Miscelánea Matemática 26 (1997) 17–27 SMM

Probabilidad y simetría: Distribuciones independientes de la noción de orden

Ricardo Berlanga Zubiaga
Departamento de Matemáticas
Instituto Tecnológico Autónomo de México

Esta breve nota surgió de uno de los planteamientos que Fabían Hernández Arellano hizo a lo largo de nuestros informales y provechosísimos coloquios.

Un escenario típico en la probabilidad combinatoria es el siguiente:

a) Se tiene un espacio de eventos finito, denotado por Ω.

b) Se le asigna a Ω la probabilidad uniforme (i.e. igual probabilidad a todos los eventos atómicos).

c) Se construye una variable aleatoria $X\colon \Omega \to \mathbb{R}$.

d) Finalmente se pregunta uno por los números

$$P\{X = x\} \qquad \forall\, x \quad \text{en la imagen de } X\,.$$

En los párrafos que siguen me gustaría replantear este esquema de manera distinta y explicar con ello un *fenómeno de regularidad* en el cálculo de probablidades.

Seguiremos pensando que nuestros espacios de eventos son finitos. Aunque cabe clarar que generalizar todo esto a espacios infinitos es tan solo un problema de lenguaje y un poco de aplomo. A saber: Apréndase la definición, técnica tal vez, pero nada que asuste a nadie, de probabilidad en el contexto de una sigma-álgebra (de Borel) y la definición de grupo topológico. Y listo. El prefijo *sigma* en *álgebra* y el adjetivo *topológico* son los ingredientes correctos para resolver todos los problemas de convergencia que a cualquier mortal en su sano juicio se le pueden presentar al enfrascarse con los conceptos de infinitud y continuidad.

1. Definiciones. Sea $f\colon\Omega \to \Lambda$ una función entre conjuntos arbitrarios. Recordemos que si $A \subseteq \Omega$ entonces la imagen (directa) de A bajo f es

$$f(A) = \{y \in \Lambda \mid \exists\, a \in A \text{ con } f(a) = y\}$$

y si $B \subseteq \Lambda$ entonces la imagen inversa de B es

$$f^{-1}(B) = \{x \in \Omega \mid f(x) \in B\}\ .$$

Es curioso, y devastadoramente importante, notar que el comportamiento algebraico de las imágenes inversas de f es impecable pero que el de las directas no:

2. Afirmación. Sean B, B_1 y B_2 subconjuntos de Λ. Entonces

a) $f^{-1}(B_1 \cap B_2) = f^{-1}(B_1) \cap f^{-1}(B_2)$.

b) En particular, si $B_1 \cap B_2 = \emptyset$ entonces

$$f^{-1}(B_1) \cap f^{-1}(B_2) = \emptyset\,.$$

c) $f^{-1}(B_1 \cup B_2) = f^{-1}(B_1) \cup f^{-1}(B_2)$.

d) En particular, f^{-1} convierte uniones ajenas en uniones ajenas.

e) $f^{-1}(\Lambda \setminus B) = \Omega \setminus f^{-1}(B)$.

El lector se podrá convencer con ejemplos de que los incisos a), b), d) y e) análogos para imágenes directas son falsos en general (ejercicio: ¿Bajo qué condiciones si se cumplen?). Por cierto: estas notas demuestran muy poco de lo que afirman. En realidad no es difícil llenar los huecos. En serio.

3. Definición. Sea $f\colon\Omega \to \Lambda$ una función y sea μ una probabilidad, no necesariamente uniforme, en Ω. Se define $f_*\mu$, *la probabilidad en* Λ *inducida por* f, mediante la fórmula

$$(f_*\mu)\,(B) = \mu\left(f^{-1}(B)\right) \qquad \forall\, B \subseteq \Lambda\,.$$

Por el contrario, si ν es una probabilidad en Λ, entonces la regla $A \mapsto \nu(f(A))$ no define, en general, una probabilidad en Ω.

4. Observación. Sea $X\colon\Omega\to\mathbb{R}$ una variable aleatoria en el espacio de probabilidad (Ω,μ), entonces

$$P\{X=x\}=(X_*\mu)(\{x\})\qquad \forall x\in\mathbb{R}.$$

He de confesar que me llevó algún tiempo perderle el miedo a las variables aleatorias y sus distribuciones. Y honestamente esto sucedió cuando me di cuenta que no eran sino funciones y sus medidas inducidas.

