

EL AZAR (versión preliminar)*

Mario Wschebor
Centro de Matemática
Facultad de Ciencias

El tema de este texto es indudablemente complicado, vasto y difícil. Siento que me desborda el desafío de decir algo sobre él que pueda ser útil para quien lo lea, porque yo mismo estoy bastante dominado por una sensación de perplejidad y de interrogación sobre ciertos aspectos sustanciales. Debo por lo tanto pedir tolerancia con respecto a algunas afirmaciones que posiblemente no sean completamente ciertas, lo cual debe atribuirse a la existencia de problemas que no he podido estudiar o entender adecuadamente.

En verdad, lo que habremos de considerar es bastante más limitado que lo que puede imaginar un lector desprevenido del título. El azar está presente en nosotros y en nuestra visión del mundo, es la manera como solemos designar lo insondable, lo que no tiene explicación. Todos los mitos registrados giran en torno a lo inexplicable de la vida o del mundo inanimado, o a los caprichos de las personas, de los reyes, de los dioses. No voy a referirme a ninguno de estos temas, ni a ningún asunto en el cual la idea de azar esté asociada con la más o menos vaga o heurística de voluntad, humana o divina. No lo haré, básicamente, porque es extraordinariamente difícil para mí incursionar en ese mundo y entender algo sobre él.

Confieso que la carencia de una formación filosófica satisfactoria me plantea una dificultad adicional para el abordaje del tema del azar. Se ha dicho mucho, y se repite ahora por parte de algunos filósofos, que nunca la vida y la sociedad humanas han estado tan sumergidas en la incertidumbre sobre el futuro como en la actualidad. En un plano puramente sentimental, yo no lo creo; la existencia humana ha estado y está rodeada por esa cantidad de cosas inexplicables que suscitan las mismas preguntas que están registradas en la Historia, y no creo que esta época sea particularmente incierta desde ese punto de vista. Las filosofías que ponen el acento en esto no me satisfacen ni me gustan, y empleo ex profeso una terminología totalmente subjetiva para referirme a este asunto.

Estas precisiones acerca de lo que no voy a hablar, tiene que ver con las extrapolaciones fáciles que se hacen de las visiones de la actividad científica o de la reflexión del mundo académico. Bueno, en realidad yo no sé bien cuales son las reglas o las características especiales de ese mundo, no sé que es eso que algunos denominan “método científico”, pero en todo caso, no tengo más remedio que reconocer que existe un conjunto de personas que se definen a si mismas como hacedores de ciencia. Lo que también es cierto es que, con frecuencia, afirmaciones que se hacen desde el medio – o desde el conocimiento – científico, son motivo de extrapolaciones fáciles. El tema de la incertidumbre es particularmente propicio para sustentar en presuntos hechos comprobados o en teorías serias, toda clase de improvisaciones, y se prolonga en discursos que con frecuencia carecen de todo fundamento. Por eso, antes que nada, es mi deber prevenir contra las extrapolaciones fáciles de lo que aquí se dice.

Una última reserva es que reduciré al mínimo todo recurso técnico. Dicho de otro modo, cualquier persona con una educación básica y un interés por el tema debería poder leer más que el 90% de este texto, aunque no tuviera una formación matemática. Esto es siempre un compromiso difícil, no quiero decir que imposible: el no recurrir a las verdaderas cuestiones técnicas suele esconder los auténticos

* Versión corregida por el autor.

problemas y edulcorar las dificultades en exceso. En el otro extremo, está la intención de que los temas lleguen más allá del público de iniciados. Digamos que me sentiría muy satisfecho si los lectores no matemáticos encontraran en los problemas que aquí se mencionan una razón para estudiar algo más de Matemática que la que conocen actualmente, y que ello les permitiera acceder a una literatura especializada más profunda que estas páginas, que intentan ser de divulgación.

La arqueología registra la existencia de juegos de azar desde la Antigüedad. Los seres humanos han jugado a la taba con huesos de animales, probablemente desde la Prehistoria; hay indicios que respaldan esta afirmación. Mucho más próximamente, se ha afirmado con fundamento que los caldeos y los egipcios jugaban a los dados, al mismo tiempo que los adivinos trataban de predecir el futuro de las sociedades y del universo. Pero las personas que jugaban a los dados entendían el azar de otra manera que las que predecían el futuro. Se han descubierto dados cargados en tumbas egipcias, ésa es una manera elemental de hacer cálculos con el azar. Quiere decir que había quienes especulaban con el azar hace muchos siglos, hace 4.000 años, probablemente desde mucho antes. Lo sorprendente, y acá entramos en tema es que, probablemente hasta fines del siglo XVIII, la visión del azar que no tuviera que ver con los misterios de la voluntad divina o humana, se reducía esencialmente a los juegos.

En el siglo XVIII, en Inglaterra se jugaba a los dados. Quizá también se hiciera en los salones franceses de la misma época, donde se apostaban grandes fortunas sobre la sobrevida de los grandes personajes del reino. ¿Con qué criterio se jugaban esas fortunas?

Uno puede imaginar sin mucha dificultad que, en el juego de los dados, cargados u honestos, hay una manera de calcular cuál debería ser la ganancia o la pérdida derivada de una apuesta. Supongamos que el dado es honesto y que apostado un peso a que sale un 3; si me pagan cinco pesos si sale el 3 y se quedan con mi peso si no sale, entonces el juego es equitativo. Si, en cambio, en el mismo juego me pagan sólo 4 pesos cuando sale el 3 y se quedan con el peso de mi apuesta si el 3 no sale, entonces, una de dos: o el dado es honesto y me están perjudicando, o el dado no es honesto.

¿Pero cómo hace uno para apostar que el Ministro de Fomento de Luis XV va a vivir 5 años más, o 10 años más o cosas parecidas? ¿Cuánto se apuesta? Este es un aspecto un poco gracioso de la cuestión, pero si el lector reflexiona, verá que se trata de un problema muy real, porque si bien nosotros creemos que sabemos bastante acerca del comportamiento de los dados con los que jugamos, la verdad es que en muchos de los problemas en los que el Cálculo de Probabilidades habrá de aplicarse, aparecen situaciones más bien parecidas a las del Ministro de Fomento. Es decir, tenemos una primera dificultad que es elegir un modelo de comportamiento de las variables que intervienen y una segunda, que consiste en tener un criterio para determinar si el modelo considerado es aceptable para hacer los cálculos.

Comienzos del cálculo de probabilidades

Algunos de los grandes matemáticos de fines del siglo XVIII y de la primera mitad del XIX, pienso en Laplace, en Lagrange, en Cauchy o en Gauss, escribieron sobre el Cálculo de Probabilidades textos que son extremadamente elementales y simples, lo cual no puede dejar de llamar la atención a los matemáticos actuales, cuando se compara esos trabajos con el enorme aparataje, el ingenio y la teoría con que los mismos matemáticos atacaban problemas de Análisis, de Álgebra o de Geometría. Hasta fines del siglo XIX, hasta Poincaré que murió en 1912 y que también escribió un tratado de Cálculo de

Probabilidades cuya última edición se publicó pocos meses antes de su muerte, lo que escriben los matemáticos sobre este tema es extremadamente ingenuo y elemental. Hay algunos casos de elaboraciones más complicadas, pero no hay una verdadera teoría.

El Cálculo de Probabilidades, la Matemática del azar, es una rama de la Matemática propia del siglo XX y se puede pensar que su comienzo real data de alrededor de 1925. Como siempre, se trata sólo de una fecha aproximada para dar una idea de lo que queremos decir; como en toda ciencia, en la Matemática interactúan la continuidad y el cambio, mezclándose de una manera a veces difícil de precisar. Aún a posteriori, suele no ser trivial reconstruir el itinerario intelectual, las influencias y el hilo real de las ideas, a pesar de que hay momentos que aparecen como de salto en el que nuevas concepciones entran en la escena. No estamos diciendo, entonces, que no hubiera cálculos de probabilidades antes de esa fecha. En particular, encontramos esos cálculos en problemas que prestaron atención a necesidades sociales y económicas y que fueron interesantes desde el siglo XVIII.

En ese período inicial, la mayoría de estas manifestaciones fueron más bien de tipo empírico. La primera gran aplicación fue la construcción de tablas de mortalidad. Las tablas de mortalidad se desarrollaron básicamente en el siglo XVIII y fueron impulsadas en Inglaterra por las compañías de seguros. Inventaron el seguro de vida, mediante el cual una persona al morir recibe una cantidad determinada de dinero. ¿Cuánto tenía que pagar, para después recibir una suma determinada? ¿En función de qué? En función del riesgo de morir. Esa cosa azarosa que había rodeado a la historia de la especie humana, que era el misterio de la muerte, había que medirla de alguna manera. ¿Qué hacían los actuarios? Construían tablas de mortalidad en base a ciertas reglas, para determinar la probabilidad de que, por ejemplo, una persona que tenía 48 años viviera 10, o 12, o 15 años más. Distintos métodos fueron inventados - o comenzados a inventar - en esa época y utilizados por las compañías de seguros para establecer la prima que tenía que pagar una persona en función de su edad (y posiblemente también de otros factores), a los efectos de que su familia tuviera un premio cuando ella muriera. Es decir que estos seguros son más bien seguros de muerte que seguros de vida.

En un plano completamente distinto, a fines del siglo XVIII aparecen algunos problemas serios vinculados con la aplicación de la Ley de Newton a la determinación de las trayectorias de los astros. Laplace, que fue posiblemente uno de los más grandes analistas de la historia, un gran calculista, tenía dificultades severas en esta materia. Aplicaba la Ley de la Gravitación a la determinación de las trayectorias de los astros, lo que implicaba que tenía que resolver las ecuaciones diferenciales que regulan su movimiento. Pero sólo lo podía hacer en forma aproximada, en virtud de una serie de razones.

En primer lugar, si bien la Ley de la Gravitación es de expresión general sencilla, su aplicación es bastante complicada, dado que se debe hacer intervenir la interacción entre los distintos cuerpos celestes. Por lo pronto, estaba obligado a reemplazar aquel sistema complicadísimo donde tenía 10 o 12 objetos celestes que interactuaban entre sí, por un conjunto de 3 o 4, digamos el Sol, Saturno, Júpiter, y algunas lunas, y esto ya era muy complicado para lo que estaba en condiciones de hacer. Esta es una primera aproximación, sustituir el sistema solar real que era el motivo de su interés, por uno más pequeño.

Segundo, para realizar el cálculo efectivo estaba obligado a reemplazar ciertas funciones que figuran en las ecuaciones, por aproximaciones manejables, dadas por desarrollos aproximados válidos dentro de ciertos límites.

