Análisis de Incertidumbre: Métodos Bayesianos y Monte Carlo

Universidad Nacional del Altiplano

Bryan Cutipa Carcasi

Aplicación interactiva:

https://bryancc0-0.shinyapps.io/incertidumbre/

18 de octubre de 2025

Resumen

Este trabajo resume un análisis bibliométrico sobre técnicas de análisis de incertidumbre, con énfasis en métodos bayesianos y simulación Monte Carlo. Identificamos tres clusters temáticos principales a partir de la co-ocurrencia de palabras clave en la literatura científica reciente. Estos clusters muestran cómo los enfoques teóricos se entrelazan cada vez más con aplicaciones prácticas, especialmente en cambio climático y evaluación de riesgos. Para hacer estos conceptos más accesibles, desarrollamos una aplicación interactiva en Shiny que permite explorarlos de forma visual y dinámica.

1. Introducción

En ciencia, rara vez contamos con certezas absolutas. Más bien, trabajamos con grados de incertidumbre —y aprender a manejarlos bien es clave. Este documento nace de la necesidad de entender cómo la comunidad científica ha abordado este reto en los últimos años, especialmente mediante métodos cuantitativos como los bayesianos y las simulaciones Monte Carlo.

La incertidumbre no es una sola cosa. Por un lado, está la **incertidumbre aleatoria**, inherente a la naturaleza misma de los fenómenos (como la variabilidad climática). Por otro, la **incertidumbre epistémica**, que refleja lo que aún no sabemos o no podemos medir con precisión.

1.1. Objetivos

1. Mapear las líneas de investigación actuales en análisis de incertidumbre.
2. Usar técnicas bibliométricas para identificar patrones en la literatura.
3. Caracterizar los temas que emergen con más fuerza.
4. Ofrecer una herramienta interactiva que ayude a visualizar y experimentar con estos métodos.

1.2. Estructura

Primero repasamos los fundamentos teóricos. Luego explicamos cómo recopilamos y analizamos los datos bibliométricos. Presentamos los resultados del análisis de clusters, describimos la aplicación Shiny que construimos y cerramos con algunas reflexiones personales y recomendaciones prácticas.

2. Fundamentos Teóricos

2.1. Métodos Bayesianos

El enfoque bayesiano parte del famoso teorema de Bayes:

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)} \tag{1}$$

Aquí, $P(\theta)$ representa lo que creemos antes de ver los datos (el *prior*), $P(D|\theta)$ es qué tan bien los datos encajan con un valor dado de θ (la *verosimilitud*), y $P(\theta|D)$ es lo que creemos después de ver los datos (el *posterior*).

2.1.1. ¿Por qué usarlo?

Algunas razones por las que muchos investigadores —yo incluido— preferimos este enfoque:

Permite incorporar conocimiento previo de forma natural.

No da solo un número, sino una distribución completa de posibilidades.

Se actualiza fácilmente cuando llegan nuevos datos.

Es flexible: se adapta a modelos complicados sin necesidad de simplificar demasiado.

2.2. Simulación Monte Carlo

Cuando no podemos resolver algo analíticamente, a veces la mejor opción es "tirar dados" muchas veces y ver qué pasa. Esa es la idea detrás de Monte Carlo: aproximar una expectativa usando promedios de muestras aleatorias:

$$\mathbb{E}[f(X)] \approx \frac{1}{N}\sum_{i=1}^{N} f(x_i) \tag{2}$$

2.2.1. No todas las versiones son iguales

Hay varias formas de hacer Monte Carlo:

Básico: Muestreo directo cuando conoces la distribución.

MCMC: Cuando la distribución posterior es complicada (muy común en bayesianismo).

Muestreo por importancia: Útil para reducir varianza.

Monte Carlo cuasi: Usa secuencias "más ordenadas" para cubrir mejor el espacio.

