La paradoja de Stein en estadística

La predicción óptima de un suceso futuro se elabora promediando los sucesos pasados. La paradoja de Stein determina las condiciones de existencia de estimadores que son mejores que la media aritmética

Bradley Efron y Carl Morris

Hay ocasiones en que un resultado matemático, cuya demostración es a todas luces válida, está en abierta contradicción con el sentir general. En 1955, Charles Stein, de la Universidad de Stanford, descubrió una de estas paradojas en estadística. Su resultado socavó siglo y medio de trabajos en teoría de la predicción, que se remontan hasta Karl Friedrich Gauss y Adrien Marie Legendre. Tras largo tiempo de resistencia a las ideas de Stein, salpicado por discusiones frecuentes y a veces ásperas, parece que la sensación paradójica ha disminuido, y dichas ideas comienzan a introducirse en estadística pura y aplicada.

La paradoja de Stein se refiere al empleo de promedios observados al objeto de estimar el valor de magnitudes no observadas. Entre las operaciones de la estadística, la obtención de promedios ocupa, por orden de importancia, el segundo lugar, estando el primero reservado al simplicísimo acto de contar. De un jugador de béisbol que consiga siete tantos en 20 ensayos con el bate durante un partido, se dice que tiene una "habilidad" de 0,350. Al calcular este valor se efectúa una estimación de la verdadera habilidad del jugador, usando para ello el promedio de sus tantos pasados. Si se nos preguntara qué resultados obtendrá el jugador en sus próximos 100 ensayos con el bate, probablemente diríamos que 35 tantos más. En la teoría estadística tradicional se puede demostrar que ninguna otra regla de estimación es uniformemente superior al promedio de observaciones.

Lo paradójico del resultado de Stein es que a veces contradice esta ley elemental de teoría estadística. Si se tienen tres o más jugadores de béisbol, y se desea estimar la habilidad futura de cada uno de ellos, hay un procedimiento mejor que la pura extrapolación a partir de sus promedios personales. Y el calificativo "mejor" es aquí muy significativo. Un estadístico que aplique a este problema el método de Stein puede esperar predicciones más exactas de los respectivos promedios futuros, independientemente de cuáles sean las habilidades verdaderas de los jugadores.

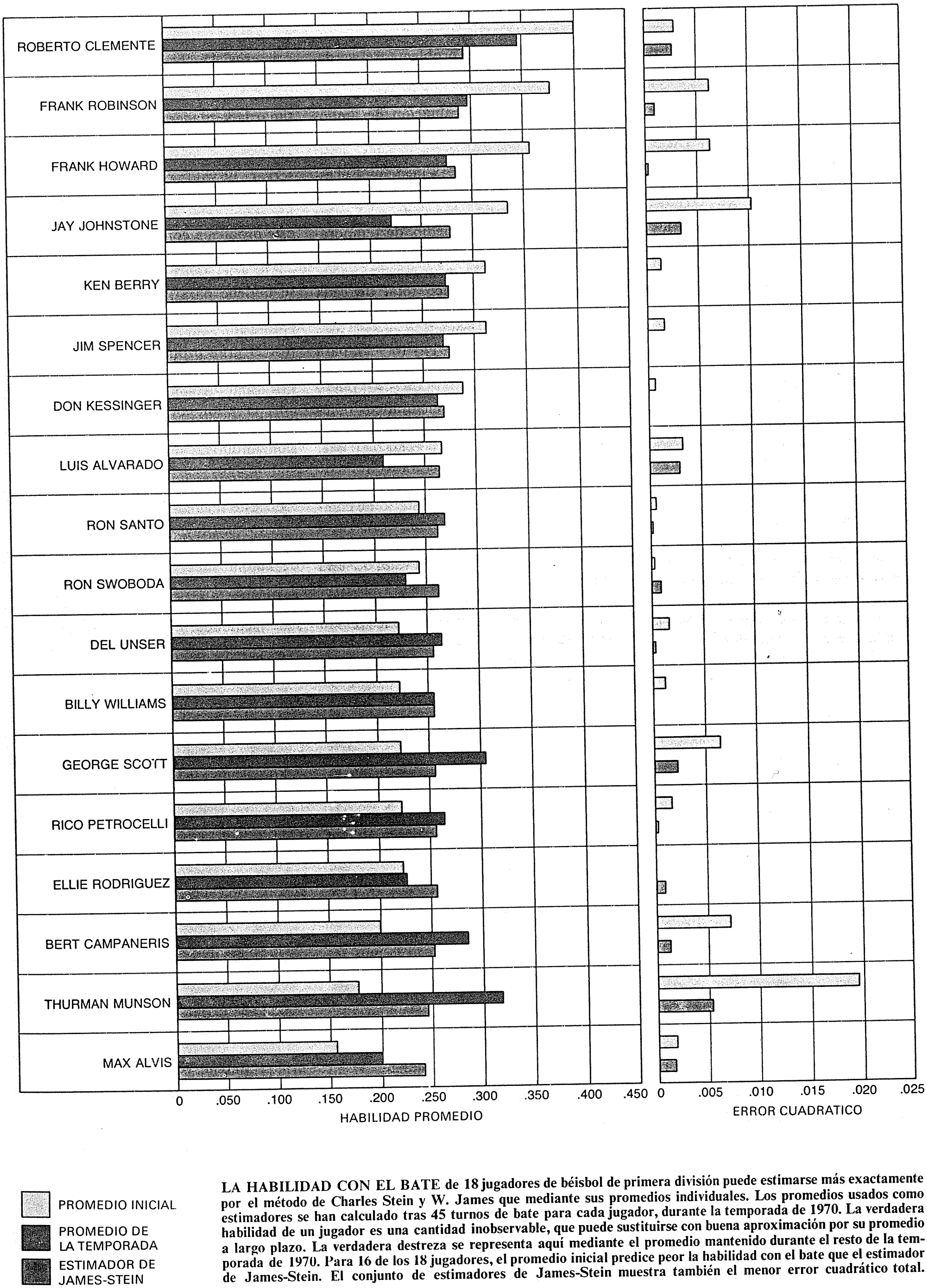
El béisbol es un deporte en que se han elaborado cuidadosamente largas series estadísticas, y que proporciona material muy adecuado para exponer el funcionamiento del método de Stein. Como datos primarios tomaremos los promedios de 18 jugadores de liga de primera división, computados tras sus primeros 45 turnos de bate, durante la temporada de 1970. Estos jugadores fueron todos los que completaron 45 turnos el mismo día que se hizo la tabulación. Los promedios se calculan, sencillamente, dividiendo el número de tantos por el número de turnos de bate; siempre es un número comprendido entre 0 y 1. Cada promedio se representará por la letra y.

El primer paso para aplicar el método de Stein es calcular el promedio de los promedios. Evidentemente, este "gran promedio", que representaremos por $\bar{y}$ estará también comprendido entre 0 y 1. El proceso esencial del método de Stein consiste en "contraer" todos los promedios individuales hacia este gran promedio. Si la marca de un jugador es superior al gran promedio, se debe disminuir dicha marca; si no alcanza el gran promedio, se deberá aumentar el tanteo previsto para el jugador. Designaremos por z al valor contraído resultante para cada jugador. Este valor se llama estimador de James-Stein de la habilidad del jugador, en honor de Stein y de W. James, quienes, en 1961, propusieron una versión especialmente sencilla del método. La paradoja de Stein consiste, simplemente, en que los valores z, es decir, los estimadores de James-Stein, dan mejores estimaciones de la verdadera habilidad de los jugadores que sus promedios simples individuales.

El estimador James-Stein de cada jugador se halla mediante la siguiente fórmula: $z=\bar{y}+c\,(y-\bar{y})$. El término $(y-\bar{y})$ expresa la diferencia entre el promedio de cada jugador y el gran promedio. La fórmula expresa que el estimador de James-Stein, z, se diferencia del gran promedio en esta misma cantidad $(y-\bar{y})$ multiplicada por una constante, c. La constante c se llama "coeficiente de constricción". Si fuese igual a 1, la fórmula diría que el estimador de James-Stein de un jugador dado es idéntico al promedio simple de dicho jugador; es decir, que y es igual a z. El teorema de Stein enuncia que c siempre es menor que 1. Su valor real viene determinado por la colección de todos los promedios observados.

