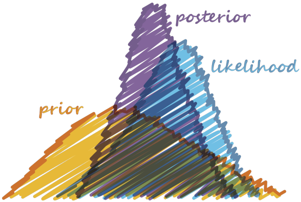
Introducción a la **Inferencia Bayesiana** con **RStan**



Ignacio Evangelista
evangelistaignacio@gmail.com

Grupo de Usuaries de R en Rosario – 5$^{\text{to}}$ Encuentro

Agosto 2019

Contenido

Sobre apuestas...

Les ofrezco la siguiente tarjeta:

> Páguese $100 al portador de esta tarjeta si en esta sala **hay** alguien que se llama Wenceslao

¿Cuánto estarían dispuestos a pagar (como máximo) por ella?

> Páguese $100 al portador de esta tarjeta si en esta sala **no hay** alguien que se llama Wenceslao

¿Cuánto estarían dispuestos a pagar (como máximo) por ella?

Nota: el máximo precio a pagar es también el mínimo precio por el que venderían

Sobre apuestas...

Les ofrezco la siguiente tarjeta:

> Páguese \$100 al portador de esta tarjeta si en esta sala **hay** alguien que se llama Pablo

¿Cuánto estarían dispuestos a pagar (como máximo) por ella?
Supongo que pagarían más que por la de Wenceslao...

Todos pagaríamos un valor $p \cdot \$100$ con $0 < p < 1$

Sobre apuestas...

Si esta charla es buena, páguese al portador \$100

Si esta charla es mala, páguese al portador \$100

Por una de ellas pagarían $p \cdot \$100$ y por la otra pagarían $q \cdot \$100$, con $q = 1 - p$.

Si no, por ejemplo: pagan \$60 por cada una. Pierden \$20 si la charla es buena y pierden \$20 si la charla es mala. **No es racional**.

Argumento del *Dutch Book*

Decidimos cuánto apostar en función de nuestra incertidumbre en la ocurrencia de un evento (de lo plausible que lo consideremos). Decidimos apostar p·\$100 en favor de un evento, porque le asignamos una plausibilidad de grado p.

El Argumento del *Dutch Book*

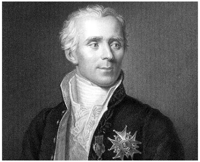
Si queremos cuantificar nuestra incertidumbre y tomar decisiones racionalmente, lo mejor que tenemos para utilizar son **probabilidades**. Si no, somos susceptibles de caer en un *Dutch Book*.

En el mundo bayesiano, las probabilidades representan nuestro estado de conocimiento del mundo (probabilidades epistémicas, más que subjetivas).

Regla de Bayes

Richard Price publicó el resultado del Teorema de la Probabilidad Inversa de **Thomas Bayes** en *An Essay Towards Solving a Problem in the Doctrine of Chances* (1763)

Pierre-Simon Laplace llegó al mismo resultado que Bayes y lo publicó en *Memoire sur la Probabilité des Causes par les Évenements* (1774). En 1814 publicó *Essai Philosophique sur les Probabilités* donde desarrolló una estructura de probabilidad para razonar con la incertidumbre (lo que hoy se conoce como **inferencia bayesiana**).



"La Teoría de las Probabilidades no es más que sentido común llevado al cálculo" – Pierre-Simon Laplace

Regla de Bayes

La mejor forma de actualizar nuestro grado de creencia sobre alguna hipótesis $\mathcal{H}$ frente a nueva información E es utilizar la regla de Bayes:

$$P(\mathcal{H}|E) = \frac{P(E|\mathcal{H})P(\mathcal{H})}{P(E)}$$

Idea central

Las probabilidades son la mejor herramienta disponible para cuantificar la incertidumbre y las leyes de la probabilidad, la mejor herramienta para operar con ella.

La probabilidad en el mundo bayesiano

- ¡Es razonable utilizar probabilidades para modelizar incertidumbre!
- Las probabilidades representan nuestro conocimiento actual del mundo
- Los parámetros tienen distribuciones de probabilidad que se actualizan con los datos
- Los datos son fijos

Inferencia bayesiana

La inferencia se basa en la Regla de Bayes:

$$\underbrace{p(\theta|D)}_{\textit{posterior}} = \frac{\overbrace{p(D|\theta)}^{\textit{likelihood}}\;\overbrace{p(\theta)}^{\textit{prior}}}{\underbrace{p(D)}_{\textit{marginal likelihood}}}$$

$$p(D) = \int_{\theta} p(D|\Theta = \theta)p(\Theta = \theta)d\theta$$

En forma simple:

$$\text{posterior} \propto \text{likelihood} \times \text{prior}$$

Inferencia bayesiana

Un modelo bayesiano para inferencia consiste en:

