

Métodos de Montecarlo basados en cadenas de Markov

Los métodos de Montecarlo basados en cadenas de Markov (MCMC) se utilizan principalmente para calcular aproximaciones numéricas de integrales múltiples. En los modelos bayesianos deseamos calcular la probabilidad a posteriori definida como

$$\Pr(\boldsymbol{\theta} \mid \text{datos}) = \frac{\Pr(\text{datos} \mid \boldsymbol{\theta}) \Pr(\boldsymbol{\theta})}{\Pr(\text{datos})}, \quad \boldsymbol{\theta} = (\theta_1, \dots, \theta_r),$$

con $\Pr(\text{datos})$ caracterizada como

$$\Pr(\text{datos}) = \int_{\theta_1} \cdots \int_{\theta_r} \Pr(\text{datos} \mid \theta_1, \dots, \theta_r) \Pr(\theta_1, \dots, \theta_r) d\theta_1 \cdots d\theta_r.$$

Los métodos de Montecarlo basados en cadenas de Markov (MCMC) integran el concepto de cadena de Markov con la simulación de Montecarlo:

- Cadena. Cada iteración depende de los resultados anteriores. Es decir, los resultados en la iteración (s) se obtienen a partir de los resultados previos de las iteraciones $(1), (2), \dots, (s-1)$:

$$p(\boldsymbol{\theta}^{(s)} \mid \boldsymbol{\theta}^{(1)}, \dots, \boldsymbol{\theta}^{(s-1)}).$$

- Cadena de Markov. De todas las posibles formas de definir una cadena, empleamos las cadenas de Markov, que son aquellas donde el resultado de la iteración (s) solo depende del resultado de la iteración anterior ($s-1$):

$$p(\boldsymbol{\theta}^{(s)} \mid \boldsymbol{\theta}^{(1)}, \dots, \boldsymbol{\theta}^{(s-1)}) = p(\boldsymbol{\theta}^{(s)} \mid \boldsymbol{\theta}^{(s-1)}).$$

- Montecarlo. La esperanza, que es una integral, se aproxima muy bien mediante un muestreo aleatorio:

$$E_{p(\boldsymbol{\theta} \mid x)}(g(\boldsymbol{\theta})) \approx \frac{1}{S} \sum_{s=1}^S g(\boldsymbol{\theta}^{(s)}).$$

- Montecarlo basado en cadenas de Markov: el resultado de la iteración (s) es un valor aleatorio generado por la función de densidad determinada en el resultado del paso anterior ($s-1$).

Uno de los trucos del MCMC consiste en eliminar la verosimilitud media ($\Pr(\text{datos})$) en nuestros cálculos. El otro truco conlleva muestrear para evitar la autocorrelación markoviana.

Supongamos que deseamos comparar dos posibles valores de θ , θ_1 y θ_2 . Entonces

$$\Pr(\theta = \theta_1 \mid \text{datos}) = \frac{\Pr(\text{datos} \mid \theta = \theta_1) \Pr(\theta = \theta_1)}{\Pr(\text{datos})},$$

$$\Pr(\theta = \theta_2 \mid \text{datos}) = \frac{\Pr(\text{datos} \mid \theta = \theta_2) \Pr(\theta = \theta_2)}{\Pr(\text{datos})}.$$

Si dividimos, resulta que la expresión de verosimilitud media $\Pr(\text{datos})$ ya no aparece en las cuentas:

$$R = \frac{\Pr(\theta = \theta_1 \mid \text{datos})}{\Pr(\theta = \theta_2 \mid \text{datos})} = \underbrace{\frac{\Pr(\text{datos} \mid \theta = \theta_1)}{\Pr(\text{datos} \mid \theta = \theta_2)}}_{\text{Factor de Bayes}} \cdot \frac{\Pr(\theta = \theta_1)}{\Pr(\theta = \theta_2)}.$$

El algoritmo MCMC calcula el ratio de probabilidades

$$R = \frac{\Pr(\theta = \theta_1 \mid \text{datos})}{\Pr(\theta = \theta_2 \mid \text{datos})}$$

que indica qué valor resulta más probable, verosímil o creíble.

La decisión de cambiar de $\theta^{(i-1)}$ y seleccionar el nuevo valor candidato $\theta^{(*)}$ como el valor de la siguiente iteración $\theta^{(i)}$ en vez de mantenerse con $\theta^{(i-1)}$ se realiza al azar ponderando las probabilidades del candidato y del valor actual. Existen diversas aproximaciones basadas en este procedimiento. Nosotros estudiaremos tres de ellas: el algoritmo Metropolis, el método de Gibbs y el Montecarlo hamiltoniano.

En el algoritmo Metropolis el valor candidato $\theta^{(*)}$ para la próxima iteración es propuesto mediante una función de salto simétrica (habitualmente la distribución Uniforme o la Normal), mientras que en el método de Gibbs el valor candidato se obtiene a partir de la función de densidad condicionada (lo que requiere saber calcularla analíticamente), y finalmente, el método Montecarlo hamiltoniano (Hamiltonian Montecarlo, HMC) simula un sistema físico ideal (sin pérdida de energía) involucrando movimientos y posiciones acordes a los datos disponibles y los valores candidatos son los que maximizan la energía del movimiento.

En el caso de Metropolis, el azar determina si el valor candidato resulta seleccionado para la siguiente iteración, mientras que en el método de Gibbs el valor candidato siempre se convierte en el valor para la siguiente iteración y el HMC cuenta una tasa teórica de aceptación del 100 % pero debido a las aproximaciones numéricas, suele mostrar tasas de aceptación en torno a un 90 % de la veces.

El método más eficiente a nivel práctico es el método de Gibbs, pero conlleva conocer las distribuciones condicionadas y esto solo resulta posible en algunas situaciones. Actualmente el HMC se ha convertido en el método estándar para la Estadística Bayesiana.