Miscelánea Matemática **52** (2011) 1–29 SMM

Andrey Kolmogorov: El último gran matemático universal

Evgueni Gordienko
gord@xanum.uam.mx
Universidad Autónoma Metropolitana
Unidad Iztapalapa

Durante estas últimas décadas el interés hacia la ciencia en general y las matemáticas en particular se está reduciendo. Disminuyen el nivel y la calidad de la educación matemática, en primer lugar en los países desarrollados. Esta situación se debe sobre todo al uso excesivo de computadoras, calculadoras y, especialmente, del Internet que convierten el aprendizaje creativo en un proceso utilitario de búsqueda de "información necesaria" presentada frecuentemente en "forma masticada". Se está gestando una actitud puramente consumidora e indiferente hacia las teorías científicas que frecuentemente contienen ideas que emocionan la imaginación (cuando uno la tiene). Además se incrementa el escepticismo y la actitud negativa hacia la ciencia considerándola como una actividad destructora de la naturaleza, que está en contradicción con una vida armónica "natural". Sin duda, el medio ambiente patológico en las grandes megalópolis combina mal con una vida digna, pero ¿qué tiene que ver con esto el desarrollo de la tecnología y la ciencia? Sin sombra no hay luz y la aplicación de los logros científicos y técnológicos a veces dan frutos destructivos. No obstante, al abordar un avión estamos "casi seguros" de que aterrizaremos felizmente en nuestro punto de destino. Poco se piensa en la cantidad enorme de resultados e ideas científicos y en particular matemáticos que han sido necesarios para crear los sistemas que garantizan la seguridad del vuelo y aterrizaje. Para construir tales sistemas es necesario aplicar resultados obtenidos en diferentes áreas de la teoría de probabilidad y procesos estocásticos al desarrollo de las cuales hizo una aportación fundamental el matemático ruso y soviético Andrey Kolmogorov (1903 - 1987). A. N. Kolmogorov fue uno de los creadores más sobresalientes de las matemáticas contemporáneas las cuales, si dejamos de lado el pragmatismo arriba mencionado, son una gran conquista de la cultura del siglo

XX, y poseen su propia belleza y poesía proveyendo fuentes inagotables para el desarrollo de las capacidades creativas de la juventud.

1. El universalismo matemático

Tal vez no es casualidad que la vida adulta de A. N. Kolmogorov prácticamente transcurrió durante la época de la existencia de Unión Soviética. En este lapso históricamente muy corto, fue creada una de las escuelas matemáticas más fuertes y la mejor enseñanza matemática en el mundo (según las evaluaciones de muchos expertos). Kolmogorov fue el representante más sobresaliente y creador de esta escuela. Contribuyó además en el desarrollo de la educación matemática del país. Por un lado, su creatividad se manifiesta en el enfoque profundamente innovador en las áreas tradicionales de matemáticas y en la fundación de teorías matemáticas contemporáneas importantes como la teoría axiomática de la probabilidad y de los procesos estocásticos y la teoría algorítmica de aleatoriedad. Por otro lado, sorprende la diversidad de sus intereses científicos y logros importantes en diferentes ramas de las matemáticas contemporáneas. En el artículo [13] están citadas las 21 áreas en las cuales trabajó Kolmogorov y donde obtuvo (en ocasiones con la colaboración de sus alumnos) resultados fundamentales. Mencionemos solamente una parte de las áreas en matemáticas donde trabajó con el mayor éxito:

- fundamentos de la teoría de probabilidad y teoremas límite;
- procesos de Márkov y procesos aleatorios estacionarios;
- lógica constructiva matemática;
- sistemas dinámicos y mecánica celeste;
- mecánica estadística;
- estadística matemática;
- teoría de funciones de variable real;
- teoría algorítmica de aleatoriedad;
- teorías de aproximación;
- probabilidad aplicada.

Este universalismo y la diversidad de los logros son sorprendentes para el siglo XX cuando se produjo un desarrollo exponencial de las investigaciones en matemáticas puras y aplicadas estimuladas por las necesidades de la física, la técnica y el desarrollo de armamentos. Las matemáticas se volvían cada año más extensas, ramificadas y especializadas. Es natural que esto ocasionara, casi inevitablemente, la especialización (a veces muy exagerada) entre los matemáticos y especialistas en las aplicaciones. Sin duda, en este sentido el "último matemático universal" Kolmogorov puede ser adscrito al gremio de los matemáticos célebres del pasado tales como L. Euler, K. Gauss, B. Riemann y, la excelencia de sus logros lo pone al nivel del gran innovador y fundador de las matemáticas contemporáneas, Henri Poincaré. Vale la pena subrayar que, a diferencia del gigante solitario Poincaré, Kolmogorov donó parte de su talento a sus múltiples alumnos que se convirtieron en notables matemáticos soviéticos. Como resultado, A. N. Kolmogorov fundó vigorosas escuelas matemáticas en diferentes áreas de las matemáticas contemporáneas, pero sobre todo, en la teoría de probabilidad y procesos estocásticos.

Se estima que el número de trabajos publicados por Kolmogorov (incluyendo los artículos de investigación, libros, manuales, memorias de congresos, artículos educativos y de divulgación científica) es cercano a los 500.

En sus memorias, los colegas y alumnos señalan su generosidad al compartir sus ideas originales y productivas, así como el formidable talento de A. N. Kolmogorov para descubrir relaciones profundas entre fenómenos matemáticos aparentemente muy diferentes y encontrar "soluciones sencillas" a problemas muy difíciles. Algunos autores señalan con sorpresa "¿por qué nadie antes se dió cuenta de esto, si es tan fácil?". En la última sección trataremos de ilustrar algunos descubrimientos "naturales" pero fundamentales de A. N. Kolmogorov.

2. Biografía comentada

2.1. Los años de formación

Andrey Nikolaevich Kolmogorov nació en abril de 1903 en Tambov (ciudad rusa situada al sureste de Moscú), en la época del florecimiento prerrevolucionario del arte y la ciencia en Rusia. Desafortunadamente su madre, M. Ya. Kolmogorova, descendiente de la nobleza rusa, murió durante el parto. Su padre, N. M. Kataev, agrónomo, que no estuvo formalmente casado con M. Ya. Kolmogorova, no participó en la educación de su hijo (en parte no por su voluntad). Andrey fue adoptado por la hermana de su madre, V. Ya. Kolmogorova, quien la sustituyó e hizo todo lo posible para darle una buena educación. En 1910 la familia se mudó a Moscú donde Andrey Kolmogorov ingresó a uno de los nuevos colegios de "orientación democrática". Para los alumnos de primarias y secundarias de hoy sería instructivo leer los siguientes recuerdos de A. N. Kolmogorov sobre la atmósfera reinante en este colegio, que propiciaba la afición a los estudios[1]:

"Los grupos eran pequeños (de 15 a 20 alumnos). Una parte considerable de los maestros eran apasionados de la ciencia. Había algunos maestros universitarios. Nuestra maestra de geografía participaba en expediciones interesantes. Muchos alumnos competían entre sí en el estudio independiente del material extra, a veces con las intenciones insidiosas de usar los conocimientos para avergonzar a los maestros.... En matemáticas fui uno de los primeros de mi clase, pero mis pasiones principales y más serias en la escuela fueron inicialmente la biología y después la historia rusa".

En el año 1920, ya en los tiempos difíciles de ruina postrevolucionaria, A. N. Kolmogorov ingresa simultáneamente a la Facultad de Matemáticas de la Universidad de Moscú y al Instituto (Superior) Químico - Tecnológico D.I. Mendeleev. Pronto los intereses matemáticos ocupan el lugar principal y deja el Instituto. Para sostenerse tiene que realizar periódicamente trabajos rutinarios (un tiempo fue cobrador de tranvía, por ejemplo). Sin embargo, además de sus brillantes estudios en la universidad, Andrey Nikolaevich encontraba tiempo para participar en el trabajo del círculo de aficionados de historia donde realizó su primer trabajo de investigación dedicado a la historia antigua de Rusia (el manuscrito fue encontrado después de su muerte y publicado). Es interesante señalar que, en sus años de madurez, Kolmogorov volvió a realizar investigaciones en ciencias humanas, elaborando métodos probabilísticos en lingüística (y aplicándolos, por ejemplo, para el análisis de las estructuras lingüísticas del gran poeta ruso Alexander Pushkin).

