo es utilizado para clasificación, la curva ROC y el AUC son medidas comunes para evaluar el desempeño del modelo.

  Esta curva muestra cómo varía la tasa de verdaderos positivos frente a la tasa de falsos positivos cuando se modifican los umbrales de decisión del modelo. Es decir, permite observar el equilibrio entre detectar correctamente los casos positivos y evitar clasificar erróneamente los negativos como positivos. Cuanto más se acerque la curva al vértice superior izquierdo, mejor será su desempeño. Concretamente, está formado por el eje X, que es la tasa de falsos positivos, y el eje Y, la tasa de verdaderos positivos.

  Mientras, asociado a esta curva está el AUC, que se trata del área inferior de la curva ROC. Resume en un solo valor la capacidad del modelo para distinguir entre clases. Un AUC de 1 indica un modelo perfecto, mientras que un valor de 0.5 sugiere un rendimiento equivalente al azar.

- K-fold Cross-Validation: La validación cruzada es una técnica estándar para evaluar, a partir de la muestra de entrenamiento, el error esperado de un modelo para muestras nuevas. Es decir, trata de evaluar la capacidad de generalización del modelo.

  Para este estudio en concreto solo nos interesa la validación con k-fold, que consiste en dividir el conjunto de datos en k subconjuntos (folds), entrenar el modelo en $k-1$ y probarlo en el subconjunto restante. Esto se repite para todos los subconjuntos, obteniendo una predicción para toda la muestra de entrenamiento, y se promedia el error para obtener una evaluación más robusta del rendimiento del modelo ante muestras no consideradas en el proceso de aprendizaje del modelo.

Capítulo 4

Aplicación práctica

En este capítulo, después de haber revisado los fundamentos conceptuales de las redes bayesianas, nos enfocaremos en su aplicación en un contexto real.
El objetivo es aplicar estos modelos probabilísticos para obtener las relaciones de dependencia de la variable objetivo "muerte por ahogamiento" en España. Para ello, se describen los pasos necesarios para construir una red bayesiana funcional, considerando criterios de selección que permiten elegir entre las redes generadas por distintos métodos. Además, analizamos las conclusiones que se pueden obtener a partir de las inferencias realizadas. El capítulo se realiza siguiendo el mecanismo expuesto en el libro [22] y realizando la programación en este Jupyter Notebook.

4.1. Interés del estudio

El ahogamiento en entornos acuáticos representa una de las principales causas de muerte no intencional a nivel mundial, y en España constituye un problema de salud pública que afecta especialmente a zonas costeras y turísticas. La prevención y respuesta ante estos incidentes no sólo requiere una intervención rápida, sino también una comprensión profunda de los factores que influyen en su ocurrencia y gravedad. En este contexto, los datos recopilados por organizaciones de socorrismo desde 2013 a 2025 ofrecen una oportunidad única para identificar patrones y perfiles de riesgo asociados a los ahogamientos, tanto mortales como no mortales.

La presente investigación se basa en el análisis de una base de datos proporcionada por la Escuela Segoviana de Socorrismo [23], que recopila información sobre ahogamientos en España a lo largo de más de diez años. Este dataset incluye multitud de variables que comprenden diversos ámbitos, relacionados con la información personal como la edad, el género, y la nacionalidad; relacionados con las circunstancias del entorno durante la ocurrencia del accidente como la actividad realizada, la presencia de vigilancia; o relacionadas con la referencia espacio-temporal como el dia y la hora del incidente o el entorno geográfico. Esto permite un enfoque multidimensional del fenómeno. Por ello, para modelar las relaciones de dependencia entre estas variables y entender mejor los factores de riesgo que influyen en la probabilidad de fallecimiento, se propone el uso de Redes Bayesianas.

Su aplicación en este estudio ofrece varias ventajas, permite explorar relaciones causales

entre los distintos factores y realizar inferencias predictivas útiles para la toma de decisiones. A diferencia de otros modelos estadísticos tradicionales, las Redes Bayesianas no solo predicen, sino que también explican cómo influyen las variables de entorno a un evento de interés, en este caso, la muerte por ahogamiento. Este enfoque puede proporcionar a las autoridades competentes una herramienta poderosa para identificar perfiles de riesgo, optimizar los recursos de vigilancia y mejorar las campañas de prevención, con el propósito de reducir este tipo de muertes mediante un análisis riguroso y probabilísticamente fundamentado. Así, las Redes Bayesianas no se limitan a ofrecer un resultado numérico, sino que promueven una sencilla interpretabilidad del problema.

4.1.1. Datos del estudio

En la Figura 4.1 se muestra el modelo relacional de la base de datos utilizada para el desarrollo del presente estudio. La base de datos fue proporcionada en ficheros de Microsoft Access. La comprensión de este modelo relacional resulta esencial, ya que sobre esta estructura se construyen los procesos de extracción, transformación y carga de datos, que posteriormente alimentan la Red Bayesiana implementada. Asimismo, dicho modelo sirve como punto de partida para identificar las variables relevantes y establecer dependencias lógicas entre ellas.

El modelo relacional está compuesto por un conjunto de tablas organizadas de manera jerárquica, entre las cuales se destacan dos tablas principales, denominadas tablas de hechos, y un conjunto de dieciocho tablas auxiliares, conocidas como tablas maestras. Las tablas maestras contienen información complementaria sobre diversas características clave que son utilizadas de forma repetitiva en las tablas de hechos. Esta estructura normalizada permite minimizar la redundancia de datos, mejorar la consistencia de la base de datos y facilitar la integridad referencial a través del uso de claves primarias y foráneas que definen las relaciones entre las distintas entidades.

La primera tabla de hechos, denominada *ListaAhogamientos*, recoge información detallada de cada caso de ahogamiento registrado en el sistema. Cada registro cuenta con una clave primaria única que identifica el evento, y con múltiples atributos asociados, tales como la fecha y hora del suceso, la localidad donde tuvo lugar, el tipo de vigilancia presente en la zona, la ubicación exacta del incidente, la actividad que se realizaba en ese momento, la persona u organismo que detectó el incidente, quién intervino inicialmente, el tipo de entorno físico donde ocurrió y la existencia de factores de riesgo ambientales identificados.

La segunda tabla de hechos, *PersonasAhogadas*, proporciona información centrada en las personas directamente involucradas en los eventos de ahogamiento. Cada registro posee una clave primaria individual que identifica a la persona afectada, así como una clave foránea que permite vincular ese registro con el evento correspondiente en la tabla *ListaAhogamientos*. Esta tabla incluye, además, datos como el género y la edad de la persona, su nacionalidad, la causa específica del suceso, la presencia de factores de riesgo personales como enfermedades previas, quién fue el primer interviniente, la persona que la rescató del agua, el resultado vital del suceso, es decir, si sobrevivió o falleció, y si se realizaron maniobras de reanimación.

