

EL CONTROL DE LA INCERTIDUMBRE: EL CÁLCULO DE PROBABILIDADES Y LA TEORÍA DE LA UTILIDAD

FRANCISCO JAVIER GIRÓN Y JOSÉ MIGUEL BERNARDO

Real Academia de Ciencias

INTRODUCCIÓN

Los físicos nos han enseñado a modelar la realidad mediante abstracciones como son las magnitudes físicas tales como la longitud, la velocidad, la masa, etc. y, lo que es más importante, nos han proporcionado métodos y patrones precisos para su medición. Este paso fue fundamental para convertir la Física en una ciencia. Sin medidas precisas no es sencillo describir fenómenos deterministas mediante ecuaciones que expresan la relación entre ciertas magnitudes.

Por otra parte, en el estudio de los fenómenos aleatorios y en la mayoría de los problemas de decisión, aparece de forma natural la incertidumbre. Lo que observamos, y podemos medir, es solo una posibilidad entre muchas —como en el lanzamiento de una moneda— y, de algún modo, necesitamos de una escala que nos represente la verosimilitud de lo realmente observado en relación con el resto. En otras ocasiones, ni siquiera existe la posibilidad de observar la realización de un suceso. Así, si nos preguntamos sobre la posibilidad de que el hombre ponga el pie en el planeta Marte antes del año 2020, éste es un suceso incierto de tipo no repetitivo, pero al que podemos asignar una mayor o menor verosimilitud, de modo puramente personal o subjetivo y que dependerá, en gran medida, de lo versado que esté uno en el tema. Un experto de la N.A.S.A. daría una respuesta muy distinta que la que nos ofrecería una persona ajena al tema.

Una de las escalas para medir la incertidumbre, quizás la más utilizada, es la probabilidad. Esta escala debe servir, igualmente, para medir probabilidades asociadas tanto a los sucesos de tipo repetitivo como a los de carácter no repetitivo, como lo son, a poco que reflexionemos, la mayoría de los sucesos que ocurren a lo largo de nuestras vidas, y de los que dependen nuestras decisiones más importantes.

¿Por qué se debe medir la incertidumbre en términos de probabilidad y no de alguna otra medida alternativa, como pudieran ser las funciones de credibilidad o la lógica difusa?

La pregunta anterior afecta a los fundamentos mismos de la inferencia inductiva, de la estadística y de la teoría de la decisión. Mucho esfuerzo se ha dedicado al tema de los fundamentos de estas disciplinas y a la justificación de que las medidas de la incertidumbre deben expresarse en términos de cantidades que obedezcan a las reglas del cálculo de probabilidades. La mayor parte de los desarrollos axiomáticos de la teoría de la decisión y de la inferencia estadística concluyen con la existencia de una medida de probabilidad sobre el álgebra de los sucesos inciertos, que representa el grado de credibilidad que un individuo determinado asigna a esos sucesos.

Lo más interesante de este tipo de resultados es que cada individuo, dependiendo de la información que tenga sobre los sucesos inciertos, cuantifica su incertidumbre con una medida de probabilidad personal o subjetiva. Este hecho es el que diferencia a la probabilidad de otras magnitudes: ¡que es personal! Nos explicaremos: si a dos individuos distintos se les pide que estimen la longitud de una determinada barra y, para ello, se les da la misma cinta métrica, ambos acabarán, en condiciones experimentales similares, con el mismo número, dentro de un margen de error presumiblemente pequeño. Sin embargo, en el caso de medir la incertidumbre de un mismo suceso, dos sujetos pueden producir respuestas muy distintas porque sus instrumentos de medida —*sus cintas métricas*— son distintas. La explicación de esta aparente paradoja es que, en el caso de medir la incertidumbre, el instrumento de medición se ve afectado por la distinta información que cada uno de ellos tiene. Si ambas personas tuviesen exactamente la misma información, la probabilidad que asignarían al suceso incierto sería exactamente la misma. La conclusión que se sigue es que *todas las probabilidades son siempre condicionadas a un cierto estado de información*. No hay, por consiguiente, probabilidades absolutas.