En la universidad, el estudiante capaz pronto fue detectado por el fundador de la escuela rusa de funciones de variable real N. N. Luzin. A. N. Kolmogorov se convirtió en uno de sus discípulos. A la edad de 19 años, siendo estudiante de segundo año de licenciatura, A. N. Kolmogorov construyó y publicó (ver [5]) un contraejemplo inesperado en la teoría de funciones. A saber, construyó una función $f : [0, 2\pi) \rightarrow$

[1] Traducción del inglés de la cita de [4]

$\mathbb{R}$, tal que $\int_0^{2\pi} |f(t)|dt < \infty$, y para la cual serie de Fourier:

$$\sum_{-\infty}^{\infty} \widehat{f}(n)e^{int}, \quad t \in [0, 2\pi), \tag{$*$}$$

donde $\widehat{f}(n) := \frac{1}{2\pi}\int_0^{2\pi} f(t)e^{-int}dt, \quad n \in \{\dots, -2, -1, 0, 1, 2, \dots\}$, *diverge* en “casi todos” los puntos del intervalo $[0, 2\pi)$ (y por eso no representa la función f). Más tarde Kolmogorov perfeccionó el ejemplo demostrando la divergencia en todos los puntos del intervalo $[0, 2\pi)$. Las integrales involucradas en las fórmulas se entienden en el sentido de Lebesgue (puesto que la función en este ejemplo no puede ser continua). Cabe señalar que en 1966, Carleson demostró que si $p > 1$ y

$$\int_0^{2\pi} |f(t)|^p dt < \infty,$$

entonces la serie de Fourier (*) *converge* a la función f en “casi todos” los puntos del intervalo. Es decir,

$$f(t) = \sum_{-\infty}^{\infty} \widehat{f}(n)e^{int}$$

para la “mayoría” de puntos de $[0, 2\pi)$. Gracias a su descubrimiento, el estudiante Kolmogorov adquirió en tan sólo algunos meses fama internacional en el mundo de las matemáticas.

Al terminar los 5 años de universidad, en 1925, A. N. Kolmogorov ingresó al doctorado bajo la asesoría de N. N. Luzin y durante los siguientes cuatro años publicó 14 trabajos de investigación en teoría de funciones, en teoría de probabilidad y en lógica constructiva matemática. En particular, realizó una serie de trabajos fundamentales en lógica constructiva que le dieron amplio renombre internacional.

2.2. La profesión

Durante el doctorado y los siguientes años de trabajo en la Universidad de Moscú, la colaboración con el gran especialista en la teoría de probabilidad A. Khinchine y el conocimiento de los trabajos de las escuelas de probabilidad rusa y francesa (Márkov, Lyapunov, Bernstein, Borel, Levy) alimentaron el interés de Kolmogorov hacia los problemas de la teoría de probabilidad. Demostró una serie de teoremas (actualmente clásicos) relacionados con las leyes de grandes números y la ley del logaritmo iterado. En 1933 publicó en alemán su libro sobre fundamentos de la teoría de probabilidad (ver su traducción al inglés [6]) en el cual construyó la axiomática de la teoría de probabilidad y demostró algunos teoremas fundamentales que actualmente son la base de la teoría de probabilidad y procesos estocásticos. La aparición de este libro, así como la fuerte influencia de las nuevas tecnologías en formación (por ejemplo, la radiolocalización), llevaron al desarrollo intensivo de los métodos de la teoría de probabilidades y procesos estocásticos en la Unión Soviética, Francia, EEUU y otros países.

En 1941 A. N. Kolmogorov creó la teoría de alisamiento y predicción de las sucesiones aleatorias estacionarias, que fue redescubierta (se afirma que en forma independiente) en 1949 por el científico norteamericano N. Wiener. Esta teoría, que encontró importantes aplicaciones en radiolocalización y otras ramas de la tecnología, fue llamada "método de Wiener". Solamente en años posteriores comenzó a llamarse "método de Wiener - Kolmogorov".

Un poco antes, Kolmogorov publicó un artículo en el cual estableció las relaciones profundas entre las teorías de procesos de Márkov y de ecuaciones diferenciales que a primera vista no tienen nada en común.

Durante la Segunda Guerra Mundial la mayoría de los científicos soviéticos trabajó sobre problemas aplicados a la defensa del país. A. N. Kolmogorov no fue la excepción. Un poco antes de la guerra obtuvo y publicó en 1941 resultados importantes en el estudio de turbulencia. Durante mucho tiempo el fenómeno de turbulencia (en cierto sentido el "movimiento caótico de líquidos o gases") ha sido un obstáculo para construir aviones de reacción aerodinámicamente estables. En este tiempo, Kolmogorov también se ocupó del análisis y la optimización de distribución de la caída de proyectiles de artillería.

2.3. Formación de la Escuela

En la historia de las matemáticas soviéticas, y de posiblemente de toda la ciencia soviética, no ha habido un científico como Kolmogorov, con la cantidad de alumnos que se convirtieron en especialistas famosos a nivel mundial. Citemos, entre otros, a matemáticos como V. Arnold (sistemas dinámicos, ecuaciones diferenciales y otros), A. Maltzev (álgebra), A. Shiryaev (procesos aleatorios y matemáticas financieras), V. Uspensky (teoría de algoritmos), Yu. Prokhorov, A. Borovkov, V. Zolotarev (teoría de probabilidad), I. Gelfand (análisis funcional), P. Martin-Löf (aleatoriedad algorítmica), y Ya. Sinay (teoría ergódica).

Muchos colegas y alumnos señalan [11] la enorme capacidad de Kolmogorov para generar una multitud de ideas nuevas, interesantes e inesperadas, y compartirlas generosamente con los que lo rodeaban, especialmente con sus alumnos, de los que apreciaba, por encima de cualquier otra virtud, la independencia creativa y los impulsaba para que trataran de resolver problemas matemáticos muy difíciles e importantes.

Kolmogorov prodigaba sus ideas científicas no sólo por su carácter generoso, sino porque entendía perfectamente que el tiempo es limitado, y que solo no podría llevar a cabo sus múltiples proyectos. Además, le gustaba buscar las soluciones principales de problemas que presentaban retos, pero no le agradaba ocuparse de detalles técnicos, de la investigación de todas las inferencias y de todas las aplicaciones posibles. Trataba de encargar esta parte de trabajo a otros. A veces formulaba los teoremas sin ofrecer su demostración. Un ejemplo ilustrativo es la creación de las bases de la teoría famosa de mecánica clásica conocida actualmente como KAM-teoría (Kolmogorov, Arnold, Moser). En 1954 Kolmogorov publicó un artículo corto en el cual expuso un nuevo método del estudio de perturbaciones y estabilidad de las trayectorias de los sistemas dinámicos analíticos (definidos por las ecuaciones diferenciales). Este método fue un avance serio en la solución del problema (presente desde los tiempos Newton y Laplace) de la estabilidad de movimiento de tales sistemas como el sistema solar. El teorema principal del artículo fue presentado por Kolmogorov sin demostrar. En 1962 el matemático suizo J. Moser generalizó el teorema de Kolmogorov del caso de las funciones analíticas al caso de funciones 333 veces diferenciables ($\frac{666}{2}$!!!).

Como escribe Arnold [1]: "Esto fue un logro formidable porque el mismo Kolmogorov consideraba que ni siquiera bastaría con un número infinito de derivadas..."

En 1963 V. Arnold publicó la demostración del teorema de Kolmogorov del 1954 y en los años subsiguientes continuó el desarrollo de la teoría de estabilidad cualitativa de sistemas dinámicos. Al mismo tiempo, a mediados de los años 60, aparecieron los artículos norteamericanos que contenían las aplicaciones y generalizaciones del "método de Moser". Arnold (completamente apoyado por Moser) se pronunció enérgicamente en contra de los intentos de atribuir la autoría del método de Kolmogorov a Moser. La teoría KAM se hizo clásica en mecánica y encontró múltiples aplicaciones en física. Aunque la teoría KAM no puede aplicarse directamente al análisis de la estabilidad del sistema n cuerpos, como el sistema solar, V. Arnold y M. Sevriuk demostraron posteriormente que para "casi todas" las condiciones iniciales, el movimiento de estos sistemas es estable y que la "probabilidad" de las evoluciones que llevan a las catástrofes (como, por ejemplo, las colisiones entre planetas) es pequeña.

Otro resultado de la colaboración entre estos grandes matemáticos fue la solución del problema 13 de Hilbert[2] que podemos formular mediante la siguiente pregunta ¿es posible representar toda función continua de tres variables como la composición de funciones continuas de dos variables? (D. Hilbert suponía que no).

Los esfuerzos conjuntos de Kolmogorov y Arnold llevaron a la solución de un problema más general que el problema propuesto por Hilbert (con el resultado opuesto a la hipótesis del último). En 1956 A. N. Kolmogorov demostró que cualquier función continua de n ($n \geq 3$) variables reales puede ser representada como una composición de funciones continuas de 3 variables. Con esto elaboró el método que permitió a su alumno V. I. Arnold demostrar lo mismo utilizando solamente composiciones de funciones de dos variables. El acorde final fue el resultado de Kolmogorov, obtenido en 1957 (véase [7]), donde estableció que toda función continua de n ($n \geq 2$) variables puede ser expresada mediante sumas y composiciones de funciones continuas de una variable.

No es nuestro propósito hacer una descripción de todos los logros matemáticos de Kolmogorov, aunque mencionaremos algunos en la sección final.

2.4. Las controversias

Kolmogorov, un matemático brillante y generoso en muchos aspectos, se vió envuelto sin embargo en algunas situaciones que arrojan algo de sombra sobre su personalidad. La vida de una persona y el retrato de esta vida como aparece en los artículos conmemorativos y memorias suelen diferir.