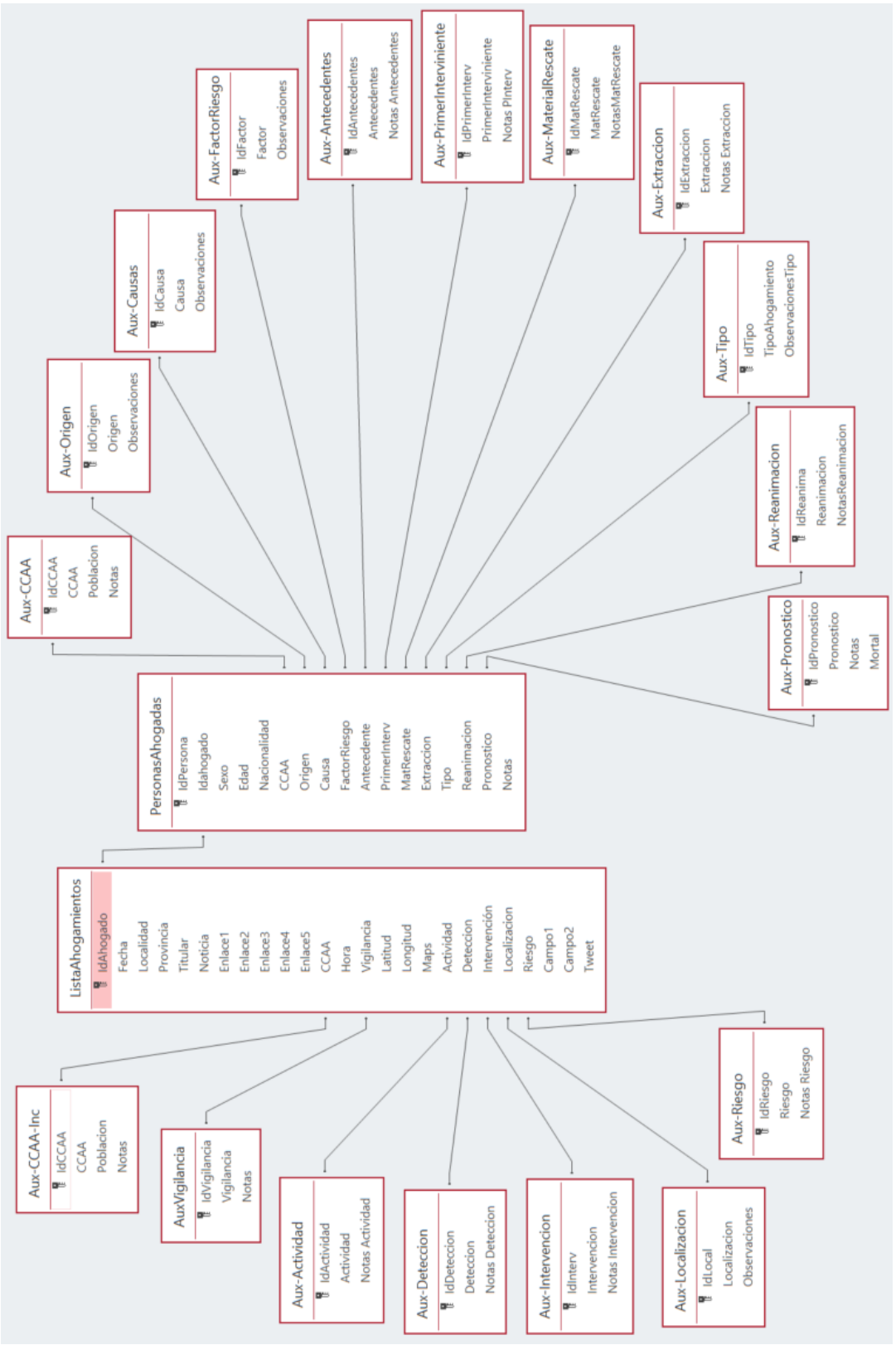


Figura 4.1: Estructura relacional de la base de datos

4.2. Introducción a la metodología

La construcción y análisis de una red bayesiana requiere una metodología estructurada. Para alcanzar estos objetivos, se ha empleado la metodología extraída del libro Análisis de Indicadores de Rendimiento Mediante Redes Bayesianas [22], que consta de las siguientes fases:

- **Recolección y estudio de los datos:**
  Primero, es necesario hacer un estudio de la base de datos para selecionar los criterios que van a ser importantes. Esos criterios serán tomados como nuestras variables de interés y las que estudiaremos si aplican una relación de dependencia con el objetivo final del estudio.

- **Limpieza y discretización de los datos:**
  Una vez tengamos las variables a estudiar seleccionadas, será necesario limpiarlas y posteriormente habrá que hacer una clasificación por su tipo de naturaleza, pudiendo ser cualitativas, cuantitativas discretas o cuantitativas continuas.

  Para agrupar las variables continuas se pueden emplear varios mecanismos. Uno de los métodos consiste en dividir los datos en intervalos de igual frecuencia, es decir, en grupos que contengan aproximadamente el mismo número de observaciones. Este enfoque busca distribuir los datos de manera equilibrada entre las categorías. Otro método utiliza el algoritmo de clustering K-medias, una técnica que agrupa los datos según su similitud, tratando de minimizar la variabilidad dentro de cada grupo. A diferencia del método anterior, aquí los intervalos no se definen por la cantidad de datos, sino por la cercanía entre ellos en función de su valor. Debe notarse que, al utilizar K-medias, es necesario definir previamente el número de grupos k, o bien probar con diferentes valores hasta encontrar el que mejor se ajuste a la estructura de los datos. Por último, con otro enfoque también se puede realizar un agrupamiento de los datos en función de la experiencia del analista, es decir, considerando los rangos que le interesan.

  Comparar los tipos de discretización permite evaluar cómo influyen estas transformaciones en la construcción y el rendimiento de la red bayesiana, y así elegir la opción que represente mejor la estructura real de los datos.

- **Aprendizaje de la estructura, de los parámetros y elección de la red que mejor se ajuste a los datos:**
  Para esta fase del estudio, despues de haber seleccionado nuestras variables, se pueden emplear dos enfoques para la construcción de la red bayesiana, uno manual y uno automatizado por algoritmos. Ambos enfoques se pueden implementar utilizando la librería pgmpy [24], especializada en modelos gráficos probabilísticos en Python.

  Una vez que se tiene la estructura de la red bayesiana dada por su grafo, hay que estimar las tablas de probabilidad de cada nodo dados sus padres. Esto se hace mediante las tablas de probabilidad condicional CPT, que nos permitirán posteriormente realizar inferencias a través de la red bayesiana construida.

  De cara a establecer la confianza de la red construida para estimar los cambios en la probabilidad de ahogamiento en base a las variables explicativas consideradas en

la red, consideramos un proceso inicial previo de validación cruzada. Esta técnica permite evaluar la capacidad de predicción del modelo al estimar su rendimiento sobre subconjuntos no vistos durante el entrenamiento. El objetivo es asegurarse de que el modelo obtenido tenga capacidad de generalización, es decir, que su poder predictivo se mantenga cuando se aplica a muestras nuevas que no fueron utilizadas en el proceso de entrenamiento. Si bien hay diferentes aproximaciones típicamente consideradas en la literatura, dado el volumen de la base de datos considerada, en este trabajo consideraremos un 15-fold, el cual divide la muestra en 15 subconjuntos disjuntos, de modo que estima la submuestra correspondiente a un fold a partir del modelo entrenado con los otros 14. Este proceso se repite para los 15 folds obteniendo una predicción de toda la muestra, obtenida uniendo las predicciones de los 15 folds. Al realizar este proceso, en la salida por consola se muestran los parámetros de evaluación obtenidos en los distintos folds y en el modelo completo. Esto permite comprobar si los modelos aprendidos para cada fold son relativamente homogéneos o alguno se desvía significativamente.

Para evaluar los distintos modelos, obtenidos considerando diferentes aproximaciones para obtener el grafo de la red, es necesario determinar cuál de ellos se ajusta mejor a los datos. Para ello, se emplean métricas de evaluación basadas en la verosimilitud, que permiten cuantificar cuán bien una estructura dada explica los datos.

Como ya hemos comentado en el capítulo anterior dentro de estas métricas se encuentran el log-likelihood, que mide la probabilidad de observar los datos dados los parámetros del modelo y su estructura, se busca que este valor sea lo mayor posible. No obstante, el uso exclusivo del log-likelihood puede favorecer modelos más complejos con un número elevado de parámetros, lo que puede conducir a un sobreajuste. Para contrarrestar este problema, se recurre a medidas de evaluación que incluyen un término de penalización por complejidad (regularización), como los criterios AIC y BIC. Ambas métricas combinan el valor del log-likelihood con un término de penalización que se incrementa con el número de parámetros del modelo. En general, BIC aplica una penalización más fuerte que AIC, por lo que tiende a seleccionar modelos más simples.

- **Propagación de probabilidades:**
  Con esta técnica, podremos estudiar los efectos que tendrán las posibles líneas de actuación sobre las variables. Consiste en realizar distintas inferencias con el modelo entrenado para estimar probabilidades condicionales de la variable objetivo según diferentes combinaciones de sucesos. Además, puede evaluar la precisión del modelo comparando las predicciones probabilísticas con las proporciones reales en los datos de prueba.

- **Propuestas de estrategias:** Tras realizar todos los pasos, analizar los resultados y examinar los posibles errores o limitaciones del modelo, es fundamental plantear unas conclusiones del estudio. Estas deben reflejar no solo la eficacia del enfoque utilizado, sino también los hallazgos principales derivados del análisis.

4.3. Aplicación de la metodología

Vamos a explorar en detalle cada uno de los pasos de este procedimiento:

4.3.1. Estudio

Recolección y estudio de los datos

En esta primera fase, se ha realizado un análisis de la base, motivados por la necesidad de los proveedores de los datos. Considerando los criterios que eran de su interés se han destacado las variables relevantes para el estudio.