Tercero, en las ecuaciones figuran ciertos parámetros físicos que deben ser objeto de medidas directas o indirectas. Esas medidas sólo pueden llevarse a cabo con error. Primero, porque los aparatos que debemos utilizar tienen una cierta sensibilidad, es decir, no son capaces de detectar valores de la magnitud a medir que estén por debajo de un cierto valor. Segundo, porque la medida es necesariamente afectada por las condiciones en las que se realiza, y es por eso que si uno la repite, aunque tenga el mayor cuidado para que esas condiciones varíen lo menos posible, la experiencia muestra que el nuevo resultado no tiene porqué coincidir con el anterior. Esos errores se propagarán a las soluciones de las ecuaciones, una vez realizado el cálculo.

Y cuarto, los cálculos requieren redondeo y los errores de redondeo también se propagan a la solución final.

Se debe agregar que si uno quiere seguir la solución en el tiempo (es decir, si se desea la descripción de la posición y de la velocidad de cada cuerpo como función del tiempo), los errores previamente mencionados - de medida, de aproximación de ciertas funciones que aparecen en las ecuaciones o en las soluciones, también los errores de redondeo - se van acumulando y las cotas que podemos asegurar para el error total cometido, son cada vez más groseras a medida que aumenta el horizonte temporal en que queremos describir el movimiento, hasta el punto que, más allá de un cierto tiempo futuro, nuestro cálculo deja de tener interés para la descripción del movimiento.

En resumen, lo que ocurría es que los cálculos conducían a algunos resultados que eran distintos de los registros empíricos muy antiguos del movimiento de los astros. Los astros estudiados eran sobre todo los planetas y también los cometas.

(Casualmente, hablando de cometas, el primero que hizo una tabla de mortalidad fue Halley, el que dio nombre al cometa célebre. Él fue el primero que hizo rigurosamente una tabla de mortalidad, quizá tuviera por cliente alguna compañía de seguros, no lo sé... El segundo que hizo una tabla de mortalidad perfeccionando la de Halley, fue Huyghens, el gran físico creador de la Teoría Ondulatoria de la luz. Como se ve, contra lo que se imaginan en nuestros días algunos que lamentablemente entienden muy poco sobre la creación intelectual y al mismo tiempo quieren orientarla, los científicos más “básicos” intervenían al mismo tiempo en la resolución de cuestiones inmediatas de la vida social).

Los astrónomos, sobre todo Laplace que perfecciona los cálculos de Newton y de los astrónomos intermedios, tenían entonces que entender la razón de esas diferencias entre los cálculos hechos aplicando la Ley de la Gravitación y las observaciones que se habían hecho desde la época de los Caldeos. [Hay que señalar que los astrónomos del Renacimiento conocían las observaciones de los Caldeos, que habían sido transmitidas por Ptolomeo al mundo occidental en un registro extraordinario de lo que era conocido entonces].

Laplace no dudaba ni de la Ley de la Gravitación ni de la información empírica de que disponía. Consideraba que se trataba de un problema que se debía a los errores, errores de observación por un lado y errores de cálculo por otro. Los cálculos, decía Laplace, no representan la totalidad del fenómeno. No se trataba de poner en cuestión la Ley de Newton, sino de desarrollar una teoría de los errores, que permitiera entender que la diferencia entre el cálculo y la observación estaba dentro del límite de lo que la tal teoría permitiría deducir. Para Laplace el cálculo de probabilidades era, antes que nada, una teoría de errores. Aquí no está planteada de ninguna manera la idea de que la descripción física y la manera como el hombre ve la Naturaleza, pudieran incluir de manera esencial el azar. Es el ejemplo clásico del mecanicismo. Utiliza la Ley de la Gravitación, las ecuaciones diferenciales que rigen el movimiento, que considera leyes universales.

Sea esto dicho para nuestros actuales estudiantes de Matemática, Laplace todavía no conocía un teorema de existencia y unicidad de soluciones, es decir un teorema que diga que el conjunto de las posiciones y las velocidades de todos los astros del sistema en un instante dado, determina para siempre el movimiento en el futuro. No tenía un teorema pero estaba seguro de que era así, no lo cuestionaba y lo que Laplace decía sobre esto, era esencialmente correcto. El primer teorema de existencia y unicidad de soluciones, se debe probablemente a Cauchy, que es posterior a Laplace en algunas décadas. La preocupación por dar rigor matemático a esta idea, estuvo sin duda asociada a que la misma ya preexistía en la cabeza de los físicos, cosa que ha sido una constante en el relacionamiento entre ambas disciplinas a lo largo de la Historia. La idea es que el sistema astral, que es un sistema de partículas desde este punto de vista, partículas que tienen masa y tienen volumen también (en realidad, no son sólo partículas, son cuerpos rígidos con movimiento propio, de modo que es algo más complicado), está regido por la Ley de la Gravitación la que, junto con las posiciones y las velocidades iniciales, determina la totalidad del movimiento en el futuro. Todo lo que difiera de eso en los cálculos, se debe a que todavía no se tiene la competencia para poder hacerlos de manera adecuada, o a que las medidas no son suficientemente precisas.

¿Qué tiene que ver todo esto con el azar? Tiene que ver y mucho. Imaginemos que una buena teoría de errores permitiera determinar un horizonte temporal durante el cual se pudiera mantener bajo control los errores cometidos al calcular las trayectorias de los objetos celestes, aplicando la Ley de la Gravitación. ¿Qué pasaría después de ese horizonte? ¿Cuál iba a ser el destino del Universo cuando las órbitas reales, que diferían en cierta magnitud de las órbitas predichas calculando en forma aproximada con la Ley de Newton, fueran más allá de un cierto valor y que las aproximaciones utilizadas dejaran de ser válidas? ¿Las órbitas cambiarían mucho? ¿Habría una especie de cataclismo, es decir, no sólo errores mayores, sino un cambio cualitativo en el comportamiento de las trayectorias de los astros? El problema, y acá aparece el tema de los errores y del cálculo de las probabilidades, era la estabilidad del Sistema Planetario.

El Sistema Planetario, tal como lo conocemos, tiene órbitas elípticas o casi elípticas y por lo tanto planas o casi planas, que se mueven lentamente. ¿Este Sistema Planetario es estable? ¿O habrá una deriva hacia una situación que va a cambiar completamente la configuración astronómica y por lo tanto todo lo que hay dentro de ella? Es decir que la teoría de errores para Laplace está asociada con el destino de todo lo que hay, del Universo, del Sistema Planetario. Ese era esencialmente su problema.

El siglo XIX perfecciona las técnicas de Laplace, aunque sin cambiar la concepción, más bien las refina y las afina. El mecanicismo con que culmina el siglo XIX es el éxtasis de la Gravitación, desde el punto de vista de las ideas sobre el mundo físico, hasta llegar a Poincaré. Alrededor de 1900, son esas ideas las que dominan la manera de ver el Cálculo de Probabilidades.

El siglo XX mirado hacia atrás y hacia adelante desde 1950.

Ahora, en vez de seguir un itinerario lineal, vamos a intentar dar un salto de 50 años y colocarnos alrededor de 1950. Es decir, en lugar de mirar el siglo XX desde el año 1900, vamos a intentar hacerlo desde 1950, hacia atrás y hacia adelante. ¿Por qué? Porque en 1950 aparece una idea radical de una importancia extraordinaria, que es la idea de Shannon según la cual la incertidumbre se puede medir. Esta idea es uno de los cambios más importantes que aporta el siglo XX al tratamiento de la cuestión del azar, y es por eso que no se puede dissociar del hecho de que el Cálculo de Probabilidades sea una rama de la Matemática que corresponde al siglo XX. Algunas de estas ideas ya estaban presentes antes

de Shannon, (habremos de hacer alguna alusión a esto más adelante) aunque Shannon las potencia y las convierte en un instrumento de la ingeniería de comunicaciones, en una época en que esta disciplina toma impulso, para convertirse en uno de los hechos de mayor gravitación en la sociedad en la segunda mitad del siglo XX.

En forma harto esquemática, podemos ubicar el comienzo del Cálculo de Probabilidades (que hoy, en una diferencia sutil pero significativa, se suele llamar "Teoría de la Probabilidad") como rama moderna de la Matemática en los años de 1920, y probablemente no sea muy injusto indicar dos grandes figuras en ese período: Khintchine y Wiener, que inventan el proceso de Wiener, o dicho de otra manera, el modelo matemático del Movimiento Browniano. Los años de 1930 son una época de gran florecimiento, especialmente en base a la figura de A.N. Kolmogorov, que es en la historia de la Matemática el probabilista más importante y, en el siglo XX uno de los matemáticos más influyentes en diversas ramas de la Matemática. Entre 1930 y la década de los 50, debemos mencionar también otros nombres, como P. Levy, K. Itô y J.L. Doob, que contribuyeron a fundar la Teoría de la Probabilidad como teoría matemática actual. La explosión de las aplicaciones de esa teoría y su inserción en la técnica y en la ciencia deben considerarse posteriores a la Segunda Guerra Mundial.

[Dado el contexto de estas líneas, es útil agregar aquí una digresión anecdótica con respecto a la figura de Kolmogorov, vinculada con grandes preguntas: ¿La ciencia es la obra de los científicos y nada más? ¿O es un producto social? ¿En qué medida?. Mi impresión es que no lo sabemos. La figura de Kolmogorov es paradigmática en ese sentido, pues reúne un conjunto de circunstancias sociales e históricas con una personalidad extraordinaria. Kolmogorov nace en 1903 en Kazajstán, hijo de un ingeniero de una empresa local. Va a estudiar a Moscú en el año 1920, después de la Guerra Civil. La joven república soviética, en ese momento se encontraba totalmente devastada por la suma de la Primera Guerra Mundial (1914-1918) y de la Guerra Civil (1918-1920). El joven Kolmogorov de 17 años no sabe si estudiar Historia o Matemática, y en el primer año en la universidad en Moscú, se inscribe en ambas. A lo largo de su extraordinaria carrera científica, también escribió libros de Historia, pero fue matemático porque la escuela histórica, lo que quedaba de la época del zarismo, estaba destruída. En ese entonces existían tres clases de historiadores: los historiadores zaristas que habían desaparecido de la universidad y no se sabía donde estaban, los historiadores como Miliukov que habían sido antizaristas y habían participado en la Revolución de Febrero (de 1917), pero que habían sido antibolcheviques y estaban exiliados, y los historiadores bolcheviques, entre los cuales había algunos muy significativos, pero que estaban en el gobierno y no en la universidad, porque era muy difícil la situación y el nuevo régimen carecía de cuadros para las funciones de gobierno. No había historiadores, entonces, pero sí había algunos matemáticos. Hubo un par de maestros que venían de la época zarista que mantuvieron la continuidad de la escuela científica, uno de los cuales fue Urysohn (el autor de los teoremas sobre metrización que estudian nuestros actuales estudiantes de Matemática). Urysohn fue profesor de Kolmogorov en el primer año y esto muy probablemente influyó en su decisión de convertirse en matemático. Después, Kolmogorov hizo una gigantesca obra de Matemática, parte de la cual consistió en sentar las bases y elaborar el desarrollo de varios capítulos centrales de la Teoría de la Probabilidad y sus aplicaciones. Tiene sentido, entonces, preguntarnos si esta teoría es un fruto del azar en la vida de este joven que hizo esas opciones, en virtud de las circunstancias históricas que le tocó vivir, o si la misma disciplina hubiera tenido una evolución diferente bajo otras condiciones].