En el caso de los datos del béisbol, el gran promedio, $\bar{y}$, es 0,265, y el coeficiente de constricción, c, es 0,212. Sustituyendo estos valores en la fórmula, vemos que, para cada jugador, z es igual a $0{,}265+0{,}212\,(y-0{,}265)$. En razón de que c es aproximadamente 0,2, cada promedio se "contraerá" alrededor de un 80 por ciento de su diferencia con el gran promedio, y la dispersión total de los promedios disminuirá también en aproximadamente un 80 por ciento.

A título de ejemplo, tomemos al fallecido Roberto Clemente, que era el bateador más destacado de las ligas de primera división de la época en que se tabularon nuestros datos estadísticos. Para Clemente, y es igual a 0,400, y z puede calcularse mediante la expresión $z = 0{,}265+0{,}212(0{,}400 - 0{,}265)$. Resulta 0,294. Dicho de otro modo, el teorema de Stein afirma que la verdadera habili-



LA HABILIDAD CON EL BATE de 18 jugadores de béisbol de primera división puede estimarse más exactamente por el método de Charles Stein y W. James que mediante sus promedios individuales. Los promedios usados como estimadores se han calculado tras 45 turnos de bate para cada jugador, durante la temporada de 1970. La verdadera habilidad de un jugador es una cantidad inobservable, que puede sustituirse con buena aproximación por su promedio a largo plazo. La verdadera destreza se representa aquí mediante el promedio mantenido durante el resto de la temporada de 1970. Para 16 de los 18 jugadores, el promedio inicial predice peor la habilidad con el bate que el estimador de James-Stein. El conjunto de estimadores de James-Stein muestra también el menor error cuadrático total.

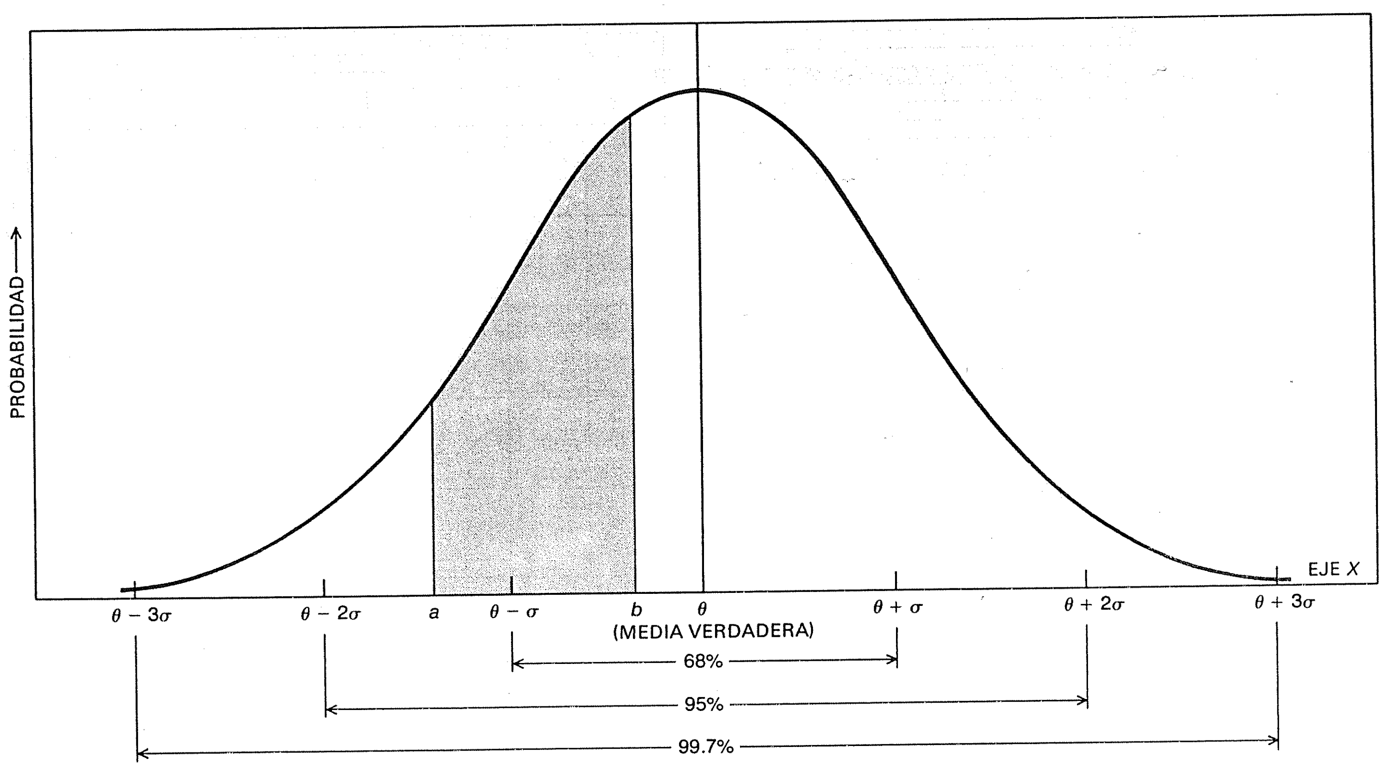
dad de Clemente puede estimarse con mayor precisión en 0,294 que en 0,400. Thurman Munson, que sufrió un brusco hundimiento al comienzo de la temporada de 1970, tuvo un promedio de tan sólo 0,178. Sustituyendo este valor en la fórmula, hallamos que su habilidad estimada es bastante mayor: el estimador de James-Stein para Munson es 0,247.

De los dos conjuntos de valores, los valores y y los valores z, ¿cuáles expresan mejor las habilidades de los 18 jugadores del ejemplo? Para poder responder con precisión a esta pregunta, tendríamos que conocer la "verdadera habilidad" de cada uno de ellos. Designaremos al verdadero promedio por Θ. Tal valor es incognoscible, y es una abstracción que representa la probabilidad de que un jugador obtenga un tanto en uno cualquiera de sus turnos de bate. A pesar de que Θ es incognoscible, tenemos una buena aproximación de su valor: los resultados obtenidos posteriormente por los bateadores. Bastará considerar el resto de la temporada de 1970, que aportó unas nueve veces más datos que los utilizados para el cálculo de los promedios iniciales. El error estadístico esperado con una muestra tal es lo suficientemente pequeño como para poder despreciarlo, y actuar como si el promedio de temporada fuese la "verdadera habilidad", Θ, del jugador. Este ha sido uno de los motivos de haber elegido promedios de béisbol para el ejemplo. En la mayor parte de los problemas, el verdadero valor de Θ no puede determinarse.

Un criterio de evaluación de ambos grupos de estimaciones puede consistir en contar sus respectivos éxitos y fracasos. Para 16 de los 18 jugadores, el estimador z de James-Stein es más próximo que el promedio observado y al valor "verdadero", o promedio de temporada, Θ. Un método más cuantitativo de comparar ambas técnicas consiste en hallar el error cuadrático total de cada estimador; para ello se determina primero el error real cometido en cada predicción, dado por $(\Theta - y)$ y $(\Theta - z)$, para cada jugador. A continuación, cada uno de estos valores se eleva al cuadrado, y se suman los valores cuadráticos obtenidos. Los promedios observados y tienen un error cuadrático total de 0,077, mientras que el error cuadrático total de los estimadores James-Stein es solamente de 0,022. Según este criterio de comparación, el método de Stein es 3,5 veces más preciso. Puede demostrarse que, para los datos de partida, el valor 3,5 está próximo a la razón esperada de los errores cuadráticos totales de los dos métodos. Así que no es que hayamos tenido pura suerte.

