Part 1. Cambiar de variables a Bernoullis independientes

Consideremos el conjunto $B \subset \{1, 2, ...M\}$ y las variables aleatorias $(\overline{X}_n)_{n\in\mathbb{N}}$ i.i.d. de manera uniforme sobre $\{1, 2, ..., M\}$. A partir de las $\overline{X}_n$, podemos definir las variables $(Y_n)_{n\in\mathbb{N}}$ de la siguiente manera: Y_n será igual a 1 cuando $\overline{X}_n$ pertenezca a B y 0 de otra manera. Claramente, las variables Y_n son i.i.d. y siguen una distribución Bernoulli con parámetro $p = \frac{|B|}{M}$. Dentro de este contexto, podemos responder la pregunta *£cuántas de las $\overline{X}_n$ cayeron dentro de B?*, simplemente considerando la suma $\sum Y_n$, que tiene distribución binomial. Sin embargo, cuando consideramos la pregunta más complicada *£cuántos valores diferentes, dentro de B, tomaron las variables $\overline{X}_n$?*, las Y_n ya no nos serán suficientes.

Para responder esta segunda pregunta, necesitaremos definir una nueva sucesión de variables, que llamaremos simplemente X_n. Ellas no tendrán distribuciones elementales; por ello estudiaremos precisamente en qué sentido podemos compararlas o intercambiarlas con las Bernoulli Y_n.

Definamos pues, la sucesión de variables $(X_n)_{n\in\mathbb{N}}$ de la siguiente manera. Pongamos $X_1 = 1_B(\overline{X}_1)$ y definamos recursivamente las siguientes según

$$X_k = 1_{B\setminus\{\overline{X}_1,\overline{X}_2,...,\overline{X}_{k-1}\}}(\overline{X}_k), k = 2, 3, ...$$

Es decir, X_1 es igual a 1 cuando $\overline{X}_1$ cae en B y 0 de otra manera; y X_k es igual a 1 cuando $\overline{X}_k$ cae en B pero no es igual a ninguno de los valores tomados hasta el momento por las variables anteriores.

Llamemos P_B a la ley del vector infinito (X_n) sobre $\{0, 1\}^{\mathbb{N}}$. Notamos que esta solo depende del tamaño de B y no de sus elementos, así que podemos escribir $\mathsf{P}_b = \mathsf{P}_B$, para todo subconjunto $B \subset \{1, 2, ..., M\}$ de tamaño b. Así, tenemos una familia de medidas de probabilidad $\{P_b : 0 \leq b \leq M\}$ sobre el espacio $\{0, 1\}^{\mathbb{N}}$. Además, observamos que X_1 tiene distribución $Ber(\frac{b}{M})$ y que X_2 se distribuye según $Ber(\frac{b}{M})$ en el caso en que $X_1 = 0$, o según $Ber(\frac{b-1}{M})$ en el caso en que $X_1 = 1$. Es decir, tenemos

$$\mathsf{P}_b(X_2 = 1\big|X_1 = z) = \frac{b - z}{M}.$$

Este hecho lo podemos interpretar de la siguiente manera. Si construimos el vector (X_n) en función a un conjunto B con $|B| = b$, la primera componente X_1 siempre tiene distribución $Ber(\frac{b}{M})$. Por la observación anterior, si sabemos que $X_1 = z$, sabremos también que $X_2 \sim Ber(\frac{b-z}{M})$. En este caso, podemos considerar X_2 como la primera componente de un vector (X'_n) definido sobre un conjunto B' de tamaño $b' = b - z$ y con ley P_{b-z}. Es decir, tenemos

$$\mathsf{P}_b(X_2 = w\big|X_1 = z) = \mathsf{P}_{b-z}(X'_1 = w).$$

Generalizamos esta observación con el siguiente resultado.

Lema. *Sean los vectores $(X_n), (X'_n)$ construidos como antes, con leyes P_b y P_{b-z}, respectivamente y escojamos $l \in \mathbb{N}$. Entonces, se cumple*

$$\mathsf{P}_b(X_2 = x_2, ..., X_{l+1} = x_{l+1}\big|X_1 = z) = P_{b-z}(X'_1 = x_2, ..., X'_l = x_{l+1}).$$

Proof. Basta expresar el término de la mano izquierda como

$$\frac{1}{P_b(X_1 = z)}\prod_{i=1}^{l+1}\mathsf{P}_b(X_i = x_i\big|\bigcap_{m=2}^{i-1} X_m = x_m \bigcap X_1 = z)$$

y efectuar el cambio de variable discutido anteriormente para obtener

$$\prod_{i=1}^{l+1}\mathsf{P}_{b-z}(X'_i = x_{i+1}\big|\bigcap_{m=1}^{i-1} X'_m = x_{m+1}).$$

□

El lemma nos dice que, si consideramos la distribución de la sucesión (X_n) a partir de un momento determinado, ella solo va a depender de cuántos elementos de B hayamos encontrado hasta ese instante. Este resultado nos da la motivación y el detalle técnico necesarios para la siguiente proposición.

