MAURICIO TORO

21-01-2022

FACULTAD DE MATEMÁTICAS
PONTIFICIA UNIVERSIDAD
CATÓLICA DE CHILE

Examen de Calificación: EPG3220 Inferencia Bayesiana

1. Sea S el espacio muestral y Θ el espacio de los parámetros de un modelo estadístico. Muestre que si ambos $(S, \mathcal{S})$ y $(\Theta, \mathcal{B})$ son espacios de Borel estándar, entonces el conjunto de distribuciones a posteriori $P(\theta \in B \mid X = x)$, donde $x \in S_0$ para un conjunto medible $S_0 \subset S$ de probabilidad marginal 1 , es dominada por una medida σ - finita si y sólo si el conjunto de medidas de probabilidad $\{P_\theta : \theta \in \Theta_0\}$ es dominada por una medida σ-finita, para algún conjunto medible $\Theta_0 \subset \Theta$, con $\Pi(\Theta_0) = 1$. Muestre además que en este caso, las distribuciones a posteriori son dominadas por la a priori. **(2.0 puntos)**

2. Considere la secuencia infinita de variables aleatorias $X_1, X_2, \ldots$, tal que

$$X_i \in \{1, 2, \ldots, k\}.$$

Demuestre que $(X_1, X_2, \ldots)$ es intercambiable si y solo si existe una distribución Π tal que para todo n, la distribución conjunta de $(X_1, \ldots, X_n)$ tiene la representación de mezcla

$$\Pr(X_1 = x_1, \ldots, X_n = x_n) = \binom{n}{h_1 \cdots h_n} \int_{\Delta_{k-1}} \left\{ \prod_{j=1}^{k} \theta_j^{h_j} \right\} d\Pi(\boldsymbol{\theta}),$$

donde Δ_{k-1} denota el simplex de dimensión $k-1, h_j = \sum_{i=1}^{n} I\left(x_i\right)_j, \boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$, Π es una función de distribución tal que

$$\Pi(\boldsymbol{\theta}) = \lim_{n \to \infty} \Pr\left(\frac{h_1}{n} \leq \theta_1, \ldots, \frac{h_k}{n} \leq \theta_k \right).$$

(2.0 puntos)

3. Suponga el modelo Bayesiano dado por $p(\boldsymbol{y} \mid \boldsymbol{\theta})$ y $\pi(\boldsymbol{\theta})$.

(a) Describa la forma general del algoritmo de Metropolis-Hastings para explorar la distribución a posteriori $\pi(\boldsymbol{\theta} \mid \boldsymbol{y})$. **(0.5 puntos)**

(b) Encuentre el kernel de transición del algoritmo de Metropolis-Hastings. **(0.5 puntos)**

(c) Muestre que el Metropolis-Hastings satisface la condición de reversibilidad. **(0.5 puntos)**

(d) Muestre que el muestreador de Gibbs es un caso particular del algoritmo de Metropolis-Hastings. **(0.5 puntos)**

Indicaciones: Durante el examen usted no puede usar ningún tipo de material del curso.

– Santiago, 21 de enero de 2022 –

P1) $(S, \mathcal{S}), (\Theta, \mathcal{B})$ espacios de Borel

PDQ las distrib. a posteriori $P(\theta \in B | X = x)$, $x \in S_0 \subset S$, es dominada por una medida σ-finita $\iff$ $\{P_\theta : \theta \in \Theta_0\}$ es dominada por una medida σ-finita para algún $\Theta_0 \subset \Theta$ con $\Pi(\Theta_0) = 1$.

Muestre además que las distrib. a posteriori son dominadas por las a priori.

Sol: Recordemos: • Π definida en $(\Omega, \mathcal{F})$ domina a λ $(\lambda \ll \Pi)$ si

$$\Pi(A) = 0 \Rightarrow \lambda(A) = 0 \quad \forall A \in \mathcal{F}.$$

• medida σ-finita si $\Pi(\Omega) < \infty$

• teorema de Bayes:

$$P(\theta \in B | X = x) = \frac{P(X = x | \theta \in B)}{P(X = x)} \cdot P(\theta \in B)$$

Que las medidas sean σ-aditivas garantiza la existencia de las conjuntas y que podamos usar Fubini. Bayes asume que las condicionales están bien definidas.