Supongamos que un estadístico hace un muestreo aleatorio de los automóviles de Chicago, y que descubre que, de los primeros 45 observados, nueve son de importación, y los restantes 36, de fabricación nacional americana. Se desea estimar la verdadera proporción de coches importados en Chicago, valor representado por Θ, y también imposible de observar. El valor observado, $9/45=0{,}200$, constituye una estimación de Θ. Puede obtenerse otra estimación englobando este problema con el de los 18 jugadores de béisbol. Introduciendo el valor 0,200 en la ecuación usada en ese problema, se obtiene para la razón de coches importados el valor de 0,251. (En realidad, la introducción de un nuevo valor modifica el gran promedio $\bar{y}$ y altera ligeramente el coeficiente de constricción c. Los cambios son, sin embargo, pequeños; el valor corregido de z es 0,249.)

En este caso, se tiene la fuerte impresión, intuitiva, de que la mejor predicción la dará el promedio observado, y no el estimador de James-Stein. Más aún, todo este segundo método parece absurdo. Pues, ¿qué relación hay entre los promedios del béisbol y la importación de automóviles? En casos así es donde mejor se aprecia la paradójica naturaleza del teorema de Stein, y donde resulta más incómoda. El teorema es tan válido para los 19 problemas como lo era para los primeros 18. Nada en su enunciado



LA DISTRIBUCION NORMAL de una variable aleatoria en torno a su valor medio fundamenta el método de estimación mediante promedios. La distribución normal queda definida por dos parámetros: la media, Θ, que sitúa el pico central de la distribución, y la desviación típica, σ, que mide el grado de dispersión de los datos. Al definir la distribución se supone que la variable x puede tomar cualquier valor del eje x. El valor más verosímil de x es, por definición, la media Θ. La probabilidad de que un valor de x se encuentre en un intervalo dado, como el de extremos a y b, es igual al área limitada entre la curva acampanada y la porción de eje comprendida entre estos puntos.

exige que exista una relación razonable entre los problemas componentes.

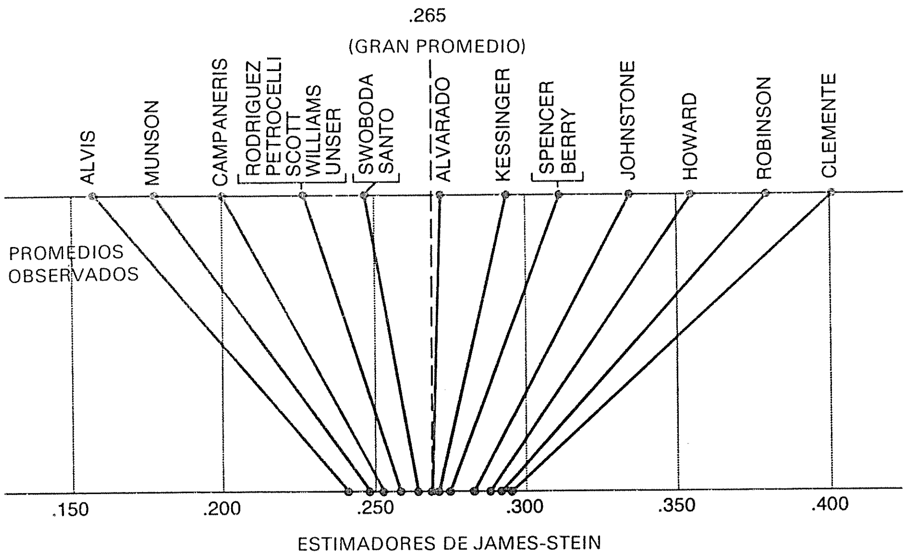
Esta desconcertante indiferencia hacia el sentido común puede ponerse de relieve de otra forma. ¿Qué tiene que ver el promedio observado de Clemente, 0,400, con Max Alvis, el peor de los 18 jugadores considerados? Si Alvis hubiese tenido una buena racha a principios de la temporada, marcando, por ejemplo, 0,444 en lugar del valor real de 0,156, el estimador de James-Stein de Clemente hubiese dado 0,325 en lugar de 0,294. ¿Por qué han de tener los éxitos o fracasos de Alvis ninguna influencia en nuestra estimación de la habilidad de Clemente? (Ni siquiera competían en la misma liga.)

Este es el tipo de objeciones presentadas por los críticos al sistema de Stein. Para poder replicarles se necesita estudiar el método con algo más de cuidado. El cálculo de promedios es un proceso fácil y muy conocido que, al parecer, no precisa justificación. En realidad, no es tan evidente por qué los promedios son a menudo tan útiles para estimar el verdadero centro de gravedad de un proceso aleatorio. La explicación se encuentra en la distribución que tienden a adoptar los valores de la variable aleatoria.

La distribución más generalizada en el trabajo científico es la distribución "normal", descrita por una curva campaniforme: fue estudiada profundamente por Gauss, y se llama en ocasiones distribución gaussiana. Para construirla, se supone que la variable aleatoria puede tomar cualquier valor a lo largo de un cierto eje; después, se hace que la probabilidad de que la variable se encuentre en un intervalo dado sea igual al área limitada entre la curva campaniforme y el intervalo en cuestión. La curva queda completamente determinada por dos parámetros: la media, Θ, que corresponde a la abscisa de la cumbre de la curva, y la desviación típica, que expresa el grado de concentración de los valores de la variable en torno a la media. Es costumbre representar la desviación típica por la letra griega σ. Cuanto mayor sea la desviación típica, tanto más dispersos se encuentran los datos.

En la teoría de la probabilidad, a partir de una media y una desviación típica conocidas se efectúan predicciones. En estadística, el problema discurre a la inversa: a partir de datos observados, el estadístico debe inferir la media Θ y la desviación típica σ.

Supongamos, por ejemplo, que la medición de cierta variable aleatoria x ha dado los cinco valores sucesivos 10,0, 9,4, 10,3, 8,6 y 9,7. Supongamos, además, que se sepa que dichos valores pertenecen a una distribución normal de



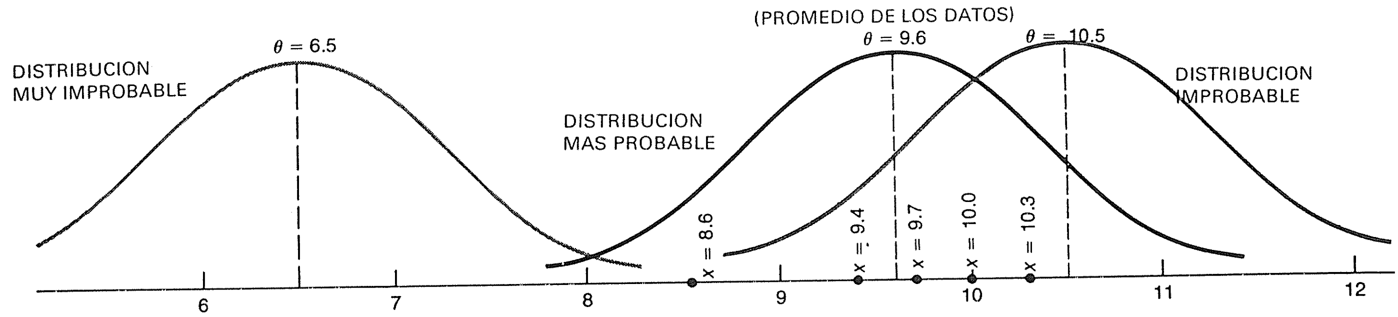
ESTIMADORES DE JAMES-STEIN para los 18 jugadores de béisbol, calculados "contrayendo" sus promedios individuales hacia el "promedio de promedios" global. En este ejemplo, el gran promedio es 0,265, contrayéndose cada promedio individual en aproximadamente el 80 por ciento de su distancia a este valor. Por lo tanto, el teorema que fundamenta el método de Stein afirma que las verdaderas habilidades de los bateadores están más densamente agrupadas de lo que parecen sugerir sus promedios de habilidad preliminares. La ilustración ha sido realizada por Gabor Kiss.

desviación típica igual a 1. ¿Cuál es el valor de la media verdadera, Θ, de la distribución? En principio, la media podría tener cualquier valor, pero ciertos valores son más verosímiles que otros. Por ejemplo, una media de 6,5 exigiría que todos los valores observados se encontraran en una cola de la curva, y ninguno cerca de su centro. Gauss demostró que, entre todos los valores posibles para la media, el valor promedio de los datos observados, $\bar{x}$ (que en este caso tiene el valor 9,6), hace máxima la probabilidad de obtener los datos observados en realidad. En este sentido, el promedio es la estimación más verosímil de la media. Gauss construyó la distribución normal de modo que tuviera esta propiedad.