Es siempre bueno recordar que en matemáticas un arenque rojo no tiene porqué ser ni arenque ni rojo (Hirsh).

Por otro lado y con todo respeto, la notación tradicional me parece intelectual y operacionalmente pobre. Por ejemplo, la relación

$$(\varphi\circ X)_*=\varphi_*\circ X_* \tag{$*$}$$

es fácil de demostrar, fácil de recordar y fácil de interpretar: El proceso de inducir probabilidades conmuta con la composición de funciones. Es decir, se cumple un principio de superposición.

Sin embargo, lo más que se puede rescatar de esto en la notación clásica es la fórmula

$$P\{\varphi(X)=y\}=\sum_{\varphi(X)=y}P\{X=x\}$$

que a mi la verdad no me dice mucho, sino hasta que la veo como un corolario algebraico de la relación funciones anterior.

Sin embargo, la tradición en ocasiones se impone (Halmos).

Seguramente al leer la igualdad en (*) el lector pensó en φ como una función real de variable real. Quiero recalcar que no debe ser así necesariamente. El siguiente teorema coloca esta ecuación fundamental en un contexto general.

5. Teorema. *Sean (Ω,μ) y (Λ,ν) espacios de probabilidad y sea $\mathbb{F}$ un conjunto. Sean π, X y Y funciones tales que el diagrama*

$$\begin{tikzcd} \Omega \arrow[r, "X"] \arrow[d, "\pi"'] & \mathbb{F} \\ \Lambda \arrow[ur, "Y"'] & \end{tikzcd}$$

conmuta (esto sólo quiere decir que $X=Y\circ\pi$). $\blacksquare$

Supongamos además que $\nu=\pi_*\mu$. Entonces $X_*\mu=Y_*\nu$.

6. Requilorio. Una manera de ver esto es como sigue:

a) Ω es un escenario para un experimento, donde el experimento consiste en tomar un elemento ω de Ω bajo condiciones no totalmente controladas por nosotros. Pero a lo mejor manipuladas perfectamente por alguien más. Quien sabe. El plan de broma, aún en los juegos de azar (del árabe *azahr*, dado para jugar), me gusta pensar que algún espíritu chocarrero ya se sabe toda la historia del partido y solamente nos deja ver la trama linealmente a lo largo del tiempo.

X por su parte, es una rigurosa forma de medir el desenlace. Si $\mathbb{F} \subseteq \mathbb{R}$ entonces las mediciones son cuantitativas, pero en general no tienen porque serlo. *X definitivamente no es algo vago que toma diversos valores con diferentes probabilidades* (Feller). Resumiendo, el *resultado* del experimento es $\omega \in \Omega$, $X(\omega)$ es el *resultado* de la medición. Luego uno se confunde.

Ya que estamos en esto, es bueno aclarar que μ también *mide*, pero habitualmente se entiende que mide algo como *probabilidad*, *grado de certidumbre*, *grado de conocimiento*, *frecuencia relativa de los eventos*, o algo por el estilo según tus inclinaciones filosóficas. Por mi parte más me convence la interpretación tangible de que Ω es un contenedor lleno de fluido o de canicas y que μ tan solo es la medida de masa.

La manivela en la urna de la figura 1, sirve para hacer girar violentamente al contenedor y así provocar una circulación turbulenta, errática, caótica e impredecible del fluido o de las canicas.

b) Λ y Y son otro escenario y otra forma de medición.

c) π es una transformación. Y también un punto de vista (o cambio de coordenadas) desde el cual podemos reinterpretar los eventos en Ω como eventos de Λ (y que a su vez transporta a μ en $\pi_*\mu$). Esta nueva perspectiva no tiene nada de aleatoria y uno como observador la controla pefectamente.

d) La conmutatividad del diagrama nos dice que las mediciones de resultados X y Y son compatibles bajo el cambio de óptica.

e) El teorema afirma que si además π también es compatible con los aparatos μ en Ω y ν en Λ (i.e. $\pi_*\mu = \nu$) entonces las probabilidades inducidas en $\mathbb{F}$ por X y Y coinciden.

