



UNIVERSIDAD NACIONAL DE SAN ANTONIO ABAD DEL CUSCO

ESCUELA DE POSGRADO

MAESTRÍA EN ESTADÍSTICA

**MODELOS JERÁRQUICOS Y REDES BAYESIANAS PARA
IMPLANTAR UN PROTOCOLO Y DETECCIÓN DE
QUERATOCONO EN EL CUSCO**

SEMINARIO DE TESIS 1

AUTOR:

Br. Kevin Heberth Haquehua Apaza

DOCENTE:

Dr. ALFREDO VALENCIA TOLEDO

CUSCO - PERÚ

2025

1. Planteamiento del problema

1.1. Situación problemática

El queratocono es una enfermedad que afecta a la córnea causando mala visión de forma progresiva y evolutiva en pacientes jóvenes. Según la Organización Mundial de la Salud (OMS) esta enfermedad puede indicar cifras epidémicas. A pesar de que se considere una afección de impacto moderado y no tan grave que afecta a la función visual y calidad de vida, esta enfermedad puede resultar más perjudicial ya que puede afectar a individuos totalmente activos [THALASSELIS et al., 1988]. El queratocono generalmente inicia en la pubertad hasta la tercera o casos de cuarta década de vida [Althomali et al., 2018], sus síntomas son variables y dependen de la etapa de progresión que se encuentre, en etapas iniciales no pueden presentarse síntomas hasta etapas más avanzadas en donde el paciente experimenta una distorsión notable en la visión acompañada de una pérdida visual severa. A pesar de eso no llegan a producir ceguera completa debido a esta enfermedad [Rabinowitz, 1998].

Los primeros estudios de prevalencia se llevó en la población de Indianapolis, USA encontró 600 casos por 100,000 (0.6 %). Un estudio poblacional más reciente informó una prevalencia de 375 por 100,000 (0.375 %) en Países Bajos [Godefrooij et al., 2017]. Mientras que en Latino América los registros de prevalencia son muy pocos, teniendo en Paraná entre Ríos-Argentina una prevalencia aproximada de 229 por cada 100,000 habitantes con edad promedio de diagnóstico a los 24.5 años [Pussetto et al., 2011], en la ciudad de Quito en Ecuador se tuvo una prevalencia de (0.75 %) [Mansfield Rojas, 2017], por otra parte en el Hospital Regional Honorio Delgado, durante el periodo 2014 - 2017 en Arequipa-Perú se evidenció que la prevalencia de queratocono fue de 1.37 por cada mil pacientes atendidos [Ramos Almirón, 2018]. Por último con respecto a la ciudad del Cusco en el trabajo de investigación de [Samaniego Mojica, 2023] reportó en el Centro de Salud Integral La Fuente una prevalencia de 3.32 % siendo esta mayor a las otras zonas estudiadas en el cual aún no se le esta dando la importancia por parte del estado peruano.

Asimismo en la clínica Oftalmo Salud del Perú indica que no se sabe con certeza porque algunas personas desarrollan Queratocono, aunque hay estudios con asociaciones a la genética o dicho hereditaria, asimismo con alergias y lesiones oculares provocado por la frotación excesiva de los ojos, el uso inapropiado de lentes de contacto y la presencia de algunas patologías como la queratoconjuntivitis vernal, la retinitis pigmentaria o la retinopatía del prematura. Múltiples estudios han desarrollado técnicas para poder realizar la detección temprana de esta enfermedad ya sea por modelos de Machine Learning que indicaron que índices de PENTACAM fueron anormales para detección y exclusión de pacientes con queratocono [Zhao et al., 2024]

En el que teniendo este enfoque nos encontramos en un marco de incertidumbre en el cual es necesario construir un modelo probabilístico del comportamiento de esta enfermedad. Los enfoques bayesianos para problemas bioestadísticos se volvieron comunes en aplicaciones epidemiológicas, el uso de la metodología bayesiana ha experimentado grandes avances, aumentando el enfoque del uso bayesiano [Lawson, 2018]. Centrandonos en la construcción de modelos para relacionar estructura de datos complicados con preguntas científicas, verificando el ajuste de dicho modelo e investigando la sensibilidad de las conclusiones a supuestos de modelos razonables. [Gelman et al., 2013] indica que una de las fortalezas del enfoque bayesiano reside en combinar información de diferentes fuentes y dar una explicación más amplia de la incertidumbre acerca de las incógnitas en un problema estadístico. Estudios de enfermedades raras ya aplican este enfoque como lo es de [Fouarge et al., 2021] en el que estudia las miopatías centronucleares con un modelo jerárquico bayesiano con distribución de probabilidad beta y una función logit de enlace, en el cual el autor menciona que este tipo de problemas pueden ser abordado bajo el enfoque bayesiano en el que se deban realizar más investigaciones con el diálogo continuo de autoridades reguladoras que permitan la aplicación de la estadística bayesiana. Asimismo en el trabajo de [Férez Rubio, 2017] indicó que una herramienta para apoyar en la tomade decisiones de la medicina basada en la evidencia son las redes bayesianas ya que tiene

una capacidad para representar las relaciones causales que tienen las variables involucradas en un problema, y que se visualice de una manera intuitiva.