Proposición 1. *Sea el vector* (X_n) *con ley* P_b *y escojamos* $l \in \mathbb{N}$. *Supongamos que* $\varphi : \{0,1\}^l \to \mathbb{R}$ *es una función medible. Escribamos* E_z *para la esperanza con respecto a la probabilidad* P_z. *Entonces, se cumple*

$$\mathsf{E}_b[\varphi(X_2, ..., X_{l+1})\big|X_1] = \mathsf{E}_{b-X_1}[\varphi(X_1, X_2, ..., X_l)],$$

donde, por abuso de notación, entendemos que las variables X_i *del término de la mano derecha definen el vector* (X'_n) *con ley* P_{b-X_1}.

Proof. Escribamos $h(x) = \mathsf{E}_{b-x}[\varphi(X_1, X_2, ..., X_l)]$ y $A_z = \{X_1 = z\}$. Queremos demostrar

$$\int\limits_{A_z} \varphi(X_2, ..., X_{l+1}) d\mathsf{P}_b = \int\limits_{A_z} h(X_1) d\mathsf{P}_b, \ z = 0, 1.$$

Pues bien, tenemos

$$\begin{aligned}\int\limits_{A_z} \varphi(X_2, ..., X_{l+1}) d\mathsf{P}_b &= \sum_{x_1, \ldots, x_{l+1}=0}^{1} 1_z(x_1)\varphi(x_2, \ldots, x_{l+1})\mathsf{P}_b(X_1 = x_1, X_2 = x_2, \ldots X_{l+1} = x_{l+1}) \\ &= \sum_{x_2, \ldots, x_{l+1}=0}^{1} \varphi(x_2, \ldots, x_{l+1})\mathsf{P}_b(X_2 = x_2, \ldots X_{l+1} = x_{l+1}\big|X_1 = z)P_b(X_1 = z).\end{aligned}$$

Ahora aplicamos el lemma y cambiamos de nombre x_i por x_{i-1} para obtener

$$\begin{aligned}\int\limits_{A_z} \varphi(X_2, ..., X_{l+1}) d\mathsf{P}_b &= \sum_{x_1, \ldots, x_l=0}^{1} \varphi(x_1, \ldots, x_l)\mathsf{P}_{b-z}(X_1 = x_1, \ldots X_l = x_l)P_b(X_1 = z). \\ &= \mathsf{E}_{b-z}[\varphi(X_1, ..., X_l)]P_b(X_1 = z) \\ &= \int\limits_{A_z} h(X_1) d\mathsf{P}_b.\end{aligned}$$

□

Esta propiedad Markoviana la usaremos pronto para probar nuestro resultado principal.

Ahora bien, consideremos los vectores $X = (X_n)$ y $X' = (X'_n)$ con leyes P_b y $\mathsf{P}_{b'}$ respectivamente. (X_n) está definido sobre $\{0,1\}^{\mathbb{N}}$ y (X'_n), en una copia posiblemente diferente del mismo espacio. Si queremos compararlos de una manera clara, primero necesitamos que ambos estén definidos en el mismo espacio de probabilidad. Esto lo lograremos mediante un acoplamiento. Consideremos, entonces, un espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$ y los conjuntos $B' \subset B \subset \{1, 2, ..., M\}$, donde $|B| = b \geq b' = |B'|$, donde definimos los vectores X, X' y ellos inducen las medidas $\mathsf{P}_b = \mathbb{P} \circ X^{-1}$ y $\mathsf{P}_{b'} = \mathbb{P} \circ X'^{-1}$ en $\{0,1\}^{\mathbb{N}}$. Ahora, tenemos que las variables $X_n, X'_n, n = 1, 2, ...$, las componentes de los vectores, están también definidas en Ω.

Sin embargo, empezamos nuestra discusión con el objetivo de comparar unas variables definidas en $\{0,1\}^{\mathbb{N}}$, no en Ω. Sean, entonces, unas variables ξ_n, ξ'_n : $\{0,1\}^{\mathbb{N}} \to \{1, 2, ...M\}$ donde los vectores $(\xi_1, \xi_2, ...), (\xi'_1, \xi'_2, ...)$ también tienen leyes $\mathsf{P}_b, \mathsf{P}_{b'}$. Compararemos estas sucesiones de variables valiéndonos de los vectores acoplados X, X'.