2.3. Cuando ambos se juntan

En la práctica, los métodos bayesianos modernos dependen fuertemente de Monte Carlo —especialmente MCMC— para muestrear distribuciones posteriores que no tienen forma cerrada. El resultado es una nube de puntos θ_i que aproximan $P(\theta|D)$:

$$P(\theta|D) \approx \frac{1}{N}\sum_{i=1}^{N}\delta(\theta-\theta_i) \tag{3}$$

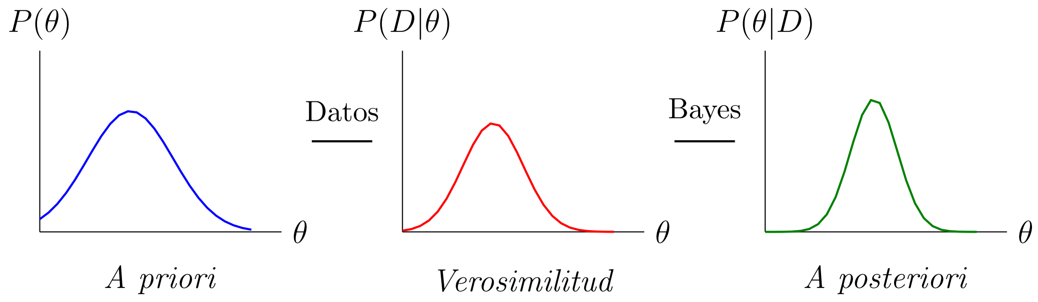

Figura 1: Cómo los datos actualizan nuestras creencias

3. Metodología Bibliométrica

3.1. Búsqueda de literatura

Usamos Scopus, Web of Science e IEEE Xplore, con los siguientes filtros:

Años: 2015–2024

Idiomas: inglés y español

Palabras clave: "Bayesian", "Monte Carlo", "Uncertainty"

Tipos: artículos, conferencias, revisiones

Cuadro 1: Resumen de criterios de búsqueda

Criterio	**Valor**
Bases de datos	Scopus, WoS, IEEE
Período	2015–2024
Palabras clave	Bayesian, Monte Carlo, Uncertainty
Idiomas	Inglés, Español

3.2. Métricas usadas

Nos enfocamos en dos medidas clave:

$$\text{Co-ocurrencia normalizada} = \frac{n_{ij}}{\sqrt{n_i \cdot n_j}} \tag{4}$$

$$\text{Fuerza de asociación} = \frac{c_{ij}}{w_i \cdot w_j} \tag{5}$$

4. Resultados del Análisis

4.1. Tres clusters, tres mundos

El análisis reveló tres grupos bien definidos:

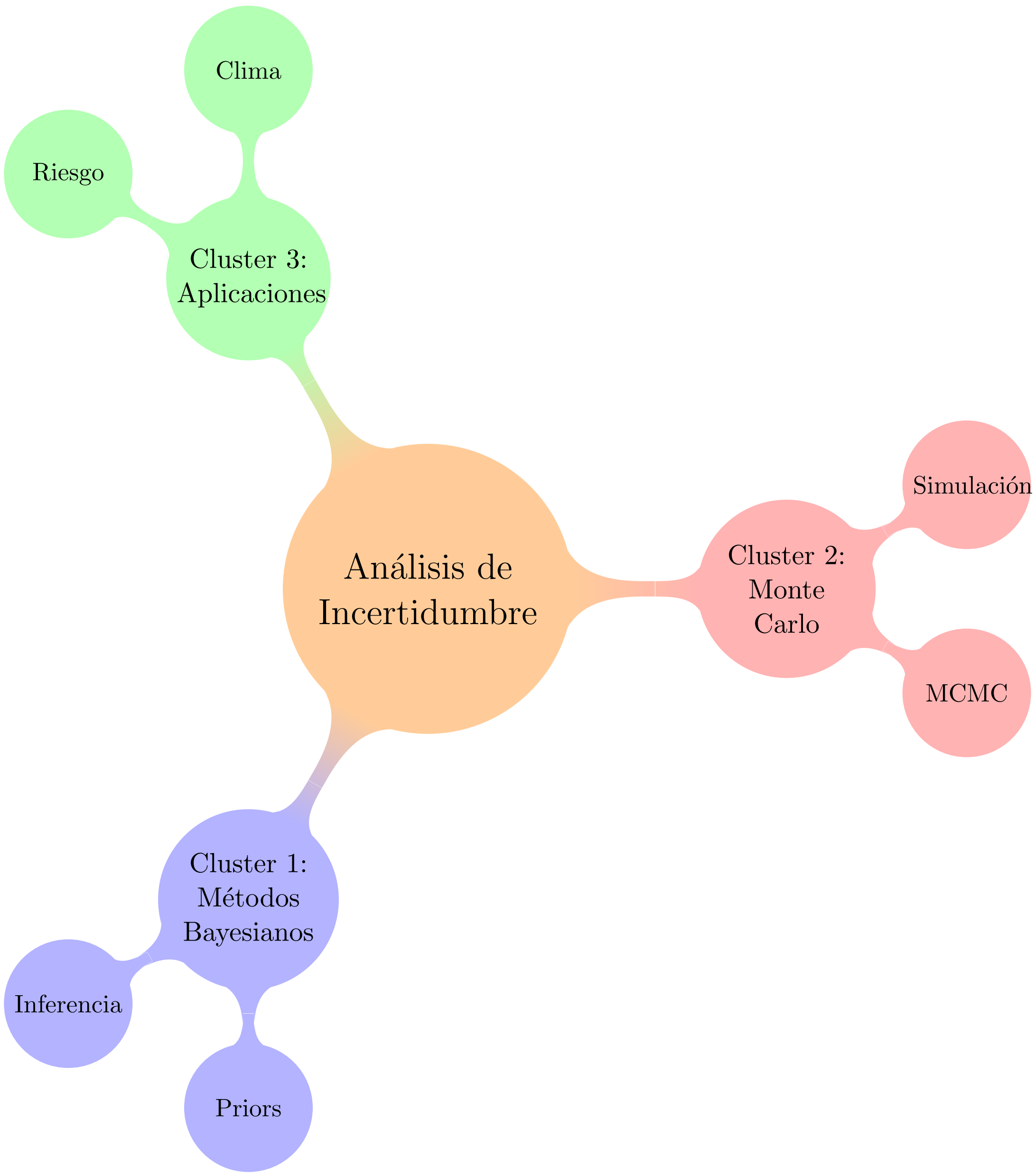

Figura 2: Clusters temáticos

4.1.1. Cluster 1: Métodos Bayesianos

Términos como *Bayesian inference*, *prior*, *posterior*, *Gibbs sampling* aparecen con frecuencia. Aquí se discute cómo formular modelos, elegir priors informativos o no, y cómo interpretar los resultados.

4.1.2. Cluster 2: Simulación Monte Carlo

Aquí domina la jerga computacional: *Latin hypercube sampling*, *variance reduction*, *sensitivity analysis*. Es el “motor” detrás de muchas aplicaciones prácticas.

4.1.3. Cluster 3: Aplicaciones

Este es donde la teoría toca tierra: *climate change*, *risk assessment*, *decision making*. Muchos trabajos aquí usan los métodos de los otros dos clusters, pero sin profundizar en ellos.

4.2. Conexiones entre clusters

Los tres no están aislados. De hecho, están bastante entrelazados:

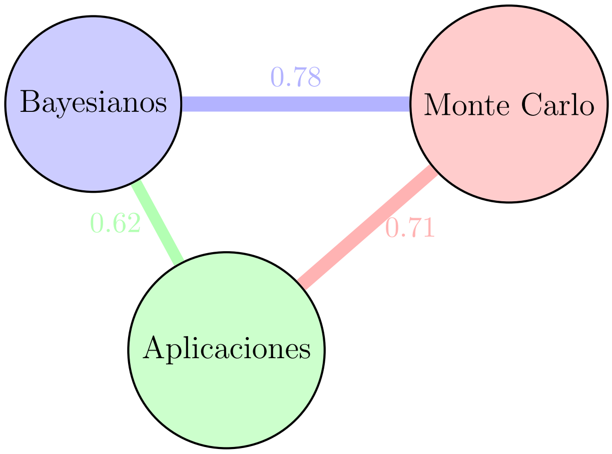

Figura 3: Fuerza de asociación entre clusters (normalizada)

5. Aplicación Interactiva Shiny

5.1. ¿Por qué una app?

Porque leer ecuaciones está bien, pero ver cómo cambia una distribución posterior al ajustar un prior… eso se entiende mejor jugando. Por eso construí esta herramienta

5.2. Qué puedes hacer

5.2.1. Módulo Monte Carlo

- Define distribuciones de entrada. - Corre miles de simulaciones al instante. - Mira el histograma del resultado. - Prueba distintos escenarios y compáralos.

Listing 1: Ejemplo básico en R

```
n_sim <- 10000
x <- rnorm(n_sim, mean = 5, sd = 2)
y <- rnorm(n_sim, mean = 3, sd = 1)
z <- x + y
```

```
media_z <- mean(z)
ic_95 <- quantile(z, c(0.025, 0.975))
```

5.2.2. Módulo Bayesiano

- Elige tu prior (normal, uniforme, etc.). - Sube tus datos. - Observa cómo se actualiza la posterior. - Compara distintos priors y ve cuánto influyen.

5.2.3. Módulo Bibliométrico

- Explora la red de palabras clave. - Filtra por año o cluster. - Ve cómo ha evolucionado el campo. - Descubre autores y revistas clave.