De modo que las probabilidades que una persona asigna a los sucesos inciertos, además de cumplir el requisito de ser compatibles con las reglas del cálculo de probabilidades, son siempre individuales o subjetivas. Esto no es óbice para que, en ciertas circunstancias, la asignación de probabilidades a sucesos inciertos sea prácticamente la misma por parte de cualquier individuo, como ocurre, p. ej., en juegos de azar, como la ruleta, o en situaciones, como suele suceder en las ciencias experimentales, en las que la información compartida constituye casi toda la información disponible.

Por otra parte, la forma en que medimos el otro ingrediente esencial en los problemas de toma de decisiones, las consecuencias de nuestras acciones que medimos mediante la denominada función de utilidad, es también de carácter subjetivo. Lo que, por otra parte, es evidente dado el distinto comportamiento que tenemos las personas ante situaciones inciertas idénticas.

La conclusión que puede extraerse de los argumentos precedentes es que para que la toma de decisiones y, en consecuencia, la inferencia estadística (al menos, la bayesiana) se haya convertido en una ciencia, que hoy conocemos como *Teoría de la decisión bayesiana* ha sido necesario aplicar las mismas ideas que se han aplicado a otras disciplinas para convertirlas en ramas de la ciencia; a saber, construir un sistema axiomático sobre sus principios y cuantificar los ingredientes que la constituyen.

Los que arguyen que la teoría de la decisión no es una disciplina científica análoga a otras bien establecidas, basados bien en argumentos de que ni la incertidumbre ni las preferencias de los individuos pueden cuantificarse de modo racional, o bien en que el comportamiento *real* de las personas no se ajuste a las teorías normativas que surgen de la teoría de la decisión, desconocen el argumento anterior.

La conclusión es clara: no debemos tomar como modelo de la teoría de la decisión el comportamiento real de las personas, primero porque la mente humana no está capacitada para procesar información sin la ayuda del cálculo de probabilidades y, segundo, porque el comportamiento humano muestra muchas incoherencias —incluso en situaciones simples, como se ha demostrado en repetidos experimentos— que deben excluirse de cualquier teoría *normativa* de la decisión.

La alternativa parece ser partir de una teoría, tal como la que ofrece el paradigma de la *maximización de la esperanza de utilidad subjetiva*, que puede y debe ser mejorable con el tiempo, como demuestran los recientes avances realizados en el campo de los fundamentos la teoría de la decisión, que se deduzca de unos principios básicos que conciten el consenso de una mayoría, al menos en el terreno de las decisiones individuales.

En las páginas que siguen queremos ilustrar, con ejemplos sencillos tomados de la realidad cotidiana, el modo de enfrentarse a situaciones inciertas utilizando el paradigma descrito.

LA PROBABILIDAD COMO MEDIDA DE LA INCERTIDUMBRE

No se puede dar una definición precisa de lo que es la Incertidumbre, del mismo modo que no se puede definir, por ejemplo en Geometría, lo que se entiende por un punto o una recta. Lo importante para construir una teoría matemática es establecer los axiomas que relacionen estos conceptos sin necesidad de definirlos. Como hemos comentado en la introducción, las relaciones entre las medidas que se asignan a los sucesos inciertos deben hacerse de acuerdo con los axiomas del cálculo de probabilidades, lo que además permite asimilar los sucesos ciertos o seguros con aquellos que tienen probabilidad uno.

Los sucesos inciertos son ciertos subconjuntos de un *espacio paramétrico* Ω que pertenecen a una cierta σ -álgebra \mathcal{B}_Ω y la incertidumbre que inicialmente asignamos a estos sucesos mediante la llamada *distribución a priori* es una medida de probabilidad $P(\theta)$ sobre el espacio probabilístico $(\Omega, \mathcal{B}_\Omega)$. En lo que sigue supondremos, sin pérdida de generalidad, que la distribución a priori $P(\theta)$ es absolutamente continua respecto de una medida $\mu(\theta)$ σ -finita definida sobre $(\Omega, \mathcal{B}_\Omega)$, de modo que la distribución a priori queda perfectamente caracterizada por la función de densidad $p(\theta)$, que es una versión de la derivada de Radon-Nikodym de $P(\theta)$ respecto de $\mu(\theta)$.