¿Qué hace Shannon hacia 1950? ¿Por qué eso es importante? Shannon era ingeniero y su problema era el de la transmisión de la información. Intuitivamente, todos sabemos que hay situaciones en las que se genera mucha información y otras en las que hay poca. Es decir, nos dice poco el saber que ocurrió algo que ya sabíamos de antemano que habría de ocurrir. En cambio, cuando ocurre una cosa

rara, eso nos informa mucho. Lo raro es muy interesante, contiene mucha información nueva; lo corriente contiene poca información, "ya lo sabemos".

En realidad, los problemas de Shannon eran bastante más complicados porque no se trataba para él solamente de definir el contenido en información de los mensajes, sino que además tenía que diseñar canales por los cuales transmitirlos y hacer eso de manera eficiente. Aunque lo primero es medir la cantidad de información. Si esto es posible, entonces también será posible medir la capacidad de un canal para transmitirla, en el mismo sentido en que un canal hidráulico transmite una cierta cantidad de un fluido de un punto a otro por unidad de tiempo. Después entender como las perturbaciones ("el ruido") afectan el funcionamiento y cómo se debe proceder para resolver diversos problemas de optimización asociados.

¿Se puede medir la información? Hay básicamente dos planos.

Primero, un plano extraordinariamente difícil sobre el cual la ciencia todavía dice muy poco, y tiene que ver con el significado de la información transmitida. ¿Cuánta información transmite un poema? ¿Cuál es el sentido de la información en la comunicación de la palabra? ¿Cómo incorporar una medida común a mensajes que tienen que ver con esferas completamente diferentes de la cultura o de la técnica, que transmiten saberes o sentimientos? ¿Nos ayudan en esto las neurociencias? Aparentemente, todavía nos dicen muy poco como para que sea posible medir significados. Ni siquiera está muy claro para mí cual es el contenido de estas expresiones, que quiere decir realmente la expresión "medir significados".

Una segunda forma de aproximación a la cuestión es comenzar despojando a la información que queremos medir de su sentido. Es un empobrecimiento, sin lugar a dudas. La gran sorpresa es que, sin embargo, lo que queda es bastante. Ocurre que una gran parte de la comunicación que hacemos se puede traducir en lenguajes canónicos, se puede "codificar", por ejemplo, bajo la forma de bits (es decir, de secuencias de 0's y 1's). Se puede organizar la información despojada de su sentido y entonces transmitirla.

Intentemos aclarar un poco más en detalle a qué nos estamos refiriendo, exponiendo un teorema sobre la misma cuestión debido al matemático húngaro Alfred Rényi, que quizá contribuya a posteriori a explicar algunas de las ideas de Shannon.

Supongamos que tenemos un experimento que tiene un cierto conjunto de n resultados posibles: R_1, R_2, \dots, R_n . Uno puede pensar en un lanzamiento de un dado ($n=6$ y los resultados posibles son las 6 caras del dado) o el lanzamiento de simultáneo de 2 dados ($n=36$ y los resultados posibles son todas las parejas formadas por las caras de cada uno de los dos dados), o en los resultados de una bolada de la ruleta ($n=37$), o en muchos otros ejemplos más complicados e interesantes. Supongamos que el resultado R_1 tiene una probabilidad p_1 de ocurrir, el R_2 una probabilidad p_2 y así sucesivamente hasta llegar al resultado R_n que tiene una probabilidad p_n . Estos números son tales que $p_j \geq 0$ y $\sum p_j = 1$, lo que corresponde al modo habitual con que encaramos de manera heurística y sin pensar mucho, lo que llamamos "modelo aleatorio" de un experimento o de un acontecimiento de la realidad. Decimos aleatorio sin precisar mayormente, simplemente aludimos a que no sabemos el resultado que va a ocurrir y le asignamos una medida o probabilidad a cada resultado: al j -ésimo resultado posible le asignamos la medida p_j .

¿Qué dice el Teorema de Rényi? Dice que existe una y una sola función (que denotamos por H y llamamos "entropía"), que en un sentido natural mide la incertidumbre que hay sobre este experimento. Cuando decimos "en un sentido natural", nos referimos a que esa función verifica las propiedades que enunciamos a continuación:

1) H es sólo función de p_1, p_2, \dots, p_n . Es decir que la incertidumbre no depende de si lo que estoy observando son los n resultados posibles de un juego de azar con probabilidades p_1, p_2, \dots, p_n , o si lo que estoy observando son los n resultados posibles de un experimento biológico, o social, o de cualquier naturaleza, con tal que las probabilidades respectivas sean las mismas. Al decir que H sólo es función de las probabilidades, se está diciendo que H no es función del significado concreto del experimento, sino tan sólo de las probabilidades individuales de los resultados posibles del mismo.

2) Llamemos E a ese experimento, que consiste en elegir al azar uno de ellos entre los n resultados posibles, con esas probabilidades. Supongamos que tenemos 2 experimentos E y F y que en algún sentido natural son independientes. ¿Qué quiere decir que son independientes? Quiere decir lo que el sentido común dice sobre independencia. Quiere decir que si el experimento E consiste en observar los resultados de una ruleta en el Casino del Parque Hotel y F consiste en observar los resultados de jugar a la taba en un juego que están haciendo dos personas en Tacuarembó, el conocimiento de lo ocurrió en uno de ellos no modifica nuestras expectativas acerca del resultado del otro. Los comunes mortales aceptamos que no hay influencia entre los resultados de ruleta de ayer a las 10:30 de la noche y lo que pasó hoy de mañana cuando las 2 personas jugaron a la taba.

La hipótesis es que la incertidumbre que genera el experimento que consiste en observar *la pareja* E, F , cuando los experimentos E y F son independientes, es igual a la suma de las incertidumbres de ambos experimentos *individuales*, o si se quiere:

$$H(E, F) = H(E) + H(F)$$

"Observar *la pareja* E, F " quiere decir, observar el resultado *conjunto* de ambos experimentos. O sea, si R_1, R_2, \dots, R_n son los resultados posibles de E y S_1, S_2, \dots, S_m son los resultados posibles de F , entonces *las parejas* (R_i, S_j) ($i = 1, \dots, n; j = 1, \dots, m$) son los resultados posibles de *la pareja* E, F . Si además p_1, p_2, \dots, p_n y q_1, q_2, \dots, q_m son respectivamente las probabilidades de los resultados posibles de E y de F , que E y F son *independientes* implica que la probabilidad de que ocurra la pareja (R_i, S_j) es el producto $p_i \cdot q_j$.

3) Se requiere además una unidad de medida. Está claro que a cualquier medida de incertidumbre, si se la multiplica por un factor fijo, se obtiene otra que tiene las mismas propiedades. Si denotamos por D al experimento canónico, que consiste en observar dos resultados posibles, 0 ó 1, cada uno de ellos con probabilidad $\frac{1}{2}$ (que no es otro que el modelo de lanzamiento de una moneda equilibrada), exigiremos que la entropía $H(D)$ del experimento D sea igual a 1.

El teorema de Rényi dice que hay una única función que cumple estas 3 propiedades (además de algunas condiciones de regularidad que no hacen al fondo de la cuestión), y que esa función es la siguiente:

$$H = - \sum_{j=1}^n P_j \log_2 P_j$$

("log₂" es el logaritmo en la base 2)

Esta función se llama la "Entropía de Shannon" o "Incertidumbre de Shannon". De modo que no solamente se puede medir razonablemente la incertidumbre que hay en un experimento en el que interviene el azar, sino que eso se puede hacer sólo de una manera, una vez fijada la unidad mencionada.

[En realidad, para ser completamente precisos, el teorema de Rényi, tal como lo hemos enunciado, dice que existen también otras funciones H - además de la entropía de Shannon - que verifican las condiciones 1), 2) y 3). Para que ésta sea *la única* función, se requiere reemplazar la condición 2) por una condición ligeramente más exigente, cuya adecuada comprensión involucra la noción de probabilidad condicional¹].

Una idea sustancial en esto es que la información que se obtiene al observar el resultado de un experimento es exactamente igual a la incertidumbre que había antes de realizarlo. Incertidumbre e información son anverso y reverso de la misma moneda. La incertidumbre sobre el experimento y la información que da el observarlo son iguales.

Es fácil probar lo siguiente. Si un experimento tiene n resultados posibles, ¿cuándo es máxima la incertidumbre? Intuitivamente, es máxima cuando todas las probabilidades son iguales, es decir cuando no se sabe nada, o dicho de otro modo, no se conoce preferencia alguna para la ocurrencia de algunos de los resultados sobre otros. Al lanzar un dado, que tiene 6 caras como resultados posibles, la máxima incertidumbre tiene lugar cuando todas las caras tienen probabilidad $1/6$; si el dado está cargado, es porque se tiene más información y alguna de las 6 caras no tiene probabilidad $1/6$. Efectivamente, uno puede verificar sin dificultad que el máximo de la entropía de Shannon se obtiene cuando p_1, p_2, \dots, p_n son iguales, y por lo tanto, cuando cada una vale $1/n$.

Volvamos entonces a Shannon en la segunda mitad de los años 40, un poco más de una década antes del teorema de Rényi. El objeto del estudio de Shannon es la transmisión de información por medio de canales, tal como ocurre en las comunicaciones más diversas. Supongamos que se quiere transmitir palabras hechas con un alfabeto de cuatro símbolos: A,B,C,D. Una palabra puede ser, por ejemplo: ABBACDDBAB, es decir, una secuencia de esos cuatro símbolos. Podríamos pensar que el lenguaje con el que hemos escrito el mensaje es un cierto lenguaje natural, como el que utilizamos para hablar o escribir, que suele tener 26 letras y 10 cifras más algunos adicionales, como puntos, comas, blancos y algunos más. Aunque para el caso sólo existen 4 símbolos (no es tan extraño un lenguaje con 4 símbolos, así es el que permite representar el ADN).