A partir de la tabla *PersonasAhogadas*, se han seleccionado las variables correspondientes al género, la edad, el origen y el pronóstico del suceso. Estas variables han sido clasificadas con los siguientes valores: en el caso del género, se consideraron las categorías "hombre" y "mujer"; para la edad, se establecieron los grupos "menor de edad", "adulto" y "anciano"; respecto al origen, se distinguió entre "español" y "extranjero"; y para el pronóstico, se diferenció si el incidente resultó en un "accidente mortal" o no.

Adicionalmente, se han incorporado al estudio variables extraídas de la tabla *ListaAhogamientos*, específicamente aquellas relacionadas con el contexto del suceso, como la presencia de vigilancia y el tipo de localización. En cuanto a la vigilancia, se tenido en cuenta si en el momento del incidente había o no personal cualificado presente. La variable localización, por su parte, distingue el entorno físico en el que ocurrió el ahogamiento, categorizado como "playa", "piscina", "río" u "otros".

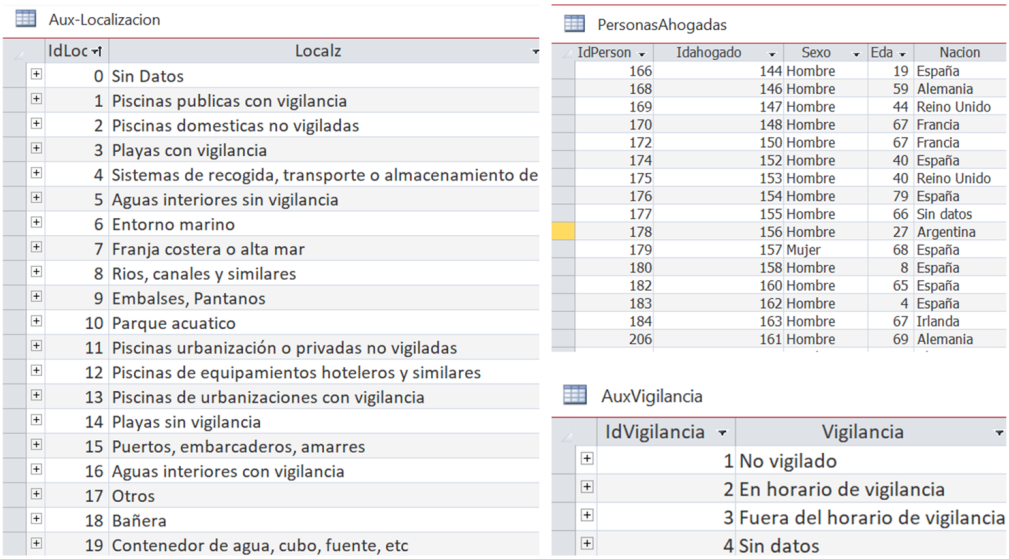
Para llevar a cabo la extracción de los datos, se ha establecido una conexión directa con la base de datos Microsoft Access utilizando el entorno de desarrollo Visual Studio Code. A través de esta conexión, se ha extraído la información contenida en las dos tablas de hechos previamente descritas. Los datos han sido organizados en dos estructuras tipo DataFrame, que es la estructura básica de datos considerada por la librería Pandas de Python, las cuales se han unido posteriormente en un solo DataFrame, a través de la clave IdAhogado, para tener todos los datos agrupados.

Limpieza y discretización de los datos

Una vez seleccionadas las variables para el estudio y con los datos ya estructurados en dataframes, se ha procedido a realizar una depuración inicial de los mismos. Esta fase consiste en la generación de nuevos dataframes que contienen exclusivamente las columnas correspondientes a las variables de interés expuestas en la fase anterior.

A continuación, se ha hecho un tratamiento de los datos, dividido en dos etapas principales. En primer lugar, una limpieza de los registros, eliminando todas aquellas filas que presentaban valores nulos en alguna de las columnas seleccionadas, así como aquellas que contenían la categoría "Sin datos", por considerarse incompletas para el análisis. En segundo lugar, se ha hecho una transformación de las variables mediante un proceso de discretización, consistente en la aplicación de una serie de mapeos. Esta transformación

ha permitido convertir variables continuas o categóricas con múltiples niveles en variables discretas con categorías homogéneas, facilitando así su integración en el modelo de red bayesiana y mejorando la interpretabilidad de los resultados. Los mapeos establecidos son los siguientes:



Aux-Localizacion

IdLoc	Localz
0	Sin Datos
1	Piscinas publicas con vigilancia
2	Piscinas domesticas no vigiladas
3	Playas con vigilancia
4	Sistemas de recogida, transporte o almacenamiento de
5	Aguas interiores sin vigilancia
6	Entorno marino
7	Franja costera o alta mar
8	Rios, canales y similares
9	Embalses, Pantanos
10	Parque acuatico
11	Piscinas urbanización o privadas no vigiladas
12	Piscinas de equipamientos hoteleros y similares
13	Piscinas de urbanizaciones con vigilancia
14	Playas sin vigilancia
15	Puertos, embarcaderos, amarres
16	Aguas interiores con vigilancia
17	Otros
18	Bañera
19	Contenedor de agua, cubo, fuente, etc

PersonasAhogadas

IdPerson	Idahogado	Sexo	Eda	Nacion
166	144	Hombre	19	España
168	146	Hombre	59	Alemania
169	147	Hombre	44	Reino Unido
170	148	Hombre	67	Francia
172	150	Hombre	67	Francia
174	152	Hombre	40	España
175	153	Hombre	40	Reino Unido
176	154	Hombre	79	España
177	155	Hombre	66	Sin datos
178	156	Hombre	27	Argentina
179	157	Mujer	68	España
180	158	Hombre	8	España
182	160	Hombre	65	España
183	162	Hombre	4	España
184	163	Hombre	67	Irlanda
206	161	Hombre	69	Alemania

AuxVigilancia

IdVigilancia	Vigilancia
1	No vigilado
2	En horario de vigilancia
3	Fuera del horario de vigilancia
4	Sin datos

Figura 4.2: Variables importantes

En el caso de la variable *Nacionalidad*, tal como se muestra en la Figura 4.2 , se observan valores asociados a distintos países como España, Alemania, Reino Unido o Argentina, entre numerosos más. No obstante, para los fines del estudio, esta variable ha sido recodificada en dos únicas categorías: "español" y "extranjero". Además, se ha aplicado un algoritmo de autocorrección basado en técnicas de agrupamiento que permitió detectar y corregir errores tipográficos frecuentes, como por ejemplo la transformación de "Sepaña" en "España". En esta variable, sin embargo, no se identificaron errores significativos tras dicha revisión.

En cuanto a la variable *Edad*, se ha procedido a su discretización en tres rangos definidos de forma cualitativa, sin tener en cuenta la frecuencia empírica de los datos. Estas categorías fueron: "menor de edad" incluyendo personas de hasta 17 años, "adulto" con un rango de 18 a 64 años y "anciano" los valores de 65 años o más. Esta clasificación permite una interpretación más clara en el contexto del análisis.

Respecto a la variable *Vigilancia*, se ha establecido una recodificación en función de los valores disponibles en la base de datos, como se muestra en la Figura 4.2. Se consideró "vigilado" todo suceso cuyo identificador se correspondía con el código 2, mientras que los códigos 1 y 3 se agruparon bajo la categoría "no vigilado".

La variable *Localización* ha requerido un tratamiento más detallado debido a la variedad de escenarios representados, visibles en la Figura 4.2. A partir de los valores presentes en la base de datos, se definieron cuatro categorías: "playa" (códigos 3, 6, 7 y 14), "piscina"

(códigos 1, 2, 11, 12 y 13), “ríos y pantanos” (códigos 5, 8, 9 y 16) y “otros” (agrupando el resto de códigos). Esta recodificación tiene como objetivo reflejar las características del entorno físico en el que ocurrieron los incidentes.

Finalmente, para la variable *Pronóstico*, se consideró únicamente si el suceso tuvo un desenlace mortal o no. Por tanto, se utilizó exclusivamente la columna “Mortal” de la tabla expuesta en la Figura 4.3, recodificándola como una variable dicotómica que distingue entre desenlace fatal y no fatal.