2.4.1. El caso Luzin

Hubo en la vida de A. Kolmogorov y algunos otros matemáticos de su generación un episodio desagradable conocido como el "proceso de Luzin" (ver [14]). En 1936 la élite partidista inició una acción contra el viejo fundador de la escuela matemática moscovita, donde se formaron, entre otros, los brillantes A. Khinchine, P. Alexandrov, A. Kolmogorov. Este "proceso", que afortunadamente no llegó hasta sus últimas consecuencias, fue emprendido en 1936 por razones puramente políticas. En

[2] V. I. Arnold considera [1] que los famosos "problemas de Hilbert" (formulados en el Congreso de matemáticos en 1900) y su solución no ejercieron una influencia considerable en desarrollo de las matemáticas en el siglo XX, a diferencia de resultados, ideas y predicciones de H. Poincaré.

vísperas de la guerra con Alemania, los mejores matemáticos soviéticos publicaban en el extranjero, en especial, en Alemania. A pesar de que se trataba de una práctica común, este argumento fue utilizado contra N. N. Luzin, para mostrar su falta de patriotismo. Con el fin de inculpar a Luzin, todavía fuera de tribunales, se organizó una comisión especial de la Academia de Ciencias de la URSS. Se atrajeron en calidad de "acusadores potenciales" a colegas y alumnos de Luzin. Y si bien algunos matemáticos conocidos de la generación mayor (S. Bernstien, A. Krylov, I. Vinogradov) se pronunciaron en su defensa justamente opinando que los "errores" reales y ficticios de N. Luzin no correspondían ni a la décima parte de las acusaciones que le imputaban, muchos de sus jóvenes alumnos (A. Khinchine, P. Alexandrov, A. Kolmogorov entre otros) y colegas (B. Sigal, S. Sobolev) apoyaron las acusaciones. Es notable que el acusador más agresivo fuera el conocido topólogo P. S. Alexandrov (amigo cercano de Kolmogorov) quien ya en aquel entonces [2] tenía serios conflictos con Luzin. La postura tan hostil y agresiva de una parte de la comunidad matemática soviética, que llegó hasta acusarlo de causar daño a la URSS y casi de espiar a favor de Alemania, hubiera podido llevar a que Luzin, ya de avanzada edad y destruido moralmente, compartiera el destino de su maestro, el famoso matemático D. Egorov que falleció en el campo de concentración. Afortunadamente, como creen los autores de la investigación [2], Stalin consideró que la "lección de patriotismo" ya había sido impartida y dio órdenes de poner fin al "proceso".

Desde aquél entonces y durante varias décadas los matemáticos soviéticos publicaron sus trabajos únicamente en la URSS. Es curioso que a pesar de todos sus méritos científicos, P. S. Alexandrov no pudo ser elegido a la Academia de Ciencias de la URSS hasta la muerte de N. Luzin en 1950 (la membresía de la Academia de Ciencias era muy prestigiosa y bien pagada.)

2.4.2. La reforma educativa

Otro episodio ensombreció los últimos años de Kolmogorov. A finales de los años 60 tanto en el medio pedagógico como, en parte, en el científico, surgió la inquietud de que la enseñanza de las matemáticas en la escuela "se encontraba en retraso respecto de las exigencias de la época". Posiblemente estas afirmaciones eran estimuladas por el proceso de la "bourbakización" de las matemáticas[3]. La "bourbakización" en las escuelas y las universidades produjo que se remplazara

[3]Terminología de V. Arnold propuesta en honor del autor colectivo francés "N. Bourbaki", conocido de sus monografías matemáticas que se distinguían por la

la enseñanza intuitiva (directa) y razonablemente formalizada (en particular, con el uso de múltiples ejemplos) por el enfoque puramente formal basado en la teoría de conjuntos. De acuerdo con la opinión de Arnold [1], en Francia (y, parcialmente, en EEUU) esto llevó con el tiempo a la caída catastrófica e irrecuperable del nivel de la educación matemática en escuelas y universidades. No obstante, en los años 60 se difundió profusamente la idea de "bourbakizar" la educación, y en la URSS se decidió llevar a cabo una reforma en la enseñanza de las matemáticas y la física en escuela (que duró hasta finales de los años 70).

Uno de los ideólogos y líderes de la reforma de la educación matemática fue A. N. Kolmogorov. Con la "bendición" de Kolmogorov y otros líderes de la reforma se escribieron nuevos manuales de álgebra elemental y de geometría, obligatorios para el uso en las secundarias y preparatorias y se editaron los materiales de apoyo para los maestros. Cabe señalar que en aquel entonces en la URSS los programas y los manuales eran estándares y obligatorios para todas las escuelas del país lo que se puede ver como reminiscencias del "totalitarismo", pero también puede argumentarse que con esto se aseguraba un nivel promedio alto de la educación.

Según la opinión de la inmensa mayoría de los matemáticos reputados de la Unión Soviética la reforma de la enseñanza matemática fracasó. Los manuales resultaron mal diseñados desde el punto de vista pedagógico, sobrecargados de definiciones formales y de demostraciones poco ilustrativas y poco intuitivas. La mayoría de los escolares y parte considerable de los maestros no podían entender mucho de lo escrito en estos libros. Esto produjo grandes problemas con la enseñanza y la comprensión. Por si esto no fuera suficiente, la exposición formal impidió que los escolares lograran dominar los procedimientos prácticos del uso de las matemáticas (lo que es importante para muchas especialidades).

A modo de ejemplo, el eminente matemático soviético L. S. Pontryagin en el libro de memorias [10], en referencia al nuevo manual de álgebra elemental para el segundo año de la secundaria (para niños de 12-13 años), escribe (traducción del ruso):

"...Después de esto en el sexto grado se introducía el concepto de función basándose en la noción de relación: Se dice que una función es la relación con la cual cada punto x del conjunto P se encuentra en la correspondencia con no más de un punto del conjunto Q. El subconjunto del conjunto P que comprende todos x que se encuentran en la relación

exposición extremadamente formal.

con algunos puntos y del conjunto Q, se llama dominio de la función...”

Como resultado, a finales de los años 70, la reforma de la educación fue detenida y la sección de las matemáticas de la Academia de Ciencias de la URSS tomó la decisión de preparar nuevos libros de texto de matemáticas, más realistas y de mejor calidad.

Una parte de la responsabilidad por la reforma fracasada de la educación matemática y la cierta disminución del nivel de la educación matemática masiva recayó en el líder de la reforma, A. N. Kolmogorov. Según la opinión de Pontryagin [10] una de las posibles razones del fracaso puede ser encontrada en el carácter de A. N. Kolmogorov que con frecuencia se entusiasmaba con un quehacer nuevo pero que no podía llevar a cabo las tareas rutinarias. Por consiguiente, el trabajo real de la preparación de libros de texto quedó en las manos de personas frecuentemente no calificadas o poco concienzudas (arribistas de la Academia de Ciencias Pedagógicas de la URSS).

Otra razón plausible del fracaso es la insuficiencia del sentido de realidad frecuentemente propia de las personalidades altamente creativas. El trato con los mejores estudiantes de la Universidad de Moscú y los escolares talentosos de su internado físico-matemático podía contribuir a la impresión errónea de Kolmogorov sobre el nivel promedio de la educación y el intelecto de los escolares y sus maestros en el país. Posiblemente Kolmogorov creía que el alto grado del pensamiento abstracto es accesible para la mayoría de la juventud. Resultó que esto no era del todo cierto, ni siquiera en la URSS que contaba con una buena educación primaria, ni qué decir de la situación en otros lugares como Francia donde, según la opinión de V. Arnold [1], el 20 % de los reclutas de las fuerzas armadas no saben leer.

Es posible suponer que la reforma fracasada de la educación en los años 70 sirvió de prototipo de la “seudo-reforma” de finales de los años 80 que comenzó cuando en la Unión Soviética realmente se presentó una crisis. En vez de encontrar los medios (que en realidad había) para superar esta crisis en los márgenes y con los métodos del socialismo, las fuerzas influyentes tanto en el interior, como en el exterior del país prefirieron, bajo la observación pasiva de la población, destruir tanto el sistema como el país. Entre las consecuencias graves de la “Perestroika” se encuentran las reformas sin fin de la escuela secundaria, preparatoria y superior, con las cuales la educación matemática (hablando de una educación real) se muere lenta, pero inexorablemente.

2.5. La labor pedagógica de A. N. Kolmogorov

A. N. Kolmogorov era generoso no solamente en ideas científicas. Ayudaba con frecuencia a sus colegas y estudiantes a resolver problemas cotidianos relacionados con el estudio y el trabajo (por ejemplo, cuando era decano de la Facultad de Mecánica y Matemáticas de la Universidad de Moscú). Otro dato relevante: Kolmogorov gastó gran parte de los recursos que obtuvo por el prestigioso Premio Bolzman (1963) en crear y mantener la biblioteca especializada en teoría de probabilidad, procesos aleatorios y estadística matemática en la Universidad de Moscú. Merece mencionar que en la Unión Soviética era común la ausencia de la aspiración de acumular riquezas entre la gente de profesiones creativas. Las condiciones de trabajo y su carácter creativo eran más importantes. Por otro lado, la gente de este medio, como regla, tenía relativamente buenos ingresos, suficientes para una vida de calidad (en las condiciones cuando, por ejemplo, el nivel de servicio médico o la accesibilidad y la calidad de educación de los niños prácticamente no dependían de nivel de ingresos).