$\boxed{\Rightarrow}$

Supongamos que $P(\theta \in B | X = x)$ es dominada por una medida σ-finita, Π. Sin pérdida de generalidad, consideremos Π medida de probabilidad (si no integra 1 se puede escalar). Si $P(X = x) \neq 0$, entonces

$$P(\theta \in B) = P(\theta \in B | X = x) \cdot \frac{P(X = x)}{P(X = x | \theta \in B)}$$

$\Rightarrow P(\theta \in B)$ es también dominada por Π, pues $P(\theta \in B) \ll P(\theta \in B | X = x)$ y $P(\theta \in B | X = x) \ll \Pi(B)$

Si $P(X = x) = 0$ (ej: X abs. continua), se procede de manera análoga pero considerando $P(X \in \Delta_{x,n})$, con $\Delta_{x,n}$ un conjunto que contiene a x y que $\Delta_{x,n} \downarrow \{x\}$. De este modo, se usa el teorema de Bayes sin que se indetermine la división $\forall n$

$$P(\theta \in B \mid X = x) = \lim_{n \to \infty} \frac{P(X \in \Delta_{x,n}, \theta \in B)}{P(X \in \Delta_{x,n})} P(\theta \in B)$$

$$\Rightarrow P(\theta \in B) = \left[\lim_{n \to \infty} \frac{P(X \in \Delta_{x,n}, \theta \in B)}{P(X \in \Delta_{x,n})} \right]^{-1} P(\theta \in B \mid X = x)$$

$\boxed{\Leftarrow}$

Si P_θ es dominada por Π σ-finita y $P(\theta \in B \mid X = x)$ es σ-finita, entonces la conjunta (θ, X) existe y podemos escribir

$$P(\theta \in B \mid X = x) = \frac{P(X = x \mid \theta \in B)}{P(X = x)} \cdot P(\theta \in B)$$

$\Rightarrow P(\theta \in B \mid X = x)$ también es dominada por Π

(Si $P(X = x) = 0$ se usa $P(X \in \Delta_{n,x})$, como se mencionó más arriba)

Más aún, notemos que $P(\theta \in B) = 0 \Rightarrow P(\theta \in B \mid X = x) = 0$.

$\therefore$ $P(\theta \in B)$ domina a $P(\theta \in B \mid X = x)$.

P2 | $X_1, X_2, \ldots$ Tales q.e $X_i \in \{1, 2, \ldots, k\}$

PDQ $(X_1, X_2, \ldots)$ intercambiables ssi

$$\exists \Pi : \forall n \quad P(X_1 = x_1, \ldots, X_n = x_n) = \binom{n}{h_1 \cdots h_n} \int_{\Delta_{k-1}} \left\{ \prod_{j=1}^{k} \theta_j^{h_j} \right\} d\Pi(\underset{\sim}{\theta})$$

con $h_j = \sum_{i=1}^{n} I(X_i = j)$ y $\Pi(\theta) = \lim_{n \to \infty} P\left(\frac{h_1}{n} \leq \theta_1, \ldots, \frac{h_k}{n} \leq \theta_k\right)$.

$\underset{\sim}{h} = (h_1, \ldots, h_k)$

<u>DEM:</u>

$\Leftarrow$ | Sean $X_1, \ldots, X_n$ v.a. y $\tau(i) = \tau_i$ permutación de $X_1, \ldots, X_n$.

$$P(X_1 = x_1, \ldots, X_n = x_n) = \binom{n}{h_1 \cdots h_k} \int_{\Delta_{k-1}} \left\{ \prod_{j=1}^{k} \theta_j^{h_j} \right\} d\Pi(\underset{\sim}{\theta})$$

Puesto que $\underset{\sim}{h}(X_1, \ldots, X_n) = \underset{\sim}{h}(X_{\tau_1}, \ldots, X_{\tau_n})$ y la expresión de la derecha solo depende de la muestra por $h_1, \ldots, h_n$, entonces la expresión de la derecha es igual a $P(X_{\tau_1} = x_{\tau_1}, \ldots, X_{\tau_n} = x_{\tau_n})$ $\forall$ permutación τ_n.