¹ La condición que debe reemplazar a 2) para que valga exactamente el enunciado que dimos del teorema de Rényi, es la siguiente:

2') Si $0 \leq t \leq 1$, entonces $H(t p_1, (1-t)p_1, p_2, \dots, p_n) = H(p_1, p_2, \dots, p_n) + p_1 H(t, 1-t)$.

Aquellos lectores que conozcan la definición de probabilidad condicional, podrán interpretar sin dificultad la condición 2').

Lo que ésta dice es que si consideramos el experimento E con resultados posibles R_1, R_2, \dots, R_n y probabilidades respectivas p_1, p_2, \dots, p_n , y el experimento E' con resultados posibles $R'_1, R''_1, R_2, \dots, R_n$, donde los últimos $(n-1)$ resultados posibles son los mismos, y los dos primeros - R'_1, R''_1 - constituyen una partición de R_1 , con probabilidades respectivas $t p_1, (1-t) p_1$, entonces, la incertidumbre asociada al experimento R' es mayor que la asociada a R y, más precisamente, lo que hay que agregar es la probabilidad de que ocurra R_1 veces la incertidumbre de que, *condicionalmente a que ocurrió R_1* , ocurra algunas de las partes R'_1, R''_1 .

Es un ejercicio elemental, la verificación de que la condición 2') implica 2); el recíproco no es cierto.

Ahora bien, supongamos que queremos transmitir ese mensaje a un destinatario utilizando un canal de comunicación. Naturalmente, en la práctica lo que estamos pensando es transmitir muchos mensajes, no uno sólo, y en hacerlo de la manera más económica posible. Para transmitir los mensajes, una manera standard de hacerlo es, primero, codificar la información que contienen los mismos en secuencias de 0's y 1's. De modo que por cada letra, por cada palabra, en general, por cada texto en ese idioma, se va a transmitir una secuencia de 0's y 1's. Hay naturalmente muchas maneras de hacer esa codificación y no habremos de ocuparnos de eso aquí; es por cierto un tema vasto.

Una primera cosa que salta a la vista es que si queremos proceder de una manera que sea económicamente eficiente, debemos tener en cuenta la frecuencia con que aparecen los símbolos - o las cadenas de símbolos - en el lenguaje en cuestión o en los textos que queremos transmitir: Por ejemplo, si se trata de elegir un código para cada uno de los 4 símbolos, es obvio que es conveniente utilizar un código corto de 0's y 1's para los símbolos más frecuentes y que podemos permitirnos usar códigos más largos para representar los símbolos menos frecuentes.

Para fijar las ideas, asignemos algunos valores (arbitrarios). Supongamos que el conocimiento previo que tenemos del lenguaje del cual habremos de transmitir mensajes, nos dice que los símbolos aparecen con las siguientes frecuencias: A con frecuencia 1/2, B con frecuencia 1/4, C con 1/8 y D con 1/8. La A de este lenguaje es como la E del Idioma Español, que es el símbolo individual que aparece con mayor frecuencia.

En una primera instancia, lo más sencillo es pensar sólo en los símbolos individuales (letras) aunque los códigos reales deben ocuparse de optimizar la transmisión de palabras, frases, etc. Como observamos más arriba, es mejor transmitir la A con expresiones cortas que transmitir la D con tales expresiones. Eso implica que conviene reservar un símbolo solo para la A, a los efectos de ser lo más económico en aquello en lo que en promedio deberemos gastar más tiempo, que es la aparición de la A.

Puede usarse - por ejemplo - el código siguiente:

A → 0	C → 110
B → 10	D → 111

Se quiere transmitir de manera que el receptor del otro lado cuando mire los símbolos 0 y 1 sea capaz de decodificar lo que se le transmitió, y reconstruir la palabra que se le dijo. El receptor va a tener algunos problemas, porque en el canal de transmisión van a existir ruidos, y existen distintas fuentes de error, que deberán ser objeto de estudio y de tratamiento por otro lado. Si como consecuencia de ello, habiendo transmitido un 1, por ejemplo, el receptor recibe un 0, eso cambia completamente la secuencia y es obvio que es una fuente de problemas. La palabra escrita anteriormente (ABBACDDBAB) quedaría representada de la siguiente manera utilizando el código: 01010011011111110010. Se le transmite al interlocutor una secuencia de 0's y 1's que tiene que tener la propiedad de que con ella, él pueda reconstruir - decodificar - la palabra original, que está escrita en el lenguaje con alfabeto A,B,C,D. Bajo la hipótesis de que no se introdujo ruido que distorsionó el mensaje, esta codificación es una buena codificación, es decir que la palabra transmitida se reconstruye de una sola manera.

Supongamos que las palabras provienen de un mecanismo de emisión de símbolos que es aleatorio y que cada símbolo es elegido del alfabeto considerado, a_1, a_2, \dots, a_n , con probabilidades respectivas p_1, p_2, \dots, p_n . Denotamos cada palabra de longitud N - es decir, cada cadena de símbolos - por

$x = (x_1, x_2, \dots, x_N)$. Codifiquemos de manera óptima la palabra x mediante una cadena de 0's y 1's de longitud $v_N(x)$.

Lo que Shannon probó es que la relación entre la longitud ($v_N(x)$) de la palabra codificada y la longitud de la palabra original (N) tiende a la entropía del alfabeto del lenguaje correspondiente cuando N tiende a infinito.

Es decir

$$v_N(x)/N \rightarrow H \text{ cuando } N \rightarrow \infty$$

Esto quiere decir que la entropía H es el número de 0's y 1's (o número de bits) por símbolo que se necesita para codificar de manera óptima una palabra larga.

Shannon probó este resultado cuando los símbolos son independientes. Intuitivamente, que los símbolos son independientes quiere decir lo siguiente: Cada uno de los símbolos es extraído al azar. Supongamos que el primero es una letra A ; cuando sorteamos para ver cuál es el segundo símbolo, es como si no se supiera nada de lo que pasó con el primero, y digamos que obtenemos una B . Después sorteamos el tercero, y el conocimiento de que los primeros dos fueron A, B no cambia nuestra incertidumbre sobre el tercero; digamos que obtenemos nuevamente B . Y así seguimos, sorteamos el cuarto símbolo y la información de que los tres primeras fueron ABB no modifica la incertidumbre acerca del cuarto. Y así sucesivamente.²

El lector ya habrá pensado, sin duda, que en lugar de suponer que los sucesivos símbolos son independientes, podríamos suponer, por ejemplo, que son independientes las sucesivas parejas, ternas o cualesquiera otras cadenas de símbolos consecutivos. Ello modificará nuestros cálculos, aunque no la base conceptual con la que los hacemos. También habrá pensado el lector que este tipo de hipótesis está bastante alejada de las realidades habituales que quisiéramos modelar. Es evidente que en general los mensajes no son de ese tipo o, más bien, que hay muy pocos mensajes que son de ese tipo. En nuestro lenguaje natural, no es cierto que cada fonema sea independiente de sus fonemas vecinos; más bien es cierto lo contrario: están muy relacionados entre si. Por ejemplo, en el idioma español, si me encuentro con una letra "S", es seguro que inmediatamente después de ella no va a haber una letra "Z". Dicho de otro modo, *si sé que en el lugar 5 del mensaje hay una "S"*, la probabilidad de que en el lugar 6 haya una "Z" es igual a 0, mientras que *si no sé que ocurre en el lugar 5*, la probabilidad de que haya una "Z" en el lugar 6 no es igual a 0. Del mismo modo, *si sé que en el lugar 5 del mensaje hay una "S"*, las probabilidades de que en el lugar 6 haya una "A", una "I" o una "T" son distintas. No es cierto que los símbolos que ocupan esos dos lugares consecutivos sean independientes, porque la incertidumbre sobre lo que ocurre en el lugar 6 *no es la misma*, si sé lo que ocurre en el lugar 5 que si no sé nada.

Nuestra ejemplificación se refiere a dos lugares consecutivos; es claro que cosas parecidas habrán de ocurrir si consideramos ternas de símbolos consecutivos y, de modo general, otros conjuntos de símbolos. Con frecuencia, en los modelos que se emplean en la práctica, se admite que si dos conjuntos

² En términos precisos y generales, utilizando la notación ut-supra, se dice que la sucesión es de símbolos independientes, cuando la probabilidad condicional de que sea a_i el símbolo que aparece en el lugar k ($k \geq 2$), dado que se conocen los símbolos que aparecieron en los lugares $1, 2, \dots, (k-1)$, es la misma que la probabilidad p_i del símbolo a_i en el conjunto del lenguaje. Dicho de otro modo, hay independencia cuando nuestra incertidumbre acerca de cual será el símbolo que aparecerá en el lugar k , *es la misma si sabemos* cuales son los símbolos que han tenido lugar "en el pasado" (es decir, hasta el lugar $k-1$) *que si no sabemos nada* sobre el mensaje transmitido.

de símbolos están suficientemente separados, entonces son independientes o, por lo menos están cerca de serlo en un sentido natural que no estamos en condiciones de precisar aquí.

En consecuencia, necesitamos un resultado como el de Shannon, que vincule la entropía con el costo (medido por el número de bits) de codificación *por símbolo*, pero en situaciones más generales que aquéllas en que los sucesivos símbolos son independientes. El teorema que da respuesta a este problema lleva el nombre de Shannon - Mac Millan – Breiman y dice que una conclusión análoga a la del teorema de Shannon sigue valiendo en condiciones más generales, por ejemplo, si el proceso de producción de símbolos es estacionario (lo que significa que la ley del mismo no cambia con una traslación en el tiempo), pero se permite que haya interacción entre los símbolos que componen el mensaje a transmitir, y se cumple además alguna condición técnica adicional que omitimos aquí. De modo que el ingeniero Shannon tuvo la idea inicial y los matemáticos Mac Millan y Breiman la perfeccionaron y la colocaron en un contexto que la volvió aplicable a los mensajes reales.

Es decir que no sólo se puede cuantificar la información contenida en un experimento aleatorio, sino también el costo unitario de codificarla y transmitirla de manera óptima y que existe una estrecha relación entre ambas cosas. Lo que acabamos de explicar se refiere a condiciones bastante ideales; sin embargo, estos mismos principios y un cuerpo de teoría bastante extenso al que no habremos de referirnos aquí, permite acercarse a situaciones muy realistas y constituye una parte del fundamento de la tecnología de comunicaciones.