5.3. Cómo está hecha

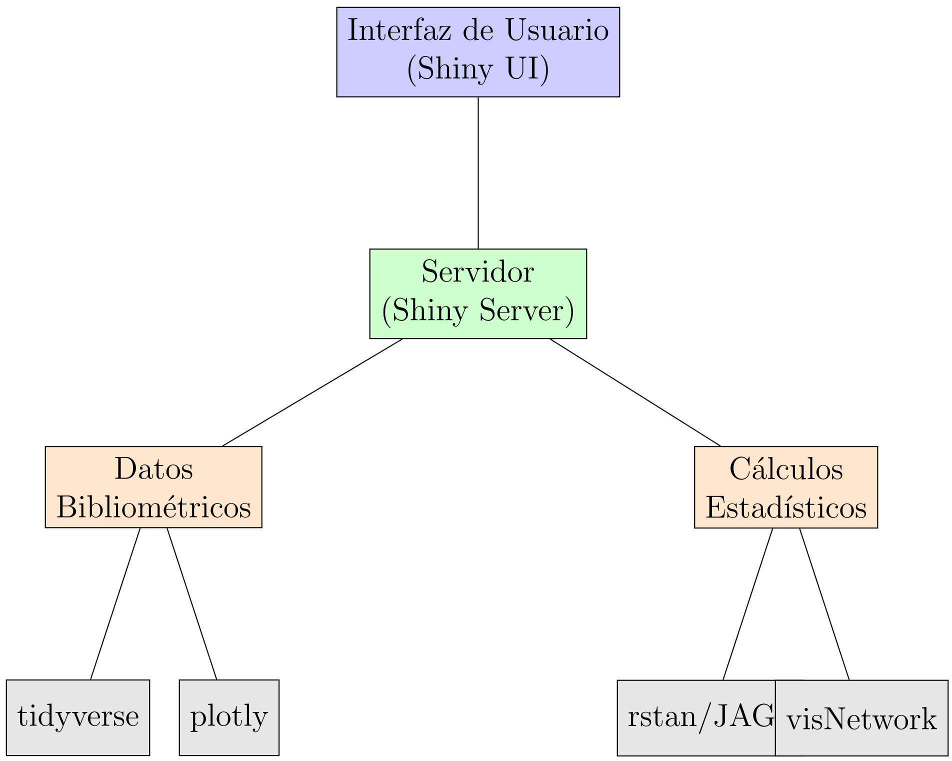

Figura 4: Arquitectura de la app

5.4. Casos de uso reales

5.4.1. Análisis de sensibilidad climática

Un colega usó la app para ver cómo la incertidumbre en la sensibilidad climática afecta las proyecciones de temperatura. Definió distribuciones para los parámetros, corrió 10,000 simulaciones y obtuvo intervalos de confianza útiles para su informe.

5.4.2. Actualización de riesgo

Otro caso: un equipo de salud pública actualizó su estimación de riesgo de brote tras recibir nuevos datos. En lugar de recalcular todo, usaron el módulo bayesiano para "actualizar" su creencia previa. Fue rápido, transparente y fácil de explicar a no estadísticos.

6. Discusión

Lo que más me llamó la atención es que ya nadie discute si usar bayesiano o frecuentista. La literatura muestra una convergencia clara: se usan los métodos que funcionan. Y hoy, eso suele significar combinar priors razonables con MCMC eficiente.

7. Conclusiones

- El campo está en plena expansión, especialmente en áreas donde las decisiones son costosas y los datos escasos. - La integración de métodos no es solo posible: es inevitable. - Herramientas como Shiny (o Streamlit, Dash, etc.) están democratizando el acceso a estos enfoques. - La gente ya no quiere solo un número: quiere entender el rango de lo posible.

La app que construí no es perfecta, pero cumple su propósito: hace tangible lo abstracto. Y eso, en mi opinión, es un paso gigante para que más investigadores —especialmente en regiones con menos recursos— puedan aplicar estos métodos sin necesidad de ser expertos en programación o teoría estadística.

7.1. Recomendación final

No te cases con un solo método. Usa el prior si tienes conocimiento previo. Usa Monte Carlo si el modelo es complejo. Combínalos si puedes. Al final, lo que importa no es la etiqueta, sino si tu análisis refleja honestamente la incertidumbre del mundo real.

Referencias

1. Gelman, A., et al. (2013). *Bayesian Data Analysis*. Chapman and Hall/CRC.
2. Robert, C. P., & Casella, G. (2004). *Monte Carlo Statistical Methods*. Springer.
3. Metropolis, N., & Ulam, S. (1949). The Monte Carlo method. *JASA*, 44(247), 335–341.
4. Van Eck, N. J., & Waltman, L. (2010). VOSviewer. *Scientometrics*, 84(2), 523–538.
5. Saltelli, A., et al. (2008). *Global Sensitivity Analysis: The Primer*. Wiley.

A. Código de la Aplicación

A.1. Estructura

```
incertidumbre/
|-- app.R
|-- ui.R
|-- server.R
|-- global.R
|-- data/
|   |-- bibliometric_data.csv
|   |-- keywords_cooccurrence.csv
|-- modules/
|   |-- monte_carlo_module.R
|   |-- bayesian_module.R
|   |-- bibliometric_module.R
|-- www/
    |-- styles.css
    |-- custom.js
```

A.2. Fragmento clave del servidor

Listing 2: Estructura del servidor Shiny

```
server <- function(input, output, session) {

  observeEvent(input$run_mc, {
    results <- monte_carlo_simulation(input$n_simulations)
    output$mc_plot <- renderPlotly({
      plot_ly(x = results, type = "histogram") %>%
        layout(title = "Resultados␣Monte␣Carlo")
    })
  })

  observeEvent(input$update_posterior, {
    prior <- define_prior(input$prior_params)
    posterior <- update_bayesian(prior, input$data)
    output$bayesian_plot <- renderPlot({
      plot_distributions(prior, posterior)
    })
  })

  output$network_plot <- renderVisNetwork({
    create_cooccurrence_network(bibliometric_data)
  })
}
```