Existe una razón más, también señalada por Gauss, para tomar el valor promedio como estimación óptima de la media, Θ, inobservable. Gauss hizo notar que el promedio de los datos es un estimador no "sesgado" de la media, en el sentido de que no favorece ningún valor especial de Θ. Con más precisión, el promedio no está sesgado porque el valor esperado de $\bar{x}$ es igual al verdadero valor de Θ, cualquiera que sea éste. Existen infinitos estimadores insesgados de Θ, ninguno de los cuales da una estimación perfecta. Gauss demostró que entre todas las funciones lineales de las observaciones que proporcionan estimaciones no sesgadas de Θ, el promedio, $\bar{x}$, es la que exhibe el error cuadrático de estimación mínimo. Hacia 1940 se demostró que ninguna otra función insesgada de los datos, sea lineal o no, puede estimar la media Θ más exactamente que el promedio, en lo que a error cuadrático se refiere. A la demostración de tal hecho contribuyó, de manera decisiva, R. A. Fisher, quien demostró hacia 1920 que toda la información referente a Θ que pueda extraerse de los datos se encuentra contenida en el promedio, $\bar{x}$.

En la década de 1930, Jerzy Neyman, Egon S. Pearson y Abraham Wald abordaron el problema de establecer la inferencia estadística sobre bases más rigurosas matemáticamente: su trabajo forma parte de la llamada teoría estadística de la decisión. Eliminaron la hipótesis de estimación insesgada, y examinaron todas las funciones de los datos que pudieran servir como estimadores de la media desconocida, Θ. Estos estimadores se comparan mediante una función de riesgo, que se define como el valor esperado del error cuadrático para cada valor posible de Θ.

Supongamos que se están examinando tres posibles estimadores: el promedio de los datos, $\bar{x}$; la mitad de dicho promedio, $\bar{x}/2$, y la mediana o valor central. Para el promedio y para la mediana, la función de riesgo es constante, lo que, dicho de otro modo, significa que el error cuadrático esperado de la predicción de la media Θ es el mismo, independientemente del verdadero valor de Θ. De las dos funciones de riesgo constantes, la del promedio es uniformemente menor en un factor de aproximadamente dos tercios; evidentemente, el estimador "promedio" es el preferido. En lenguaje de teoría de la decisión, la mediana es un estimador "inadmisible" de Θ, pues existe otro estimador de menor riesgo (error cuadrático esperado) independientemente del valor de Θ. (Es preciso decir, sin embargo, que cuando la distribución de



EN ESTADISTICA, EL PROBLEMA consiste en deducir, a partir de un conjunto de datos, la media verdadera y la desviación típica de su distribución. Aun sabiendo que la distribución es normal y que su desviación típica es 1, la media puede tener, en principio, cualquier valor. Sin embargo, algunos valores son más verosímiles que otros. Por ejemplo, los cinco puntos "dato" (x) mostrados en la figura corresponderían a una distribución normal de media 6,5, sólo si se encontraran a más de dos desviaciones típicas por encima de ella. Puede demostrarse que lo más verosímil es que estos datos hayan sido generados por una distribución de media igual al promedio de los datos, que se denota $\bar{x}$.

los datos no sea la normal, puede ocurrir que se invierta el orden de preferencia.)

Para el estimador $\bar{x}/2$, sesgado hacia el valor $\Theta=0$, la función de riesgo no es constante; este estimador es preciso cuando Θ se encuentre próxima a 0, pero el error cuadrático esperado crece rápidamente cuando la media verdadera se aleja del origen. La fundación de riesgo describe una parábola con mínimo en el punto $\Theta=0$. Si la media es efectivamente cero, la función de riesgo de $\bar{x}/2$ es cuatro veces menor que la del propio promedio. No obstante, para valores grandes de la media, el promedio $\bar{x}$ recupera su superioridad. Usando otros estimadores puede rebajarse en cualquier punto dado la función de riesgo hasta valores inferiores a los del promedio, pero siempre a costa de mayores riesgos en otros lugares.

Resta todavía la posibilidad de que algún otro estimador tenga riesgo uniformemente menor que el promedio. En 1950, Colin R. Blyth, Erich L. Lehmann y Joseph L. Hodges, Jr., demostraron la imposibilidad de la existencia de tal estimador. Dicho de otra forma, el promedio $\bar{x}$ es admisible, al menos al utilizarlo para determinar una media desconocida a partir de un conjunto finito de observaciones.

El teorema de Stein se refiere a la estimación de varias medias desconocidas. No se necesita presuponer relación alguna entre ellas; pueden ser habilidades de un bateador o proporciones de automóviles de importación. Por otra parte, se supone que dichas medias son independientes unas de otras. Al evaluar los méritos de los estimadores para estas medias conviene, de nuevo, emplear una función de riesgo definida como la suma de los valores de error cuadrático esperados de todas las medias individuales consideradas.

La elección primaria y obvia de un estimador de cada una de varias medias es tomar el promedio de los datos referentes a cada una de ellas. Todo el desarrollo histórico de la teoría estadística, comenzando por Gauss y pasando por la teoría de la decisión, viene a decir que el promedio es un estimador admisible en tanto que se trate de estimar una sola media, Θ. Stein demostró en 1955 que el promedio también es admisible para la estimación de dos medias. La paradoja de Stein consiste, simplemente, en que demostró que cuando han de estimarse más de dos medias, hacerlo a partir de sus correspondientes promedios es un procedimiento inadmisible. Sean cuales fueren los verdaderos valores de las medias, existen procedimientos de estimación con riesgo total inferior.

En 1955 Stein pudo tan sólo demostrar su proposición para casos en los que el número de medias, que denominaremos k, es muy grande. En el artículo de 1961, Stein, en colaboración con James, pudo extender el resultado para todo valor de k mayor que 2; además, lo hizo de manera constructiva. Stein y James no sólo demostraron que han de existir estimadores superiores en todo punto a los promedios, sino que además pudieron dar un ejemplo de tal tipo de estimador.

El estimador de James-Stein ya ha sido definido en nuestra exposición sobre la habilidad con el bate. Viene dado por la fórmula $z=\bar{y}+c\,(y-\bar{y})$, siendo y el promedio de un solo sistema de datos, $\bar{y}$ el gran promedio de los promedios, y c el coeficiente de constricción. Existen otras varias expresiones del estimador de James-Stein, pero difieren tan sólo en cuestiones de detalle. Todas tienen en común el coeficiente de constricción c, que es la característica fundamental del estimador de James-Stein.

En el problema del béisbol se manejó el coeficiente c como si se tratase de una constante. En realidad, está determinado por los promedios observados, y por consiguiente no es constante. El coeficiente de constricción viene dado por la fórmula

$$c=1-\frac{(k-3)\sigma^2}{\Sigma(y-\bar{y})^2}$$

En ella, k vuelve a ser el número de medias desconocidas a estimar, σ^2 es el cuadrado de la desviación típica, y $\Sigma(y-\bar{y})^2$ es la suma de los cuadrados de las desviaciones de los promedios individuales y respecto al gran promedio $\bar{y}$.