Es interesante que, desde otro ángulo que el de las comunicaciones, estas mismas ideas fueron tomadas algunos años después por Kolmogorov quien generaliza la noción de entropía y la convierte en un poderoso instrumento para atacar los problemas de la incertidumbre que venían de la Mecánica Racional y de los Sistemas Dinámicos a los que hicimos referencia anteriormente, estableciendo un puente entre ambas cosas aparentemente distantes. Para tratar de entender en forma aproximada lo que hizo Kolmogorov y su escuela, debemos volver previamente a 1900, cosa que haremos en la próxima sección.

En una versión definitiva de este texto, habremos de regresar sobre el tema de las medidas de información, haciendo referencia a otras medidas que tienen importancia en Estadística y a otras ideas que las sustentan. Asimismo, incluiremos una breve digresión sobre el desarrollo de la Teoría de Probabilidad y de la Estadística.

De Poincaré, hacia 1900, a Kolmogorov y la noción de caos.

En el año 1900 las ideas dominantes son las ideas de Laplace, perfeccionadas por un siglo de Matemática. H. Poincaré (1846?-1912) introduce modificaciones sustanciales en los cálculos que tienen que ver con la estabilidad del Sistema Planetario y con los problemas de la mecánica celeste. Recordemos que para Laplace la probabilidad se refería a errores de observación y a errores de cálculo. Poincaré no cuestiona la Ley de la Gravitación y no incorpora, salvo en uno de los últimos artículos de su vida, pocos meses antes de morir, ninguna nueva noción desde este punto de vista. Sin embargo, difiere de Laplace en dos cosas fundamentales relacionadas con el tema que estamos tratando aquí.

Primero, para Laplace, los errores son lo que ignoramos, sea lo que ignoramos desde el punto de vista experimental, sea lo que ignoramos desde el punto de vista del cálculo. Poincaré refuta este punto de vista con una argumentación que es muy ingeniosa y dice aproximadamente lo siguiente: imaginemos un físico que descubre una ley un día martes. ¿Es lógico decir que hasta el lunes el

fenómeno cuya ley descubrió era aleatorio y que en ese momento dejó de serlo? No podemos adoptar un punto de vista subjetivo. Si queremos descubrir el universo, el carácter aleatorio del fenómeno no puede depender de que ignoremos su naturaleza. Este es un argumento retórico, pero interesante. Dice además que el problema del cálculo de probabilidades no es el problema de nuestra ignorancia (por ejemplo, el problema de los errores de observación) e introduce de manera premonitrice la idea de caos que conduce a que realmente los tales errores no tienen solución.

Segundo, la idea nueva de Poincaré (posiblemente no deba considerarse enteramente nueva, porque los astrónomos acariciaban esa idea, pero no tenían una formulación precisa ni extraían las conclusiones correspondientes) es la dependencia sensible de las condiciones iniciales. Tratemos de entender los fundamentos de esta noción a través de un ejemplo sencillo; en lugar de la rotación de un planeta alrededor del Sol o del movimiento de los astros en general, vamos a considerar la situación en que estamos constreñidos a movernos en un círculo y que, en el instante inicial, partimos de uno de sus puntos.

La operación que vamos a aplicar sucesivamente es la siguiente: a cada punto del círculo le asignamos como imagen el punto que forma ángulo doble con el semieje positivo de las abscisas, medido en sentido antihorario. Es decir, al cabo de una unidad de tiempo, rotamos el punto que corresponde al ángulo x llevándolo al punto que corresponde al ángulo $2x$. Y luego, repetimos ("iteramos") el procedimiento: al cabo de una nueva unidad de tiempo, el punto que corresponde a $2x$ irá al que tiene ángulo $4x$, etc. Es decir, tenemos una función $F: S^1 \rightarrow S^1$ (S^1 es la notación habitual para el círculo), que en nuestro caso es $F(x) = 2x$ (x es el ángulo) y estamos mirando las "posiciones sucesivas" de un móvil que, al cabo de una unidad de tiempo ejecuta el movimiento indicado por F , que no es otro que duplicar el ángulo. Por ejemplo, si el punto inicial corresponde al ángulo de 30° , al cabo de una unidad de tiempo estaremos en la posición que corresponde a 60° , al cabo de dos unidades de tiempo en 120° , de tres unidades 240° y así sucesivamente. Si partimos del ángulo x_0 en el instante inicial, al cabo de una unidad de tiempo estaremos en $x_1 = F(x_0)$, al cabo de dos en $x_2 = F(x_1) = F(F(x_0)) = F^2(x_0)$ y, en general, al cabo de n unidades de tiempo, estaremos en $x_n = F^n(x_0)$. La sucesión de puntos $\{x_0, x_1, \dots, x_n, \dots\}$ es la "órbita" del punto inicial x_0 .

La primera observación de Poincaré es la siguiente: Imaginemos que medimos el punto inicial x_0 con un pequeño error δ (error en el ángulo). Entonces, al cabo de una unidad de tiempo, el error al calcular la imagen $x_1 = F(x_0)$ será 2δ , el error al calcular x_2 será 4δ y al cabo de n unidades de tiempo, el error cometido cuando se calcula el punto de la órbita a partir del valor medido, en lugar del verdadero valor x_0 , será $2^n \delta$. Obsérvese que 2^n crece muy rápidamente con n . Por ejemplo, si el error en la medida del ángulo inicial es $\delta = \pi/180.000$, que es un milésimo de grado angular, entonces la propagación de dicho error hace que el error cometido al calcular x_{16} sea mayor que $\pi/3$, es decir, 60° grados. Esto es lo que se llama un sistema con sensibilidad a las condiciones iniciales: si parto inicialmente de dos puntos que difieren poco, al cabo de tiempo n suficientemente grande, las dos órbitas pueden separarse mucho.

La otra cosa que ocurre con este tipo de sistemas, es que si el tiempo n es grande, la órbita puede encontrarse en cualquier parte y, más aún, el número de "visitas" que realiza a las diversas partes del espacio donde el movimiento tiene lugar, se comporta como si fuera aleatorio y se independiza del punto de partida. Más precisamente, si denotamos por k_n el número de veces que los puntos de la órbita hasta el instante n (son los puntos x_0, x_1, \dots, x_n) caen en un intervalo angular I , entonces k_n / n tiende a la longitud del intervalo I cuando n tiende a infinito. Esto no es exactamente cierto: observe el lector

que si, por ejemplo, el punto inicial es $\mathbf{x}_0 = 0$, la órbita es $\mathbf{x}_n = 0$ para todo tiempo n y, por lo tanto, no es cierto que la proporción de puntos de la órbita que caen en un intervalo tienda a la longitud del mismo. De hecho, si el intervalo no contiene a 0, entonces esa proporción es siempre 0 y si lo contiene es siempre 1. Lo mismo si el punto inicial es $\mathbf{x}_0 = \pi/2$, verifique el lector que ocurre algo similar y reflexione sobre otros ejemplos parecidos. Lo que sí es cierto, y es un teorema, es que salvo si una parte de un conjunto excepcional de valores iniciales que es pequeño en un cierto sentido (tiene "medida nula"), entonces se cumple que la proporción de "visitas" a I sobre el tiempo total de observación tiende a la longitud del intervalo cuando ese tiempo de observación tiende a infinito. Esto es lo que se suele llamar un "Teorema Ergódico" e intuitivamente nos dice que las órbitas se comportan como una sucesión de observaciones aleatorias, que tienen una cierta probabilidad (asintótica) de caer en cada conjunto I y que ella es independiente ("casi-independiente") de la posición inicial.

Lo que es especialmente interesante, es que esto que estamos explicando para una simple función del círculo en si mismo, que consiste en duplicar el ángulo, también ocurre para muchas leyes físicas, en particular para las que están vinculadas con el movimiento de los planetas. Claro que en este caso el tiempo es continuo, el movimiento tiene lugar en un espacio que no es un círculo y las leyes del movimiento son bastante más complicadas que nuestra duplicación del ángulo.

La visión de Poincaré es que la aleatoriedad en la dinámica proviene de la sensibilidad con respecto a las condiciones iniciales y, lo que es más importante, que ésa es la fuente de todo azar. Esto difiere de manera sustancial con la visión de Laplace que recordemos que se centraba en los errores de observación y de cálculo; por lo tanto, su programa consiste en disminuir éstos. En cambio, si adoptamos el punto de vista de Poincaré, como los errores de observación y de cálculo son inevitables, si un sistema está regido por una dinámica caótica, por más pequeños que sean los errores, necesariamente, al cabo de un cierto tiempo, vamos a ver el comportamiento de las órbitas como aleatorio y las fluctuaciones serán grandes. No hay la posibilidad de eliminar - o de disminuir en un sentido esencial - los "errores" a través de mejoras en los métodos de cálculo o de medición: en todo caso, la sensibilidad a las condiciones iniciales seguirá actuando con todas sus consecuencias. Por lo tanto, para su adecuada modelización y utilización práctica se requiere un modelo matemático diferente al sistema dinámico que indica la ley del fenómeno. Poincaré da diversos ejemplos de estas situaciones, que permiten interpretar que veía en estas dinámicas la fuente de toda aleatoriedad.

Lo característico es que entonces, la aleatoriedad aparece en sistemas que están regidos por leyes determinísticas: la duplicación del ángulo, en nuestro ejemplo elemental que guía el razonamiento, o las leyes de la Dinámica, en el caso general de la Mecánica Clásica. Si se elige el punto inicial \mathbf{x}_0 entonces la órbita $\{\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_n, \dots\}$ está totalmente determinada, ya que la función F es una función dada, la ley del fenómeno. *Sin embargo, si se deja pasar un tiempo suficientemente grande*, la órbita (la "trayectoria" en el lenguaje de la Teoría de Probabilidad), se comporta como si se tratara de las sucesivas observaciones de un proceso aleatorio.

Aquí tenemos un problema esencial y es ¿qué significa "*un tiempo suficientemente grande*"?. ¿Cuál es el horizonte de observación en el tiempo? La estimación del tiempo de vida del Sistema Planetario cambió mucho en la apreciación de los astrónomos a lo largo del siglo XIX. La unidad con la cual se contaba la vida del universo a principios del siglo XIX, en la época de Laplace, eran las decenas de miles de años. Sobre fines de este siglo aparecen los millones de años y, cuando aparece la radioactividad, la unidad de la Historia del Universo pasan a ser los miles de millones de años.

Entonces, es claro que no es lo mismo el tema de la estabilidad cuando los errores que se podían cometer en el cálculo o en la observación de las trayectorias de los objetos celestes se consideran en un horizonte de 90 mil años (que parecía la historia total del Universo a mediados del siglo XIX), que si se consideran sobre miles de millones de años. Al cabo de tiempos de esos nuevos órdenes de magnitud, la sensibilidad del sistema a las condiciones iniciales puede implicar que los errores vuelvan no predecible (aleatorio) el estado del sistema.