Uno de los problemas básicos de la inferencia o estadística inductiva es el de *reducir* la incertidumbre inicial dada por $P(\theta)$ mediante razonamientos deductivos o mediante la realización y observación de experimentos estadísticos que nos proporcionen información sobre θ , en sentido amplio. Si el experimento es concluyente, como sucede por ejemplo cuando se recurre a una demostración matemática o se examinan todos los elementos de una población finita, la incertidumbre se reduce totalmente y llegamos a la certeza.

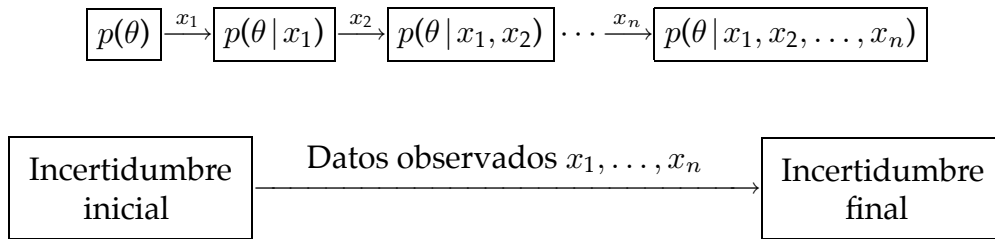
Básicamente, un experimento estadístico es un proceso de reducción o disminución de la incertidumbre que permite el tránsito de la incertidumbre a la certeza, aunque ésta solamente se alcance de manera asintótica.

Las observaciones son subconjuntos de un *espacio muestral* \mathcal{X} , que pertenecen a una cierta σ -álgebra $\mathcal{B}_\mathcal{X}$, que supuestamente nos permiten reducir la incertidumbre sobre los sucesos inciertos.

Con más precisión, un experimento estadístico es un modelo probabilístico, representado por una medida de probabilidad condicionada $P(x|\theta)$ sobre el espacio $(\mathcal{X}, \mathcal{B}_\mathcal{X})$. Supondremos que la familia de probabilidades condicionada $\{P(x|\theta); \theta \in \Omega\}$ está dominada por una medida σ -finita $\nu(x)$ sobre $(\mathcal{X}, \mathcal{B}_\mathcal{X})$, de modo que el experimento queda

caracterizado por la función de densidad condicionada $f(x|\theta)$, que relaciona lo *incierto* θ con lo *observable* x .

El proceso de aprendizaje, o reducción secuencial de la incertidumbre inicial, $p(\theta)$, a medida que se van observando los datos x_1, x_2, \dots, x_n , se realiza mediante la aplicación sucesiva del teorema de Bayes, tal como se muestra en el diagrama siguiente:



y donde la incertidumbre residual sobre el parámetro θ en la etapa i -ésima, una vez observados los datos x_1, x_2, \dots, x_i , viene dada por la *densidad a posteriori* $p(\theta|x_1, x_2, \dots, x_i)$ que se calcula por el teorema de Bayes mediante la fórmula

$$p(\theta|x_1, x_2, \dots, x_i) \propto p(\theta|x_1, x_2, \dots, x_{i-1})f(x_i|\theta, x_1, x_2, \dots, x_{i-1}),$$

donde la constante de proporcionalidad se determina con la condición de que la función $p(\theta|x_1, x_2, \dots, x_i)$ sea una función de densidad, es decir integre la unidad, con respecto de la medida $\mu(\theta)$.

Si las observaciones x_1, x_2, \dots, x_n se pueden suponer (condicionalmente) independientes, entonces la fórmula anterior se simplifica, resultando

$$p(\theta|x_1, x_2, \dots, x_i) \propto p(\theta|x_1, x_2, \dots, x_{i-1})f(x_i|\theta).$$

El proceso de aprendizaje que acabamos de describir es lo que básicamente se conoce como *inferencia bayesiana*, y su principal máxima es, como acabamos de señalar, que toda la información referente al estado de nuestra incertidumbre tras haber observado la muestra x_1, x_2, \dots, x_i del experimento estadístico viene descrita o representada por la densidad a posteriori $p(\theta|x_1, x_2, \dots, x_i)$.