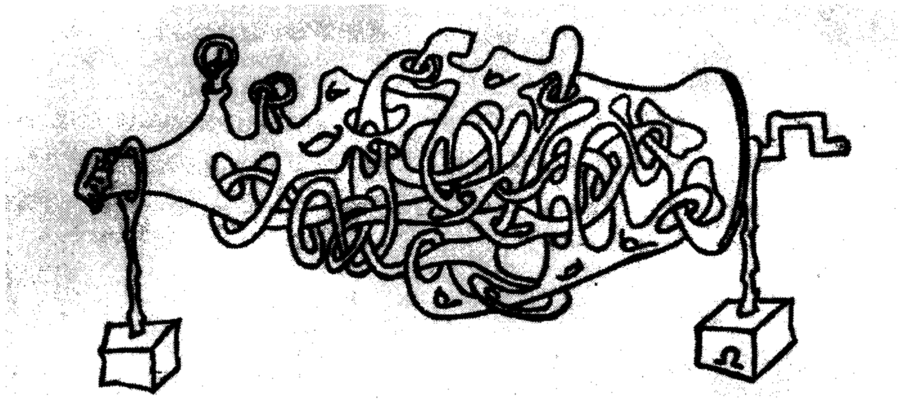

Figura 1: *Escenario Dinámico Estocástico Universal*

En particular y más imprecisamente,

$$P\{X = x\} = P\{Y = x\} \qquad \forall\, x \in \mathbb{F}.$$

7. Definición. Un caso muy especial, pero importante, es aquel cuando tenemos una proyección $\pi\colon \Omega \to \Lambda$ con la propiedad

$$|\pi^{-1}(\lambda)| = |\pi^{-1}(\zeta)| \qquad \forall\, \lambda, \zeta \in \Lambda\,.$$

En general a estas transformaciones se les conoce como fibraciones. Al conjunto $\pi^{-1}(\lambda)$ se le llama la fibra sobre λ. Aquí la idea es que Ω es un manojo de fibras *semejantes* (i.e. todas con la misma cardinalidad) atadas por Λ.

8. Afirmación. Sea $\pi\colon \Omega \to \Lambda$ una fibración. Entonces

$$\pi_* m_\Omega = m_\Lambda$$

donde m_Ω y m_Λ son las probabilidades uniformes en los espacios respectivos.

9. Acciones de grupo. Un contexto en el que las fibraciones aparecen naturalmente es el siguiente:

Sea G un grupo y sea Ω un conjunto (si el lector no se siente cómodo con la idea de grupo entonces deténgase aquí y consulte las páginas 110 a 120 del *Survey of Modern Algebra* de Birkhoff y Mac Lane. O bien, refiérase al libro del álgebra más cercano. Pero el Birkhoff-Mac Lane es muy leíble).

Una función $\alpha\colon \Omega \times G \to \Omega$ tal que $(\omega, g) \overset{\alpha}{\mapsto} \omega g$ se llama una acción de G en Ω si

a) $\omega e = \omega \qquad \forall\, \omega \in \Omega$, donde e es el elemento idéntico en G.

b) $\omega(gh) = (\omega g)h \qquad \forall\, \omega \in \Omega,\ g, h \in G.$

La acción es efectiva si dados $\omega \in \Omega$ y $g \in G$ tales que $\omega g = \omega$ entonces $g = e$.

Decimos que dos puntos $\omega_1, \omega_2 \in \Omega$ son equivalentes (bajo α) y escribimos $\omega_1 \sim \omega_2$ si existe $g \in G$ con $\omega_1 g = \omega_2$.

10. Afirmaciones. La relación entre elementos de Ω inducida por la acción de G es, en efecto, una relación de equivalencia. En consecuencia α induce una partición del espacio Ω denotada por Ω/G (en caso de duda refiérase a Birkhoff-Mac Lane páginas 145–147).

Si $[\omega]$ denota a la clase de equivalencia de $\omega \in \Omega$ entonces la regla $\omega \mapsto [\omega]$ define una suprayección $\pi\colon \Omega \to \Omega/G$. Si α es efectiva entonces π es una fibración tal que la cardinalidad de cada fibra es $|G|$ y la cardinalidad del *espacio cociente* Ω/G es $|\Omega|/|G|$. ∎

Aquí el marco visual es el de que la acción del grupo *fibra* a Ω, permitiéndole a G solamente permutar a los elementos en G dentro de cada una de sus fibras (u órbitas) correspondientes.