Aux-Pronostico

IdPron	Pron	Notas	Mortal
0	Sin datos	No data	☐
1	Rescate sin consecuencias	Rescue without consecuences	☐
2	Rescate y traslado a supervisión médica	Rescue and transport to medical supervision	☐
3	Rescate con necesidad de observación posterior	Rescue in need of further observation	☐
4	Ahogamiento no mortal con rescate y tratamiento in	Non-fatal drowning rescue and treatment on	☐
5	Ahogamiento no mortal sin secuelas	Non-fatal drowning without sequelae	☐
6	Ahogamiento no mortal con secuelas	Non-fatal drowning with sequelae	☐
7	Ahogamiento mortal	Fatal Drowning	☑
8	Ahogamiento no mortal con hospitalizacion	Non-fatal drowning and hospital care	☐
9	Traumatismo/TRM	Trauma/Spinal injury	☐
10	Lesiones	Injury	☐
11	Fallecimiento Sin Ahogamiento	Non-Drowning related death	☑

Figura 4.3: Maestro pronóstico

Aprendizaje de la estructura, de los parámetros y elección de la red

A continuación, se muestran dos imágenes, correspondientes a los grafos obtenidos considerando los algoritmos automáticos Hill-Climbing (Fig. 4.4) y PC (Peter-Clark) (Fig. 4.5).

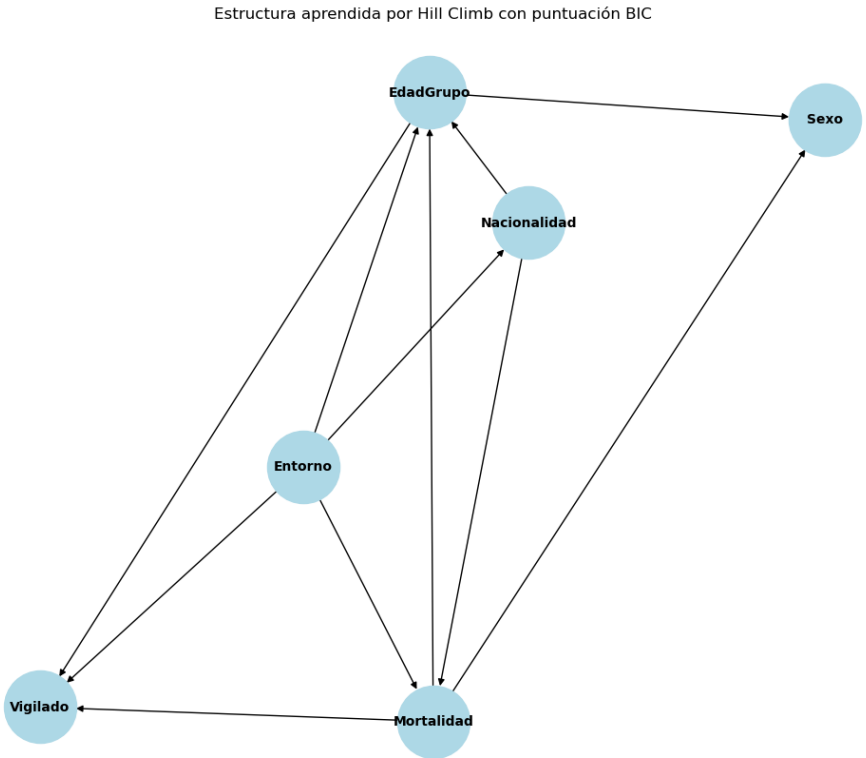


Figura 4.4: Red Hill-Climbing

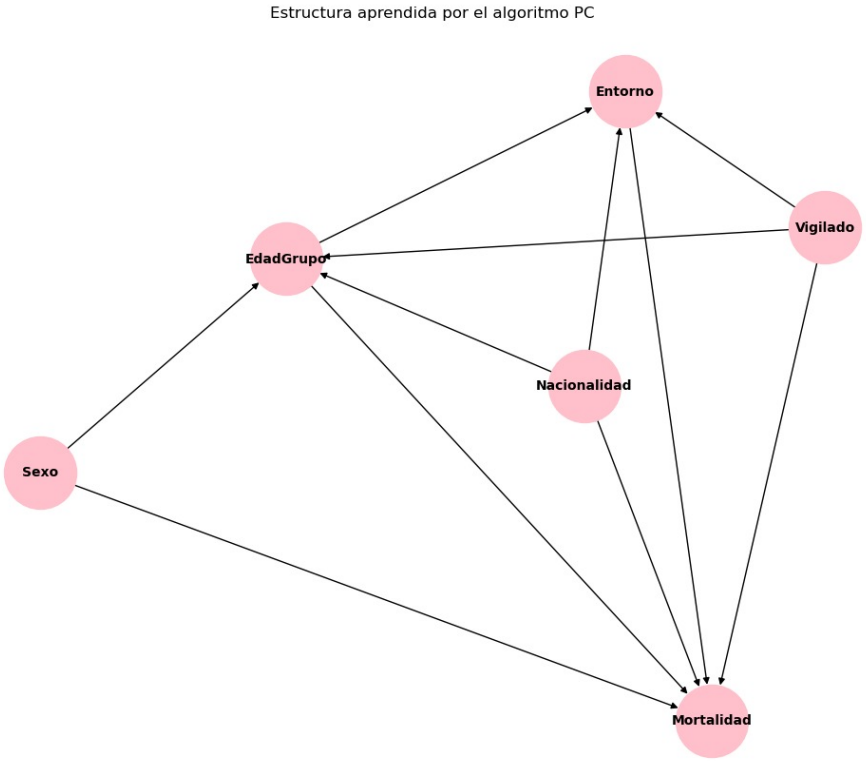


Figura 4.5: Red PC

Una vez completado el proceso de preprocesamiento y transformación de los datos, se ha procedido a la creación de la estructura de la red bayesiana del experimento. Para esta fase del estudio se han empleado dos estrategias.

Por un lado, se ha hecho una construcción manual de la estructura de la red, fundamentada en el conocimiento experto y la intuición del analista respecto a las posibles relaciones entre las variables.

Por otro lado, se han aplicado métodos automáticos de aprendizaje para obtener la arquitectura de red más adecuada. En particular, se han utilizado los algoritmos PC, basado en la detección de independencias condicionales, y Hill-Climbing, basado en la optimización de la bondad de ajuste, junto con la puntuación BIC como criterio para penalizar la complejidad.

Cada uno de estos enfoques ha generado una estructura de red distinta, las cuales deben ser evaluadas para la elección de la que mejor se ajuste a los datos observados. En esta sección se ha descartado el análisis de la red construida manualmente, quedándonos con las dos aproximaciones automáticas mostradas anteriormente en las Figuras 4.4 y 4.5, que se analizarán a continuación.

En primer lugar, se ha procedido al cálculo del número de parámetros de las redes bayesianas obtenidas, dado que este valor es imprescindible para la estimación de los criterios de información AIC y BIC. Consideramos que todas las variables del modelo son categóri-

cas con dos valores, a excepción de la variable *EdadGrupo*, que presenta tres, y *Entorno*, que posee cuatro, por ello el número teórico de parámetros sería: $(2 \cdot 2 \cdot 2 \cdot 2 \cdot 3 \cdot 4 - 1) = 191$.

A partir del grafo mostrado en la Figura 4.4 se identifican las siguientes relaciones de parentalidad entre las variables:

- Padres de EdadGrupo: Nacionalidad, Mortalidad, Entorno.
- Padres de Mortalidad: Nacionalidad, Entorno.
- Padres de Nacionalidad: Entorno.
- Padres de Vigilado: EdadGrupo, Mortalidad, Entorno.
- Padres de Sexo: EdadGrupo, Mortalidad.
- Entorno no tiene padres.

Realizando los cálculos necesarios con base en esta estructura los parámetros requeridos para cada variable son:

- EdadGrupo: 32 parámetros.
  $(k_i - 1) \prod_{X_i \in X_{\pi(i)}} k_j = (3 - 1)(2 \cdot 2 \cdot 4) = 32$
  Esto se debe a que EdadGrupo tiene k_i=3 valores, mientras que el número de valores de sus padres son respectivamente 2, 2 y 4, por ser Nacionalidad, Mortalidad y Entorno. Esto proceso se repite para el resto de variables.
- Mortalidad: 8 parámetros.
  $(k_i - 1) \prod_{X_i \in X_{\pi(i)}} k_j = (2 - 1)(2 \cdot 4) = 8$
- Nacionalidad: 4 parámetros.
  $(k_i - 1) \prod_{X_i \in X_{\pi(i)}} k_j = (2 - 1)(4) = 4$
- Vigilado: 24 parámetros.
  $(k_i - 1) \prod_{X_i \in X_{\pi(i)}} k_j = (3 - 1)(3 \cdot 2 \cdot 4) = 24$
- Sexo: 6 parámetros.
  $(k_i - 1) \prod_{X_i \in X_{\pi(i)}} k_j = (2 - 1)(2 \cdot 3) = 6$
- Entorno: 3 parámetros.
  $(k_i - 1) \prod_{X_i \in X_{\pi(i)}} k_j = (4 - 1)(1) = 3$

Por tanto, la dimensión de esta primera red es de 77 parámetros, resultado de sumar los parámetros de cada una de las variables.

