XXXI SEMANA NACIONAL DE INVESTIGACIÓN Y DOCENCIA EN MATEMÁTICAS

Una característica común de algunos Modelos de Probabilidad

Valeria Campa Morán ♦ Dra. Luz del Carmen Rosas Rosas

Licenciatura en Matemáticas, Universidad de Sonora

INTRODUCCIÓN

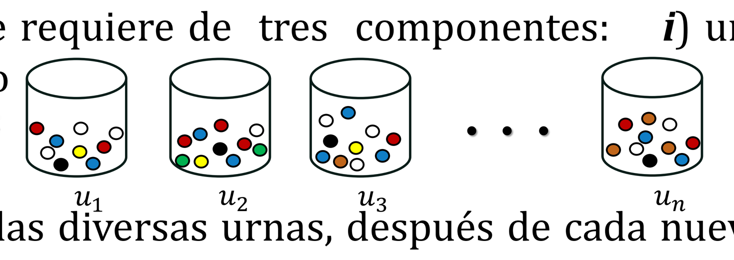
La presente es una breve reseña en la que se consideran modelos de probabilidad con estructura de urna. De hecho, para tales objetos se contemplan aquí varios escenarios con el propósito de describir y clasificar determinados modelos de probabilidad discretos involucrados en dicho contexto, así como destacar y analizar algunos resultados relevantes correspondientes.

PRELIMINARES

La implementación de modelos probabilísticos basados en estructura de urna(s) han proporcionado una útil herramienta matemática relativamente simple al analizar situaciones de incertidumbre que surgen de gran variedad de escenarios, tales como: epidemiología, genética, economía, control de calidad, estructuración de datos, etc. Además, resulta imprescindible resaltar el gran énfasis que estas estructuras han encontrado en fenómenos con naturaleza de ramificación, como es en el caso específico de la propagación de enfermedades contagiosas, así como en estructuras arborescentes en ciencias de la computación. Cabe señalar que, respecto a los múltiples autores que a lo largo de la historia han contribuido con tales objetos se encuentra: George Pòlya, F. Eggenberger [ver 1)], B. Friedman [ver 2)], R. Permantle [ver 8)], H. Mahmoud [ver 7)], K. Inoue y S. Aki [ver 4)], entre otros, así como S. Kotz y N.L Johnson [ver 3)], entre otros.

Notación. Denotaremos por $\mathbb{Z}, \mathbb{N}_0$ y $\mathbb{N}$, respectivamente, al conjunto de los números: enteros, enteros no negativos y naturales.

Elementos básicos. Usualmente para esta clase de modelos se requiere de tres componentes: ***i***) un número fijo de urnas (recipientes), ***ii***) cada urna conteniendo una cantidad dada de esferas idénticas pero al menos en dos colores distintos, ***iii***) una secuencia de reglas que especifican cómo proceder con la adición (y/o remoción) de esferas en las diversas urnas, después de cada nueva extracción aleatoria en las diferentes etapas del experimento.



Modelo clásico. Se considera solo una urna, la cual contiene determinado número de esferas en exactamente k colores distintos ($k \in \mathbb{N} - \{1\}$), y en cada etapa se extrae una esfera aleatoriamente, se observa su color, y luego el contenido de la urna evoluciona de acuerdo a la regla: *Si la esfera seleccionada es de color* $i \in \{1, \ldots. k\}$, *entonces se devuelve a la urna junto con* a_{ij} *esferas de color* $j \in \{1, \ldots. k\}$, con $a_{ij} \in \mathbb{Z}$. En consecuencia, resulta cómodo organizar tal información en lo que llamaremos **matriz de reemplazo** R (orden $k \times k$), cuyos renglones indican el color de la esfera seleccionada, mientras que sus columnas representan el color de la(s) esfera(s) añadida(s) inmediatamente después de cada extracción.

$$R := \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{kk} \end{pmatrix}$$

OBJETIVO. Observar la evolución de la urna a lo largo de la secuencia particular generada por la regla de extracciones aleatorias sucesivas dada.

DESCRIPCIÓN

En cada contexto a continuación, de acuerdo al modelo clásico, asumimos únicamente una urna conteniendo un total de $a + b$ esferas, de las cuales a son azules, y blancas las b restantes, es decir, $k = 2$; *a.s.e.* significará "antes de la siguiente etapa". Sea $n \in \mathbb{N} - \{1\}$, y considérese n extracciones aleatorias

(*a*) **Esquema 1.** Realizar una secuencia de n extracciones aleatorias de acuerdo a la regla: *En cada etapa se extrae una esfera aleatoriamente de la urna, se observa su color, se regresa a la urna y, si es de color azul se añaden* c *esferas de su mismo color junto con* d *esferas del color complementario a.s.e.* $(c, d \in \mathbb{Z})$; *mientras que, si es de color blanco se añaden* g *esferas de su mismo color junto con* h *esferas del color complementario a.s.e.* $(g, h \in \mathbb{Z})$.