El nuevo aporte de Kolmogorov y de su escuela, a partir de 1958³, es que la noción de entropía, la noción de información de Shannon, es el útil para estudiar la incertidumbre en estos sistemas y que la entropía es el invariante fundamental para estudiar la evolución de largo plazo de estos sistemas que tienen sensibilidad a las condiciones iniciales. Entonces va a ser estudiada la órbita, que es la evolución del móvil en el tiempo, recurriendo a métodos probabilísticos, lo cual es complicado, porque hay que probar previamente que hay una estructura probabilística subyacente que es compatible con la dinámica. Observe el lector que, a priori, no hay ninguna probabilidad en el modelo: se trata de las leyes de la dinámica en la Mecánica Clásica, de las ecuaciones del movimiento, y si volvemos a lo que decíamos con respecto a Laplace, dadas las posiciones y velocidades en el instante inicial, queda determinada la evolución del sistema en el futuro. En el ejemplo anterior de la duplicación del ángulo, así como en los ejemplos que vienen de la dinámica, a priori el azar no está presente. Si se sale de un punto inicial x_0 , la trayectoria está totalmente determinada. Sólo que no puedo hacer predicciones a largo plazo sobre la trayectoria, porque estoy forzado a conocer x_0 con error, y por más pequeño que éste sea, al cabo de un tiempo N suficientemente grande, el error se propaga de tal modo que genera un error grande en x_N .

³ En una breve nota de 1958, que se puede considerar el punto de partida de un gran cuerpo de teoría, Kolmogorov considera el problema siguiente: ¿cuándo dos sistemas dinámicos probabilísticos son equivalentes? Un “sistema dinámico probabilístico” es un proceso aleatorio en que el tiempo recorre todos los números enteros, $\dots, X_{-n}, \dots, X_{-1}, X_0, X_1, \dots, X_n, \dots$ y que es estacionario. Esto significa que cualquiera que sean los tiempos n y m , las distribuciones de probabilidad conjuntas de $X_{-n}, \dots, X_{-1}, X_0, X_1, \dots, X_m$ y $X_{-n+1}, \dots, X_0, X_1, X_2, \dots, X_{m+1}$, coinciden. O sea que, una traslación en el tiempo no altera la ley del fenómeno que estamos observando.

Uno puede pensar que tiene entonces un nuevo espacio, cuyos elementos son todas las órbitas (que en esta caso son sucesiones “doblemente infinitas”, desde el punto inicial X_0 podemos ir hacia adelante y hacia atrás). La transformación que define la dinámica es la traslación en el tiempo, la cual deja invariante la probabilidad en el espacio de órbitas. Dos tales sistemas son equivalentes si hay una biyección entre los espacios de órbitas, que conserva la probabilidad y conmuta con las traslaciones.

Para tales sistemas, se puede definir la entropía y si dos sistemas son equivalentes, entonces la entropía es la misma. Además, si las variables aleatorias X_n son independientes, la entropía de Kolmogorov coincide con la entropía de Shannon.

Por ejemplo, si uno considera los dos sistemas siguientes:

1) el sistema “de la moneda”, en el que cada X_n puede tomar 2 valores – digamos cara y cruz – cada uno de ellos con probabilidad $1/2$ y las X_n son independientes y;

2) el sistema “del dado”, en cada X_n puede tomar 6 valores – digamos 1,2,3,4,5,6 – cada uno de ellos con probabilidad $1/6$ y las X_n son independientes,

entonces ambos sistemas *no son equivalentes* porque la entropía del primero es 1 y la del segundo es $\log_2 6$.

[Una impresión ingenua es que esto debería ser un hecho simple, pero no lo es, porque la equivalencia que permite pasar de un sistema a otro podría ser muy complicada. En efecto, en este caso, dicha equivalencia no existe, puesto que el *invariante* que es la entropía es distinto para uno y para otro]

Una cuestión importante aquí es si el recíproco es cierto, es decir, si dos sistemas tienen la misma entropía, bajo que condiciones podemos asegurar que son equivalentes, lo cual significaría que el valor de la entropía “clasifica” a los sistemas dinámicos probabilísticos. No haremos referencia a esto aquí.

Debemos distinguir entre determinismo y determinación. Son dos nociones distintas. El sistema es determinista en el sentido de Laplace, hay una ley física que lo regula de manera única en el futuro, cuando conozco el punto de partida. *Sin embargo*, al cabo de cierto tiempo no es posible determinar la posición del móvil y el modelo adecuado para describirla es el modelo probabilístico.

Para la conexión entre los Sistemas Dinámicos y la Teoría de Probabilidad, a los efectos de estudiar el comportamiento asintótico y el problema clásico (que viene desde Poincaré) del estudio cualitativo y de la clasificación de los sistemas dinámicos, es fundamental el aporte de la escuela soviética en su época de esplendor, en primer lugar de Kolmogorov y luego de sus colaboradores, Sinai, Pesin, Anosov, etc.

Habitualmente se denomina “caos determinístico” a esta propiedad de los sistemas dinámicos que tienen sensibilidad con respecto a las condiciones iniciales, de mostrar un comportamiento de tipo aleatorio al cabo de un tiempo largo. Algunos ejemplos clásicos son el modelo de Ehrenfest (al que habremos de referirnos más abajo) y el ejemplo de Lorenz en Mecánica de Fluidos, asociadas al fenómeno de la turbulencia. El meteorólogo Lorenz, escribió un modelo simplificado de las ecuaciones de la dinámica de los fluidos en dimensión 3, y lo puso a evolucionar en un computador (un tipo de “experimentación” que el desarrollo de la informática y su enorme potencialidad han vuelto frecuente) y descubrió que las trayectorias eran caóticas, sin que entonces (1963) hubiera una explicación clara de porqué ello ocurría. El estudio de las ecuaciones de Lorenz continúa siendo un tema atractivo en la actualidad, pero no estamos en condiciones de describirlo en este texto. En cambio, vamos a mencionar el ejemplo del Movimiento Browniano.

El modelo de Ehrenfest de la Mecánica Estadística.

El modelo de Ehrenfest es un modelo clásico de la Mecánica. El modelo consiste en lo siguiente:

Tenemos un recipiente con un tabique en el medio, que lo separa en dos espacios con volúmenes iguales. En uno de esos dos espacios, que llamaremos A, están las moléculas de un gas, helio por ejemplo. Sea B el espacio complementario de A en el recipiente original.

A presión y temperatura ambiente, el número de moléculas que hay en un litro es aproximadamente $2,7 \times 10^{22}$ (esencialmente, del mismo orden que el número de Avogadro). Las moléculas están en movimiento y admitimos que no hay inconvenientes para aplicar la teoría cinética de los gases, es decir que, en función de la presión y de la temperatura ambiente, las moléculas están dotadas de un movimiento rectilíneo y uniforme, salvo que además pueden chocar contra las paredes o entre ellas y entre dos choques de éstos, recuperan su movimiento rectilíneo y uniforme. Para que la simplificación sea máxima, suponemos además que las partículas tienen diámetro nulo.

Supongamos que retiramos el tabique que separa A y B y queremos observar como evoluciona el intercambio de moléculas entre ambos espacios con el paso del tiempo. El modelo de Ehrenfest intenta dar una descripción de lo que ocurre. Por ejemplo, se mira el tiempo cada segundo, cada microsegundo, cada 10^{-10} segundos, y se supone que en cada intervalo correspondiente a la unidad de tiempo elegida, hay una molécula que pasa de un lado para el otro. En el sistema hay en total N moléculas. En el instante inicial - cuando quitamos el tabique - están las N en el espacio A y 0 en el B. Al cabo de un cierto tiempo, hay K moléculas en el espacio donde antes no había ninguna (B) y $N-K$ en el otro (A), donde K es un número entero entre 0 y N .

Para tener un modelo, tenemos que indicar la ley mediante la cual, en un intervalo de una unidad de tiempo, se produce el intercambio de moléculas entre A y B. Supondremos que en cada instante que hay una y una sola molécula que pasa de un lado a otro y la regla mediante la cual ello ocurre es que es más probable que pase una molécula desde la parte donde hay más hacia donde hay menos, que a la inversa. Más precisamente, la probabilidad de que una molécula pase de A a B es proporcional al número de moléculas que hay en A, y la de que ocurra el traspaso inverso, es el complemento a la unidad. En un instante dado si hay K moléculas en B y $N-K$ en A, entonces la probabilidad de que la molécula pase de B a A es K/N y de que vaya una en sentido contrario es $1-K/N$.

En consecuencia, en el primer intervalo de tiempo, como todas las moléculas están en el espacio A, no puede pasar ninguna de B a A, y es igual a 1 la probabilidad de que una molécula pase de A a B. En el segundo intervalo de tiempo, la primera molécula podría volver al espacio A, con probabilidad $1/N$ o haber una segunda molécula que pase de A a B, con probabilidad $1-1/N$, Y así sucesivamente. Como se ve intuitivamente, no se debe esperar que la evolución del sistema tienda a estabilizarse y lo que describen los resultados sobre el modelo de Ehrenfest es un equilibrio dinámico.

Se prueba que, si en el instante t el número de moléculas que hay en el espacio A es A_t , y si K es un número entre 0 y N , entonces $P(A_t = K)$ - la probabilidad de que A_t sea igual a K - tiende al valor

$$[N!/(N-K)!K!]1/2^N$$

cuando $t \rightarrow \infty \dots$

[Aquí, el factorial del número natural N es $N! = N(N-1)(N-2)\dots 3.2.1$]

Cuando $t \rightarrow \infty$, se establece un equilibrio dinámico entre los 2 espacios A y B y cada uno le pasa moléculas al otro con cierta distribución de probabilidad, que se llama Binomial. El comportamiento asintótico es como si, en lugar de este fenómeno de transferencia de moléculas de un espacio al otro, lo que sucediera fuera lo siguiente: Se considera cada una de las N moléculas e independientemente y con probabilidad $1/2$, se coloca cada una de ellas en uno u otro de los dos espacios del recipiente. Es decir, es como si se eligiera al azar con probabilidad $1/2$ (porque los volúmenes son iguales en ambos espacios), el espacio donde va estar esa molécula. Si los volúmenes de los espacios del recipiente no fueran iguales, entonces la probabilidad de colocar las moléculas sería proporcional al volumen. Este tipo de resultado, que permite entender el equilibrio dinámico al cabo de un tiempo largo, corresponde a un capítulo de la Teoría de Probabilidad, las cadenas de Markov, porque la evolución del modelo de Ehrenfest es lo que se llama una cadena de Markov ergódica.

Mencionemos dos observaciones sobre este modelo.