Por otra parte, a partir del grafo obtenido con el algoritmo PC, mostrado en la Figura 4.5, las relaciones de parentalidad se especifican como sigue:

- Padres de EdadGrupo: Vigilado, Sexo, Nacionalidad.

- Padres de Entorno: Nacionalidad, EdadGrupo, Vigilado.
- Padres de Mortalidad: Nacionalidad, EdadGrupo, Vigilado, Sexo, Entorno.
- Sexo, Nacionalidad y Vigilado no tienen padres.

Realizando los cálculos necesarios con base en esta estructura, el número de parámetros requeridos para cada variable son:

- EdadGrupo: 16 parámetros
  $(k_i - 1)\prod_{X_i \in X_{\pi(i)}} k_j = (3-1)(2 \cdot 2 \cdot 2) = 16$
- Entorno: 36 parámetros
  $(k_i - 1)\prod_{X_i \in X_{\pi(i)}} k_j = (4-1)(2 \cdot 3 \cdot 2) = 36$
- Mortalidad: 96 parámetro
  $(k_i - 1)\prod_{X_i \in X_{\pi(i)}} k_j = (2-1)(2 \cdot 2 \cdot 2 \cdot 3 \cdot 4) = 3$
- Sexo: 1 parámetros
  $(k_i - 1)\prod_{X_i \in X_{\pi(i)}} k_j = (2-1)(1) = 2$
- Vigilado: 1 parámetro
  $(k_i - 1)\prod_{X_i \in X_{\pi(i)}} k_j = (2-1)(1) = 1$
- Nacionalidad: 1 parámetros
  $(k_i - 1)\prod_{X_i \in X_{\pi(i)}} k_j = (2-1)(1) = 1$

Por tanto, la segunda red requiere un total de 151 parámetros, casi el doble que el grafo obtenido con el algoritmo Hill-Climbing.

Tras analizar el número de parámetros de ambas redes, podemos considerar que la segunda tiene un incremento muy alto de parámetros con respecto al número total, en proporción de 77/191 a 151/191 es en torno al doble de parámetros.

Una vez se ha aprendido la estructura de la red bayesiana, podemos realizar el aprendizaje paramétrico. En paralelo, para evaluar el modelo obtenido y su capacidad de generalización, se ha replicado el proceso de entrenamiento en un entorno de validación cruzada.

Hemos obtenido las tablas de probabilidad condicional de ambas redes. Para el algoritmo Hill-Climbing es una tabla más reducida, esto está estrechamente unido a que la variable mortalidad tiene menos padres en esta red. La tabla del algoritmo PC al ser más extensa se expone un trozo, correspondiente al grupo de edad "menor".

Entorno	**Nacionalidad**	**Mortalidad NO**	**Mortalidad SÍ**
Otro	España	0.3913	0.6087
	Extranjero	0.2121	0.7879
Piscina	España	0.5095	0.4905
	Extranjero	0.3581	0.6419
Playa	España	0.5820	0.4180
	Extranjero	0.4297	0.5703
Ríos y pantanos	España	0.3228	0.6772
	Extranjero	0.1669	0.8331

Tabla 4.1: CPD correspondiente a la red Hill-Climbing

Edad	**Entorno**	**Origen**	**Sexo**	**Vigilado**	**NO**	**SÍ**
Menor	Otro	España	Hombre	NO	0.5556	0.4444
				SI	0.7500	0.2500
			Mujer	NO	0.4444	0.5556
				SI	1.0000	0.0000
		Extranjero	Hombre	NO	0.4000	0.6000
				SI	0.5000	0.5000
			Mujer	NO	0.0000	1.0000
				SI	1.0000	0.0000
	Piscina	España	Hombre	NO	0.4500	0.5500
				SI	0.7917	0.2083
			Mujer	NO	0.6667	0.3333
				SI	0.7609	0.2391
		Extranjero	Hombre	NO	0.2632	0.7368
				SI	0.5600	0.4400
			Mujer	NO	0.2500	0.7500
				SI	0.7692	0.2308
	Playa	España	Hombre	NO	0.9294	0.0706
				SI	0.9366	0.0634
			Mujer	NO	0.9352	0.0648
				SI	0.9913	0.0087
		Extranjero	Hombre	NO	0.7500	0.2500
				SI	0.8571	0.1429
			Mujer	NO	0.7692	0.2308
				SI	0.8889	0.1111
	Ríos y pantanos	España	Hombre	NO	0.7434	0.2566
				SI	0.5000	0.5000
			Mujer	NO	0.6905	0.3095
				SI	0.5000	0.5000
		Extranjero	Hombre	NO	0.0952	0.9048
				SI	0.5000	0.5000
			Mujer	NO	0.2857	0.7143
				SI	1.0000	0.0000

Tabla 4.2: Fragmento CPD correspondiente a la red PC para menores.

Asimismo, necesitamos los valores de los log-likelihood obtenidos mediante la realización del entrenamiento por validación cruzada realizada para 15 subconjuntos de entrenamiento. Para el primer modelo, se ha obtenido el promedio de $-2652{,}17$, mientras que para el segundo el valor ha sido de $-2863{,}80$. Teniendo en cuenta la interpretación de este dato, que mide la verosimilitud del modelo, cuanto mayor sean los valores (menos negativos) indicarán un mejor ajuste. Por lo tanto, es conveniente evaluarlos para analizar las diferencias, sobre todo viendo que un modelo tiene el doble de parámetros que el otro. La diferencia observada de aproximadamente 200 unidades sugiere que el primer modelo ofrece un ajuste mejor a los datos, sin embargo, se queda en una diferencia moderada y habrá que estudiar otros criterios que consideran la complejidad del modelo en la evaluación, como los criterios AIC y BIC para la mejor selección.

Después de la obtención de todos los datos necesarios, hemos podido estimar los valores AIC y BIC para ambas estructuras, siendo para la primera los valores AIC: 5458,35 y BIC: 6005,27, mientras que para la segunda han sido AIC: 6029,60 y BIC: 7102,13.
En este caso, obtenemos una diferencia más acentuada entre ambas propuestas, en la cuál se sugiere que la primera ofrece un mejor equilibrio entre el ajuste a los datos y la complejidad.

Dado que el objetivo principal de este análisis es modelar adecuadamente la variable Mortalidad, resulta necesario priorizar la estructura de red bayesiana que evidencie relaciones significativas con dicha variable. En este sentido, comparando ambas estructuras obtenidas, se ha observado que para la primer red, la variable mortalidad está relacionada mediante padres con entorno y nacionalidad, mientras que sus hijos son género, vigilancia, y edad. Por otro lado, la segunda red, está relacionada por padres con todas las variables y no tiene ningún hijo. Teniendo en consideración la manta de Markov, que, recordando, incluye los padres del nodo, los hijos y los copadres de sus hijos, podemos afirmar que en ambas redes nuestra variable objetivo tiene como manta de Markov todas las otras variables del estudio y no queda ninguna fuera. Este hecho implica que la mortalidad está relacionada con todas las variables de nuestro estudio para ambas estructuras.

Después de estos análisis, viendo que ambas redes dan buenas relaciones, y que la primera obtiene unos valores mejores para las métricas de evaluación de la estructura, nos quedaremos con esa, la obtenida con el algoritmo Hill-Climbing con puntuación BIC.

Como parte adicional del estudio, para comprobar las dependencias obtenidas en los modelos aprendidos se han evaluado a través del test de independencia χ^2.
Para ello, se establece como hipótesis nula que las variables son independientes y como hipótesis alternativa que son dependientes. De este modo, si se obtiene un p-valor menor que 0,05, correspondiente con un nivel de significación del 95 %, rechazaremos la hipótesis nula, concluyendo que las variables son dependientes y ratificando los resultados dados por la estructura de l

En concreto hemos obtenido los siguientes resultados mostrados en la Figura 4.6:

	Variable	p-valor	¿Rechaza H_0?	¿Dependiente de Mortalidad?
0	Entorno	9.169682e-77	Sí	Sí
1	EdadGrupo	7.781449e-197	Sí	Sí
2	Vigilado	4.993633e-59	Sí	Sí
3	Sexo	3.587070e-19	Sí	Sí
4	Nacionalidad	3.702335e-22	Sí	Sí

Figura 4.6: Test independencia sobre Mortalidad

Observamos que para todas las variables el p-valor es menor que 0,05. Por ello, rechazamos que sean independientes con respecto a la mortalidad. Queda comprobado que sí existe la relación expuesta en la red.

