TECNOLÓGICO DE MONTERREY

CAMPUS GUADALAJARA

Evidencia 3. Artículo de investigación - Modelos Bayesianos

Por

Paola Enríquez Reyes

A01741055

Luis Jesús Castillo Goyenechea

A01275697

Erik Ernesto Ocegueda Sambrano

A01639729

Juan Pablo Bernal Lafarga

A01742342

Análisis de métodos de razonamiento e incertidumbre

Profesor Javier Edgardo Garrido Guillén

(Grupo 101)

03 de Septiembre de 2023

Abstract

En el presente trabajo, abordamos la utilidad y aplicabilidad de los modelos Bayesianos en el ámbito estadístico y análisis de datos, haciendo hincapié en la modelización de incertidumbre y la toma de decisiones en escenarios complejos. Utilizamos el modelo logístico de crecimiento poblacional para modelar la evolución de la población de aligátores en función del tiempo, aprovechando las capacidades de la estadística bayesiana para incorporar información previa y actualizaciones basadas en evidencia observacional. Implementamos nuestro modelo usando JAGS y técnicas de inferencia bayesiana como el muestreo de Markov Monte Carlo (MCMC). Los resultados, en armonía con los datos observados, reiteran la relevancia y precisión de los modelos Bayesianos en estudios ecológicos y otras disciplinas. La discusión subyacente resalta la importancia de la colaboración interdisciplinaria y aborda tanto las fortalezas como los desafíos inherentes a los enfoques bayesianos. En conclusión, esta investigación resalta el valor de los modelos Bayesianos en la investigación contemporánea y su capacidad para manejar problemas intrincados de manera efectiva y coherente.

Introducción

Los modelos Bayesianos se nos han presentado durante el curso como una herramienta esencial en el área de la estadística y el análisis de datos, lo cual por naturaleza va de la mano con nuestra carrera. Estos modelos se basan en la teoría de la probabilidad bayesiana, que proporciona un marco flexible y significante para la modelización de incertidumbre y la toma de decisiones en situaciones complejas. A medida que los conjuntos de datos se vuelven más grandes y complejos, y las preguntas de investigación más detalladas, los enfoques estadísticos basados en modelos Bayesianos han ganado prominencia en diversas disciplinas científicas y aplicaciones prácticas; esto lo hemos logrado ver con cada una de las entregas y desarrollos de evidencias, además de los cursos y temas vistos en clase.

La estadística bayesiana se diferencia de la estadística frecuentista tradicional en su tratamiento de los parámetros desconocidos, de ahí la palabra clave que hemos tenido presente durante todo el curso, “incertidumbre”. En lugar de considerar los parámetros como valores fijos pero desconocidos, los modelos Bayesianos los tratan como variables aleatorias, incorporando información previa (conocida como prior) y evidencia observacional (likelihood) para actualizar las creencias sobre estos parámetros a través del teorema de Bayes. Esta flexibilidad permite abordar problemas de inferencia, estimación y predicción.

En esta tercera evidencia, exploraremos los fundamentos de los modelos Bayesianos y cómo se aplican en enfoques estadísticos para resolver una amplia gama de problemas. Analizaremos cómo se formulan los modelos Bayesianos, cómo se eligen las distribuciones priors, y cómo se obtienen las distribuciones posteriors a través de técnicas de inferencia bayesiana, como el muestreo visto en anteriores cursos y en este, de Markov Monte Carlo (MCMC) y otros métodos numéricos. También nos centraremos en la solución de este tercer problema que se nos plantea una población de aligátores, destacando las aplicaciones de modelos Bayesianos en el campo de la biología y anteriormente en la epidemiología, la economía, etc. Aquí resaltamos la ventaja de incorporar la incertidumbre de manera coherente en el proceso de toma de decisiones.

Métodos

Definición del Modelo de Regresión No Lineal

Para modelar la población de aligátores a lo largo del tiempo (t), utilizaremos el modelo logístico de crecimiento poblacional, que se expresa mediante la siguiente ecuación diferencial:

$$\frac{dP}{dt} = k \cdot P \cdot (1 - \frac{P}{K})$$

Donde:

- P representa la población de aligátores en un momento dado.
- t es el tiempo.
- K es la capacidad de carga, que representa el tamaño máximo de la población sostenible dictado por limitaciones de recursos y otros factores.
- k es la tasa de crecimiento.