En los últimos 25 años de su vida A. N. Kolmogorov dedicó mucho tiempo a la educación matemática de los escolares. En 1963 organizó un internado físico-matemático adjunto a la Universidad Estatal de Moscú para los adolescentes talentosos. Por medio de las olimpiadas escolares buscaban a los jóvenes capaces (hombres y mujeres) y les proponían vivir y estudiar en este colegio especializado. Muchos de los egresados del internado se convirtieron posteriormente en científicos conocidos (no solamente en física o las matemáticas). Andrey Nikolaevich dedicaba muchas fuerzas y tiempo al internado. Elaboraba programas y daba, él mismo, conferencias y seminarios allí (hasta 10 horas por semana). Las clases que impartía no se limitaban a las matemáticas, a veces eran dedicadas al arte y literatura.

En 1970 A. Kolmogorov, junto con el conocido físico I. Kikoin, fundó y empezó a redactar una revista físico-matemática ilustrada para los escolares y maestros del magisterio. Recuerdo bien que las ediciones de la revista, como regla, contenían artículos de divulgación científica muy interesantes (escritos por científicos famosos), excelentes problemas y sus soluciones originales, buenas ilustraciones a color, etc. Esta revista jugó un papel importante en la capacitación de los maestros, el mejoramiento de la calidad de educación y la estimulación del interés hacia las ciencias exactas entre los escolares.

2.6. Algunas observaciones adicionales

No hace falta decir que laborando toda su vida en la Universidad estatal de Moscú, A. N. Kolmogorov preparó y dio una cantidad enorme de cursos generales y especializados en diversas áreas de las matemáticas. A veces ofrecía conferencias para un auditorio amplio (escolares y demás) sobre cuestiones filosóficas de las matemáticas y la cibernética. Se dice que sus conferencias especializadas eran un modelo del proceso creativo, pero que no siempre eran comprensibles para los oyentes. En aquel entonces se decía: "los estudiantes de doctorado pueden entender las conferencias de Kolmogorov para los escolares, los doctores en ciencias pueden entender las conferencias para los estudiantes de doctorado y las conferencias para los doctores en ciencias son sólo accesibles para el propio Kolmogorov". El autor de estos apuntes tuvo la suerte de poder escuchar en ocasiones conferencias de Kolmogorov en congresos matemáticos y puede confirmar lo justo de esta opinión.

En muchas memorias mencionan la perfecta condición física de Kolmogorov quien practicó activamente el deporte hasta la edad de 70 años. Hablando acerca de su personalidad, sin duda podemos creer a A. Shiryaev (quien trató a Kolmogorov a lo largo de varias décadas) en que Andrey Nikolaevich se distinguía por "la suprema honestidad científica, la objetividad, la modestia, la sensibilidad y la generosidad" [11].

A. N. Shiryaev, su alumno y famoso especialista en el área de procesos estocásticos, testifica documentalmente (ver [11]) que Kolmogorov en 1943 (cerca de cumplir los 40 años) hizo y anotó en su diario un plan bastante detallado de la actividad que desarrollaría en los siguientes cuarenta años (qué investigaciones realizar, qué monografías, manuales y artículos escribir, qué materias nuevas enseñar, etc.). En [11] fue notado que el plan se ha cumplido en su totalidad, a excepción del último punto "escribir las memorias de mi vida". Los 40 años mencionados de la vida de A. N. Kolmogorov fueron saturados con una intensa actividad científica, científico-organizacional y pedagógica que trajeron muchos resultados formidables en diversas áreas de las matemáticas (y no solamente de matemáticas) y sus aplicaciones.

3. Comentarios acerca de algunas ideas "naturales" de A. N. Kolmogorov

En esta sección expondremos, sin entrar en demasiados detalles, cuatro descubrimientos matemáticos de Kolmogorov en diferentes áreas de las matemáticas. Aunque parecen "sencillos", el primer descubrimiento sentó las bases del desarrollo de las teorías de probabilidad y procesos estocásticos en el siglo XX, el segundo resultó más tarde ser clave para formular y demostrar muchos teoremas límite para los procesos aleatorios y procedimientos estadísticos (incluyendo el procedimiento de reconocimiento de imágenes), el tercero estableció los fundamentos de la teoría algorítmica de aleatoriedad, y el cuarto resultó ser extremadamente útil en la estadística no paramétrica aplicada.

3.1. Axiomática de la teoría de probabilidad y procesos estocásticos

Los primeros cálculos probabilísticos relacionados con el análisis de los juegos de azar aparecieron a finales del siglo XV y los dos primeros teoremas serios de la teoría de probabilidad fueron demostrados al inicio del siglo XVIII por J. Bernoulli y A. de Moivre. Además del álgebra elemental, las herramientas matemáticas se reducían a la fórmula de Stirling para factoriales (que descubrió De Moivre, al parecer antes de Stirling). Después, a diferencia, por ejemplo, del cálculo o de las ecuaciones diferenciales, durante casi 250 años no se observó un desarrollo significativo de la teoría de probabilidad hasta el último cuarto del siglo XIX cuando gracias a los esfuerzos de la "escuela rusa" (Chebyshev, Márkov, Lyapunov) se inició la elaboración de métodos de la teoría de probabilidad y, a comienzos del siglo XX, A. Lyapunov demostró un teorema del límite central muy general. Paralelamente en Francia, E. Borel aplicó las nuevas ideas de la teoría de medida para demostrar una versión de la ley fuerte de grandes números y P. Levy inició el fructífero desarrollo de la teoría de convergencia débil de sumas de variables aleatorias independientes.

A pesar de estos éxitos evidentes completados por el desarrollo intenso de la estadística matemática, en los años 20 del siglo pasado, la teoría de probabilidad, al igual que la teoría de procesos aleatorios y la estadística matemática, no estaba basada en un fundamento sólido (axiomática). Contrariamente a lo sucedido, por ejemplo en el análisis, donde la definición de función continua había sido establecida de manera lógicamente irreprochable basada en la teoría de conjuntos, en la teoría de probabilidad, aparte de algunos casos particulares, no era claro cómo precisar y definir formalmente conceptos básicos tales como "evento", "probabilidad", "independencia", "promedio", "esperanza condicional", etc. No había ni siquiera una definición general satisfactoria del objeto central de la teoría de probabilidad: variable aleatoria (o de vector aleatorio).

Por otra parte, en esa época ya existía la teoría de la medida y Lebesgue había desarrollado la teoría la integración en espacios abstractos con respecto a una medida. Aprovechando estas dos teorías y partiendo de los descubrimientos iniciales de E. Borel, A. N. Kolmogorov construyó al inicio de los años 30 la axiomática de la teoría de probabilidad (y procesos estocásticos) que publicó en Alemania en el año 1933 en un libro pequeño (véase la traducción al inglés [6]).

La noción base de esta axiomática es el *espacio de probabilidad* $(\Omega, \mathcal{F}, P)$, donde Ω es un conjunto arbitrario pero fijo, $\mathcal{F} = \{A, B, \dots\}$ es la familia dada de subconjuntos de Ω que satisface ciertas condiciones (σ-álgebra). Los elementos de $\mathcal{F}$ se denominan *eventos*. Finalmente, $P : \mathcal{F} \to [0, 1]$ es la función que define una medida de probabilidad en $\mathcal{F}$, es decir, para cada evento $A \in \mathcal{F}$ se define su *probabilidad* $P(A) \in [0, 1]$. La probabilidad debe satisfacer las siguientes condiciones

$$P(\Omega) = 1;\ P(A_1 \cup A_2 \cup \dots) = P(A_1) + P(A_2) + \dots \tag{1}$$

para cualesquiera eventos $A_1, A_2, \dots$ disjuntos. La propiedad (1) se llama σ-*aditividad* de la probabilidad.

Un *vector aleatorio* X, se define como una función $X : \Omega \to \mathbb{R}^n$ que satisface ciertas condiciones muy generales (de "medibilidad"). Es decir, para cada $w \in \Omega$ se pone en correspondencia el vector $X(w) \in \mathbb{R}^n$ que representa uno de los valores posibles de X. Cuando $n = 1$, X se llama *variable aleatoria*. A partir de las probabilidades

$$P(X \in B) := P(\omega \in \Omega : \quad X(\omega) \in B), \quad B \subset \mathbb{R}^n,$$

se define la *distribución del vector aleatorio* X.

La *esperanza matemática* EX (valor promedio) de una variable aleatoria X se define como el siguiente número real

$$EX := \int_\Omega X(\omega) P(d\omega). \tag{2}$$

En la parte derecha de la igualdad (2) se utiliza la integral de Lebesgue, la cual se calcula, por ejemplo, como: $EX = \sum_k X(\omega_k) P(\omega_k)$, si el espacio muestral Ω es numerable o finito, o bien se reduce a la integral habitual de Riemann cuando $\Omega = \mathbb{R}^k$, y X es una función "suficientemente buena" (por ejemplo, continua). Además de la definición de promedio, la integral de Lebesgue se utiliza eficientemente para otros objetivos en la teoría construida sobre la base de axiomática de Kolmogorov. La noción de la integral de Riemann es insuficiente para realizar estos objetivos.