4.3.2. Resultados y conclusiones

La Figura 4.10 muestra una comparación entre los dos modelos estudiados mediante sus respectivas curvas ROC, centrándose especialmente en el valor del AUC. El modelo Hill-Climbing alcanza un AUC de 0,7416, mientras que el modelo PC obtiene un AUC de 0,7392. Aunque la diferencia es pequeña, el valor más alto del AUC en el modelo Hill-Climbing indica que distingue mejor entre clases positivas y negativas. Ambos valores, sin embargo, se sitúan en un rango moderado (entre 0,7 y 0,8), lo que sugiere un desempeño razonablemente bueno pero aún con margen de mejora. Esta similitud en los resultados también indica que ambos enfoques tienen un comportamiento parecido en términos de precisión y sensibilidad, aunque siendo mejor nuestro algoritmo escogido.

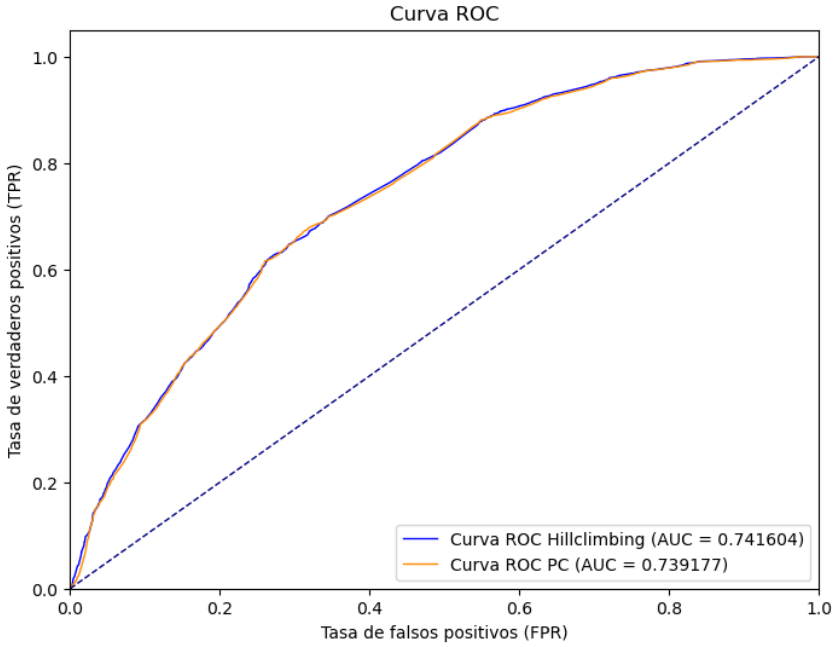


Figura 4.7: Curvas ROC

Veamos en las Tablas 4.3 y 4.4 las matrices de confusión para realizar una comparativa en profundidad sobre la capacidad similar de predicción.

	Predicho: No	**Predicho: Sí**
Real: No	2826	1469
Real: Sí	1401	3285

Tabla 4.3: Matriz de Confusión mortalidad Hill-Climbing

	Predicho: No	**Predicho: Sí**
Real: No	2826	1469
Real: Sí	1401	3285

Tabla 4.4: Matriz de Confusión mortalidad PC

Observamos en las Figuras 4.8 y 4.9 que ambas matrices de confusión para los dos algoritmos obtienen los mismos valores (Tablas 4.3 y 4.4). Podemos observar que en estas impresiones en pantalla de los valores predichos por los algoritmos, empleados para la construcción de la matriz de confusión de cada uno, se obtienen algo distintos a pesar de clasificar igual. Esto se debe a que las probabilidades salen valores por encima del 0.5 o por debajo de igual manera. Esto respalda lo observado con las curvas ROC.

```
inferencia = VariableElimination(model_final_hill)

y_pred_probs = []
for _, fila in df_completo_categoricas_limpio.iterrows():
    evidencia = fila.drop('Mortalidad').to_dict()  # Todo menos la variable objetivo
    # Obtenemos la distribución posterior de la variable objetivo
    dist = inferencia.query(variables=['Mortalidad'], evidence=evidencia)
    prob_positivo = dist.values[1]  # Probabilidad de clase positiva (por ejemplo, clase 1)
    y_pred_probs.append(prob_positivo)

y_true = df_completo_categoricas_limpio['Mortalidad'].values
print(y_pred_probs)
```

✓ 3.8s

```
[0.6793168500129687, 0.6097242497480097, 0.6603295763935402, 0.6422915409881359, 0.6422915409881359,
```

Figura 4.8: Comparación valores predichos Hill-Climbing

```
y_pred_probs_pc = []
for _, fila in df_completo_categoricas_limpio.iterrows():
    evidencia = fila.drop('Mortalidad').to_dict()  # Todo menos la variable objetivo
    # Obtenemos la distribución posterior de la variable objetivo
    dist = inferencia.query(variables=['Mortalidad'], evidence=evidencia)
    prob_positivo = dist.values[1]  # Probabilidad de clase positiva (por ejemplo, clase 1)
    y_pred_probs_pc.append(prob_positivo)

y_true_pc = df_completo_categoricas_limpio['Mortalidad'].values
print(y_pred_probs_pc)
```

✓ 3.1s

```
[0.6917637197025094, 0.7761795166858458, 0.6599761940187472, 0.6702322388742706, 0.6702322388742706,
```

Figura 4.9: Comparación valores predichos PC

Así, de este modo, a través del entrenamiento considerando la validación cruzada, de la curva ROC, del AUC y de las matrices de confusión hemos podido analizar su capacidad

predictiva. Estableciendo una confianza moderada a la hora de aplicar la propagación de probabilidades de nuestros modelos.
Por otro lado, cabe destacar la relevancia de las relaciones modeladas por los modelos ya que, en ambos, nuestra variable objetivo se relaciona con todas las variables, que forman la Cobertura de Markov. Por ello, tiene sentido plantearnos los siguientes experimentos para evaluar la sensibilidad de los cambios según qué valores tomen las variables. Primero planteemos las relaciones directas con la variable mortalidad:

- Sexo: el ser hombre o mujer tendrá un impacto importante en la variable objetivo, siendo uno de estos dos grupos más propenso a tener accidentes de este tipo.
- Edad: habrá un grupo de edad con más número de accidentes registrados.
- Nacionalidad: el ser local o extranjero tambien tiene una influencia directa sobre los ahogamientos.
- Entorno: hay entornos en los que se darán más accidentes.
- Vigilancia: no tener personal cualificado supervisando los comportamientos para prevenir y para remediar los accidentes tiene un impacto importante sobre la incidencia de estos casos.

Asímismo, también se dan relaciones indirectas con respecto a la mortalidad. Estas relaciones indirectas muestran como distintas variables a través de sus relaciones con otras variables, pueden afectar a la mortalidad.

Veamos primero que conclusión sacamos con las probabilidades condicionadas a una sola variable, mostradas en la Tabla 4.5. Cabe destacar, que los valores subrayados han sido seleccionados personalmente para poder reflejar los dos mejores valores y los dos peores valores, considerando como peor los mayores porcentajes.

Tabla 4.5: Inferencias de mortalidad para todos los valores de las variables.

Variable Analizada	**Valor**	**Predicción mortalidad**
Género	Mujer	43.69 %
	Hombre	54.67 %
Nacionalidad	Español	49.28 %
	Extranjero	62.78 %
Vigilancia	Con vigilancia	37.68 %
	Sin vigilancia	57.22 %
Edad	Menor	22.68 %
	Adulto	51.34 %
	Anciano	70.92 %
Entorno	Playa	44.89 %
	Piscinas	52.19 %
	Ríos y pantanos	69.83 %
	Otros	63.62 %

Tras analizar las probabilidades condicionadas de mortalidad en función de diversas variables, podemos afirmar que la mortalidad tiene mayores predicciones para los valores ser hombre, tener edad avanzada, ser extranjero, estar en entornos de mayor riesgo, como los ríos y pantanos, y la ausencia de vigilancia. Estas variables representan factores de riesgo que deberían considerarse prioritarios en estrategias preventivas y de intervención para reducir la mortalidad en el contexto estudiado.