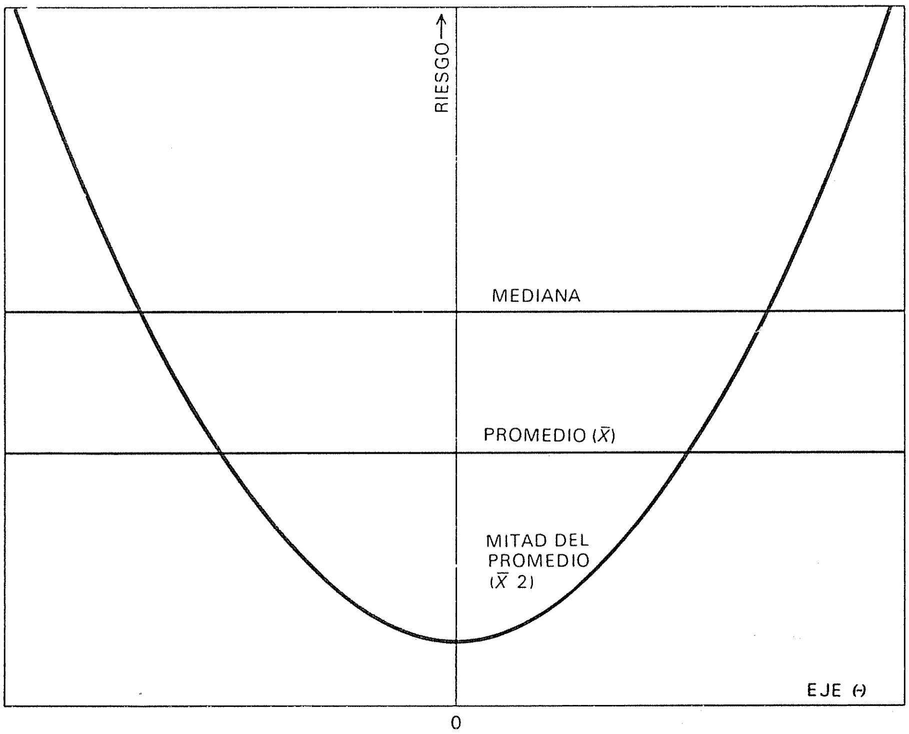
Examinaremos brevemente el significado de esta fórmula de aspecto poco accesible. Una vez fijados k y σ^2, vemos que el coeficiente de constricción disminuye (con lo que las predicciones de las medias se verán tanto más afectadas) al disminuir la expresión $\Sigma(y-\bar{y})^2$. Por otra parte, c aumenta, tendiendo hacia la unidad, cuando la expresión $\Sigma(y-\bar{y})^2$ aumenta, haciéndose menos drástica la contracción de las estimaciones.

¿Qué revelan estas fórmulas sobre el comportamiento del estimador? De hecho, el procedimiento James-Stein comienza conjeturando que las medias inobservables toman valores cercanos al gran promedio $\bar{y}$. Si los datos observados confirman esta conjetura, es decir, si los promedios de las observaciones no se desvían demasiado de $\bar{y}$, las estimaciones se contraen aún más hacia el gran promedio. Si la conjetura se ve infirmada, no se realiza apenas contracción. Estos ajustes del coeficiente de constricción se efectúan en virtud de la influencia que la distribución de los promedios en torno al gran promedio tiene en la fórmula que determina el valor de c. También influye en el coeficiente de constricción el número de medias a estimar, en virtud del factor $(k-3)$ que aparece en esta misma fórmula. Cuando hay muchas medias, la contracción impuesta por la fórmula es más drástica, pues en tal caso es menos verosímil que las variaciones observadas representen puras fluctuaciones aleatorias.

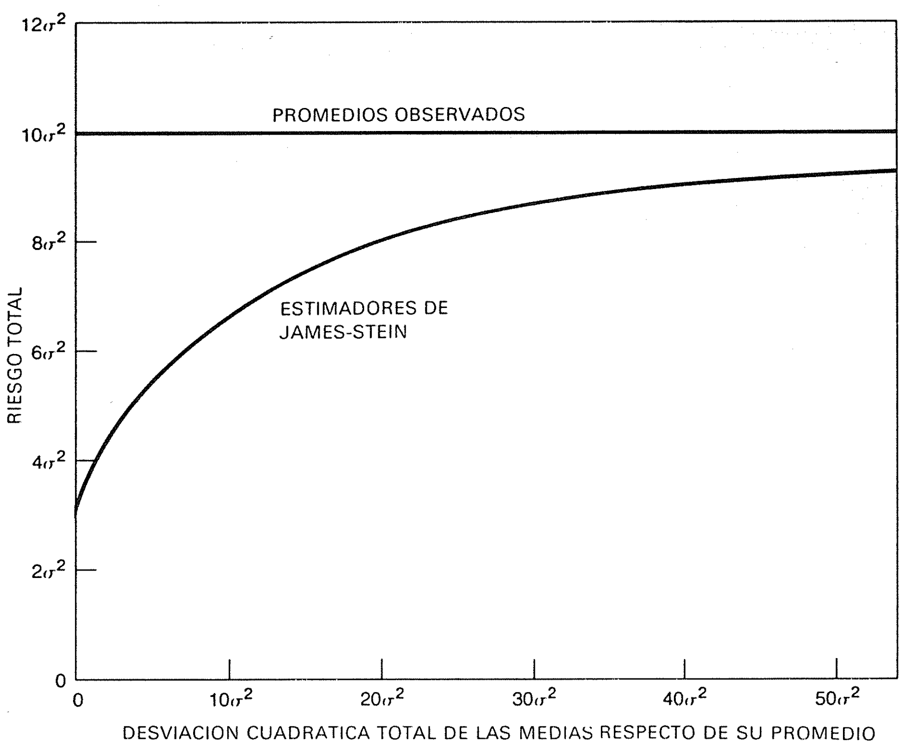
Calculado así el valor de c, la función de riesgo del estimador de James-Stein es inferior a la de los promedios de las muestras, independientemente de los verdaderos valores Θ de las medias. El riesgo puede disminuir de modo apreciable, especialmente si el número de medias excede de cinco o seis. A diferencia de lo que ocurre con los promedios observados, la función de riesgo no es constante para todo valor de Θ. El riesgo del estimador de James-Stein es mínimo cuando todas las medias verdaderas son iguales. Conforme las medias van alejándose unas de otras, va aumentando el riesgo del estimador, tendiendo hacia el de los promedios observados, aunque sin alcanzarlo jamás del todo. El estimador de James-Stein tan sólo se muestra verdaderamente superior a los promedios cuando las medias verdaderas están próximas entre sí, con lo que la conjetura inicial resulta confirmada. Lo que de todos modos resulta sorprendente es que el estimador se muestre siempre superior a los promedios, aunque sea marginalmente, con independencia de los verdaderos valores de las medias.

En la versión del estimador James-Stein aquí expuesta, se comparan con el gran promedio $\bar{y}$ todos los promedios observados. Este proceder no es el único posible; en otras expresiones del estimador se prescinde por completo de $\bar{y}$. Pero no es posible prescindir de cierta hipótesis inicial, arbitraria en mayor o menor grado, que sirva como origen o punto de referencia del estimador. Como es evidente, los promedios observados no dependen de la elección del origen. Antes de que Stein descubriese su método se tenía la opinión de que los estimadores "invariantes", como los anteriores, habrían de ser preferibles a los que dan predicciones dependientes del origen elegido. La teoría de invariancia, de la que Stein había sido uno de los principales defensores, se vio conmovida en sus cimientos por el contraejemplo de James-Stein. Desde el punto de vista matemático, éste es sin duda el aspecto más perturbador del teorema de Stein. De hecho, la paradoja no se descubrió antes por el prejuicio fuertemente arraigado de que, al estar planteado el problema de estimación sin referencia a ningún origen, debería resolverse de manera análoga, es decir, sin referencia a ningún origen.