Una de las nociones centrales en la teoría de probabilidades –la noción de *independencia*– se define en los términos del producto de distribuciones. Usando los conceptos de la teoría de la medida y la teoría de funciones, se introducen diferentes tipos de convergencia de vectores aleatorios y variables aleatorias y otras nociones fundamentales para la teoría de probabilidad, la teoría de procesos estocásticos y la estadística matemática.

Es importante subrayar que en el libro de Kolmogorov [6] junto con algunos teoremas básicos de la teoría de probabilidad (tales como la versión general de la ley fuerte de los grandes números) están demostrados dos teoremas fundamentales que predeterminaron el desarrollo de la teoría contemporánea de procesos estocásticos:

(a) El teorema de existencia y unicidad de la distribución infinito-dimensional con distribuciones finito-dimensional dadas.

(b) El teorema general de existencia y unicidad de la esperanza condicional.

Para demostrar el teorema mencionado en el inciso (b), Kolmogorov utilizó la integral de Lebesgue (del tipo (2)) y el concepto de la "derivada" de una medida con respecto a otra. En ciertos casos la derivada usual también puede ser considerada como la derivada de la medida μ definida por la relación

$$\mu(B) := \int_B f(t)dt, \quad B \subset \mathbb{R}$$

con respecto a la *medida de Lebesgue* ℓ generada por $\ell([a,b]) := b - a$ (la longitud del intervalo). En efecto, tomando $B = [x, x + \Delta x]$ y suponiendo que f es continua en x, obtenemos:

$$\begin{aligned} \mu([x, x+\Delta x]) &= \int_x^{x+\Delta x} f(t)dt \approx f(x)\Delta x, \quad \text{lo que implica:} \\ \frac{d\mu}{d\ell}(x) &:= \lim_{\Delta x \to 0} \frac{\mu([x, x+\Delta x])}{\ell([x, x+\Delta x])} = \lim_{\Delta x \to 0} \frac{f(x)\Delta x}{\Delta x} = f(x). \end{aligned}$$

El teorema en el inciso (a) en particular, establece la existencia de espacios de probabilidad apropiados para el estudio de fenómenos aleatorios muy diversos. Además, este teorema ha tenido muchas aplicaciones fuera de la teoría de probabilidad.

El axioma (1) de la σ-*aditividad* lleva a "paradojas" del siguiente tipo. Supongamos que "arrojamos al azar" un punto (matemático, sin tamaño) al intervalo (0,1) de tal manera que $P(\text{caída en } B) := \ell(B)$ (la medida de Lebesgue que se reduce a longitud cuando B es un intervalo). Sea $Q \subset (0,1)$ el conjunto de todos los números racionales de (0,1) (numerable y denso en (0,1)). Entonces por (1),

$$P(\text{el punto caerá en } Q) = \sum_{x \in Q} P(\{x\}) = \sum_{x \in Q} \ell([x,x]) = \sum_{x \in Q} 0 = 0,$$

y, por otra parte

$$P(\text{el punto caerá en el conjunto de números irracionales}) = 1 - 0 = 1.$$

Sin embargo, al realizar cualquier medición física (o cualquier otro tipo) siempre obtenemos un número racional. La naturaleza de esta "paradoja" ilusoria podría ser relacionada con la definición de velocidad instantánea $V(t)$ en física

$$V(t) = \frac{dS(t)}{dt} = \lim_{\Delta t \to 0} \frac{S(t + \Delta t) - S(t)}{\Delta t}, \tag{3}$$

donde $S(t)$ es la posición del cuerpo en el instante t. Para realizar "cálculos reales" (o "mediciones reales") nadie puede tomar en cuenta los intervalos de tiempo Δt, digamos menores que $10^{-10^{100}}$ (tal vez, intervalos de tiempo de esta magnitud ni siquiera existen). Es decir, el límite en (3) "carece de sentido físico". No obstante, el uso de límites abstractos como él que aparece en (3), está en la base de una gran parte de las ciencias exactas contemporáneas, sin las cuales la civilización no sería lo que es ahora.

La axiomática propuesta en [6] fue apreciada en su justo valor por la inmensa mayoría de los investigadores de fenómenos aleatorios y desde entonces sirve como base para el desarrollo de la teoría de probabilidad, procesos estocásticos y la estadística matemática (aunque existen las axiomáticas alternativas). El papel excepcional del libro [6] en la historia de la ciencia es señalada por los clásicos de la teoría de procesos aleatorios D. L. Doob (EEUU) y K. Ito (Japón). El hecho de que la axiomática de Kolmogorov es adecuada para la construcción y el desarrollo de las teorías mencionadas se demuestra por sus múltiples aplicaciones exitosas casi en todas las ciencias exactas, ingenierías, economía, medicina, industria, agricultura, etc.

3.2. ϵ-entropía de conjuntos compactos

En la recta real $\mathbb{R} \equiv (-\infty, \infty)$ con la distancia (métrica) $d(x,y) := |x - y|$; $x, y \in \mathbb{R}$, la "esfera" cerrada B_ϵ con radio $\epsilon > 0$ y centro en el punto x_0 es el conjunto

$$B_\epsilon \equiv B_\epsilon(x_0) := \{x \in \mathbb{R} :\ d(x_0, x) \leq \epsilon\} \equiv [x_0 - \epsilon, x_0 + \epsilon].$$

Llamaremos ϵ-*esfera* al conjunto B_ϵ.

Para cualquier intervalo *acotado cerrado* (i.e. compacto) $K_1 = [a, b] \subset \mathbb{R}$ $(a < b)$, el número mínimo de ϵ-esferas $(\epsilon > 0)$ que cubren $[a, b]$, es igual a

$$N_\epsilon(K_1) = \left\{\frac{b-a}{2\epsilon}\right\} \leq C\epsilon^{-1}, \tag{4}$$

donde $\{y\}$ denota el mínimo número entero mayor o igual a y, y $C = C[K_1]$ es una constante que depende solamente de a y b. En la desigualdad (4), y en lo que sigue, consideramos $\epsilon > 0$ "pequeños", es decir nos interesamos en el caso cuando $\epsilon \to 0$.

En el plano $\mathbb{R}^2$ con la métrica

$$d(x, y) := \text{máx}\{|x_1 - y_1|, |x_2 - y_2|\};\ x = (x_1, x_2),\ y = (y_1, y_2) \in \mathbb{R}^2, \tag{5}$$

la "ϵ-esfera" cerrada B_ϵ con centro en $\bar{x} = (x_0, y_0)$ será el conjunto:

$$B_\epsilon \equiv B_\epsilon(x_0, y_0) := \{x \in \mathbb{R}^2 : d(x, \bar{x}) \leq \epsilon\} = [x_0 - \epsilon, x_0 + \epsilon] \times [y_0 - \epsilon, y_0 + \epsilon]$$

que es, en realidad, un cuadrado.

No es difícil entender porqué el número mínimo $N_\epsilon(K_2)$ de ϵ-esferas que cubren el rectángulo $K_2 = [a, b] \times [c, d]$, se estima mediante

$$N_\epsilon(K_2) \leq C\epsilon^{-2}, \tag{6}$$

donde C es una constante que depende solamente de K_2.

Consideremos ahora el espacio euclideano n-dimensional $\mathbb{R}^n$ con la métrica euclidiana (o cualquier otra métrica equivalente, como por ejemplo la generalización de la dada en (5)). Sea K_n cualquier compacto en $\mathbb{R}^n$ (conjunto acotado y cerrado, como una esfera o un paralelepípedo). Generalizando (4) y (6), observamos que el número mínimo $N_\epsilon(K_n)$ de ϵ-esferas con las cuales se puede cubrir el conjunto K_n satisface la desigualdad

$$N_\epsilon(K_n) \leq C\epsilon^{-n}, \tag{7}$$

con alguna constante C que depende de K_n. Notemos que cuando el compacto K_n es "verdaderamente n-dimensional", es imposible disminuir el exponente n en (7). Sin embargo, aún en la recta $\mathbb{R}$ hay subconjuntos compactos no numerables para los cuales N_ϵ crece más lento que $1/\epsilon$ cuando $\epsilon \to 0$ (los llamados "conjuntos de dimensión fractal").

Por ejemplo, para el conocido compacto de Cantor $K_c \subset [0,1]$ se tiene que:

$$N_\epsilon(K_c) \leq C\epsilon^{-d}, \tag{8}$$

donde $d = \dfrac{\ln 2}{\ln 3} \approx 0.631$ (el valor de d se interpreta como la dimensión fractal del conjunto K_c).