Veamos en las tablas a continuación una serie de combinaciones de más variables que pueden tener cierto valor para el estudio, que se tendrán en cuenta para después realizar unas conclusiones sobre los valores obtenidos:

Tabla 4.6: Combinaciones de inferencias de mortalidad con edad y nacionalidad.

Edad	**Nacionalidad**	**Predicción mortalidad**
Menor	Español	18.09 %
	Extranjero	45.78 %
Adulto	Español	47.50 %
	Extranjero	65.41 %
Anciano	Español	71.53 %
	Extranjero	67.42 %

Tabla 4.7: Combinaciones de inferencias de mortalidad con edad y género.

Edad	**Género**	**Predicción mortalidad**
Menor	Mujer	19.20 %
	Hombre	24.26 %
Adulto	Mujer	39.22 %
	Hombre	54.89 %
Anciano	Mujer	66.30 %
	Hombre	72.88 %

Tabla 4.8: Combinaciones de inferencias de mortalidad con edad y vigilancia.

Edad	**Vigilancia**	**Predicción mortalidad**
Menor	Vigilado	12.53 %
	No vigilado	29.27 %
Adulto	Vigilado	33.01 %
	No vigilado	56.04 %
Anciano	Vigilado	60.93 %
	No vigilado	76.19 %

Tabla 4.9: Combinaciones de inferencias de mortalidad con género y vigilancia.

Genero	**Vigilancia**	**Predicción mortalidad**
Mujer	Vigilado	31.79 %
	No vigilado	48.84 %
Hombre	Vigilado	40.02 %
	No vigilado	60.06 %

Tabla 4.10: Combinaciones de inferencias de mortalidad con entorno y vigilancia.

Entorno	**Vigilancia**	**Predicción mortalidad**
Playas	Vigilado	37.97 %
	No vigilado	48.75 %
Ríos y pantanos	Vigilado	57.92 %
	No vigilado	69.92 %
Piscinas	Vigilado	35.37 %
	No vigilado	66.41 %
Otros	Vigilado	39.42 %
	No vigilado	64.75 %

Tabla 4.11: Combinaciones de inferencias de mortalidad con nacionalidad y vigilancia.

Nacionalidad	**Vigilancia**	**Predicción mortalidad**
Español	Vigilado	35.27 %
	No vigilado	54.64 %
Extranjero	Vigilado	48.06 %
	No vigilado	68.48 %

Tabla 4.12: Combinaciones de inferencias de mortalidad con entorno y nacionalidad.

Entorno	**Nacionalidad**	**Predicción mortalidad**
Playas	Español	41.80 %
	Extranjero	57.03 %
Ríos y pantanos	Español	67.72 %
	Extranjero	83.31 %
Piscinas	Español	49.05 %
	Extranjero	64.19 %
Otros	Español	60.87 %
	Extranjero	78.79 %

Tabla 4.13: Combinaciones de mortalidad con edad, entorno, nacionalidad y vigilancia.

Edad	**Entorno**	**Nacionalidad**	**Vigilancia**	**Predicción mortalidad**
Menor	Playas	Español	Si	4.11 %
			No	7.16 %
		Extranjero	Si	15.13 %
			No	24.30 %
	Rios y pantanos	Español	Si	6.04 %
			No	27.90 %
		Extranjero	Si	42.04 %
			No	81.37 %
	Piscinas	Español	Si	20.25 %
			No	49.35 %
		Extranjero	Si	41.21 %
			No	72.91 %
Adulto	Playas	Español	Si	28.36 %
			No	44.76 %
		Extranjero	Si	45.17 %
			No	62.77 %
	Rios y pantanos	Español	Si	58.43 %
			No	64.75 %
		Extranjero	Si	79.60 %
			No	83.61 %
	Piscinas	Español	Si	34.01 %
			No	73.78 %
		Extranjero	Si	57.53 %
			No	88.09 %
Anciano	Playas	Español	Si	61.12 %
			No	70.32 %
		Extranjero	Si	61.97 %
			No	71.07 %
	Rios y pantanos	Español	Si	65.67 %
			No	85.80 %
		Extranjero	Si	62.81 %
			No	84.22 %
	Piscinas	Español	Si	58.73 %
			No	80.26 %
		Extranjero	Si	49.71 %
			No	73.85 %

Destacar el hecho de que se han resaltado en color verde los valores obtenidos menores del 10 %, en rojo débil los comprendidos entre un 70-80 % y los rojos fuerte los comprendidos entre 80-90 %, con el mero propósito de resaltar los valores de ambos extremos.

Como conclusión de estas combinaciones, la red muestra que la variable género no tiene un peso alto en la mortalidad. Se puede observar que los hombres suelen ser algo más temerarios que las mujeres, pero no tiene un impacto significativo. Si que se puede ver claramente en la Tabla 4.9 que al disponer de un servicio de vigilancia se obtiene un cambio trascendente en los resultados llegando a disminuir las predicciones alrededor de un 20 %.

En cuanto a la vigilancia, se observa que es un factor muy importante. Fijándonos en las tablas 4.8, 4.9, 4.10, 4.11 podemos hacer una apreciación que en todas las combinaciones de las variables con vigilancia tienen una disminución alta del porcentaje de mortalidad entorno al 15-20 %. Tras observar esto, se puede hacer la recomendación de realizar programas de vigilancia, muy necesarios, y de hecho, cabe resaltar que sería necesario un refuerzo especialmente en piscinas, donde se tiene el mayor impacto, obteniendo una disminución del 31 %. Sin embargo, sería de provecho incluirla en los entornos populares más rurales de las localidades. Además, hay que replantearse si es necesario hacer un incremento de la vigilancia con más personal en las playas, o que realicen más chequeo de actividades temerarias por parte de los bañistas.

Asimismo, podemos observar claramente que los turistas no suelen tomar las suficientes medidas de precaución o que reducen la atención ante los peligros estando de vacaciones. En las tablas 4.6 y 4.12 se pueden sacar varias conclusiones. La primera, es necesario concienciar a los familiares adultos extranjeros de que, a pesar de relajarse en las vacaciones, es necesario tener máximo nivel de atención con sus familiares menores, dado que tenemos un incremento de mortalidad del 27 % para los niños extranjeros con respecto a los españoles. Asimismo, en general, se presentan valores altos para todos los entornos, esto puede deberse al desconocimiento del lugar, o de las indicaciones y advertencias de las señales y personal de las playas o piscinas.

Con respecto al rango de edad, observamos una diferencia considerable entre grupos, en concreto del 48 %, siendo los menores menos víctimas que los ancianos. Esto nos lleva a plantearnos en centrar campañas de concienciación para estar más atentos de los mayores, más vulnerables en estos entornos. Cabe destacar viendo la tabla 4.6 que con respecto a este grupo de edad, hay un mayor valor para los locales. Entre medias, tenemos el grupo de adultos, siendo aún un valor bastante alto. Como línea de mejora se podría haber segmentado este grupo para poder observar si el valor es homogéneo o solo se debiera a los adultos más jóvenes.

La tabla grande 4.13 refuerza nuestras conclusiones, siendo los adultos y ancianos el grupo más propenso a estos accidentes, las playas el entorno más seguro, y los turistas los más temerarios. Todo esto con una reducción glogal muy significativa de la predicción de muerte si hay presencia de vigilancia.

Para contrastar todas estas conclusiones obtenidas, se han aplicado unos z-test de dos proporciones, con el propósito de evaluar si la mortalidad en ciertos valores observados difiere significativamente de la mortalidad general. Esta comparación se realizó sobre los datos reales, de forma complementaria al análisis probabilístico realizado con la red

bayesiana. La concordancia o discrepancia entre los resultados observados en estos test y los inferidos por la red permite validar la fidelidad del modelo respecto a la realidad empírica. Se ha realizado utilizando la librería statsmodels [25] del siguiente modo:

```python
import pandas as pd
from statsmodels.stats.proportion import proportions_ztest

# Conteo de éxitos (muertes) en ambos grupos
muertes_total = df_completo_categoricas_limpio['Mortalidad'].sum()
muertes_extranjero = df_completo_categoricas_limpio[(df_completo_categoricas_limpio['Mortalidad'] == 1)
                                                    & (df_completo_categoricas_limpio['Nacionalidad'] == 'Extranjero')].shape[0]

# Tamaños de muestra
n_total = df_completo_categoricas_limpio.shape[0]
n_extranjero = df_completo_categoricas_limpio[df_completo_categoricas_limpio['Nacionalidad'] == 'Extranjero'].shape[0]

# Número de éxitos en cada grupo (muertes)
count = [muertes_extranjero, muertes_total]
# Tamaño de muestra de cada grupo
nobs = [n_extranjero, n_total]

# Aplicar z-test de dos proporciones
stat, pval = proportions_ztest(count, nobs)

# Mostrar resultados
print(f"Z-statistic: {stat:.4f}")
print(f"P-valor: {pval:.16f}")

# Interpretación
alpha = 0.05
if pval < alpha:
    print("La diferencia en proporciones es estadísticamente significativa.")
else:
    print("No hay diferencia estadísticamente significativa en las proporciones.")
```

✓ 0.0s

```
Z-statistic: 8.0449
P-valor: 0.0000000000000009
La diferencia en proporciones es estadísticamente significativa.
```

Figura 4.10: Z-test

Anteriormente, hemos observado que con respecto a la variable género los hombres tienen mayor predisposición según el modelo, lo cual se confirma al haber realizado el z-test obteniendo un p-valor de $5x10^{-4}$, mientras que al agregar la presencia de vigilancia obtenemos $5x10^{-15}$, un p-valor mucho inferior, lo que sugiere la veracidad de lo concluido por la red sobre esta variable en ambos valores.