En las aplicaciones del teorema de Stein suelen intervenir grandes conjuntos de datos con muchos parámetros desconocidos. Algunas de las dificultades de tales problemas, así como la potencia práctica del método, pueden ilustrarse mediante un ejemplo: el análisis de la distribución de la enfermedad llamada to-



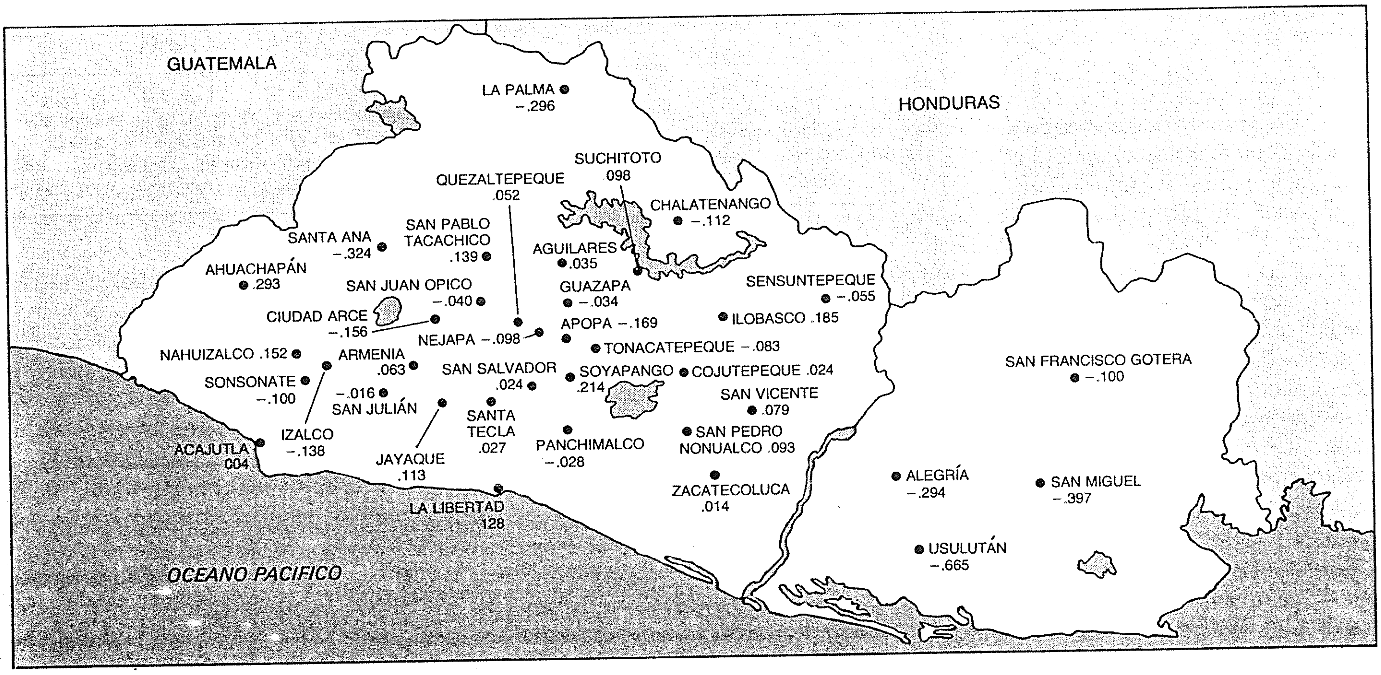
DIFERENTES ESTIMADORES de una única media verdadera, Θ, pueden evaluarse mediante una función de riesgo. El riesgo se define como valor esperado del error cuadrático de la estimación, considerado como función de la media Θ. El promedio de los datos, $\bar{x}$, es un estimador cuya función de riesgo es constante: independientemente del verdadero valor de la media, el valor esperado del error cuadrático es el mismo. La mediana, o valor central, de los datos también tiene riesgo constante, pero su riesgo es mayor en todo punto (por un factor de 1,57) al del promedio. La mitad del promedio ($\bar{x}/2$) es un estimador cuyo riesgo depende del valor real de la media; el riesgo es mínimo si la media se encuentra próxima a cero y aumenta rápidamente al alejarse del origen.



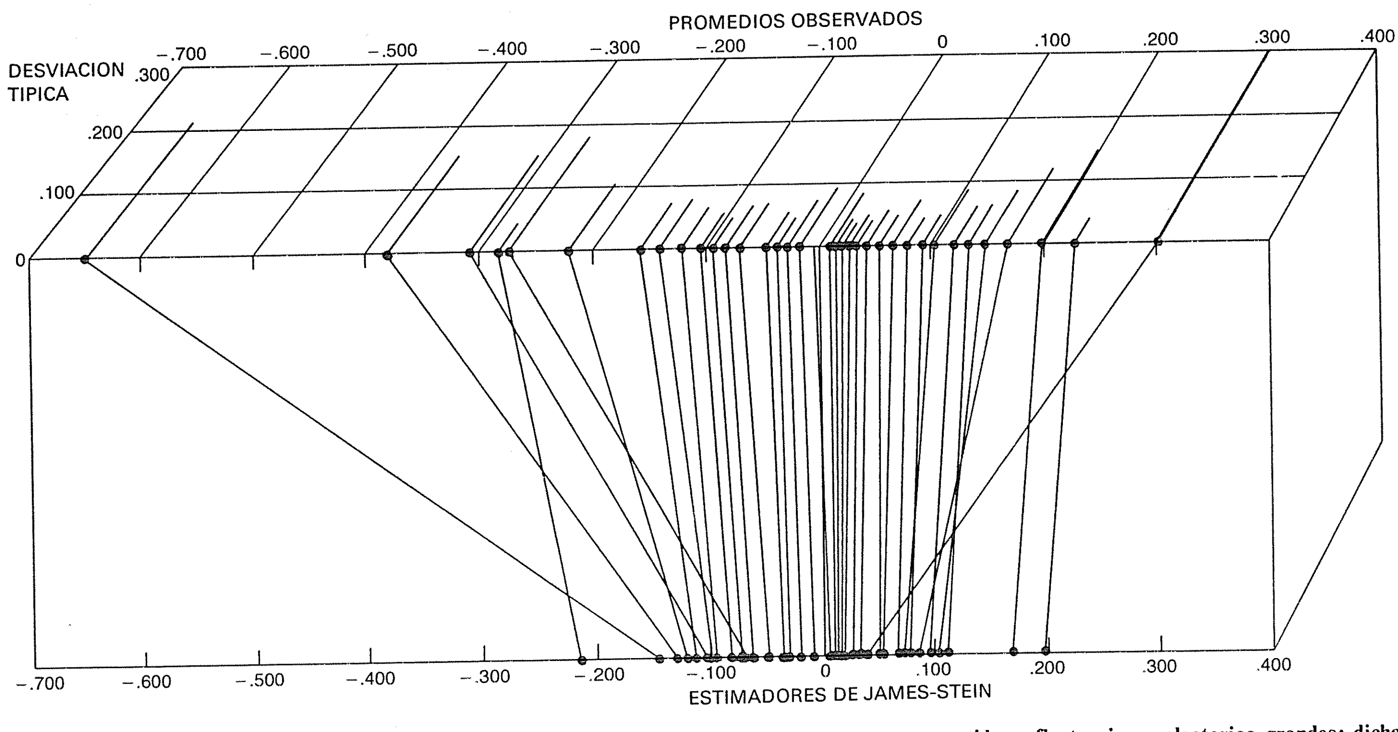
LA FUNCION DE RIESGO TOTAL del estimador de James-Stein es en todo punto menor que la de los promedios individuales observados, tan pronto como el número de medias a estimar sea mayor que dos. En este ejemplo hay 10 medias desconocidas. El riesgo es mínimo cuando todas las medias se apiñan en un solo punto. Conforme las medias se separan unas de otras el riesgo de los estimadores de James-Stein aumenta, tendiendo hacia el de los promedios, aunque sin alcanzarlo.

xoplasmosis en la república salvadoreña.