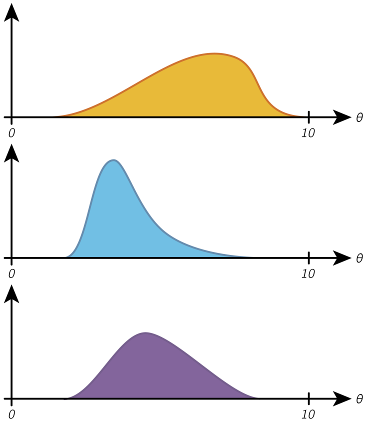
- Parámetros: que conocemos o que deseamos estimar
- *Prior*: plausibilidad inicial para cada parámetro desconocido
- *Likelihood*: el número de formas en que cada conjunto de parámetros puede producir una observación
- *Posterior*: credibilidad actualizada luego de obsrvar los datos

Hay dos elecciones que hacer:

- Un modelo de generación de datos (que depende de ciertos parámetros)
- Una probabilidad a priori para cada parámetro del modelo

Ejemplo conceptual

Supongamos que ustedes evalúan mi capacidad para explicar lo que es la inferencia bayesiana...



Observaciones

- Si el *prior* o el *likelihood* vale 0 para algún valor de θ, lo mismo ocurrirá con el *posterior*. *Una persona que está negada*.
- Si el *prior* es muy angosto, el *posterior* se acercará al *prior*. *Difícil cambiar la opinión de un terco*.
- Si el *likelihood* es muy anogosto, el *posterior* se acercará a la estimación del *likelihood*. *Evidencia abrumadora*.

"A Bayesian is someone who, vaguely expecting a horse, and catching a glimpse of a donkey, strongly believes he has seen a mule"

Idea central

La inferencia bayesiana es la realocación de la credibilidad del conjunto de parámetros de un modelo, una vez observado un conjunto de datos.

Crítica de la estadística *clásica*

Lo que escandaliza a los estadísticos *clásicos* u *ortodoxos* es la necesidad de escoger un *prior* (elección arbitraria o subjetiva)

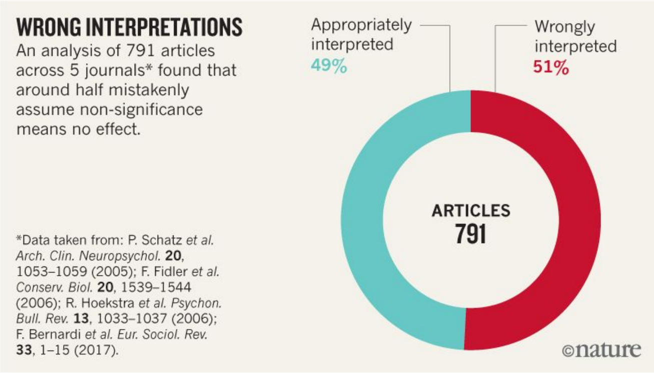
- Todos los modelos parten de algún supuesto y las conclusiones son condicionales a esos supuestos
- La inferencia bayesiana obliga a un investigador a explicitar la elección de su *prior* para que su modelo sea confrontado por otros
- Es posible comparar modelos que partan de diferentes *priors*
- Con suficientes datos, todas las inferencias llegan a los mismos resultados

Puntos interesantes de la inferencia bayesiana

- Incorporación de conocimiento previo (*prior*)
- Cuantificación de la incertidumbre en las inferencias y predicciones realizadas
- La inferencia bayesiana arroja resultados a las preguntas que los investigadores hacen: $P(\mathcal{H}|D)$
- La interpretación de los resultados de la inferencia bayesiana es intuitiva
- Sólo utiliza los datos observados (no imagina infinitas repeticiones)

"Dentro de cada no bayesiano hay un bayesiano luchando por salir"
– Dennis Lindley

Amhrein, V. et al. 2019. *Scientists rise up against statistical significance*. Nature.



¿Cómo se hace inferencia bayesiana?

- Exacta (lápiz y papel). Aprovechando distribuciones conjugadas.
- Aproximación *grid*
- Simulación

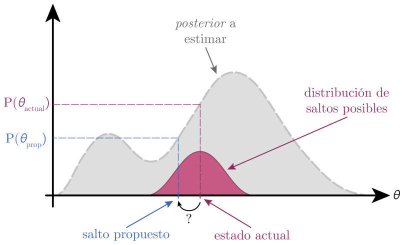
La evidencia (*likelihood* marginal) puede ser imposible de calcular analíticamente (espacio de gran dimensión). Las aproximaciones grid también se vuelven imposibles...

Markov chain Monte Carlo

Monte Carlo es un término acuñado entre John von Neumann, Nicholas Metropolis y Stanislav Ulam (en un artículo de 1949)

MCMC permite aproximar el *posterior* sin calcular la integral de la Regla de Bayes.