Si la distribución a posteriori es degenerada en un punto, digamos θ_0 , entonces la incertidumbre ha desaparecido y tenemos certeza (con probabilidad 1)¹.

Incertidumbre en situaciones deterministas

Muchas situaciones puramente deterministas, presentan incertidumbre, precisamente por no tener información suficiente o por no disponer de medios de cálculo adecuados que suplan nuestra incapacidad para procesar datos. La valoración de dicha incertidumbre no es fácil y, en muchas ocasiones, la intuición resulta ser contradictoria con la lógica. De aquí la necesidad de recurrir, cuando se pueda, a las matemáticas o a los argumentos rigurosos para reducir, e incluso eliminar, la incertidumbre inherente a ciertas situaciones deterministas en las que está presente.

¹ El lector interesado en una descripción elemental más detallada del paradigma bayesiano puede consultar Bernardo (2001).

El ejemplo siguiente, adaptación del problema que apareció por vez primera en el libro de Acton (1970), es sorprendente pues parece contradecir la intuición de la mayoría de las personas, incluidos los matemáticos.

Ejemplo 1. Se tienen dos varillas de 1 m de longitud. A una de ellas se le añade 1 cm y se curva de modo que sus extremos coincidan con los de la otra varilla y forme un arco de circunferencia (véase la figura 1). ¿Cuál es la deflexión máxima d entre ambos?

El problema es puramente determinista; no hay nada de incierto en su solución, que se obtiene resolviendo un sistema *no lineal* de tres ecuaciones con tres incógnitas, d , r y θ , pues $y = r - d$, al que se llega por sencillos razonamientos geométricos y trigonométricos. Pero si no dispusiésemos de métodos de cálculo, y nos pidiesen una estimación subjetiva de dicha distancia, ¿en cuanto estimaríamos el valor de d ?

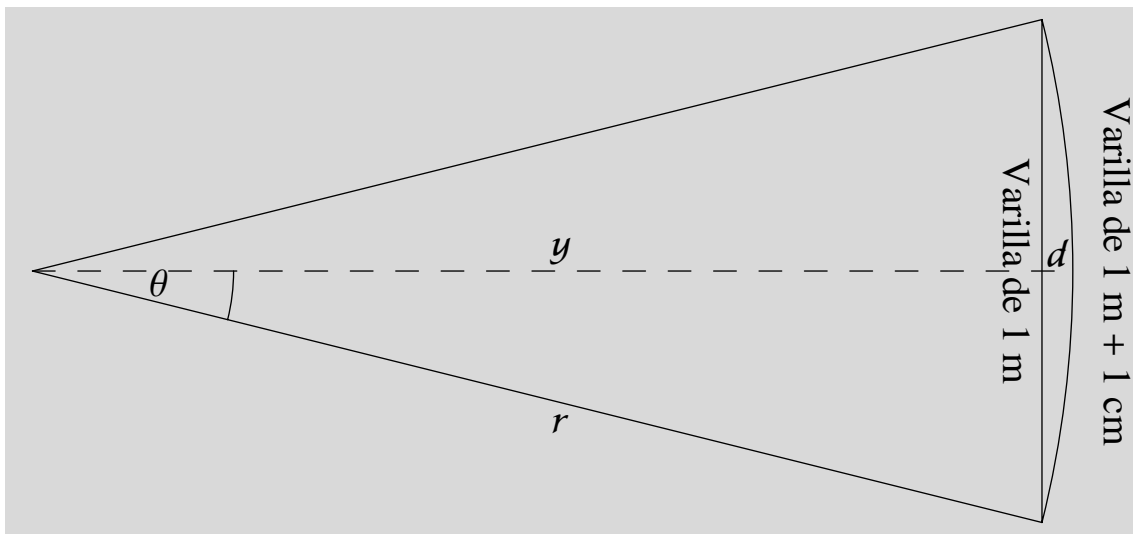


Figura 1. Representación gráfica del problema de las varillas.