$\therefore$ $X_1, X_2, \ldots$ son intercambiables

$\Rightarrow$ | Sea n cualquiera. Suponga que las v.a. $X_1, \ldots, X_n$ son tales q.e $\underset{\sim}{h}(X_1, \ldots, X_n) = (h_1, \ldots, h_k)$. Por de Finetti, sabemos que existe una medida G tal que $\forall n$

$$X_1, \ldots, X_n \mid G \overset{iid}{\sim} G$$

y la distribución empírica $G_n \to G$. Como $X_i \in \{1, \ldots, k\}$, entonces la distribución empírica para cada categoría es $G_n(\underset{\sim}{\theta}) = P\left(\frac{h_1}{n} \leq \theta_1, \ldots, \leq \frac{h_n}{n} \leq \theta_n\right)$ para algún $\underset{\sim}{\theta}$. Esta

DISTRIBUCIÓN ESTÁ DEFINIDA EN Δ_{K-1} (SINO, HAY PROBLEMAS DE IDENTIFICABILIDAD).

COMO SE COMENTÓ MÁS ARRIBA, PARA TODAS LAS PERMUTACIONES DE $(x_1, \dots, x_n)$ los valores de $h_1, \dots, h_K$ son iguales. $(x_1, \dots, x_n)$ PUEDE SER PERMUTADO DE $n!$ FORMAS POSIBLES. ADEMÁS, SE SABE QUE n bolitas de K colores, con $h_1, \dots, h_K$ bolas por color, SE PUEDEN ORDENAR DE $\binom{n}{h_1 \cdots h_K}$ formas. A SU VEZ, CADA GRUPO SE PUEDE ORDENAR de $h_1!, \dots, h_K!$ FORMAS, respectivamente.

Luego, la probabilidad de la suma es

$$P(h_1, \dots, h_K) = \underbrace{P(x_1, \dots, x_n) + \dots + P(x_{\tau_1}, \dots, x_{\tau_n})}_{\binom{n}{h_1 \cdots h_K} \text{ formas posibles de obtener la misma suma.}}$$

$$\Rightarrow P(h_1, \dots, h_K) = \binom{n}{h_1 \cdots h_K} P(x_1, \dots, x_n)$$

$$\Rightarrow P(x_1, \dots, x_n) = \frac{1}{\binom{n}{h_1 \cdots h_K}} P(h_1, \dots, h_K)$$

La distribución de $x_1, \dots, x_n$ depende a su vez de la distribución de las sumas POR CATEGORÍAS. PARA UN n fijo, es intuitivo PENSAR EN QUE, COMO x_i tiene K categorías, $\exists\, \underset{\sim}{\theta}$ TAL q.e SE PUEDA MODELAR EL CONTEO DE CATEGORÍAS COMO

$$h_1, \dots, h_K \mid \underset{\sim}{\theta} \sim \text{Multinomial}(n, \underset{\sim}{\theta}), \quad \underset{\sim}{\theta} \in \Delta_{K-1}$$

$$\Rightarrow P(h_1, \dots, h_K \mid \underset{\sim}{\theta}) = \binom{n}{h_1 \cdots h_K} \prod_{j=1}^{K} \theta_j^{h_j}$$