En el trabajo de Kolmogorov y Tikhomirov (ver [8]) fueron investigadas las tasas de crecimiento (cuando $\epsilon \to 0$) de los tamaños de recubrimientos mínimos finitos por ϵ-esferas de compactos en los diferentes espacios métricos cuyos elementos son funciones (espacios funcionales). Examinemos un ejemplo sencillo. Para un segmento dado $[a,b]$ en $\mathbb{R}$, sea $C[a,b]$ el conjunto de todas las funciones continuas $\varphi : [a,b] \to \mathbb{R}$. Para cada par de funciones $\varphi, f \in C[a,b]$ se define la métrica uniforme

$$d(\varphi, f) := \max_{x\in[a,b]} |\varphi(x) - f(x)|.$$

La ϵ-esfera cerrada B_ϵ (con centro en $f_0 \in C[a,b]$) que corresponde a esta distancia es el siguiente subconjunto de $C[a,b]$:

$$B_\epsilon \equiv B_\epsilon(f_0) := \{\varphi \in C[a,b] : \max_{x\in[a,b]} |\varphi(x) - f_0(x)| \leq \epsilon\}. \tag{9}$$

La desigualdad en esta definición significa que $\varphi \in B_\epsilon(f_0)$ si y sólo si el *valor* $\varphi(x)$ se encuentra en el intervalo $[f_0(x) - \epsilon,\ f_0(x) + \epsilon]$ para cada $x \in [a,b]$. Consideremos el subconjunto *compacto* K de $C[a,b]$ que se compone de todas las funciones φ derivables en $[a,b]$ tales que $|\varphi'(x)| \leq 1$ para todo $x \in [a,b]$. Denotemos por $N_\epsilon(K)$ el número mínimo de ϵ-esferas que cubren K. De la compacidad K se sigue que $N_\epsilon(K)$ es finito para cualquier $\epsilon > 0$. Los resultados obtenidos en [8] implican que existen constantes positivas b_1 y b_2, tales que para cada $\epsilon > 0$,

$$N_\epsilon(K) \leq \exp\left\{b_1[2 + (b-a)]/\epsilon\right\}, \tag{10}$$

$$N_\epsilon(K) \geq \exp\left\{b_2[2 + (b-a)]/\epsilon\right\}. \tag{11}$$

A diferencia de (7) donde la cantidad de ϵ-esferas que cubren el compacto aumenta, cuando $\epsilon \to 0$, como una función de potencias $\left(\dfrac{1}{\epsilon}\right)^n$, en (10) la tasa del crecimiento es exponencial. Esto refleja el hecho de que el espacio métrico $C[a,b]$ tiene “dimensión infinita”.

Las desigualdades del tipo (10) y (11) tienen aplicaciones útiles, sobre todo en la teoría de la aproximación de funciones (la cual, a su vez, se usa en matemáticas numéricas). Para $\epsilon > 0$ dado, escogiendo en cada una de las ϵ-esferas B_ϵ que cubren al conjunto K una función suave conveniente f_i (por ejemplo un spline), se construye un conjunto *finito* $M = \{f_i\}$ que *aproxima* a K en el sentido de que para cada función $\varphi \in K$ es posible encontrar una función $f_i \in M$, tal que $|\varphi(x) - f_i(x)| \leq \epsilon$ para todo $x \in [a, b]$.

Las desigualdades (10) y (11) proporcionan la estimación del número mínimo de funciones aproximantes f_i (los elementos del conjunto M).

Considerando un espacio métrico arbitrario (X, d) y un compacto $K \subset X$, para cada $\epsilon > 0$, el número $\ln[N_\epsilon(K)]$ se denomina *ϵ-entropía* de K. De las desigualdades (8), (10) y (11) se sigue que para los compactos n-dimensionales la ϵ-entropía es proporcional a $\ln(1/\epsilon)$ (aumenta muy lentamente con el crecimiento de $1/\epsilon$ cuando $\epsilon \to 0$), mientras que la ϵ-entropía de $K \subset C[a, b]$ en el último ejemplo crece, cuando $\epsilon \to 0$, con velocidad proporcional a $1/\epsilon$. Existen espacios compactos y conjuntos compactos en ellos donde la ϵ-entropía aumenta mucho más rápido.

Entre 10 y 15 años después de la publicación del trabajo [8] quedó claro que el estudio de tasas del crecimiento de ϵ-entropía de los compactos en algunos espacios funcionales (no necesariamente con la métrica uniforme como en (9)) es decisivo para el desarrollo de las siguientes áreas importantes de la teoría de probabilidad contemporánea:

- investigación de las propiedades de trayectorias de procesos estocásticos gaussianos generales;

- los teoremas límite (del tipo del teorema límite central) para los elementos aleatorios en espacios funcionales;

- los teoremas límite uniformes para los procesos aleatorios empíricos (construidos con base en funciones de distribución empíricas, ver sección 3.4) y sus aplicaciones en la estadística.

En efecto, fue posible demostrar algunos teoremas importantes en estas teorías solamente bajo la condición de que la ϵ- entropía de ciertos compactos en espacios funcionales no aumente, cuando $\epsilon \to 0$, "demasiado rápido". Aparecieron muchos trabajos que desarrollan la investigación aparecida en [8] y muchos resultados que utilizan el concepto de ϵ-entropía. Con respecto a esto el lector puede consultar el libro [12] o el artículo [3] (y encontrar allí otras referencias). En su tiempo (a mediados de los años 80) el autor de estas notas utilizó ϵ-entropía de los compactos y estimaciones de tasa de su crecimiento para construir procedimientos de control adaptado para procesos de Márkov a tiempo discreto.

3.3. Aleatoriedad algorítmica

La axiomatización matemática de conceptos básicos tales como aleatoriedad, probabilidad, independencia, etc. propuesta por A. N. Kolmogorov llevó al desarrollo de una teoría profunda y con aplicaciones exitosas. No obstante, en el contexto más amplio, los conceptos arriba mencionados son filosóficos (y, en un sentido, hasta cotidianos). Por eso las discusiones sobre su naturaleza y las interpretaciones no han cesado de manifestarse a través de los siglos.

Está muy difundido el enfoque "matemático-pragmático" en el cual, independientemente de la interpretación, la teoría de probabilidad se utiliza para modelar tales fenómenos aleatorios para los cuales ésta demuestra su eficiencia y adecuación. Con esto se considera (justificadamente) importante la posibilidad de verificar los resultados de aplicaciones, y evitar las aplicaciones especulativas, como el cálculo de la probabilidad de existencia de civilizaciones extraterrestres (ver comentarios al respecto en [9]).

Por otro lado, muchas mentes extraordinarias (H. Poincaré, por ejemplo) emprendieron intentos serios para entender la naturaleza de la aleatoriedad y su relación con la "determinabilidad". En las últimas décadas el interés hacia estos problemas se incrementó en relación con el descubrimiento y la investigación de "conducta caótica" de algunos sistemas dinámicos deterministas.

Hace falta mencionar algunas de las interpretaciones de aleatoriedad y probabilidad que se formaron y se difundieron en el siglo XX.

(1) La interpretación "objetivista" de los fenómenos aleatorios como particularidad inmanente y dada objetivamente de algunos procesos en la naturaleza y la sociedad. Este enfoque es típico, por ejemplo, para la física cuántica donde los cambios (la dinámica) del estado cuántico tienen naturaleza probabilística. (Aunque los partidarios de la tal llamada "hipótesis de parámetros ocultos" no lo consideran así).

(2) La aleatoriedad como característica de fenómenos "masivos", por ejemplo de las "pruebas aleatorias" repetidas múltiples veces en las cuales puede realizarse uno u otro "evento aleatorio". La frecuencia límite de la aparición del evento se interpreta como su probabilidad. Basándose en la existencia de frecuencias límites de eventos (como un postulado) R. von Mises construyó al inicio del siglo XX una axiomática de la teoría de probabilidad. Como se vió después, ésta resultó inconsistente (es decir, tenía contradicciones internas). En la teoría de probabilidad basada en la axio-

mática de A. Kolmogorov la coincidencia de la frecuencia límite de un evento con su probabilidad *se demuestra* en los teoremas, como leyes fuertes de grandes números. (Por ejemplo, aumentando ilimitadamente el número de lanzamientos de un dado simétrico la frecuencia de "6" se aproxima a 1/6.)

(3) El enfoque "termodinámico" con respecto a la aleatoriedad se da cuando, sin tener la posibilidad de la descripción detallada de un sistema complejo (un gas, por ejemplo), se supone un comportamiento estocástico del "conjunto total de moléculas" (con los promedios correspondientes y las fluctuaciones alrededor de los promedios). Este enfoque resultó exitoso en la física estadística, la economía matemática y otras ciencias, pero llevó a dificultades serias relacionadas con la compatibilidad entre la irreversibilidad termodinámica (de la cual se analiza la aleatoriedad de las moléculas, por ejemplo, en la difusión) y la reversibilidad en tiempo de las ecuaciones que describen la dinámica de moléculas que forman parte del conjunto.

(4) El enfoque "subjetivo" hacia la interpretación de la probabilidad como grado de desconocimiento de algo o la inseguridad en algo. Este enfoque se realiza en las declaraciones, como: "Es muy probable que esta persona haya cometido este crimen".

Independientemente de la opinión mencionada de los "pragmáticos", los problemas de la interpretación de los conceptos básicos de la teoría de probabilidad: aleatoriedad, probabilidad, valores promedio, independencia - dependencia, influyen frecuentemente en la eficiencia de unas u otras aplicaciones y la adecuación de la interpretación de los resultados obtenidos. Con frecuencia los métodos de la teoría de probabilidad y procesos estocásticos se aplican incorrectamente con los "resultados" que llevan a conclusiones erróneas. Digamos que algunas "verdades" obtenidas con ayuda de los métodos estadísticos ya se convirtieron en el tema de anécdotas (como, por ejemplo, la "prueba" de que el consumo de los pepinos es mortalmente peligroso puesto que la gran mayoría de las personas que murieron comían pepinos).