Por otro lado, al aplicarlo para la vigilancia en general obtenemos el p-valor $3x10^{-31}$, lo cual refuerza contundentemente la significancia estadística. Al evaluarlo en entornos como la playa obtenemos $7x10^{-28}$, en la piscina $3x10^{-11}$ que indican una fuerte evidencia de que la diferencia de las proporciones es estadísticamente significativa. Por otro lado, para ríos y pantanos se ha obtenido el p-valor 0.66, reflejando que están asociadas de manera aleatoria.

Continuando con la variable nacionalidad, para los extranjeros tenemos el valor $9x10^{-16}$, evaluando para los menores, tenemos 0.13, lo cual indica que no hay evidencia suficiente para afirmar que ser extranjero y menor esté asociado con un cambio real en la mortalidad. Sin embargo, cuando se trata de un menor español, tenemos $1x10^{-118}$ que indica todo lo contrario. Con respecto a los entornos, obtenemos para playas $7x10^{-27}$, para pantanos y

ríos $8x10^{-29}$, por lo tanto hay una diferencia no atribuible al azar entre ambos grupos. Pero para piscinas no ocurre esto pues tenemos el p-valor 0.14.

Por último, para la variable edad, tenemos para menores $9x10^{-98}$, para ancianos $1x10^{-77}$ y para adultos 0.57, aceptándose esta última hipótesis nula de que no hay diferencia entre ambas proporciones.

Finalmente, como línea de mejora para ese margen presente en la Red Bayesiana elegida, se pueden considerar distintos frentes:

Considerando el primero, realizar un desglose aún mayor en las variables propuestas, sobre todo en los rangos de edad y en los entornos físicos, puesto que se ha hecho una agrupación bastante extensa y puede haber llevado a una generalización de los resultados obtenidos.

Por otro lado, sería beneficioso considerar nuevas variables en el estudio, que pueden aportar otros puntos de vista, como por ejemplo qué tipo de actividad que estaba realizando el sujeto, si estaba acompañado, si tenía enfermedades considerables, si fue socorrido por personal cualificado o si se le realizaron maniobras de reanimación exitosas o no.

Bibliografía

[1] D. Obmann, P. Münch, B. Graf, and collaborators. Comparison of different ai systems for diagnosing sepsis, septic shock, and cardiogenic shock: a retrospective study. *Sci Rep*, 15:15850, 2025.

[2] K. Nithya, Mary Selvan, V. K. Gnanavel, P. Solainayagi, and K. Venkatesh. Leveraging bayesian networks for enhanced transparency in financial data provenance tracking. In *2024 International Conference on Cybernation and Computation (CYBERCOM)*, pages 671–676, 2024.

[3] Yan Cheng. Study on the construction of learner profile model and the recommendation path of personalized innovation and entrepreneurship learning strategies. In *Proceedings of the 2024 4th International Conference on Artificial Intelligence, Automation and High Performance Computing*, AIAHPC '24, page 362–367, New York, NY, USA, 2024. Association for Computing Machinery.

[4] Chaohui Zhang, Jiyuan Liu, and Shichen Zhang. Online purchase behavior prediction model based on recurrent neural network and naive bayes. *Journal of Theoretical and Applied Electronic Commerce Research*, 19(4):3461–3476, 2024.

[5] Hammond James and Smith V. Anne. Bayesian networks for network inference in biology. *J. R. Soc. Interface*, page 2220240893, 2025.

[6] G. Gogoshin and A.S. Rodin. Minimum uncertainty as bayesian network model selection principle. *BMC Bioinformatics*, 26(100), 2025.

[7] Mohammad Najafzadeh, Sajad Basirian, and Zhiqiang Li. Vulnerability of the rip current phenomenon in marine environments using machine learning models. *Results in Engineering*, 21:101704, 2024.

[8] Javier García-Alba, Javier F. Bárcena, Luis Pedraz, Felipe Fernández, Andrés García, Marcos Mecías, Javier Costas-Veigas, María Luisa Samano, and David Szpilman and. Soseas web app: An assessment web-based decision support tool to predict dynamic risk of drowning on beaches using deep neural networks. *Journal of Operational Oceanography*, 16(2):155–174, 2023.

[9] C. Houser, J. Lehner, N. Cherry, and P. Wernette. Machine learning analysis of lifeguard flag decisions and recorded rescues. *Natural Hazards and Earth System Sciences*, 19(11):2541–2549, 2019.

[10] Salman Jalalifar, Andrew Belford, Eila Erfani, Amir Razmjou, Rouzbeh Abbassi, Masoud Mohseni-Dargah, and Mohsen Asadnia. Enhancing water safety: Exploring recent technological approaches for drowning detection. *Sensors*, 24(2), 2024.

[11] Xinshan Xie, Zhixing Li, Haofeng Xu, Dandan Peng, Lihua Yin, Ruilin Meng, Wei Wu, Wenjun Ma, and Qingsong Chen. Non-fatal drowning risk prediction based on stacking ensemble algorithm. *Children*, 9(9), 2022.

[12] Han-Chung Yang Yi-Jun Lin Meng-Tsung Lee, Yang-Chi Chang. Assessing risk associated with recreational activities in coastal areas by using a bayesian network. *ScienceDirect*, 9(9), 2023.

[13] Chistopher M. Bishop. *Pattern recognition and machine learning*. Springer, Cambrigde, 8 edition, 2009.

[14] Kevin P. Murphy. *Machine Learning: A Probabilistic Perspective*. MIT Press, Cambridge, MA, 1 edition, 2012.

[15] Mario Fiovaranti. Apuntes matemática discreta. Universidad de Cantabria, 2022. Apuntes de clase.

[16] J.A Cuesta Albertos y A. Tuero Díaz. Apuntes estadística. Universidad de Cantabria, 2022. Apuntes de clase.

[17] Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein. *Introduction to Algorithms*. MIT Press, Cambridge, MA, 3th edition, 2009.

[18] Daphne Koller and Nir Friedman. *Probabilistic Graphical Models: Principles and Techniques - Adaptive Computation and Machine Learning*. The MIT Press, 2009.

[19] Maialen Iturbe y Javier Díez Sixto Herrera. Apuntes machine learning 2. Universidad de Cantabria, 2025. Apuntes de clase.

[20] Peter Flach. *The Art and Science of Algorithms that Make Sense of Data*. Cambrigde University Press, Cambrigde, 1ra edition, 2012.

[21] George Casella and Edward I. George. Explaining the gibbs sampler. *The American Statistician*, 46(3):167–174, 1992.

[22] Antonio Slamerón Cerdán y Carmelo Rodriguez Torreblanca María Morales González. *Análisis de Indicadores de Rendimiento Mediante Redes Bayesianas*. Universidad de Almería, Almería, España, 1ra edition, 2007.

[23] Escuela Segoviana de Socorrismo. Ahogamientos: Blog sobre prevención de ahogamientos y seguridad acuática. `https://ahogamientos.blogspot.com/`, 2024. Consulta realizada en mayo de 2025.

[24] Ankur Ankan, Abinash Panda, and contributors. pgmpy: Probabilistic graphical models using python. `https://pgmpy.org`, 2015–2024. Versión usada: 1.0.0.

[25] Skipper Seabold, Josef Perktold, and contributors. statsmodels: Statistical modeling in python. `https://www.statsmodels.org`, 2010–2024. Versión usada: 0.14.0.