A este esquema se le conoce como ***modelo de Pòlya***, y en tal situación, nótese que la matriz de reemplazo $R_{(1)}$ está dada por

$$R_{(1)} := \begin{pmatrix} c & d \\ h & g \end{pmatrix}$$

(*b*) **Esquema 2.** En la secuencia de n extracciones aleatorias se pone de manifiesto la siguiente regla: *En cada etapa se extrae una esfera aleatoriamente de la urna, se observa su color, se regresa a la urna y además, se añaden* c *esferas de su mismo color junto con* d *esferas del color complementario a.s.e.* $(c, d \in \mathbb{Z})$.

De modo que, para este caso la matriz de reemplazo $R_{(2)}$ queda como sigue

$$R_{(2)} := \begin{pmatrix} c & d \\ d & c \end{pmatrix}.$$

caso al que se le conoce como ***modelo de B. Friedman***.

(**c**) **Esquema 3.** Supongamos ahora que en el modelo de B. Friedman se llevan a cabo n extracciones aleatorias bajo las restricciones de la regla a continuación: *En cada etapa se extrae una esfera aleatoriamente de la urna, se observa su color, se regresa a la urna y además, únicamente se añaden* c *esferas de su mismo color a.s.e.* $(c \in \mathbb{Z})$.

De tal escenario se deduce que la matriz de reemplazo $R_{(3)}$ obtenida toma la forma

$$R_{(3)} := \begin{pmatrix} c & 0 \\ 0 & c \end{pmatrix}$$

y en cuyo contexto se incluye un análisis con mayor detalle a continuación.

ANÁLISIS ESPECÍFICO / ESQUEMA 3

Aquí nos concentramos en un breve análisis del último de los esquemas previos, lo cual es en función de la variación del parámetro c, específicamente cuando $c \geq -1$, de donde se observa lo siguiente:

i. Si $c < 0$, entonces "agregar c esferas" significa remover (retirar) c esferas.
ii. El caso específico en que $c = -1$, da lugar a lo que se conoce como *muetreo simple sin reemplazo*.
iii. Mientras que, cuando $c = 0$ se obtiene el *muestreo simple con reemplazo.*
iv. Y en general,
   - Si $c < 0$, entonces el proceso de extracción se verá interrumpido en algún momento.
   - En caso contrario, si $c \geq 0$, entonces el proceso de extracción continuará de forma indefinida.

Dado el espacio de probabilidad $(\Omega, \mathfrak{F}, P)$ correspondiente al experimento descrito en el Esquema 3, sea la v. aleat. X: *número de esferas azules obtenidas en* n *extracciones aleatorias sucesivas*, de modo que

$$f_X(x) := P[X = x], \qquad x \in \mathbb{R},$$

denotará la función de probabilidad de X. Y como consecuencia, no es difícil obtener cada uno de los resultados a continuación.

Proposición 1. (Sin reemplazo) Si $c = -1$ y $n \leq a + b$, entonces la v.a. X tiene **distribución hipergeométrica** con parámetros $a + b$, a y n; rango $R_X = \{x \in \mathbb{N}_0 : m \leq x \leq M\}$ donde $m := max\{0; n - b\}$ y $M := min\{a, n\}$; y

$$f_X(x) = \begin{cases} \dfrac{\binom{a}{n}\binom{b}{n-a}}{\binom{a+b}{n}} & si\ x \in R_X \\ 0 & en\ otro\ caso \end{cases}$$

Proposición 2. (Con reemplazo) Si $c = 0$, entonces la v.a. X tiene **distribución binomial** con rango $R_X = \{0,1, \ldots n\}$; parámetros n y $p := a/(a + b)$; por lo cual

$$f_X(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & si\ x \in R_X \\ 0 & en\ otro\ caso \end{cases}$$

Proposición 3. (Generalización) Si $a, b, c \in \mathbb{N}$, entonces la v.a. X tiene **distribución de Pòlya** con rango $R_X = \{0,1, \ldots n\}$; parámetros $a + b, a, n$ y c, y por consiguiente

$$f_X(x) = \begin{cases} \binom{n}{x} \dfrac{a^{[x,c]} b^{[n-x,c]}}{(a+b)^{[n,c]}} & si\ x \in R_X \\ 0 & en\ otro\ caso \end{cases}$$

donde

$$a^{[x,c]} := \prod_{j=0}^{x-1} (a + jc)\,; \quad b^{[n-x,c]} := \prod_{j=0}^{n-x-1} (b + jc) \quad y \quad (a+b)^{[n,c]} := \prod_{j=0}^{n-1} (a + b + jc)$$