En los años 60 A. N. Kolmogorov propuso un nuevo enfoque con respecto a *aleatoriedad de una sucesión de símbolos* (digamos, ceros y unos). Sus trabajos y, también, los esfuerzos de sus seguidores (Chaitin, Levin) llevaron a la creación de la teoría de aleatoriedad algorítmica donde la última se define en los términos de la complejidad del algoritmo (o del programa) que genera (o describe) la sucesión, la aleatoriedad de la cual se analiza. Otra teoría de aleatoriedad (cercana) fue creada por el alumno de Kolmogorov, el científico sueco P. Martin-Löf. No teniendo la posibilidad de describir aquí ni siquiera en la forma más breve la teoría de aleatoriedad algorítmica, nos limitaremos a hacer unas observaciones. Antes que todo en la teoría de Kolmogorov se define la aleatoriedad de una sucesión finita de ceros y unos que contienen n símbolos (para un n arbitrario pero fijo).

Consideremos las tres siguientes sucesiones ($n = 12$):

$$\{0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\} \tag{12}$$

$$\{0\ 1\ 0\ 1\ 0\ 1\ 0\ 1\ 0\ 1\ 0\ 1\} \tag{13}$$

$$\{1\ 0\ 0\ 1\ 0\ 1\ 1\ 1\ 0\ 1\ 0\ 0\} \tag{14}$$

Intuitivamente, vemos que las sucesiones (12) y (13) son “regulares”, mientras que la (14) “se asemeja” a una sucesión aleatoria. Imaginemos un “experimento aleatorio” que consiste en 12 lanzamientos independientes de una moneda simétrica, en los cuales la salida del “águila” se marca con el símbolo “1”, y del “sol” con el símbolo “0”. Entonces las sucesiones en (12) - (14) representan tres posibles resultados de este experimento. Alguien podría decir que los resultados en las (12) y (13) son “poco probables” mientras que el resultado en la (14) “tiene mayor probabilidad”.

Utilizando la axiomática de Kolmogorov (véase la sección 3.1), los resultados del experimento se describen por el espacio de probabilidad $(\Omega, \mathcal{F}, P)$, donde el espacio muestral finito Ω se compone de todos 2^{12} posibles resultados, es decir de todos los arreglos posibles de ceros y unos que forman un vector con 12 componentes. Con esto, la probabilidad de cada vector equivale (por razón de la independencia de los lanzamientos) a $\frac{1}{2} \cdot \frac{1}{2} \cdots \frac{1}{2} = \left(\frac{1}{2}\right)^{12} = 1/4096 \approx 0.00024414$.

Por eso, las probabilidades de los resultados de los lanzamientos de la moneda representadas en (12), (13) y (14) *son iguales* y equivalen a 1/4096. De tal manera, la definición “clásica” de la probabilidad en el “conjunto de todos los resultados posibles del experimento” (es decir, en el espacio muestral Ω) no le atribuye una “mayor o menor aleatoriedad” a *diferentes elementos* $\omega \in \Omega$ (o sea, de diferentes resultados del experimento como, por ejemplo, las presentadas en (12) - (14)).

Por otro lado está claro que para la descripción de las sucesiones en (12) y (13) (o para la descripción de un programa que genera estas sucesiones) son suficientes las cuatro palabras: “repite “0” 12 veces” o “repite “01” 6 veces”. Al mismo tiempo para una descripción de la sucesión (14) son necesarias muchas más palabras, y una cantidad mínima de tales palabras (tamaño mínimo del programa o algoritmo) no se diferencia mucho de la cantidad de los símbolos $n = 12$. Esta particularidad de las sucesiones "no regulares" o "no descriptibles brevemente" fue puesta por Kolmogorov en la base de la definición de aleatoriedad. Simplificando la idea, una sucesión finita se llama aleatoria si la longitud mínima de un programa (algoritmo) que genera esta sucesión es "casi igual" a la longitud de la misma sucesión. En caso contrario la sucesión se interpreta como no aleatoria ("regular", ver (13)). Kolmogorov extendió esta definición de la aleatoriedad algorítmica a las sucesiones de símbolos infinitos. Esta generalización resultó inconsistente porque pronto Martin-Löf demostró que no existen las sucesiones de símbolos infinitas que satisfacen dicha definición.

Unos años después Levin y Chaitin propusieron utilizar en la definición de Kolmogorov una variante modificada de la complejidad. Esto permitió superar la dificultad del paso al límite cuando $n \to \infty$. Finalmente resultó exitosa la teoría de aleatoriedad basada en la complejidad algorítmica.

Retornando a la cuestión general de naturaleza de la aleatoriedad, notemos que probablemente no exista su solución universal. Por ejemplo, el enfoque de complejidad algorítmica no concuerda con el fenómeno de aleatoriedad (o de "caoticidad") en algunos *sistemas dinámicos deterministas*, a veces muy sencillos.

Consideremos, por ejemplo, las famosas ecuaciones recurrentes *logísticas*. Supongamos que el estado inicial $x_0 \in (0, 1)$ es un número irracional dado. Definimos una sucesión de números reales $x_0, x_1, \ldots, x_n, \ldots$ de la siguiente manera

$$x_1 = 4x_0(1-x_0), \quad x_2 = 4x_1(1-x_1), \ldots, x_{n+1} = 4x_n(1-x_n), \ldots . \tag{15}$$

Tomando, por ejemplo, $x_0 = 1/\pi$ y calculando en computadora o calculadora, digamos, los primeros 100 términos de la sucesión $\{x_n,\ n \geq 0\}$ en (15), es fácil descubrir su conducta "extremadamente irregular" y que la sucesión de los números $x_0, x_1, \ldots, x_n, \ldots$ se asemeja a la realización de la trayectoria de un proceso aleatorio. Por ejemplo, para $x_0 = 1/\pi$ una calculadora sencilla proporciona los siguientes valores para los primeros quince términos de la sucesión (redondeados hasta los cinco signos decimales):

n	0	1	2	3	4	5	6
x_n	0.31831;	0.86795;	0.45844;	0.99309;	0.02745;	0.10678;	0.38152
7	8	9	10	11	12	13	14
0.94385;	0.21200;	0.66823;	0.88679;	0.10039;	0.36126;	0.92300;	0.28427

Las trayectorias que presentan tal comportamiento se suelen llamar "caóticas" y, en realidad, ellas satisfacen algunos criterios "estadísticos" que se utilizan para diferenciar las sucesiones “aleatorias” de las “regulares”. Por otra parte, **teóricamente** la trayectoria $x_0, x_1, \ldots, x_n, \ldots$ es **absolutamente predecible**, es decir, conociendo el valor **exacto** del estado inicial de x_0 es posible calcular con precisión el valor de x_n para cualquier número $n \geq 1$.

Es bien conocido que la caoticidad arriba mencionada es debida a la sensibilidad extremadamente alta de los valores x_n en (15) (para n suficientemente grande) con respecto a variaciones infinitesimales del valor inicial x_0. En la práctica aunque se use la mejor computadora en el mundo, se introducirá el valor inicial x_0 (¡un número irracional!) con un error inevitable, posiblemente muy pequeño, pero no nulo (la computadora opera solamente con números racionales). Aunque sea mínimo, el error en x_0 implicará desviaciones grandes en los valores x_n para n suficientemente grande. Es decir, la trayectoria $x_0, x_1, \ldots, x_n, \ldots$ del sistema dinámico (15) es **absolutamente impredecible** (caótica). Por ejemplo, al escoger en vez del estado inicial $x_0 = 1/\pi$, el estado $x_0' = 1/\pi + 0.001$, los cálculos dan la siguiente tabla ($x_{n+1}' = 4x_n'(1 - x_n')$, $n = 0, 1, 2, \ldots$):

n	0	1	2	3	4	5	6
x_n'	0.31931;	0.86940;	0.454162;	0.99160;	0.03334;	0.12890;	0.44913
7	8	9	10	11	12	13	14
0.98965;	0.04097;	0.15717;	0.52987;	0.99643;	0.01422;	0.05607;	0.21170

Comparando las dos tablas presentadas arriba, vemos que desde $n = 8$ los valores x_n y x_n' difieren notoriamente. No hace falta decir que el algoritmo (15) que *genera* $\{x_n,\ n \geq 1\}$ *no es complejo* en ningún sentido.

Los “sistemas complejos” en contextos más amplios, como los económicos o financieros, muestran con frecuencia alta sensibilidad con respecto a cambios pequeños en algunos de los parámetros. Esto puede ser fuente de aleatoriedad (manifestada en “caos”, impredecibilidad, hasta “catástrofes” o “crisis económicas globales”). Por otra parte y bajo ciertas condiciones, en algunos sistemas complejos (no lineales e inequilibrados) que inicialmente manifiestan una “conducta puramente aleatoria” pueden surgir espontáneamente fenómenos de “auto-organización”: la aparición de un movimiento ordenado con alto grado de predecibilidad. Tal autoorganización a veces surge a cuenta del fortalecimiento de algunas *fluctuaciones aleatorias*, o sea, se puede decir que en estos casos la aleatoriedad ocasiona unos procesos ordenados que “matan” la aleatoriedad. Experimentos recientes en biofísica muestran que, probablemente, la generación de formas primarias de la vida fue resultado de estos procesos de auto-organización espontánea.