Y QUE SI SE CONSIDERA UNA PRIORI $\pi(\underset{\sim}{\theta})$ SE PUEDE marginalizar

$$P(h_1,\dots,h_k) = \int_{\Delta_{k-1}} P(h_1,\dots,h_k \mid \underset{\sim}{\theta})\, d\Pi(\underset{\sim}{\theta}) = \int_{\Delta_{k-1}} \binom{n}{h_1 \cdots h_k} \prod_{j=1}^{k} \theta_j^{h_j}\, d\Pi(\underset{\sim}{\theta})$$

$$\Rightarrow P(x_1,\dots,x_n) = \int_{\Delta_{k-1}} \prod_{j=1}^{k} \theta_j^{h_j}\, d\Pi(\underset{\sim}{\theta})$$

lo que claramente está malo

IDEA: Como $x_1, x_2, \dots$ es intercambiable, ocupar también $x_1,\dots,x_n, x_{n+1},\dots,x_{2n}$, que también es intercambiable. Luego, calcular la probabilidad de $P(h_1(X_1^{2n}) = h_1,\dots,h_k(X_1^{2n}))$ y mostrar que $P(h_1,\dots,h_k)$ puede ser aproximada a $\left[\binom{n}{h_1 \cdots h_k}\right]^2 \int_{\Delta_{k-1}} \{\prod \theta_j^{h_j}\}\, d\Pi$.

Mi idea anterior estaba mala porque no aprovechó el hecho que la intercambiabilidad es $\forall n$.

Escribí $x_1,\dots,x_n, x_{n+1},\dots,x_{2n}$ porque me imagino que marginalizando se puede llegar a algo y se me acabó el tiempo :(

P3 | Modelo dado por $p(\underset{\sim}{y}|\underset{\sim}{\theta})$ y $\pi(\underset{\sim}{\theta})$

a) Algoritmo M-H para explorar $\pi(\underset{\sim}{\theta}|\underset{\sim}{y})$

- Se escoge distribución propuesta con densidad $q(\cdot)$ y el mismo dominio que $\underset{\sim}{\theta}$. $q(\cdot)$ no necesariamente simétrica.

Para la iteración $k+1$:

i) Se muestrea candidato $\underset{\sim}{\theta}^*$ desde $q(\underset{\sim}{\theta}|\underset{\sim}{\theta}^k)$. k-ésimo valor

ii) Calculamos la probabilidad de aceptación:

$$\alpha(\underset{\sim}{\theta}^k, \underset{\sim}{\theta}^*) = \min\left\{ \frac{p(\underset{\sim}{\theta}^*|\underset{\sim}{y})}{p(\underset{\sim}{\theta}^k|\underset{\sim}{y})} \cdot \frac{q(\underset{\sim}{\theta}^k|\underset{\sim}{\theta}^*)}{q(\underset{\sim}{\theta}^*|\underset{\sim}{\theta}^k)}, 1 \right\}$$

iii) Muestreamos $u \sim U(0,1)$:

- Si $u < \alpha(\underset{\sim}{\theta}^k, \underset{\sim}{\theta}^*) \Rightarrow \underset{\sim}{\theta}^{k+1} = \underset{\sim}{\theta}^*$ (se acepta $\underset{\sim}{\theta}^*$)
- Si $u > \alpha(\underset{\sim}{\theta}^k, \underset{\sim}{\theta}^*) \Rightarrow \underset{\sim}{\theta}^{k+1} = \underset{\sim}{\theta}^k$ (se rechaza $\underset{\sim}{\theta}^*$ y se conserva valor anterior).

b) Kernel de transición del M-H:

El kernel de transición de moverse de $\underset{\sim}{\theta} \equiv \underset{\sim}{\theta}^k$ a $\underset{\sim}{\phi} \equiv \underset{\sim}{\theta}^*$ está dado por las dos posibilidades:

- Moverse de $\underset{\sim}{\theta}$ a $\underset{\sim}{\phi}$ $(\underset{\sim}{\theta}^{k+1} = \underset{\sim}{\phi})$
- Quedarse pegado en $\underset{\sim}{\theta}$ $(\underset{\sim}{\theta}^{k+1} = \underset{\sim}{\theta})$

Si se mueve de $\underset{\sim}{\theta}$ a $\underset{\sim}{\phi}$: $K(\underset{\sim}{\theta}, \underset{\sim}{\phi}) = \alpha(\underset{\sim}{\theta}, \underset{\sim}{\phi})\, q(\underset{\sim}{\phi}|\underset{\sim}{\theta})$