La primera es que resulta que este “como si” (tomar las moléculas independientemente y colocarlas con probabilidad $1/2$ cada una en A o en B) puede extenderse de la manera siguiente. Si se observa el recipiente al que se le quitó el tabique (el tabique imaginario sirve simplemente para que, al principio, todas las moléculas estén en uno de los espacios, el A), y se se considera una parte cualquiera del recipiente, cuando pasa un tiempo suficientemente grande, ocurre algo enteramente similar, sólo que en lugar de $1/2$ debemos tomar el volumen relativo de esa parte del recipiente con respecto al volumen total. Ésta es una manera de enunciar la llamada hipótesis ergódica de Boltzmann. Ello es independiente del estado inicial: si en lugar de empezar con todas las moléculas en A, se hubieran puesto de otra manera, el comportamiento asintótico sería el mismo.

La segunda observación tiene que ver con la reversibilidad. Las ecuaciones microscópicas de la Física son reversibles en el tiempo; sin embargo, la realidad nos presenta muchos fenómenos que son irreversibles desde el punto de vista de la observación. Esto está ligado a nuestra experiencia corriente de que cuando hablamos de incertidumbre, nos estamos refiriendo al futuro. El tiempo de la incertidumbre y del azar están flechados y conscientemente o no, estamos pensando en un mundo irreversible con relación al tiempo, a pesar de que las ecuaciones que conocemos y lo regulan, son reversibles. Por cierto, es éste un arduo tema clásico que escapa al alcance de este texto. Sin embargo, para tirar de la punta de este complejo ovillo, en el caso del modelo de Ehrenfest tendemos a pensar que, una vez que el sistema está en régimen al cabo de un cierto tiempo (recordemos, un equilibrio dinámico con fluctuaciones), habiendo partido con todas las moléculas en A, no habrá de ocurrir que en algún instante futuro todas las moléculas vuelvan a estar en A y, por lo tanto, ninguna en B. Es decir, tenemos más bien la impresión de que si bien el sistema fluctúa, lo hace en configuraciones que no están lejos de la configuración media, y aproximadamente la mitad de las moléculas estarán en A y la otra mitad en B.

Sin embargo, el modelo permite que eso sea posible, ya que la probabilidad de que A_t (con t grande, un instante lejano) sea igual a N no es cero, sino que cuando t tiende a infinito tiende al número positivo $(1/2)^N$. Éste no es cero, pero es astronómicamente pequeño, porque el exponente N es del orden del número de Avogadro. Otra manera de ver el mismo problema, es preguntarse cuanto tiempo (es claro que es un tiempo aleatorio) debe esperar el observador para que, partiendo con todas las moléculas en A, las fluctuaciones reproduzcan esa configuración. Tiene lugar entonces el siguiente fenómeno: hay una cierta probabilidad positiva de volver en poco tiempo; pero una vez que el sistema se aleja en una cierta medida de esa configuración, entonces, con una probabilidad muy próxima a 1, demora un tiempo astronómico en regresar. En realidad, los resultados son más fuertes: no sólo es muy pequeña la probabilidad $P(A_t=N)$ sino que también es pequeña la probabilidad $P(|A_t - N/2| > \delta N)$ donde δ es un número positivo. Es decir, es muy poco probable que al cabo de un tiempo lejano t , el número de moléculas que hay en el recipiente A_t difiera de la mitad del número total de moléculas N en una pequeña fracción δ del mismo. Una manera posible de estimar estas probabilidades finamente – en el caso presente y en otros esencialmente más complicados técnicamente - es utilizar un capítulo de la Teoría de Probabilidad que es la Teoría de las Grandes Desviaciones, iniciada por H. Cramér en 1942 y que ha alcanzado un gran desarrollo en virtud de su vasto dominio de aplicación.

Lo que va a ocurrir para un observador que pueda medir alguna magnitud indirecta (no puede contar el número de partículas, pero sí ciertas magnitudes físicas que dependen del número de partículas, como la presión, a temperatura y volumen constantes) es que va a deducir, pasado un tiempo largo, que aproximadamente la mitad de las moléculas están en A y la otra mitad en B y que la probabilidad de estar algo distante de ese equilibrio, va a ser astronómicamente pequeña. No sólo no se vacía el espacio B, sino que al cabo de un cierto tiempo, A y B van a contener cada uno aproximadamente la mitad, fluctuando mucho en términos absolutos y fluctuando cantidades astronómicamente pequeñas en términos relativos. Es decir, todo ocurrirá para el observador como si el fenómeno fuera irreversible, no se vuelve al estado inicial. Esta es una interpretación posible de la irreversibilidad observable en este modelo.

El movimiento browniano.

El segundo ejemplo que quiero mencionar es el Movimiento Browniano. Brown fue un botánico inglés, que en la primera mitad del siglo XIX observaba partículas en suspensión en los líquidos, como lo habían hecho otros antes que él. Es la época del desarrollo de la Microscopía; los microscopios y los aparatos de óptica le daban a las personas que trabajaban en los laboratorios una gran cantidad de información de todo tipo. Brown observó que las partículas tenían un movimiento que no podía explicar. Ese movimiento había sido visto anteriormente muchas veces y la interpretación aceptada de la existencia de ese movimiento, era que esas partículas eran partículas con vida. Lo que hizo Brown fue, primero, poner partículas inertes del mismo tamaño, usando sus técnicas de microscopista y de laboratorista y observó que también se movían. Entonces, hizo una gran cantidad de experimentos, cambiando las condiciones (la temperatura, la composición química, etc.) y las partículas siempre se movían, importando poco las condiciones en las que realizaba el experimento. Estas observaciones de Brown, acerca de las agitaciones inexplicables de las partículas en el seno de los líquidos, conservaron este carácter inexplicable durante todo el siglo XIX. La contribución de Brown fue señalar que el problema no era ni biológico ni químico, sino físico, pero no fue capaz de explicar el fenómeno. El modelo físico se debe a Einstein y Smoluchovsky en 1905, que simultáneamente hicieron un intento de explicación. En cierto modo, se puede decir que es el modelo básico de azar que hoy utilizamos.

¿Por qué este modelo adquirió tanta importancia? Porque es el modelo canónico del ruido. Desde principios del siglo XX, físicos y químicos estaban interesados en entender que pasaba dentro de los fluidos. Para estudiar el movimiento de una partícula en suspensión, tenemos una dinámica descrita por una ecuación diferencial de la forma siguiente:

$$\frac{dx}{dt} = G(x) + F(x, t)$$

donde $x = x(t)$ representa la pareja posición-velocidad como función del tiempo, dx/dt es su derivada con respecto al tiempo, $dx/dt = G(x)$ es la ecuación de escurrimiento del fluido en ausencia de otras fuerzas y F es una fuerza externa. G es una "buena" función diferenciable, con derivadas regulares.

Un caso muy importante es el caso en que la fuerza externa F es aleatoria. En ese caso, las partículas suspendidas en el fluido se mueven aunque el fluido esté en reposo. El propósito de la ecuación ut supra es superponer el efecto general de arrastre del fluido sobre la partícula, con ese ruido y estudiar el movimiento resultante.

¿Cuál es el problema? Con respecto al cálculo diferencial tradicional - llamémosle newtoniano - esto no tiene ninguna novedad si la función F también fuera una función regular, tuviera derivadas o, por lo menos, fuera una función de variación acotada⁴. En ese caso, hay métodos clásicos para resolver estas ecuaciones diferenciales, entender las propiedades de las soluciones e incluso resolverlas en forma aproximada. Esto es lo que hacían Laplace, Poincaré y muchos otros antes y después de ellos. En el caso que nos interesa, el modelo de la fuerza externa que Einstein construye es el movimiento browniano y lo que se prueba algunos años después (Khinchin, Wiener, Lévy) es que, como función del tiempo, esa fuerza externa no tiene esas propiedades de regularidad y que en consecuencia, no se puede

⁴ Una función $F: [a, b] \rightarrow \mathbb{R}$ es "de variación acotada", si se puede escribir como la diferencia de dos funciones monótonas crecientes, es decir, si existen dos funciones A y B monótonas crecientes, con el mismo dominio $[a, b]$, tales que $F(t) = A(t) - B(t)$. En nuestro caso, los valores de F son vectoriales y decir que es de variación acotada, es decir que cada una de sus coordenadas lo es.

estudiar esta ecuación diferencial por los métodos clásicos. El problema de estas ecuaciones es que tienen una fuerza externa que oscila mucho, es muy irregular.

El gran salto para resolver este problema se asocia, alrededor de 1940, a los nombres de Paul Lévy poco después de Kyoshi Itô, y también de A.N. Kolmogorov, que inventan un nuevo cálculo, que modifica el cálculo diferencial newtoniano y que se llama hoy "cálculo estocástico" (a veces se le llama también "cálculo diferencial de Itô") que permite resolver esas ecuaciones y estudiar la mecánica aleatoria. Es decir, tenemos un sistema dinámico con una fuerza externa que es ruido que se superpone a otras fuerzas externas tradicionales que también pueden estar presentes. El ejemplo canónico es el ruido que genera la agitación de las partículas en los fluidos, que observaba Brown y del cual no tenía la explicación.

Esto a su vez genera un vasto desarrollo, sobre todo después de la Segunda Guerra Mundial y es el eje de la vinculación entre la Teoría de Probabilidad, los Sistemas Dinámicos y otras ramas tradicionales de la Matemática, como la Teoría de las Ecuaciones Parabólicas.

Un caso particular de estas ecuaciones es la difusión del calor, cuyas extensiones son centrales en la comprensión de los fenómenos físicos. El origen de la teoría que conecta las ecuaciones parabólicas - "de evolución" - y las soluciones de las ecuaciones diferenciales estocásticas - "procesos markovianos", en los casos más importantes - se debe esencialmente a Kolmogorov. Existen antecedentes ya conocidos de los físicos, las llamadas ecuaciones de Fokker–Planck, que en probabilidad se suelen llamar de Chapman–Kolmogorov. La teoría física de la difusión aparece conectada con esto y, más importante aún, con la ecuación de Schrödinger, que es la ecuación que verifican las funciones de onda en la mecánica cuántica, es una ecuación de difusión pero con coeficientes complejos, lo que establece una dificultad técnica esencial. Es el estudio de la mecánica cuántica, usando los métodos que provienen de esta teoría, que es lo que se llama la integración funcional, uno de los grandes aportes de R.Feynman.

En una versión final de este texto volveremos sobre estos temas, destacando su relación con la concepción del azar y comparando con las visiones preexistentes, especialmente la que proviene de Poincaré y del caos determinístico.

Estos temas, que aparecen originalmente en el estudio de un sistema dinámico cuando uno lo perturba con ruido, se convierten en una cuestión focal central de la Física y la Matemática a lo largo de todo el siglo XX.