3.4. Prueba estadística de Kolmogorov-Smirnov

Un resultado puramente matemático, publicado por A. N. Kolmogorov en 1933 (y complementado por N. V. Smirnov en 1939), encontró un amplio uso en la estadística aplicada.

Supongamos que, como resultado de las observaciones estadísticas, se obtuvieron n valores independientes $X_1, X_2, \dots X_n$ de una variable aleatoria X con la *función de distribución F_X continua*:

$$F_X(x) := P(X \leq x\}, \qquad x \in \mathbb{R}. \tag{16}$$

Supongamos que F_X es desconocida para el estadístico. ¿De qué manera utilizando los datos $X_1, X_2, \dots X_n$ es posible verificar la hipótesis estadística de que F_X coincide con cierta función de distribución F dada? Introducimos la *función de distribución empírica* $\widehat{F}_n$ que se define como sigue

$$\widehat{F}_n(x) := \frac{1}{n} \sum_{k=1}^{n} I_{\{X_k \leq x\}}, \quad x \in \mathbb{R}, \tag{17}$$

donde $I_{\{X_k \leq x\}} = \begin{cases} 1 & \text{si} \quad X_k \leq x, \\ 0 & \text{si} \quad X_k > x. \end{cases}$

(Aquí $I_{\{X_k \leq x\}}$ es la indicadora del evento $\{X_k \leq x\}$.) Para cada x, $\widehat{F}_n(x)$ en (17) representa la *frecuencia* de los eventos $\{X_k \leq x\}$ (el número de realizaciones de estos eventos dividido entre el tamaño de la muestra n). De acuerdo con la ley *fuerte* de grandes números la frecuencia $\widehat{F}_n(x)$ converge cuando $n \to \infty$ a la probabilidad del evento $\{X_k \leq x\}$, es decir al número $P(X_k \leq x) = P(X \leq x) = F_X(x)$ (véase (16)). Entonces para cada $x \in \mathbb{R}$, $\widehat{F}_n(x) \to F_X(x)$ con probabilidad 1 cuando $n \to \infty$, (o bien, cuando el tamaño de la muestra $X_1, X_2, \dots, X_n$ crece ilimitadamente).

Se demuestra (el teorema de Glivenko-Cantelli) que la desviación máxima de $\widehat{F}_n$ con respecto a F_X también se acerca al cero con probabilidad 1 cuando $n \to \infty$. (Subrayemos que $\widehat{F}_n$ es una función de distribución aleatoria porque depende de las variables aleatorias $X_1, X_2, \dots, X_n$ (ver (17)). De tal modo tenemos, con probabilidad 1, que

$$D_n := \max_{x \in \mathbb{R}} |\widehat{F}_n(x) - F_X(x)| \to 0, \quad \text{cuando } n \to \infty. \tag{18}$$

La variable aleatoria D_n en (18) se llama la *estadística de Kolmogorov.*

Si la hipótesis $F_X = F$ es cierta, entonces al sustituir en (18) F_X por F obtenemos que los valores de $D_n = \max_{x \in \mathbb{R}} |\widehat{F}_n(x) - F(x)|$ son cercanos al cero para n suficientemente grandes. Si la hipótesis no es cierta, es decir, $F_X \neq F$, entonces la variable aleatoria $\max_{x\in\mathbb{R}} |\widehat{F}_n(x) - F(x)|$ no se acerca al cero a medida que $n \to \infty$. La prueba de *Kolmogorov* usa estas propiedades de D_n, rechazando la hipótesis $F_X = F$ siempre y cuando

$$\max_{x\in\mathbb{R}} |\widehat{F}_n(x) - F(x)| \geq K_\alpha, \tag{19}$$

donde $K_\alpha > 0$ es un *valor crítico* asignado. ¿Cómo se escoge K_α? Está claro que debido al carácter aleatorio de la función de distribución empírica $\widehat{F}_n$ la desigualdad (19) puede cumplirse con una probabilidad positiva aún si la hipótesis $F_X = F$ es verdadera. En este caso la prueba (19) rechaza erróneamente la hipótesis correcta. Usualmente los valores de K_α se escogen a partir de la condición

$$P(\max_{x\in\mathbb{R}} |\widehat{F}_n(x) - F(x)| \geq K_\alpha) \equiv P(D_n \geq K_\alpha) \approx \alpha, \tag{20}$$

donde α es un nivel aceptable del error (por ejemplo, $\alpha = 0.05$). En (20) $\widehat{F}_n$ es la función de distribución empírica definida en (17) bajo la condición de que la hipótesis $F_X = F$ es cierta (es decir, cuando las variables aleatorias $X_1, X_2, \ldots X_n$ tienen la función de distribución F).

El siguiente resultado de Kolmogorov permite calcular aproximadamente (para valores grandes de n) las probabilidades en la parte izquierda de (20), y por lo tanto escoger el valor crítico K_α, que satisface (20).

Para cualquier $z > 0$,

$$\lim_{n\to\infty} P(\sqrt{n} D_n > z) = 2\sum_{j=1}^{\infty} (-1)^{j-1} \exp(-2j^2 z^2). \tag{21}$$

En la demostración de (21) se usa un proceso aleatorio relacionado con el movimiento browniano. El hecho admirable consiste en que la distribución límite en (21) no depende de la función de distribución concreta F (aunque F participa en la definición de D_n en (20)).

La *estadística bimuestral de Kolmogorov-Smirnov*:

$$D_{n,m} := \max_{x\in\mathbb{R}} |\widehat{F}_n(x) - \widehat{G}_m(x)|$$

compara dos funciones de distribución empíricas obtenidas a partir de dos diferentes muestras $X_1, X_2, \ldots X_n$ e $Y_1, \ldots, Y_m$. La estadística $D_{n,m}$ se usa para verificar la hipótesis de que la distribución de las *variables aleatorias* $X_k \quad (1 \leq k \leq n)$ coincide con la distribución de las *variables aleatorias* $Y_k \quad (1 \leq k \leq m)$. Para dicha estadística, Kolmogorov y Smirnov también encontraron una distribución límite universal (cuando $n, m \to \infty$).

El autor agradece a la Dra. Elena Zaitseva y al Dr. Andrey Novikov por su gran ayuda con la redacción del artículo. Un agradecimiento especial a la Dra. Shirley Bromberg por su importante apoyo en el mejoramiento del estilo y la estructuración de este artículo.

Referencias

[1] V. I. Arnold. "Qué son las matemáticas". Edición MCNMO, Moscú, 2004 (en ruso).

[2] S. S. Demidov, V. D. Esakov. "El proceso del académico N. N. Lusin" en la memoria colectiva de la comunidad científica" http://russcience.euro.rn/papers/dees99dl.htm (en ruso).

[3] R.M. Dudley, Universal Donsker classes and metric entropy. Annals of Probability. 15 (1987), 1306-1326.

[4] J. James, Remarkable Mathematics . From Euler to von Neumann. Cambridge University Press, 2002.

[5] A. Kolmogorov, Une série de Fourier-Lebesgue divergente presque partout, Fundamenta Math. 4 (1923), 324-328.

[6] A. N. Kolmogorov, Foundations of the Theory of Probability. Chelsea Publisting Company, New York, 1950.

[7] A. N. Kolmogorov. Sobre la representación de las funciones continuas de varias variables por superposición de las funciones continuas de una variable y la adición. Dokl. Akad. Nank SSSR, 114:5 (1957), 953-956 (en ruso).

[8] A. N. Kolmogorov, V. M. Tikhomirov. ϵ-entropia y ϵ-capacidad de los conjuntos en los espacios funcionales. Uspehy Matematicheskih Nauk, 14:2 (1959), 3-86 (en ruso).

[9] R. Meester, A Natural Introduction to Probability Theory. Birkhauser, 2000.

[10] L. S. Pontryagin. La biografía de Lev Semenovich Pontryagin, matemático, compilada por él mismo. Nacimiento 1908. Moscú,

`http://ega-math.narod.ru/LSP/ch1.htm#c` (en ruso).

[11] A. N. Shiryaev. La vida en la búsqueda de la verdad. Para los 100 años desde el natalicio de Andrey Nikolaevich Kolmogorov. Priroda 4, (2003) (en ruso).

(ver: “Andrei Nikolaevich Kolmogorov. A Biographical Sketch of His Life and Creative Paths” by A. N. Shiryaev, en *Kolmogorov in perspective.* AMS.)

[12] A. W. van der Vaart, J. A. Wellner, Weak Convergence of Empirical Processes. With Applications to Statistics, Springer, 1996.

[13] P. M. Vitanyi. Andrey Nikolaevich Kolmogorov, Scholarpedia, 2 (2): 2798, 2007.

[14] S. S. Demidov. *The Moscow School of the Theory of Functions.* En Golden Years of Moscow Mathematicas, 2.ª ed. Smilka Zdravkovks, Peter L. Duren Edts, AMS-LMS, 2007.

[15] G. Shafer y V. Vovk. *The origins and legacy of Kolmogorov’s Grundbegriffe*

`http://www.probabilityandfinance.com/articles/04.pdf`

`http://www.pdmi.ras.ru/~arnsem/Arnold/arn-papers.html`