La mayoría de las personas suele estimar el valor de d en menos de 1 centímetro. Sin embargo, la respuesta correcta es $d = 6,13$ centímetros, $r = 206,89$ centímetros y el ángulo $\theta = 14^\circ 0' 13''$.

Si en vez de 1 cm solamente se añadiese 1 mm, la respuesta sería aún más sorprendente, pues $d = 1,94$ centímetros, $r = 646,37$ centímetros y el ángulo $\theta = 4^\circ 26' 12''$.

El ejemplo 2 muestra un comportamiento habitual en las personas, incluidos los expertos en cálculo de probabilidades, que generalmente tienden a infravalorar o sobrevalorar probabilidades si las asignamos de modo intuitivo —como se hace con frecuencia en el diagnóstico médico— en vez de computarlas de acuerdo con las reglas del cálculo de probabilidades.

Ejemplo 2. Se tienen dos urnas con 100 bolas cada una. La primera, U_1 , contiene 70 bolas rojas y 30 azules y la segunda, U_2 , 30 bolas rojas y 70 azules. Se extraen sin reemplazamiento 10 bolas de una urna elegida al azar (sin saber de cual se trata) y resultan 7 rojas y 3 azules. ¿Cuántas veces es más probable que la urna elegida sea la primera que la segunda?

Está claro que es más probable que la urna elegida sea la primera que la segunda, pero determinar cuántas veces es más probable que sea la primera que la segunda no es inmediato. La respuesta correcta es 43,67 veces, que es equivalente a que la probabilidad de que la urna sea la U_1 es 0,978 y la de que sea la U_2 es 0,022, por lo que podemos estar *muy seguros* de que la urna es la U_1 . Sin embargo, la mayoría de las personas a las que se les plantea el problema anterior, dan como respuesta un número inferior a 10.

Ejemplo 3: Paridad de los dígitos del número π . ¿Cuál es la probabilidad de que el dígito que ocupa la posición p de la parte decimal de π sea par?

Este ejemplo, sobre la presencia de la incertidumbre en una situación determinista, se ha comentado con detalle en Girón (2001). Aquí solamente nos limitamos a unos breves comentarios, ilustrados con las figuras 2 y 3.

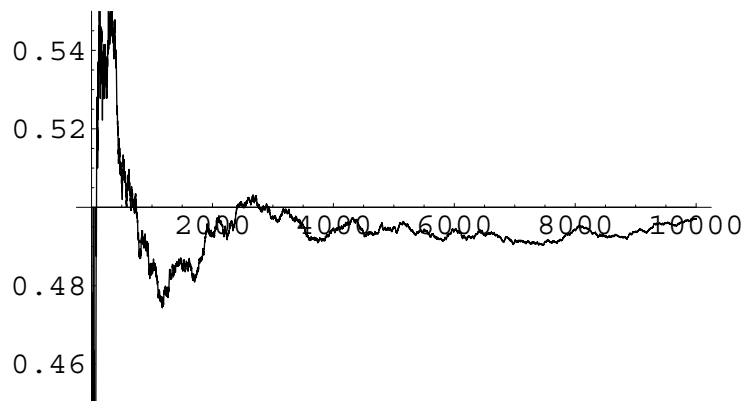


Figura 2. Evolución secuencial de la proporción empírica de los dígitos pares correspondiente a los 10.000 primeros dígitos de la parte decimal de π .

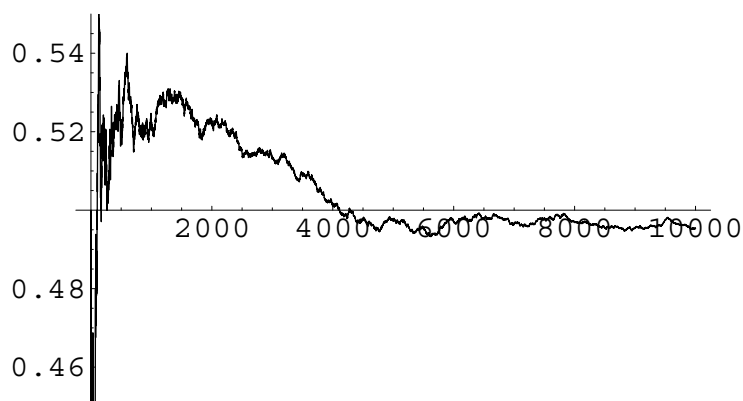


Figura 3. Evolución secuencial de la proporción empírica de éxitos de una muestra aleatoria de tamaño 10.000 de un modelo de Bernoulli con probabilidad de éxito 0,5.