Observaciones finales.

Simultáneamente con la construcción de Einstein y de Smoluchovsky del movimiento browniano, hay una tercera construcción que se debe a Bachelier. Bachelier fue un matemático dedicado a los temas económicos, interesado hacia 1900 en las fluctuaciones de la bolsa. Hizo su tesis en Matemática y uno de los miembros del jurado, fue el propio Poincaré.

Bachelier construye el mismo proceso de oscilaciones rápidas con el propósito de modelizar la evolución de las cotizaciones de la bolsa. Sus trabajos no fueron muy apreciados a lo largo del siglo XX, hasta aproximadamente 1970, cuando se extiende y se generaliza el fenómeno de la mundialización de las bolsas y de las monedas. Aparece la fiebre de la especulación financiera y los bancos convierten el problema de la modelización de las fluctuaciones del mercado financiero, en el

muy corto plazo, en un problema central. Los estadísticos, los probabilistas y los matemáticos habían hecho antes una experiencia de trabajo para los bancos - o los particulares - para modelizar las fluctuaciones del mercado financiero, contribuyendo a los criterios para la selección las carteras de inversión, con el propósito de optimizar sus ganancias. Se trataba de la resolución de problemas de optimización en condiciones de incertidumbre, optimización sobre un futuro cuya evolución está descrita por una ecuación donde hay ruido.

Lo que se vuelve urgente es que la modelización sea de muy corto plazo, porque la introducción de la informática y de las telecomunicaciones, ha cambiado el orden de magnitud de la velocidad de las fluctuaciones. Los modelos anteriores de las llamadas "series de tiempo", que distinguían entre las observaciones según los intervalos de tiempo entre ellas, pero que eran básicamente modelos con tiempo discreto, que habían sido muy exitosos e interesantes en una gran variedad de aplicaciones, se vuelven inoperantes en los mercados financieros porque las fluctuaciones son demasiado rápidas. Se vuelve a los trabajos de Bachelier, quien modelizaba las fluctuaciones de la bolsa en tiempo continuo. Ese es el trabajo de Black y Scholes (1973) que recibieron el premio Nobel en los años de 1990 por un trabajo de optimización que posiblemente los matemáticos no aprecian mucho, pero parece que los economistas y los banqueros sí. Encontraron una fórmula, que se deriva del cálculo estocástico de Itô, para optimizar una cartera de valores, cuando se modela la tendencia del mercado mediante la solución de una cierta ecuación diferencial, cuya fuerza externa es un ruido en tiempo continuo. Los resultados de ese cálculo se convierte en un insumo más que utiliza el manejador del portafolio de un inversor para tomar sus decisiones.

Retomemos aquí el tema de las leyes de grandes números. En una de sus versiones clásicas, debida a Kolmogorov, lo que dice es que si tenemos una sucesión de variables aleatorias independientes X_1, X_2, \dots, X_n ... todas con la misma distribución y valor medio igual al número m entonces la sucesión de los promedios

$$S_n = (X_1 + X_2 + \dots + X_n) / n$$

tiende a m cuando el número n de observaciones tiende a infinito. Más precisamente, la convergencia tiene lugar para "casi toda posición del azar" o "salvo un evento de probabilidad nula". Es decir, los promedios en el tiempo - pensando que las observaciones tienen lugar en tiempos sucesivos - tienden al valor medio teórico m . Este tipo de teoremas es lo que se denomina una "ley fuerte de grandes números". Quien aparentemente inventó las Ley de los grandes números fue Bernoulli, a principios del siglo XVIII, aunque en una situación muy simple, que es aquélla en que cada X_i vale 1 ó 0 según que ocurra o no un evento determinado A . En ese caso, S_n es la frecuencia relativa de ocurrencias de A hasta la n -ésima observación y m es la probabilidad $P(A)$. Por otra parte, el teorema de Bernoulli es un teorema "débil", la noción de convergencia que interviene es diferente.

En situaciones generales Kolmogorov y Doob extendieron este tipo de resultado, para entender cómo era el comportamiento medio de estos sistemas fluctuantes. Un punto importante, esencial en todas las aplicaciones, es una vez que uno sabe que el promedio de las observaciones a medida que pasa el tiempo tiende al valor medio teórico, poder estimar la velocidad de esa convergencia en función de n . No nos alcanza con saber a qué tiende, necesitamos saber también la velocidad. Los teoremas de velocidad más corrientes son los llamados Teoremas de convergencia en Ley o teoremas de desviaciones (grandes, medianas, etc.). De ese modo, tenemos una idea de cuán lejos estamos del promedio teórico, o dicho de otro modo, de cuál es la probabilidad de que el apartamiento del promedio teórico sea mayor que un número dado.

Pero, la vida ¿va por el promedio? Quién sabe. Aquí aparecen los problemas que constituyen el estudio de la dinámica fuera del equilibrio, que son muy corrientes y clásicos en Física y en Biología.

Esto puede ilustrarse con el ejemplo de una epidemia. ¿Cómo se propaga una epidemia? La forma más simple de propagación es por contagio. ¿Cuál es el modelo más simple de contagio que se puede imaginar? El modelo más simple consiste en una red de individuos inmóviles. Supongamos que en el instante inicial, tenemos un individuo enfermo. Éste, con una cierta probabilidad p , en el instante 1 contagia a cada uno de sus vecinos y con probabilidad $(1-p)$ no lo contagia, con comportamiento independiente los mismos. En el modelo más simple, en el instante 2, el contagio se propaga aún más, porque cada uno de los que se enfermó en el instante 1 puede a su vez contagiar a uno de sus vecinos con probabilidad p y los distintos individuos contagian independientemente. Este modelo de contagio puede continuarse ad infinitum.

¿Hay epidemia? La idea es que si p es un número muy pequeño, el primer individuo enfermo va a contagiar a algunos de sus vecinos, pero la enfermedad no se propaga mucho, se forma una isleta en torno a él y en algún instante futuro no hay contagio; no hay una epidemia. Este es el ejemplo más simple, pero sin embargo, no es nada sencillo desde el punto de vista matemático. La pregunta de si existe epidemia puede formularse matemáticamente de la siguiente manera: ¿Hay mucha gente que se enferma? Dicho de otra manera: ¿Existe una línea que conecta al enfermo inicial con el infinito sin interrupción?, ¿Al cabo de un tiempo muy grande, tiende a infinito el número de enfermos? Lo que se puede probar es que existe un valor p ($0 < p < 1$) que se llama p crítico: cuando $p < p_{\text{crítico}}$ no hay epidemia, y si $p > p_{\text{crítico}}$ entonces hay epidemia con una probabilidad positiva. Este es el modelo más simple de epidemia que se puede imaginar, en el que los individuos se colocan en lugares fijos y no se mueven, imposibilitando la interacción entre ellos. La existencia del $p_{\text{crítico}}$ es un teorema de los años 60 debido a H. Kesten. Existe una amplia literatura sobre este tema y aspectos relacionados. Por ejemplo, ¿qué ocurre si justo $p = p_{\text{crítico}}$? La probabilidad $\phi(p)$ de que haya epidemia es una función del número p , que se anula si $p \leq p_0$ y es positiva si $p > p_0$. Cuanto mayor es p , mayor es la función $\phi(p)$. ¿Cómo es la función $\phi(p)$? Se sabe poco del azar epidémico.

Esta situación se parece mucho a lo que pasa en otras áreas. Podemos encontrar fenómenos muy similares al contagio también en la Física Estadística. En un modelo representado por partículas, estas también se contagian, interactúan entre ellas, tienden a colocarse en promedio de cierta manera. Pero hay también fenómenos de dinámica fuera del equilibrio, que hace que surjan como lenguas de fuego, fenómenos que no son patológicos y no se sabe si la evolución del mundo natural no es así. Nadie ha dicho que la evolución real en este mundo azaroso en el que estamos, no es fruto justamente de una dinámica fuera del equilibrio.

Existen otros ejemplos de dinámica fuera del equilibrio que son muy importantes y tienen que ver con lo que se llama valores extremos. ¿Qué es lo importante, el comportamiento medio o los valores extremos? Hay algunos casos de la técnica, donde claramente son los valores extremos. Por ejemplo, si se quiere construir un dique, lo que más importa no es el comportamiento medio del río, sino la crecida máxima. ¿Cómo se hace para calcular la crecida máxima? Es complicado, porque la información estadística de que se dispone es muy corta y se pueden hacer predicciones sobre períodos muy cortos de tiempo con respecto al horizonte que se desea prever. Lo mismo sucede a la hora de construir estructuras metálicas sometidas a los esfuerzos generados por ciertos fenómenos naturales, vientos, olas, terremotos. Por ejemplo, las estructuras que están en el mar, o la estructura de un avión, o de un satélite. ¿Cuál es el horizonte de ruptura que se puede determinar de esa estructura? Ella va a ser diseñada para la ruptura. De modo que es un problema de valores extremos y de estudio de dinámica fuera del equilibrio.

ALGUNAS REFERENCIAS BIBLIOGRAFICAS.

- Cramér, H.; Leadbetter, M.R. "Stationary and Related Stochastic Processes", John Wiley & Sons, 1967.
- Doob, J.L. "Stochastic Processes", John Wiley, 1953.
- Feller, W. "An Introduction to Probability Theory and its Applications", Vol 1, 3d ed.,1968, Vol II, 2d ed, 1971, John Wiley & Sons.
- Kendall, M.; Pearson, E.S. "Studies in the history of statistics and probability", Vol I, Griffin, 1978.
- Kendall, M.; Plackett, R.L. "Studies in the history of statistics and probability", Vol II, Griffin, 1980.
- Kullback, S. "Information theory and statistics", J. Wiley, 1959.
- Laplace, P.S. "Théorie analytique des probabilités", Oeuvres complètes, t7, Gauthier-Villars, 1886.
- Laplace, P.S. "Essai philosophique sur les probabilités", I-II, Gauthier-Villars, 1921.
- Lévy, P. "Processus Stochastiques et Mouvement Brownien", Gauthier-Villars, 1948.
- Lévy, P. "Théorie de l'addition des variables aléatoires", Gauthier-Villars, 1954.
- Pearson, K. "The History of Statistics in the 17th and 18th Centuries", Griffin, 1978.
- Poncaré, H. "Calcul des probabilités", Carré-Naud, 1912.
- Rényi, A. "On measures of entropy and information", Proc 4th Berkeley Symp, Vol I, 547-561, Univ. Calif. Press, 1961.
- Yaglom, A.M.; Yaglom, I.M. "Probabilité et Information", Dunod, 1969.

31 de agosto de 1999.