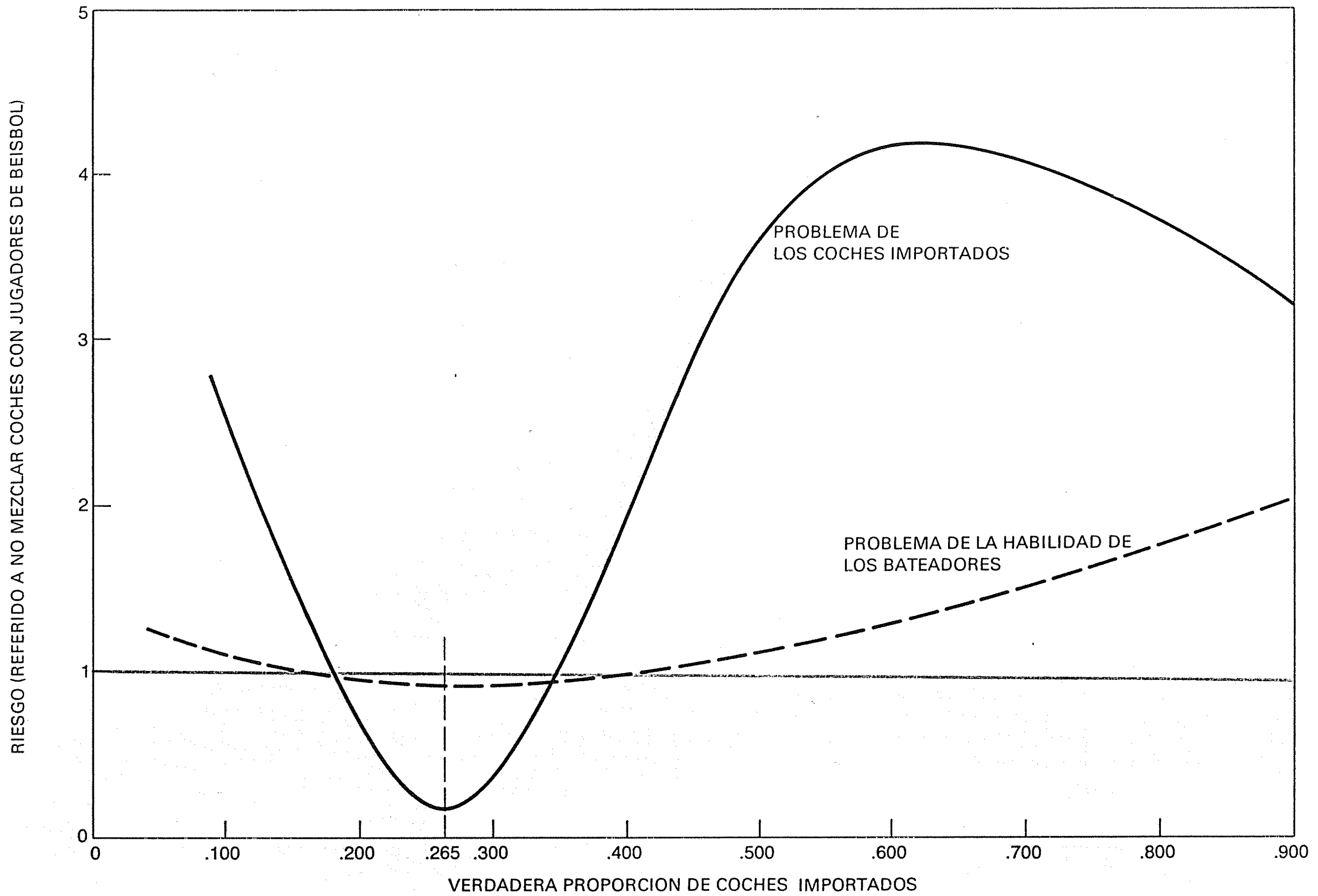
La toxoplasmosis es una enfermedad de la sangre, endémica en muchos países de América Central y, en general, en regiones tropicales. En El Salvador se examinó a aproximadamente 5000 personas, procedentes, en número variable, de 36 ciudades. Conviene expresar la morbosidad por ciudad comparándola con el valor nacional (es decir, con el gran promedio $\bar{y}$). Así, por ejemplo, una indicación de 0,050 representa una ciudad en la que la incidencia de la enfermedad es un 5 por ciento superior al promedio nacional. Las tasas medidas tienen una distribución aproximadamente normal. Las desviaciones típicas de dichas distribuciones son conocidas, pero varían de ciudad a ciudad, dependiendo inversamente del tamaño de la muestra examinada en cada una. La tarea del estadístico con-



MORBOSIDAD DE TOXOPLASMOSIS, una enfermedad de la sangre, medida en 36 ciudades de la República de El Salvador. La incidencia observada en cada ciudad puede tomarse como estimación de la verdadera incidencia, que es inobservable. Esta morbosidad observada tiene distribución normal, y su desviación típica está determinada por el número de personas examinadas en esa ciudad. Las mediciones se expresan en forma de diferencias respecto a la morbosidad nacional (que es el promedio de las morbosidades observadas en todas las ciudades). Así pues, una ciudad cuya incidencia observada sea −0,040 tendrá una morbosidad un cuatro por ciento menor que la global de todo el país.



CONTRACCION de los porcentajes observados de toxoplasmosis de acuerdo con la técnica de estimación de James-Stein. La distribución aparente de la enfermedad queda muy alterada. El coeficiente de contracción no es el mismo para todas las ciudades, sino que depende de la desviación típica del porcentaje medido en cada ciudad. Una desviación típica grande significa que la medida está basada en una muestra pequeña, y, por tanto, sometida a fluctuaciones aleatorias grandes; dicha medida se comprime más que las otras. En los datos referentes a El Salvador, las observaciones extremas tienden a estar en correlación con las desviaciones típicas máximas, lo que hace temer su escasa fiabilidad. Puede demostrarse que los estimadores de James-Stein implican, en comparación con los promedios observados, un menor error total.



PROBLEMAS SIN RELACION MUTUA pueden mezclarse en el método de análisis de Stein, pero a riesgo de aumentar el error. Por ejemplo, podría añadirse al conjunto de 18 promedios de béisbol un decimonono número que indicase la proporción de coches de importación observada en Chicago. Se podría calcular entonces nuevos estimadores de James-Stein para el sistema mixto de jugadores y automóviles, basados en el gran promedio de estos 19 números. En el enunciado del teorema de Stein nada impide proceder así, aunque su evidente arbitrariedad ha sido justificablemente criticada. En realidad, la inclusión de datos sin relación sólo podrá reducir la función de riesgo si la proporción de coches importados resulta hallarse cercana al promedio, 0,265, de habilidades de los bateadores; en caso contrario, el error de estimación esperado aumento tanto para los coches de importación en Chicago como en el caso de que se trate de los dieciocho jugadores.

siste en estimar la media verdadera Θ de la distribución de cada ciudad, a partir de la incidencia observada, y.

En este caso, la forma adecuada del estimador James-Stein es $z=cy$. La simplificación, que hemos introducido nosotros, se posibilita por la forma en que se ha elegido expresar las observaciones y. Se las define de modo que el gran promedio $\bar{y}$ sea cero, y todos los términos en que intervenga el gran promedio $\bar{y}$ desaparezcan, por tanto, de la fórmula. Por otra parte, el proceso de estimación es ahora más complicado, al ser el coeficiente de constricción distinto para cada ciudad, variando en proporción inversa a la desviación típica de y de cada ciudad. La dependencia del coeficiente de constricción respecto de la desviación típica tiene una justificación sencilla e intuitiva. Una desviación típica grande implica un alto grado de azar e incertidumbre en las correspondientes observaciones. Si la morbosidad observada es insólitamente grande, quizá sea más razonable atribuirla a fluctuaciones aleatorias dentro de la distribución normal que a un valor genuinamente grande de la media verdadera Θ. En tal caso se justifica aplicar un coeficiente de constricción pequeño, que reduzca drásticamente el valor anormal.

Para reforzar aún más estas consideraciones, volvamos por un instante al béisbol. En la temporada de 1893, Frank O'Connor fue "pitcher" del equipo de Filadelfia. En su carrera como jugador de primera división hubo de tomar el bate en dos ocasiones, ambas con éxito. Su promedio observado es, por tanto, 1,000. La regla de James-Stein para los 18 jugadores del ejemplo inicial estima que la verdadera habilidad de O'Connor es igual a $0{,}265+0{,}212\ (1{,}000 - 0{,}265) = 0{,}421$ (despreciando el efecto de los nuevos datos sobre el coeficiente de constricción). Esta estimación es absurda, pero no lo es tanto como 1,000. Alcanzar un promedio perfecto en dos ensayos no está en contradicción, en modo alguno, con un verdadero valor comprendido entre 0,242 y 0,294, que es el estimado para los otros jugadores. El coeficiente c aplicado a O'Connor debería ser todavía menor, para compensar el menor número de datos referentes a él.