El siguiente teorema, todavía abstracto, estructura un frecuente *patrón de regularidad* en el cálculo de probabilidades.

11. Teorema. *Sea $\alpha\colon \Omega \times G \to \Omega$ una acción y sea $X\colon \Omega \to \mathbb{F}$ una variable aleatoria invariante bajo la acción del grupo. Es decir, $X(\omega) = X(\omega g) \qquad \forall\, \omega \in G,\ g \in G$. Entonces existe una única variable $\widetilde{X}\colon \Omega/G \to \mathbb{F}$ tal que el diagrama*

$$\begin{tikzcd}
\Omega \arrow[r, "X"] \arrow[d, "\pi"'] & \mathbb{F} \\
\Omega/G \arrow[ur, "\widetilde{X}"'] &
\end{tikzcd}$$

conmuta. Si α es efectiva y consideramos a Ω y Ω/G como espacios de probabilidad con la probabilidad uniforme, entonces X y $\widetilde{X}$ están igualmente distribuidas.

Demostración. Es fácil ver que $\widetilde{X}: \Omega/G \to \mathbb{F}$ tal que $\widetilde{X}([\omega]) = X(\omega)$ $\forall [\omega] \in \Omega/G$ está bien definida. Ahora el resultado se sigue de los párrafos (5), (8) y (10). ∎

Aquí la idea es que X restringida a cada fibra es una constante (haga dibujitos. Piense quien podría ser el *Fibrado Estocástico Universal*).

12. Ejemplo. Sea Ω el conjunto de palabras con letras distintas de longitud k en el alfabeto A. Otra manera de decir esto es que Ω es el conjunto de permutaciones de A tomadas de k en k; o el conjunto de muestras ordenadas de elementos de A, sin repetición, de tamaño k. En fín, aún otra manera es pensar que los elementos de Ω son realmente las funciones inyectivas de $[1, k] = \{1, 2, \ldots, k\}$ en el conjunto A.

Sea S_k el grupo simétrico en k símbolos. Es decir, S_k es el grupo de permutaciones de $[1, k]$ en sí mismo siendo la operación en el grupo la composición de funciones.

No es difícil vér que la función $\alpha: \Omega \times S_k \to \Omega$ dada por la regla $(\omega, g) \mapsto \omega \circ g$ define una acción efectiva.

Como $|S_k| = k!$ entonces $k!$ es la cardinalidad de la fibra de la proyección natural $\pi: \Omega \to \Omega/S_k$.

Afirmación. $\omega \sim \rho$ si y solamente si $\omega([1, k]) = \rho([1, k])$.

Demostración. Si ω y ρ están relacionados, entonces es obvio que tienen la misma imagen. Supongamos ahora que ω y ρ tienen la misma imagen. Por la inyectividad de ρ y la hipótesis, se tiene que para toda $i \in [1, k]$ existe una única $j \in [1, k]$ tal que $\omega(i) = \rho(j)$. Por la inyectividad de ω, la correspondencia $i \mapsto j$ define un elemento g en S_k. Por definición, $\omega = \rho \circ g$. ∎

Corolario. *Ω/S_k está en correspondencia biyectiva de manera natural con el conjunto Λ de subconjuntos de orden k en A, de tal modo que a cada clase $[\omega] \in \Omega/S_k$ se le asocia el conjunto $\omega([1, k]) \subseteq A$.*

Debido a este corolario uno se olvida de que Ω/S_k es un conjunto de clases de equivalencia, que para cualquiera es un tanto engorroso de visualizar y mejor se piensa como si Λ realmente fuese Ω/S_k. En esta situación, la igualdad $|\Omega/S_k| = |\Omega|/|S_k|$ se traduce en la vieja identidad

$$\binom{n}{k} = \frac{(n)_k}{k!}$$

donde $|A| = n$ y $(n)_k = n(n-1)(n-2)\dots(n-k+1)$.

No quiero que el lector piense que he colocado un montón de resultados fáciles en un lenguaje difícil sólo por molestar. La redacción de estas notas intenta estar en el espíritu de la evolución de ideas poderosas que se han venido desarrollando en el curso de los últimos 100 años para atacar un sinnúmero de problemas, al principio sobre todo de geometría, que involucran los conceptos, otrora vagos, de *simetría*, *transformación natural* e *invariancia*. Es mi opinión que estas nociones están entre las más profundas e importantes de la ciencia en general.