Metropolis-Hastings



1. Dado un estado actual, proponer un salto según una determinada distribución
2. Computar $f(\theta_{prop}) \propto P(\theta_{prop})$
3. Saltar al estado propuesto con probabilidad:

$$\min\left(\frac{P(\theta_{prop})}{P(\theta_{actual})}; 1\right)$$

4. Repetir

- El algoritmo de **Metropolis-Hastings** converge a una distribución límite estacionaria para cualquier estado inicial y para cualquier distribución de saltos posibles.
- Para que la convergencia sea en un *tiempo razonable*, es necesario *tunear* la distribución de saltos posibles.
- En muchas dimensiones se puede generalizar la distribución de saltos posibles. De todas formas, esta distribución está centrada en la posición actual, lo cual puede originar ineficiencias. Se complica todo... (el algoritmo de **Gibbs** ayuda un poco)

Hamiltonian Monte Carlo

Donde la magia comienza...

¿De qué se trata?:

- Pensar el *negative log-posterior* como una función energía potencial. Para proponer un salto, se le da un impulso a una bolita imaginaria (energía cinética) y se ve dónde termina al cabo de un tiempo.
- Hay que resolver las Ecuaciones de Hamilton (un sistema de ecuaciones diferenciales)...
- Muy eficiente para todo tipo de distribuciones
- ***No free lunch***

¿Cómo hacemos?

- Programar el algoritmo de Metropolis-Hastings o el de Gibbs es una tarea realizable (no es trivial su optimización)
- Hamiltonian Monte Carlo...
- Otras variantes: inferencia variacional, *expectation-propagation*

Por suerte existen lenguajes de programación probabilísticos como PyMC, Julia, Edward o **Stan**.

Idea central

La inferencia bayesiana exacta (*fully bayesian*) es difícil (imposible) salvo en casos sencillos. Existen métodos para aproximar el *posterior*: los algoritmos de MCMC.

¿Qué es Stan?

- **Stan** es un lenguaje de programación probabilístico imperativo
- Está desarrollado sobre C++
- Los programas de **Stan** definen un modelo probabilístico (datos + parámetros) y realizan inferencias sobre él
- Es robusto: en general funciona. Si no funciona; dice por qué
- Tiempos de cálculo razonables
- Capacidad de procesamiento multinúcleo
- *Open source*

`https://mc-stan.org/`

¿Qué es RStan?

- **RStan** es una interfaz de Stan en R
- Permite compilar y ejecutar modelos de Stan directamente en R
- Además, incluye la clase `stanfit` y funciones para operar con ella

`https://mc-stan.org/users/interfaces/rstan`

Pasos a seguir

1. Escribir un programa en Stan (archivo `.stan` o `string` en una variable)
2. Usando la función `rstan::stan_model`, generar el código C++, compilarlo y generar un objeto `stanmodel`
3. Usando la función `rstan::sampling`, obtener muestras del *posterior* dado el `stanmodel` y los datos. Las muestras quedan en un objeto `stan_fit`
4. Procesar el `stan_fit` y sacar conclusiones

Estructura de un modelo de Stan

```
data{
}
transformed data{
}
parameters{
}
transformed parameters{
}
model{
}
generated quantities{
}
```

Lecturas sugeridas

- Ghahramani, Z. (2015). *Probabilistic machine learning and artificial intelligence*. Nature, 521(7553), 452.
- Gelman, A. y Shalizi, C. R. (2013). *Philosophy and the practice of Bayesian statistics*. British Journal of Mathematical and Statistical Psychology, 66(1), 8-38.
- Lindley, D. V. (2000). *The philosophy of statistics*. Journal of the Royal Statistical Society: Series D (The Statistician), 49(3), 293-337.
- Lindley, D. V. (2013). *Understanding uncertainty*. John Wiley & Sons.
- DeGroot, M. H. (2005). *Optimal statistical decisions (Vol. 82)*. John Wiley & Sons.
- O'Hagan, T. (2004). *Dicing with the unknown*. Significance, 1(3), 132-133.

Lecturas sugeridas

- Gelman, A. y otros. (2013). *Bayesian Data Analysis*. Chapman and Hall/CRC.
- Reich, B. J. y Ghosh, S. K. (2019). *Bayesian Statistical Methods*. Chapman and Hall/CRC Press.
- McElreath, R. (2015). *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. CRC Press

Lecturas sugeridas

- Carpenter, B., Gelman, A. y otros. (2017). S*tan: A probabilistic programming language*. Journal of statistical software, 76(1).
- Stan Development Team. (2018). *Stan Modeling Language Users Guide and Reference Manual, Version 2.18.0.*

RStan

¡Allá vamos!