Observaciones. Respecto a la Proposición 3, nótese que:

- Las Proposiciones 1 y 2 son casos particulares, pues cuando $c = -1$ se obtiene el modelo hipergeométrico, mientras que, $c = 0$ conduce al modelo binomial.
- Además, para el caso especial en que $a = b = c$ se obtiene que la v.a. X es tal que $R_X = \{x_1, \ldots, x_n\}$, y tiene una distribución uniforme discreta, es decir,

$$f_X(x) = \begin{cases} \dfrac{1}{n} & si\ x \in R_X \\ 0 & en\ otro\ caso \end{cases}$$

ÚTIL HERRAMIENTA / EJEMPLOS ILUSTRATIVOS

Intercambio de calor entre dos cuerpos aislados / Modelo de Erhenfest

Se considera solamente una urna conteniendo $a + b$ esferas (a son negras, y blancas las b restantes; $a, b \in \mathbb{N}$), de forma tal que en cada etapa las restricciones son: seleccionar una esfera aleatoriamente, se observa su color, luego se regresa a la urna, pero además se añade únicamente una esfera del color complementario. (Este modelo debe su nombre a que el problema que le dio origen fue planteado por el físico austriaco Paul Erhenfest).

Campaña de seguridad

Se trata de urna única, la cual contiene una cantidad conocida de esferas en dos colores (blancas y negras). Y bajo tal contexto lo que plantea el modelo es que, a raíz de que ocurre un accidente, es decir, la esfera extraída resultó ser negra, suele lanzarse una campaña de prevención de accidentes; mientras que, de no ocurrir accidente alguno, entonces ese tipo de campañas caen en un período de relajamiento, lo cual a su vez conduce eventualmente a un incremento en la probabilidad de que ocurra lo contrario (un accidente).

COMENTARIO FINAL

En este documento se muestra brevemente la relación que guarda el modelo clásico de estructura de urnas con cada uno de tres modelos de probabilidad discretos: hipergeométrico, binomial y uniforme discreto, razón por la que coloquialmente se hace referencia a sus respectivas distribuciones como distribuciones de Pòlya. Cabe mencionar que aquí se enfatiza el hecho de que dicha relación depende directamente de la variación del parámetro de remoción de esferas c en el modelo clásico. No obstante, análogamente resulta de interés analizar además la situación en que se considera más de una urna en sus variantes de acuerdo a la regla de reemplazo preestablecida, así como también cuando esta se hace depender del tiempo o etapa de extracción [ver por ejemplo: 3), 4), 8)], y en contextos de aplicación diversa [ver por ejemplo 9)].

REFERENCIA BIBLIOGRÁFICA

1) Eggenberger,F., Pòlya,G. *Calcul des probabilités sur l'interprétation the certaines courbes de frequence*. ComptesRendus, Académie des Sciences, Paris; 187, 870-872. (1928)
2) Friedman,B. *A simple urn model*. Commun.of Math.Statist.Probab.Lett.; 2, 59-70. (1928)
3) Gouet,R. *A martingale approach to strong convergence in a generalized Pòlya-Eggenberger urn model*. Statist.Probab.Lett.; 8, 225-228. (1989)
4) Inoue,K.; Aki,S. . *Pòlya urn models under a general replacement schemes*. J.Japan Statistic.Soc.; 31:2, 193-205. (2001)
5) Janardan,K.G.; Schaeffer,D.J. *A generalization of Markov-Pòlya distribution its extentions and applications*. Biom.J.; 19, 87-106. (1977)
6) Johnson,N.L.;Kotz,S. *Urn Models and their Applications: an approach to Modern Discrete Probability Theory*. Editorial Wiley; NewYork. (1977)
7) Mahmoud,H.M. *Drawing multisets of balls from tenable balanced linear urns*. Probab.Engrg.Inform.Sci.; 27: 147-162. DOI: 10.1017/S0269964812000381 (2013)
8) Permantle,R. *A time-dependent version of Pòlya's urn*. J.Theoretic.Probab.; 3, 627-637. (1990)
9) Wei,L.J.; Durham, S. *The randomized play the winner rule in medical trials*. J.Amer.Statist.Soc.; 73:364, 840-843. (1978)