En el caso de las observaciones de El Salvador, la mayoría de coeficientes de constricción son muy suaves; fluctúan en general entre 0,6 y 0,9, aunque unos cuantos oscilan entre 0,1 y 0,3. ¿Qué conjunto de cifras deberíamos elegir: las estimaciones de James-Stein, o las morbosidades observadas? La respuesta depende mucho de la aplicación que vaya a dárseles.

Si el Ministerio de Sanidad de El Salvador se propusiese construir una red de hospitales locales para enfermos de toxoplasmosis, las estimaciones de James-Stein ofrecerían probablemente una orientación más segura, pues el valor esperado del error cuadrático total es, en este ejemplo, unas tres veces menor para estas estimaciones. Es importante tener presente que el error previsto se ha calculado sumando los de todas las ciudades. Un hospital determinado podría estar mal emplazado o no tener la capacidad apropiada, pero la suma de todos los desajustes sería menor para los estimadores de James-Stein que para las morbosidades observadas.

También es verosímil que resulten preferidas las estimaciones del método James-Stein a la hora de determinar la ordenación de las medias verdaderas. Destaca, a este respecto, que la ciudad de máxima morbosidad (de acuerdo con los promedios medidos, y) ocupe el duodécimo lugar en el método de James-Stein. La drástica reducción de la estimación se debió a que en esa ciudad la muestra fue muy reducida. Esta información podría ser útil si solamente se dispusiera de fondos para un hospital.

Supongamos que un epidemiólogo desease investigar la correlación entre la morbosidad verdadera de cada ciudad y factores tales como la pluviosidad, la temperatura, la densidad de población, la altitud, etc. Volverían a ser preferibles los estimadores de James-Stein. Un cálculo somero mostraría que darían mejores aproximaciones en alrededor del 70 por ciento de los casos.

Hay una finalidad para la cual la morbosidad observada puede muy bien ser superior al estimador de James-Stein: cuando se estudia aisladamente una sola ciudad. Como ya hemos visto, el método de James-Stein da mejores resultados cuando hay muchas ciudades, y hace menor el error total de estimación de la suma correspondiente a todas las ciudades. No puede demostrarse, sin embargo, que el método de Stein sea superior al aplicarlo a una ciudad concreta; en realidad, la estimación de James-Stein puede ser francamente peor.

Pueden cometerse graves errores al estimar por el método de Stein la media verdadera de una ciudad aislada cuando dicha media tenga valor atípico. La idea subyacente en el método es reducir el riesgo global suponiendo que las medias verdaderas son más parecidas entre sí que los datos observados. Esta suposición puede degradar las estimaciones de medias genuinamente atípicas. Ahora podemos comprender por qué no deben mezclarse los automóviles importados en los mismos cálculos que los 18 jugadores de béisbol: hay una gran probabilidad de que los automóviles formen una población atípica.

Supongamos que se decide correr el riesgo y agrupar en uno solo los 19 problemas. Podemos calcular entonces el error cuadrático total esperado en función del verdadero porcentaje de coches importados. Resulta que el riesgo, tanto para los jugadores como para los autos, se reduce solamente si ocurre que el porcentaje de importación cae dentro de la gama de valores de habilidad de los bateadores; de lo contrario, el riesgo de error aumenta para ambos problemas.

Decidir si una media concreta es o no "típica" es una cuestión sutil, cuyas implicaciones aún no se comprenden del todo. Volvamos al problema de la toxoplasmosis en El Salvador, y fijémonos en la ciudad de Alegría, que muestra la quinta cifra más baja de incidencia observada de la enfermedad: $-0{,}294$. Se trata de una de las cuatro ciudades situadas al este del río Lempa incluidas en la encuesta; todas ellas tienen valores netamente negativos de la morbosidad medida. Es válido suponer que no se trata de coincidencias, y que la proporción de toxoplasmosis al este del río Lempa es verdaderamente inferior. Un estimador de James-Stein que dé información consolidada para todo el país puede ser mucho menos que óptimo para estas ciudades. Nosotros hemos desarrollado técnicas para aprovechar este tipo de información extra, pero la teoría que sustenta dichas técnicas sigue siendo rudimentaria.

Con un poco de malicia, un aficionado al béisbol podría caer en la cuenta de que así como la habilidad individual de cada jugador puede representarse mediante una curva de Gauss, también las verdaderas habilidades de todos los jugadores de primera división tienen una distribución aproximadamente normal. La media de esta distribución es 0,270 y la desviación típica 0,015. Con esta valiosa información adicional, que los estadísticos llaman distribución previa, se puede construir una estimación mejor de la verdadera habilidad de los jugadores. Este nuevo estimador, que denotaremos Z, se define por la ecuación $Z=m+C(y-m)$, siendo y la habilidad observada de cada jugador, aunque aquí el gran promedio y se ha sustituido por m, que es la media de la distribución previa, cuyo valor conocido es 0,270. Hay además un coeficiente de constricción distinto, C, que depende de modo sencillo de la desviación típica de la distribución previa (igual a 0,015).

Este procedimiento no es un refinamiento del método de Stein; por el contrario, es 200 años anterior, y es la formulación matemática de un teorema publicado (póstumamente) por el reverendo Thomas Bayes en 1763. Bayes pudo demostrar que este estimador hace mínimo el error cuadrático previsto asociado al carácter aleatorio tanto de los promedios observados (y) como de las medias verdaderas, (Θ).

El parecido de las fórmulas de los estimadores de Bayes y de James-Stein llama poderosamente la atención. De hecho, cuando el número de medias consideradas tiende a infinito, ambas ecuaciones se hacen idénticas. Los dos coeficientes de constricción c y C convergen hacia un mismo valor, y el gran promedio $\bar{y}$ se hace igual a m precisamente cuando se hace intervenir en el cálculo a todos los jugadores. El procedimiento de James-Stein tiene, sin embargo, una importante ventaja sobre el método de Bayes: es aplicable sin necesidad de conocer la distribución previa. En realidad, ni siquiera es preciso suponer que las medias que se quiere estimar tengan una distribución normal. Por otra parte, la ignorancia tiene un precio, que se paga con una precisión inferior de las estimaciones. Nosotros hemos podido demostrar que el método de James-Stein aumenta la función de riesgo en proporción a $3/k$, siendo k, como antes, el número de medias por estimar. El riesgo extra es despreciable cuando k pasa de 15 o 20, y es tolerable para valores tan bajos como $k=9$.

En este contexto histórico, el estimador de James-Stein puede considerarse como una "regla de Bayes empírica", según la expresión acuñada por Herbert E. Robbins, de la Universidad de Columbia. En unos trabajos que comenzaron en 1951, Robbins demostró que puede alcanzarse el mismo riesgo mínimo asociado a la regla de Bayes, aun sin conocer la distribución previa, siempre que el número de medias por estimar fuese muy grande. La teoría de Robbins se reconoció de inmediato como una aportación fundamental; mucho más lenta ha sido la aceptación del resultado de Stein, con el que guarda estrecha relación.

El estimador de James-Stein no es el único del cual se sepa que es mejor que los promedios muestrales. En realidad, el propio estimador de James-Stein es inadmisible. Su defecto consiste en que el coeficiente de constricción c puede tomar valores negativos, y entonces puede ocurrir que las medias, en vez de contraerse hacia el gran promedio, se dispersen. En tales casos, basta cambiar c por cero para obtener una estimación mejor. A su vez, este estimador también es inadmisible, aunque todavía no se conoce un estimador uniformemente superior.

Continúa la búsqueda de nuevos estimadores. Los esfuerzos más recientes se están concentrando en la obtención de resultados análogos a los de Stein para problemas con distribuciones distintas de la normal. Varios métodos de investigación, entre los que se encuentran los de Stein, Robbins y métodos bayesianos más formales, parecen estar convergiendo hacia una potente teoría general de estimación de parámetros.

Número 10
Julio 1977

INVESTIGACION Y
CIENCIA
Edición en español de SCIENTIFIC AMERICAN

EDITA Prensa Científica, S.A.
Calabria, 235-239
Barcelona-15
ESPAÑA