En el cual el objetivo es determinar un modelo probabilístico basado en los modelos bayesianos jerárquicos que permitan cuantificar la incertidumbre de la detección de la enfermedad del queratocono en la ciudad del Cusco, tomando en cuenta los diferentes estudios de esta enfermedad ya sea en la asociación de variables clínicas, tomográficas, tests, etc. Con el fin de modelar una distribución probabilística que permita cuantificar la incertidumbre de esta enfermedad y asimismo optimizar los costos en pruebas más costosas en base a la información disponible para la mayor parte de centros de salud.

1.2. Formulación del problema de investigación

1.2.1. Problema general

¿Cómo aplicar la metodología bayesiana en la detección del queratocono?

1.2.2. Problemas específicos

1. ¿Qué variables clínicas, demográficas y tomográficas apoyan en la detección del queratocono?
2. ¿Qué modelo jerárquico representa la relación entre las variables clínicas, demográficas y tomográficas en la detección del queratocono?
3. ¿Cómo obtenemos la inferencia de probabilidad en la presencia o ausencia de queratocono a partir del modelo?
4. ¿Cómo implementar las inferencias de riesgo a nivel individual?

1.3. Objetivos de investigación

1.3.1. Objetivo general

Desarrollar un protocolo de detección de queratocono bajo un enfoque metodológico bayesiano.

1.3.2. Objetivo específicos

1. Determinar las variables clínicas, demográficas y tomográficas que apoyen en la detección del queratocono.
2. Diseñar un modelo jerárquico que represente la relación entre las variables clínicas, demográficas y tomográficas en la detección del queratocono.
3. Construir una red bayesiana estructurada que infiera la probabilidad de presencia o ausencia del queratocono a partir de las variables observadas.
4. Implementar un prototipo computacional que visualice las inferencias de riesgo de queratocono a nivel individual.

2. Marco teórico conceptual

2.1. Bases teóricas

2.1.1. Axiomas de probabilidad y notaciones básicas

Definición 2.1. Notaciones [Koski y Noble, 2011]

- El conjunto de todos los posibles resultados de un experimento aleatorio es denotado por Ω
- El resultado de un posible experimento aleatorio o un conjunto del resultado es representado por X
- El espacio de parámetros es denotado por $\tilde{\Theta}$, tal que $\Omega = X \times \tilde{\Theta}$
- Si A y B son dos conjuntos, luego $A \cup B$ representa la unión, además que si A_1, A_2, \dots, A_n son una colección finita de conjuntos, entonces $\bigcup_{i=1}^n A_i$ representa la unión de conjuntos.
- Si A y B son dos conjuntos, luego $A \cap B$ representa la intersección, además que si A_1, A_2, \dots, A_n son una colección finita de conjuntos, entonces $\bigcap_{i=1}^n A_i$ representa la intersección de conjuntos.
- $A \subset B$ denota que A es un subconjunto de B . $A \subseteq B$ denota que A es un subconjunto de B , posiblemente igual a B
- El conjunto vacío puede ser denotado por \emptyset
- A^c denota el complemento de A

Definición 2.2. Probabilidad [Koski y Noble, 2011] La distribución de probabilidad P es una función que satisface los axiomas de Kolmogorov

1. $P(\emptyset) = 0$ y $P(\Omega) = 1$

2. Si A_1, A_2, \dots, A_n es una colección finita tal que para cada A_i como eventos que satisface $A_j \cap A_k = \emptyset$ para todo $j \neq k$, entonces

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i)$$

3. $0 \leq P(A) \leq 1$ para todo A que pertenece a Ω

Definición 2.3. Distribución de probabilidad sobre X [Koski y Noble, 2011] Si X contiene un número finito de elementos x , una distribución de probabilidad sobre X satisface