En el contexto de este ejemplo, decir que una variable aleatoria X en Ω es invariante bajo la acción del grupo simétrico, es decir, simplemente que *la definición de X no depende del orden de la muestra.* Si este es el caso, entonces la teoría nos garantiza que existe una variable $\widetilde{X}$ en Λ, *el espacio de muestras desordenadas*, con la misma distribución que X.

En la práctica X y $\widetilde{X}$ tienden a parafrasearse del mismo modo. Ejemplos específicos de esta situación son los siguientes:

Ejemplo Concreto 1. Sean $0 < k < n$ números naturales. Sea $A = [1, n]$, $\mathbb{F} = [k, n]$ y $\Omega = \{\omega : [1, k] \to A \mid \omega \text{ es inyectiva}\}$. Por lo tanto $\Lambda = \{H \subseteq A \mid |H| = k\}$. Si definimos $X : \Omega \to \mathbb{F}$ tal que

$$X(\omega) = \max\{\omega_1, \omega_2, \dots, \omega_k\} \qquad \forall\, \omega \in \Omega$$

entonces es obvio que X es invariante y que $\widetilde{X} : \Lambda \to \mathbb{F}$ se define exactamente por la misma fórmula, cambiando simplemente la expresión "$\forall\, \omega \in \Omega$" por "$\forall\, \omega \in \Lambda$". Ahora ya no es sorpresa que se cumpla que

$$\frac{k \cdot (x-1)_{k-1}}{(n)_k} = P\{X = x\} = P\{\widetilde{X} = x\} = \frac{\binom{x-1}{k-1}}{\binom{n}{k}}$$

Ejemplo Concreto 2. Supongamos que A es un conjunto con n_j individuos del tipo j para $j = 1, 2, 3, \ldots, M$. Es decir,

$$A = \bigcup_{j=1}^{M} \{j\} \times [1, n_j]$$

Sea Ω el conjunto de k-permutaciones en A, pero definamos $\mathbb{F}$ igual a $\mathbb{R}^M$ (el lector tal vez quisiera tomar una $\mathbb{F}$ *más chica*). Finalmente, sea $X\colon \Omega \to \mathbb{F}$ tal que, dado $\omega \in \Omega$

$$\begin{aligned} X(\omega) &= \Big(|(\rho \circ \omega)^{-1}(\{1\})|, \ldots, |(\rho \circ \omega)^{-1}(\{M\})|\Big) \\ &= (|\omega([1,k]) \cap \{1\} \times [1, n_1]|, \ldots, |\omega([1,k]) \cap \{M\} \times [1, n_M]|) \end{aligned}$$

donde $\rho(a, b) = a \quad \forall\, (a, b) \in A$.

La situación se repite: X es invariante y $\widetilde{X}$ se determina con la regla

$$\widetilde{X}(\omega)\,(|\omega \cap \{1\} \times [1, n_1]|, \ldots, |\omega \cap \{M\} \times [1, n_M]|) \qquad \forall\, \omega \in \Lambda$$

que es *casi* la misma que para X. En palabras, ambas variables le asocian a cada muestra de k individuos en A, la lista de los números de individuos de cada tipo.

De aquí que, sin más cálculo

$$P\{X = (k_1, k_2, \ldots, k_M)\} =$$

$$= \frac{\left[\binom{n_1}{k_1}\binom{k}{k_1} k_1!\right]\left[\binom{n_2}{k_2}\binom{k-k_1}{k_2} k_2!\right]\cdots\left[\binom{n_M}{k_M}\binom{k_M}{k_M} k_M!\right]}{(n)_k}$$

sea igual a

$$P\{\widetilde{X} = (k_1, k_2, \ldots, k_M)\} = \frac{\binom{n_1}{k_1}\binom{n_2}{k_2}\cdots\binom{n_M}{k_M}}{\binom{n}{k}}$$

donde $n = \sum n_j, \; k = \sum k_j$.