Parámetros

Para las distribuciones prior de K y k, asumimos que ambos son números positivos, ya que representan cantidades físicas positivas en el contexto del crecimiento de la población. Así podemos utilizar distribuciones no informativas o basarnos en conocimiento previo si está disponible.

Resolución de la ecuación diferencial

Para poder implementar el modelo bayesiano, necesitamos resolver la ecuación diferencial del modelo logístico de crecimiento poblacional, la cual queda de la siguiente manera:

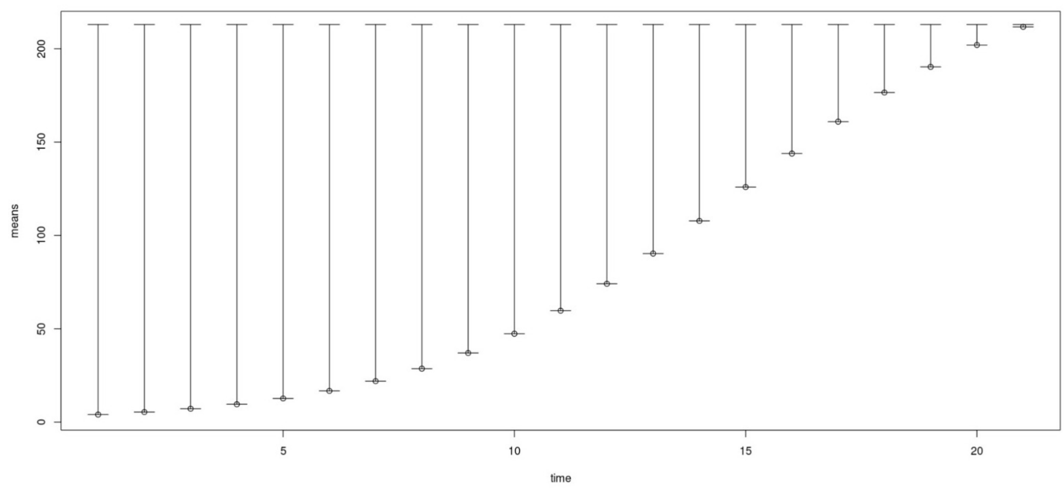
K/(exp(-k*t[i]+c_5)+1)

Implementación del Modelo Bayesiano

Usamos JAGS para implementar el modelo bayesiano. Esto incluye la definición del modelo logístico, la especificación de las distribuciones prior para K y k, y la formulación de las distribuciones likelihood basadas en una distribución normal para los datos observados. Continuo a esto, ejecutamos cadenas MCMC para muestrear la distribución posterior de los parámetros.

Al sustituir los valores de K, k y C_5 proporcionados por el modelo bayesiano, en la diferencial del modelo logístico de crecimiento poblacional, obtuvimos resultados bastante similares a los de la base de datos.

Por último realizamos un scatterplot en el qué representamos el crecimiento poblacional de los alligators, así como unas barras de error, las cuales representan la incertidumbre de la función. Las barras de error fueron generadas tomando los cuantiles 2.5 y 97.5.



Descriptivas y Gráficos

Calculamos estadísticas descriptivas, como medias, desviaciones estándar, para los parámetros K y k. Estas estadísticas nos proporcionan información sobre la estimación y la incertidumbre en los parámetros del modelo. Para visualizar los datos y la función de crecimiento estimada, generamos gráficos que muestran la población de aligátores a lo largo del tiempo junto con la función de crecimiento media estimada. Utilizaremos intervalos de confianza basados en cuantiles para representar la incertidumbre en esta función.

Aplicación

El modelo logístico de crecimiento poblacional se ha utilizado en una amplia gama de aplicaciones en diferentes campos, que se han mencionado al inicio de esta evidencia, pero se hablará más de sus aplicaciones:

Ecología y Biología

-Estudio de poblaciones animales y vegetales para comprender su crecimiento en función de factores ambientales. También se puede aplicar en la evaluación de la capacidad de carga de un ecosistema y su impacto en la conservación de especies.

Economía y Negocios

-Ayuda a la predicción del crecimiento de empresas y productos en función de la adopción del mercado.

Epidemiología

-Modelado de la propagación de enfermedades infecciosas en poblaciones humanas y animales. Además de la estimación de la capacidad de carga de sistemas de atención médica.

Ciencia de Datos e Inteligencia Artificial

-Modelado de la adopción de tecnologías y productos en redes sociales y la utilización de algoritmos de machine learning basados en modelos logísticos para análisis de datos.