En los siguientes resultados, escribimos $\mathbb{E}$ para la esperanza en Ω respecto de la medida $\mathbb{P}$ y E_c para la esperanza tomada en $\{0,1\}^{\mathbb{N}}$ respecto de P_c.

Proposición 2. *Fijemos $k \in \mathbb{N}$. Sea $g : \{0,1\}^k \to \mathbb{R}$ una función creciente y supongamos $b \geq b'$. Entonces,*

$$\mathsf{E}_b[g(\xi_1, ..., \xi_k)] \geq \mathsf{E}_{b'}[g(\xi'_1, ..., \xi'_k)].$$

Proof. Observamos que pada cada $\omega \in \Omega$ se cumplen las desigualdades $X_k(\omega) \geq X'_k(\omega), \forall k$. Luego, tenemos $g(X_1, ..., X_k) \geq g(X'_1, ..., X'_k)$ por ser g creciente y tenemos,

$$\mathsf{E}_b[g(\xi_1, ..., \xi_k)] = \mathbb{E}[g(X_1, ..., X_k)] \geq \mathbb{E}[g(X'_1, ..., X'_k)] = \mathsf{E}_{b'}[g(\xi'_1, ..., \xi'_k)].$$

□

Observación. Ahora se hace obvia la necesidad del acoplamiento. Si pudiéramos decir $\xi \geq \xi'$, no habría necesidad de acoplamiento. Sin embargo, estas familias de funciones están definidas en espacios diferentes y compararlas directamente no tiene sentido.

Regresemos a nuestro objetivo principal. Queremos relacionar las variables $\xi_1, \xi_2, ...$ con unas variables $Y_n \sim Ber(p)$, que podemos considerarlas definidas en Ω. Usando las proposiciones anteriores, tenemos el siguiente resultado.

Teorema 3. *Sea $1 \leq k \leq b \leq M$ y g una función de k variables. Entonces, si $p \leq \frac{b-k+1}{M}$, entonces se cumple*

$$\mathsf{E}_b[g(\xi_1, ..., \xi_k)] \geq \mathbb{E}[g(Y_1, ..., Y_k)],$$

donde (ξ_i) son i.i.d. construidas como anteriormente, y (Y_i) son i.i.d. de distribución $Ber(p)$.

Proof. Procedemos por inducción en k. Para $k = 1$, observamos que ξ_1 tiene distribución $Ber(\frac{b}{M})$, igual que Y_1. Entonces, claramente $\mathsf{E}_b[g(\xi_1)] = \mathbb{E}[g(Y_1)]$. Ahora, suponemos que el teorema está probado para $k = l$ y queremos probar $\mathsf{E}_b[g(\xi_1, ..., \xi_{l+1})] \geq \mathbb{E}[g(Y_1, ..., Y_{l+1})]$.

Pues bien, escribamos

$$\mathsf{E}_b[g(\xi_1, ..., \xi_{l+1})] = \mathsf{E}_b[\mathsf{E}_b[g(\xi_1, ..., \xi_{l+1})\big|\xi_1]] = \mathsf{E}_b[h(\xi_1)],$$

donde $h : \{0,1\} \to \mathbb{R}$ es una función medible que cumple

$$h(z) = \mathsf{E}_b[g(z, ..., \xi_{l+1})\big|\xi_1] = \mathsf{E}_b[g_z(\xi_2, ..., \xi_{l+1})\big|\xi_1], z = 0, 1.$$

Por la proposición 1, podemos escribir

$$h(z) = \mathsf{E}_b[g_z(\xi_2, ..., \xi_{l+1})\big|\xi_1] = \mathsf{E}_{b-z}[g_z(\xi_1, ..., \xi_l)].$$

Por la hipótesis de inducción, para $p \leq \frac{b-l}{M}$ tenemos

$$h(z) \geq \mathbb{E}[g_z(Y_1, ..., Y_l)] = \mathbb{E}[g(z, Y_2, ..., Y_{l+1})] = \overline{h}(z), z = 0, 1,$$

donde $\overline{h}$ es creciente de una variable y las variables Y_n son i.i.d. según $Ber(p)$ y las $Y_n \sim Ber(p - \frac{z}{M})$. Aplicamos el Teorema con $k = 1$ para $\overline{h}$ para obtener

$$\mathsf{E}_b[g(\xi_1, ..., \xi_{l+1})] = \mathsf{E}_b[h(\xi_1)] \geq \mathsf{E}_b[\overline{h}(\xi_1)] \geq \mathbb{E}[\overline{h}(Y_1)] = \mathbb{E}[g(Y_1, Y_2, ..., Y_{l+1})].$$

□