1. $P(x) \geq 0$ para todo $x \in X$
2. $\sum_{x \in X} P(x) = 1$

2.2. Marco conceptual (palabras clave)

2.3. Antecedentes empíricos de la investigación

2.3.1. Antecedentes Internacionales

El estudio de [Waddell, 2023] aplicó las redes bayesianas para el apoyo en la toma de decisiones, tomando datos de diferentes fuentes para explorar fenotípicos específicos y resultados clínicos, en la cual se pueden tener deficiencias debido al conjunto de datos pequeños y cohortes desequilibrados llegando a realizar inferencias falsas probabilísticas. Algo importante de destacar es como el uso de redes bayesianas ayuda en la detección de enfermedades especialmente en incertidumbre compleja, asimismo el autor indicar de que a pesar que la medicina actualmente se encuentra en competencia con el desarrollo de la inteligencia artificial, esta a su vez debe ser poder interpretada por los médicos en el cual el verdadero potencial de las redes bayesianas viene a ser la capacidad de hablar el lenguaje clínico, modelar preguntas relacionales causales y responder preguntas al explorar fenotipos individuales llegando a plantear preguntas cruciales

como ¿por qué? y ¿que pasaria sí?.

El trabajo de [Férez Rubio, 2017] presenta una mejor evidencia de la medicina basada en la evidencia en el cual se debe tener un uso consciente, explícito y juicioso de la evidencia científica disponible para tomar la decisión de los pacientes, en el cual comprender la realidad y expresarla de manera sistemática, inteligible y sintética. Integrar la experiencia de los profesionales con la evidencia científica disponible para mejorar la toma de decisiones clínicas. Para conseguir este propósito se tienen diferentes entes matemáticas en la que su estudio enfatizó a las redes bayesianas como una de las herramientas más valiosas en el proceso de toma de decisiones. El trabajo planteo conceptos probabilísticos enfocado al análisis bayesiano en los cuales se fundamentan las redes bayesianas, para posteriormente crear dos redes a ejemplos prácticos, en la cual la red bayesiana tiene una capacidad para representar la relación causa entre variables involucradas en un problema y visualizarlas de una manera muy intuitiva, el cual es el modelo gráfico probabilístico más utilizado en el marco de la toma de decisión diagnóstica.

2.3.2. Antecedentes Nacionales

La investigación de [Lopez Esquivel, 2023] para optar el título de maestro en Estadística de la Pontífice Universidad Católica del Perú, se basó acerca de las infecciones respiratorias, en la cual buscó establecer una relación entre la incidencia de infecciones respiratorias agudas (IRA) y la incidencia de neumonía en el Perú en la cual además de ver si estas variables están correlacionadas quiso dar un enfoque espacial a nivel provincial en la cual se esperaría que la incidencia de (IRA) y neumonía sea mayor en provincias vecinas. Por lo cual se estudió la distribución espacial entre la incidencia de (IRA) y neumonía a nivel provincial en el Perú a través de un modelo espacial multivariado con efectos aleatorios condicionales autoregresivos utilizando inferencia bayesiana en el modelo jerárquico espacial multivariado a través del método de integración anidada de Laplace (INLA). El cual mostró una bondad de ajuste del modelo y

su eficacia en la simulación realizada teniendo una optimización de tiempo menor, asimismo se vio una correlación positiva directa entre las personas con (IRA) sin neumonía y neumonía. Además de que se tuvo evidencia estadística entre las precipitaciones y la neumonía, indicando que el aumento de precipitaciones en una provincia reduce significativamente la incidencia de IRA sin neumonía, mientras el aumento de precipitación en una provincia aumenta la incidencia de neumonías en dicha provincia, teniendo una autocorrelación espacial moderada con respecto a las tasas de enfermedad entre provincias vecinas, recomendado que se puede trabajar con más covariables y factores que expliquen mejor la causalidad del modelo, de la misma forma incluir un modelo espacio-temporal empleando INLA que tiene una mejor eficiencia computacional al trabajar con más datos.