Discusión

En esta evidencia, hemos explorado en profundidad los fundamentos de los modelos Bayesianos y su aplicación en la resolución de problemas reales, centrándonos en el crecimiento poblacional de aligátores como ejemplo. Durante todo el proceso, hemos destacado la importancia de incorporar la incertidumbre de manera coherente en el proceso de toma de decisiones, lo cual es una característica distintiva de la estadística bayesiana en comparación con enfoques estadísticos frecuentistas tradicionales.

Los modelos Bayesianos ofrecen una herramienta poderosa para abordar problemas complejos y dinámicos, como el crecimiento de poblaciones en el tiempo. En nuestro caso, utilizamos el modelo logístico de crecimiento poblacional, que se basa en principios ecológicos sólidos y se ha aplicado en una variedad de contextos, desde ecología hasta economía y epidemiología.

Una de las ventajas clave de la estadística bayesiana es su capacidad para incorporar información previa (las distribuciones prior) y actualizarla con nueva evidencia observacional (las distribuciones likelihood) para obtener distribuciones posteriores más precisas. Esto se traduce en estimaciones de parámetros más informadas y en una comprensión más profunda de la incertidumbre inherente en los datos.

Nuestro enfoque se basó en la implementación del modelo bayesiano utilizando JAGS, una herramienta que simplifica la tarea de especificar modelos bayesianos complejos y realizar cadenas de Markov Monte Carlo (MCMC) para muestrear la distribución posterior de los parámetros. Los resultados obtenidos mostraron una concordancia notable entre los valores estimados por nuestro modelo y los datos observados, lo que respalda la validez del modelo logístico en este contexto particular.

Además, al calcular estadísticas descriptivas y generar gráficos que representan la incertidumbre en la función de crecimiento, pudimos proporcionar una visión completa de la población de aligátores a lo largo del tiempo. Esta información es valiosa para la gestión y conservación de especies, así como para la comprensión de los factores que influyen en el crecimiento de la población.

Hemos discutido las aplicaciones de los modelos Bayesianos en una variedad de campos, desde ecología y biología hasta economía, epidemiología, ciencia de datos e inteligencia artificial. Estos modelos tienen la capacidad de abordar preguntas complejas y proporcionar información valiosa para la toma de decisiones en diversas disciplinas.

Es importante destacar que, si bien hemos obtenido resultados prometedores en este estudio, la estadística bayesiana también tiene sus limitaciones y desafíos. La elección de las distribuciones prior puede influir en gran medida en los resultados, y es fundamental considerar cuidadosamente la especificación del modelo. Además, la ejecución de cadenas MCMC puede ser computacionalmente intensiva en casos de modelos altamente complejos o conjuntos de datos masivos.

En última instancia, esta evidencia subraya la importancia de la colaboración interdisciplinaria, ya que la aplicación de modelos Bayesianos en contextos biológicos y ecológicos requiere una comprensión profunda de los dominios específicos, en este caso, la biología de los aligátores.

Referencias

-The Editors of Encyclopaedia Britannica. (2011, 19 septiembre). *Bayesian Analysis | Probability Theory, Statistical Inference*. Encyclopedia Britannica. https://www.britannica.com/science/Bayesian-analysis

-Nuzzo, R. (2017). An introduction to Bayesian Data Analysis for Correlations. *Pm&r*, *9*(12), 1278-1282. https://doi.org/10.1016/j.pmrj.2017.11.003

-van de Schoot R, Kaplan D, Denissen J, Asendorpf JB, Neyer FJ, van Aken MAG. A gentle introduction to bayesian analysis: applications to developmental research. Child Dev. 2014 May-Jun;85(3):842-860. doi: 10.1111/cdev.12169. Epub 2013 Oct 9. PMID: 24116396; PMCID: PMC4158865.

-Van De Schoot, R., Kaplan, D. H., Denissen, J. J. A., Asendorpf, J. B., Neyer, F. J., & Van Aken, M. A. G. (2013). A gentle introduction to Bayesian Analysis: Applications to Developmental Research. *Child Development*, *85*(3), 842-860. https://doi.org/10.1111/cdev.12169

-Hammond, P. S., Francis, T. B., Heinemann, D., Long, K. J., Moore, J. E., Punt, A. E., Reeves, R. R., Sepúlveda, M., Sigurðsson, G. M., Siple, M. C., Víkingsson, G. A., Wade, P. R., Williams, R., & Zerbini, A. N. (2021). Estimating the abundance of marine mammal populations. *Frontiers in Marine Science*, *8*. https://doi.org/10.3389/fmars.2021.735770