La probabilidad de quedarse pegado está dada por

$$r(\underset{\sim}{\theta}) = 1 - P(\text{moverse a cualquier } \underset{\sim}{\phi} \text{ posible}) = 1 - \int_{\Theta} \alpha(\underset{\sim}{\theta}, \underset{\sim}{\phi})\, q(\underset{\sim}{\phi}|\underset{\sim}{\theta})\, d\underset{\sim}{\phi}$$

Sea $B \in \Theta$. La probabilidad de partir en $\underset{\sim}{\theta}$ y llegar a $\phi \in B$ es la mezcla

$$P(\underset{\sim}{\theta}, B) = \underbrace{\int_B \overbrace{\alpha(\underset{\sim}{\theta}, \underset{\sim}{\phi})\, q(\underset{\sim}{\phi}|\underset{\sim}{\theta})}^{K(\underset{\sim}{\theta},\phi)}\, d\underset{\sim}{\phi}}_{\text{continuo}} + \underbrace{r(\underset{\sim}{\theta})\, I(\underset{\sim}{\theta} \in B)}_{\text{discreto}}$$

Si $\underset{\sim}{\theta} \notin B$, entonces no se puede quedar pegado.

c) CONDICIÓN DE REVERSALIDAD DEL M-H.

SE CUMPLE SI $\int_A p(\underset{\sim}{\theta}, B)\, d\underset{\sim}{\theta} = \int_B p(\underset{\sim}{\phi}, A)\, d\underset{\sim}{\phi}$

i.e. SI LA PROBABILIDAD DE MOVERSE DE B A A ES IGUAL A LA PROBABILIDAD DE MOVERSE DE A A B, CON $A, B \subset \Theta$.

EN EFECTO:

$$\int_A p(\underset{\sim}{\theta}, B)\, d\underset{\sim}{\theta} = \int_A \left[\int_B K(\underset{\sim}{\theta}, \underset{\sim}{\phi})\, d\underset{\sim}{\phi} + r(\underset{\sim}{\theta})\, I(\underset{\sim}{\theta} \in B) \right] d\underset{\sim}{\theta}$$

$$= \int_A \int_B K(\underset{\sim}{\theta}, \underset{\sim}{\phi})\, d\underset{\sim}{\phi}\, d\underset{\sim}{\theta} + \int_A r(\underset{\sim}{\theta})\, I(\underset{\sim}{\theta} \in B)\, d\underset{\sim}{\theta}$$

$$= \int_A \int_B K(\underset{\sim}{\theta}, \underset{\sim}{\phi})\, d\underset{\sim}{\phi}\, d\underset{\sim}{\theta} + \int_\Theta r(\underset{\sim}{\theta})\, I(\underset{\sim}{\theta} \in A \cap B)\, d\underset{\sim}{\theta}$$

$$= \int_A \int_B K(\underset{\sim}{\theta}, \underset{\sim}{\phi})\, d\underset{\sim}{\phi}\, d\underset{\sim}{\theta} + \int_B r(\underset{\sim}{\theta})\, I(\underset{\sim}{\theta} \in A)\, d\underset{\sim}{\theta}$$

SI SE CUMPLE LA CONDICIÓN DE BALANCE $\int_A \int_B K(\underset{\sim}{\theta}, \underset{\sim}{\phi})\, d\underset{\sim}{\phi}\, d\underset{\sim}{\theta} = \int_B \int_A K(\underset{\sim}{\phi}, \underset{\sim}{\theta})\, d\underset{\sim}{\theta}\, d\underset{\sim}{\phi}$

$$\Rightarrow \int_A p(\underset{\sim}{\theta}, B)\, d\underset{\sim}{\theta} = \int_B \int_A K(\underset{\sim}{\phi}, \underset{\sim}{\theta})\, d\underset{\sim}{\theta}\, d\underset{\sim}{\phi} + \int_B r(\underset{\sim}{\theta})\, I(\underset{\sim}{\theta} \in A)\, d\underset{\sim}{\theta}$$

$$= \int_B \left[\int_A K(\underset{\sim}{\phi}, \underset{\sim}{\theta})\, d\underset{\sim}{\theta} + r(\underset{\sim}{\phi})\, I(\phi \in A) \right] d\phi$$

$$= \int_B p(\underset{\sim}{\phi}, A)\, d\underset{\sim}{\phi}$$

$\therefore$ LA CONDICIÓN DE BALANCE DEL M-H IMPLICA REVERSIBILIDAD.

d) Gibbs es un caso particular DE M-H.

