

LA ALEATORIEDAD Y EL DISEÑO DE EXPERIMENTOS COMPUTACIONALES

Carlos Polanco, Jorge Alberto Castañón González y Lourdes Tavera Sierra

El diseño de experimentos es un componente muy importante en el quehacer científico, de su reproducibilidad dependerá su validez. Su planteamiento e implementación son testigos de la creatividad de quienes lo diseñan; sin embargo, su reproducibilidad debe seguir reglas específicas, una de ellas implica no incurrir en ningún sesgo al momento de seleccionar las variables involucradas. ¿Cómo evitar que no ocurran los sesgos? La respuesta es simple, aunque no su cumplimiento: dando a todas las variables igual oportunidad de ser seleccionadas. Si el experimento posee esta cualidad se dice que el experimento es aleatorio.

Mencionamos que su implementación no es tarea simple, nada mejor que algunos ejemplos para exponer en qué radica la dificultad de su cumplimiento. Considere una encuesta en vía pública donde el encuestador da preferencia en sus entrevistas a determinado grupo poblacional, lo cual inducirá un sesgo de género; o piense en un experimento que requiera la selección de números aleatorios, y para llevar a cabo ello se opta por usar un dado, pero se desconoce que el dado está cargado, lo que provocará un sesgo constructivo. Estos dos experimentos bien se pueden enmendar entrenando al encuestador, y cambiando el dado.

Pero cuando un experimento implica números astronómicos el tema no es simple. Quien diseña experimentos computacionales algunas veces ignora lo simple que es incurrir en un sesgo, ello se debe a que las rutinas generadoras de números aleatorios son de periodo finito, i.e. una rutina de números aleatorios de periodo $10^8=100,000,000$, significa que la rutina generará una sucesión de 10^8 números aleatorios, y luego se repetirá la misma sucesión. Entonces el número 10^8+1 corresponderá al primer número de la sucesión inicial, el 10^8+2 , será el segundo número de la sucesión inicial, y así

sucesivamente. Seguramente el lector debe pensar que 10^8 números aleatorios es una cifra muy grande inclusive para una computadora, y que es difícil alcanzarla, pero no es así. Si uno emplea una laptop para llevar a cabo estos cálculos en 24 horas habrá consumido esos 10^8 números aleatorios, o sea cada cálculo lo desarrollará en sólo una milésima de segundo. A una supercomputadora le tomaría únicamente cinco segundos ese mismo proceso.

¿Debería siempre verificarse la “calidad” de los números aleatorios producidos por estas rutinas? Así es, y para poder efectuarla es conveniente implementar computacionalmente una prueba estadística, particularmente la prueba denominada *Prueba de Rachas de una muestra*, una prueba eficiente y simple. Debe aplicarse sobre la totalidad de los números aleatorios producidos por la rutina generadora antes de emplearse en todo experimento. ¿Qué tan importante es llevar a cabo esta prueba? La respuesta les sorprenderá: casi nunca se efectúa, a pesar de que la validez del experimento depende de la aleatoriedad de las variables involucradas. La razón en buena medida se debe a que los usuarios que emplean las computadoras para realización de experimentos, asumen que las rutinas están debidamente probadas y resultan prácticamente infalibles. Nada más alejado de la realidad, las rutinas fallan. ☒

Carlos Polanco (Ciudad de México, 1958). Mexicano, matemático y profesor de la Facultad de Ciencias de la Universidad Nacional Autónoma de México. Es miembro del Sistema Nacional de Investigadores. Desde el año 2007 desarrolla modelos estocásticos predictivos orientados a epidemiología matemática, proteómica computacional y origen de la vida.

Jorge Alberto Castañón González (Ciudad de México, 1957), Mexicano, médico por la Universidad Nacional Autónoma de México, profesor de Medicina e investigador en los Departamentos de Medicina Crítica e Investigación Biomédica del Hospital Juárez de México. Es miembro del Sistema Nacional de Investigadores.

Lourdes Tavera Sierra (Ciudad de México, de 1968). Mexicana, matemática y profesora de la Facultad de Ciencias de la Universidad Nacional Autónoma de México. Desde el año 2010 estudia la dinámica de bifurcaciones de funciones de Markov.