3. Matriz de consistencia

Problemas	Objetivos	Hipótesis	Variables	Metodología
General			Dependientes	En proceso
¿Cómo aplicar la metodología bayesiana en la detección del queratocono?	Desarrollar un protocolo de detección de queratocono bajo un enfoque metodológico bayesiano	En proceso	<ul style="list-style-type: none"> ○ Probabilidad de presencia de queratocono ○ Clasificación de presencia o ausencia de queratocono 	
Específicos			Independientes	
<ul style="list-style-type: none"> ○ ¿Qué variables clínicas, demográficas y tomográficas apoyan en la detección del queratocono? ○ ¿Qué modelo jerárquico representa la relación entre las variables clínicas, demográficas y tomográficas en la detección del queratocono? ○ ¿Cómo obtenemos la inferencia de probabilidad en la presencia o ausencia de queratocono a partir del modelo? ○ ¿Cómo implementar las inferencias de riesgo a nivel individual? 	<ul style="list-style-type: none"> ○ Determinar las variables clínicas, demográficas y tomográficas que apoyen en la detección del queratocono. ○ Diseñar un modelo jerárquico que represente la relación entre las variables clínicas, demográficas y tomográficas en la detección del queratocono. ○ Construir una red bayesiana estructurada que infiera la probabilidad de presencia o ausencia del queratocono a partir de las variables observadas. ○ Implementar un prototipo computacional que visualice las inferencias de riesgo de queratocono a nivel individual. 	En proceso	<ul style="list-style-type: none"> ○ Variables demográficas. ○ Variables clínicas. ○ Variables tomográficas. 	

Referencias

- [Althomali et al., 2018] Althomali, T. A., Al-Qurashi, I. M., Al-Thagafi, S. M., Mohammed, A., y Almalki, M. (2018). Prevalence of keratoconus among patients seeking laser vision correction in taif area of saudi arabia. *Saudi Journal of Ophthalmology*, 32(2):114–118.
- [Férez Rubio, 2017] Férez Rubio, J. A. (2017). Redes bayesianas aplicadas a la medicina.
- [Fouarge et al., 2021] Fouarge, E., Monseur, A., Boulanger, B., Annoussamy, M., Seferian, A. M., De Lucia, S., Lilien, C., Thielemans, L., Paradis, K., Cowling, B. S., et al. (2021). Hierarchical bayesian modelling of disease progression to inform clinical trial design in centronuclear myopathy. *Orphanet Journal of Rare Diseases*, 16:1–11.
- [Gelman et al., 2013] Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., y Rubin, D. B. (2013). *Bayesian Data Analysis*. Chapman and Hall/CRC.
- [Godefrooij et al., 2017] Godefrooij, D. A., De Wit, G. A., Uiterwaal, C. S., Imhof, S. M., y Wisse, R. P. (2017). Age-specific incidence and prevalence of keratoconus: a nationwide registration study. *American journal of ophthalmology*, 175:169–172.
- [Koski y Noble, 2011] Koski, T. y Noble, J. (2011). *Bayesian networks: an introduction*. John Wiley & Sons.
- [Lawson, 2018] Lawson, A. B. (2018). *Bayesian disease mapping: hierarchical modeling in spatial epidemiology*. Chapman and Hall/CRC.
- [Lopez Esquivel, 2023] Lopez Esquivel, M. A. (2023). Modelamiento bayesiano espacial multivariado para datos de áreas.
- [Mansfield Rojas, 2017] Mansfield Rojas, N. (2017). El queratocono en pacientes de una institución privada de la ciudad de quito, ecuador en el período de enero de 2015 a octubre de 2016. B.S. thesis, Quito: Universidad de las Américas, 2017.

- [Pussetto et al., 2011] Pussetto, A. P., Rossi, M. L., Ciani, J. L., y Magurno, M. G. (2011). Alta prevalencia de pacientes con queratocono en la población de la ciudad de paran , entre r os. *Oftalmol Clin Exp*, 4(4).
- [Rabinowitz, 1998] Rabinowitz, Y. S. (1998). Keratoconus. *Survey of ophthalmology*, 42(4):297–319.
- [Ramos Almir n, 2018] Ramos Almir n, V. A. (2018). Prevalencia y factores asociados al queratocono en pacientes atendidos en el hospital regional honorio delgado, arequipa, 2014-2017.
- [Samaniego Mojica, 2023] Samaniego Mojica, A. C. (2023). Prevalencia y factores asociados al queratocono en pacientes entre los 10 a 30 a os de edad atendidos en el centro de salud integral la fuente, cusco-per  durante el a o 2019.
- [THALASSELIS et al., 1988] THALASSELIS, A., TAIE, H. F., ETCHEPAREBORDA, J., y SELIM, A. (1988). Keratoconus, magnesium deficiency, type a behavior, and allergy. *Optometry and Vision Science*, 65(6):499–505.
- [Waddell, 2023] Waddell, T. H. (2023). *Applying Bayesian-networks to model multi-organ pathologies*. Tesis de doctorado, University of Oxford.
- [Zhao et al., 2024] Zhao, S. H., Berkowitz, C., Ralay Ranaivo, H., Laurenti, K., Bohnsack, B. L., Basti, S., y Rossen, J. L. (2024). Evaluation of parameters for early detection of pediatric keratoconus. *BMC ophthalmology*, 24(1):463.