Si $\underset{\sim}{\theta}$ tiene p PARÁMETROS, EL GIBBS TIENE p funciones DE TRANSICIÓN

$$q_j(\underset{\sim}{\theta}^* | \underset{\sim}{\theta}^k) = \begin{cases} p(\theta_j | \theta^k_{(j)}, \underset{\sim}{y}) & \text{si } \theta^*_{(j)} = \theta^k_j \\ 0 & \text{e.o.c.} \end{cases}$$

Donde $\theta_{(j)} = \{\theta_1, \dots, \theta_{j-1}, \theta_{j+1}, \dots, \theta_p\}$. DE ESTE MODO, CADA $q_j(\cdot|\cdot)$ está modificando un solo PARÁMETRO.

ADEMÁS, GIBBS ES UN M-H QUE SIEMPRE ACEPTA. ESTO OCURRE PORQUE SIEMPRE $\alpha = 1$. EN EFECTO:

$$r = \frac{p(\underset{\sim}{\theta}^* | \underset{\sim}{y})}{p(\underset{\sim}{\theta}^k | \underset{\sim}{y})} \cdot \frac{q(\underset{\sim}{\theta}^k | \underset{\sim}{\theta}^*)}{q(\underset{\sim}{\theta}^* | \underset{\sim}{\theta}^k)} = \frac{p(\underset{\sim}{\theta}^* | \underset{\sim}{y})}{p(\underset{\sim}{\theta}^k | \underset{\sim}{y})} \cdot \frac{p(\theta_j^k | \theta_{(j)}^*, \underset{\sim}{y})}{p(\theta_j^* | \theta_{(j)}^k, \underset{\sim}{y})}$$

$$= \frac{p(\underset{\sim}{\theta}^* | \underset{\sim}{y})}{p(\underset{\sim}{\theta}^k | \underset{\sim}{y})} \cdot \frac{p(\theta_j^k, \theta_{(j)}^* | \underset{\sim}{y})}{p(\theta_{(j)}^* | \underset{\sim}{y})} \cdot \frac{p(\theta_{(j)}^k | \underset{\sim}{y})}{p(\theta_j^*, \theta_{(j)}^k | \underset{\sim}{y})}$$

Y COMO $\theta_{(j)}^* = \theta_{(j)}^k \Rightarrow \begin{cases} p(\theta_{(j)}^* | \underset{\sim}{y}) = p(\theta_{(j)}^k | \underset{\sim}{y}) \\ \{\theta_j^k, \theta_{(j)}^*\} = \{\theta_j^k, \theta_{(j)}^k\} = \underset{\sim}{\theta}^k \\ \{\theta_j^*, \theta_{(j)}^k\} = \underset{\sim}{\theta}^* \end{cases}$

ENTONCES $r = \frac{p(\cancel{\underset{\sim}{\theta}^*} | \underset{\sim}{y})}{\cancel{p(\underset{\sim}{\theta}^k | \underset{\sim}{y})}} \cdot \frac{p(\cancel{\theta_j^k, \theta_{(j)}^*} | \underset{\sim}{y})}{p(\cancel{\theta_{(j)}^*} | \underset{\sim}{y})} \cdot \frac{p(\cancel{\theta_{(j)}^k} | \underset{\sim}{y})}{p(\theta_j^*, \cancel{\theta_{(j)}^k} | \underset{\sim}{y})} = 1$

$\Rightarrow \alpha = \min\{r, 1\} = 1.$