13. Contraejemplo. Sea Ω el conjunto de palabras, no necesariamente con letras distintas, de longitud k en el alfabeto A. Es decir, Ω es el conjunto de muestras ordenadas de elementos de A, con repetición, de tamaño k. Formalmente, Ω es el conjunto de *todas* las funciones de $[1, k]$ en A. También, Ω es el espacio de distribuciones de k partículas (o bolas) distinguibles en $|A|$ estados (o celdas).

Nuevamente, el grupo simétrico S_k actúa en Ω por composición.

Afirmación. Si $\omega, \rho \in \Omega$ entonces

$$\omega \sim \rho \iff |\omega^{-1}(\{a\})| = |\rho^{-1}(\{a\})| \qquad \forall a \in A .$$

Corolario. Si $A = [1, n]$ entonces Ω/S_k está en correspondencia biyectiva con el conjunto

$$\Lambda = \left\{ (k_1, k_2, \ldots, k_n) \in [0, k]^n \mid \sum_{k=1}^{n} k_j = k \right\}$$

de tal modo que a cada clase $[\omega]$ se le asocian sus *números de ocupación* $(|\omega^{-1}(\{1\})|, |\omega^{-1}(\{2\})|, \ldots, |\omega^{-1}(\{n\})|)$.

Λ se puede pensar como el espacio de distribuciones de k partículas indistinguibles en n estados.

De aquí

$$|\Omega/S_k| = \binom{k+n-1}{k}$$

■

La acción de S_k en Ω no es efectiva. El siguiente resultado general nos dice como varían las cardinalidades de las distintas clases de equivalencia.

Afirmación. Sea $\alpha \colon \Omega \times G \to \Omega$ una acción arbitraria y sea $\omega \in \Omega$. Entonces

$$|[\omega]| = |G|/|\mathrm{Stab}\,(\omega)|$$

donde $\mathrm{Stab}\,(\omega) = \{g \in G \mid \omega g = \omega\}$. ■

A $\mathrm{Stab}\,(\omega)$ se le llama el *grupo de isotropía o estabilizador de* ω. Desde esta perspectiva, una acción es efectiva cuando todos los estabilizadores son triviales.

Regresando a la situación particular de nuestro *contraejemplo*, tomemos $\omega \colon [1, k] \to [1, n]$ y definamos $k_j = |\omega^{-1}(\{j\})|\ \forall j \in [1, n]$. Entonces no es difícil ver que

$$|[\omega]| = \binom{k}{k_1, k_2, \ldots, k_n} .$$

(Sugerencia: Calcule el grupo de isotropía de ω).

Por lo tanto, si m_Ω y m_Λ son las probabilidades uniformes en los espacios respectivos, entonces $\pi_* m_\Omega$ es distinta a m_Λ.

En particular,

$$\pi_* m_\Omega([\omega]) = \binom{k}{k_1, k_2, \ldots, k_n} / n^k$$

pero

$$m_\Lambda([\omega]) = 1 \Big/ \binom{k+n-1}{k} .$$

Sea $X: \Omega \to \mathbb{F}$ una variable aleatoria invariante ante la acción del grupo simétrico. Entonces la teoría nos garantiza que existe $\widetilde{X}: \Lambda \to \mathbb{F}$ con $\widetilde{X} \circ \pi = X$. Sin embargo, sus distribuciones serán en general distintas.

A manera de ilustración, el lector puede analizar la situación para $X(\omega) = \max\{\omega_1, \omega_2, \ldots, \omega_k\}$.

Para concluir, quisiera mencionar que en el contexto de la mecánica estadística $(\Lambda, \pi_* m_\Omega)$ es el espacio de Maxwell-Boltzmann y (Λ, m_Λ) el de Bose-Einstein. Este último sirve para modelar fotones, núcleos y átomos que contienen un número par de partículas elementales. Sin embargo el primero, considerado como *el intuitivamente correcto*, ya fracasó en la teoría moderna (ver el primer volumen de Feller, pags. 39–41).

Creo que la maquinaria bosquejada en estas notas clarifica las diferencias estruoturales entre estos enfoques: Maxwell-Boltzmann es un espacio cociente bajo la acción inefectiva del grupo simétrico, pero Bose-Einstein es un espacio uniforme.

También en este lenguaje, *la estadística* de Fermi-Dirac, válida para neutrones, protones y electrones, es tan solo el espacio cociente, bajo la acción efectiva del grupo simétrico, del ejemplo en el párrafo 12.