Parece claro que a medida que p aumenta, y nos alejamos en el desarrollo de la parte decimal de π , nuestra incertidumbre sobre la la paridad de los dígitos aumenta, si no disponemos de métodos cálculo adecuados; de hecho, si p es demasiado grande probablemente nunca podamos saber cual es el dígito correpondiente. Pero, por otra parte, la proporción de los dígitos pares o impares, no solo parece que tiende a estabilizarse sino que lo hace en torno al valor 0,5 tal como muestra la figura 2 para los 10.000 primeros dígitos. Además, la sucesión de las frecuencias empíricas de los dígitos pares presenta un aspecto similar al de la figura 3 obtenida por simulación de una muestra de tamaño 10.000 de una distribución de Bernoulli de parámetro 0,5.

No solamente se estabilizan las frecuencias sino que la sucesión de los dígitos pasa las pruebas habituales de aleatoriedad e independencia estocástica.

Si en vez de analizar los 10.000 primeros dígitos se analizan los 100.000 primeros, el comportamiento de estos aparece como puramente aleatorio, con distribución uniforme, y lo mismo se puede concluir sobre el desarrollo decimal del número transcendente e .

El comportamiento límite de la sucesión de dígitos no se conoce, pero el afirmar que la probabilidad de que el dígito que ocupa la posición p -sima de la parte decimal de π sea par, cuando p es grande, es aproximadamente $1/2$ parece una respuesta totalmente razonable.

EL PROCESO DE APRENDIZAJE: EL TEOREMA DE BAYES

El ejemplo siguiente muestra una aplicación sencilla del teorema de Bayes en un caso de diagnóstico médico, que pone de manifiesto el importante papel que juega la probabilidad a priori a la hora de emitir un diagnóstico.

Diagnóstico probabilístico de una enfermedad

A un grupo de personas con trastornos respiratorios se les aplicó un test de tuberculosis para determinar si padecían dicha enfermedad. Del test se conoce, por estudios previos, su *sensibilidad* y su *especificidad*, dadas por:

- *Sensibilidad* = $\Pr\{+ | \text{Enfermo}\} = 0,98$,
- *Especificidad* = $\Pr\{- | \text{Sano}\} = 0,95$,

de modo que la probabilidad de falsos positivos es $= \Pr\{+ | \text{Sano}\} = 1 - 0,95 = 0,05$, y la de falsos negativos es $= \Pr\{- | \text{Enfermo}\} = 1 - 0,98 = 0,02$.

Los sucesos inciertos son $\theta_1 = \text{Enfermo}$ y $\theta_2 = \text{Sano}$, mientras que las observaciones son los resultados del test $x_1 = \text{Positivo (+)}$ y $x_2 = \text{Negativo (-)}$. Las probabilidades condicionadas del experimento se resumen en la tabla siguiente:

	Positivo (+)	Negativo (-)
Enfermo	0,98	0,02
Sano	0,05	0,95

Si e resultado del test, aplicado a una persona de ese grupo de riesgo, fue Positivo (+): ¿Cuál es la *evidencia* a favor de que esa determinada persona estuviese realmente enferma?

La solución correcta, es decir el diagnóstico probabilístico, lo da la probabilidad de que la persona esté enferma *condicionada* a que el resultado del test es positivo, es decir, la cantidad $\Pr(\text{Enfermo} | +)$.

Muchas personas, desconocedoras de las reglas del cálculo de probabilidades, cometen un grave error —lo que en el contexto legal se conoce como *paradoja del fiscal*—, que consiste en confundir, y por consiguiente igualar, la probabilidad condicionada de un suceso A supuesto que B ha ocurrido con la probabilidad condicionada de un suceso B supuesto que A ha ocurrido. En nuestro caso esto conduciría a que

$$\Pr(\text{Enfermo} | +) = \Pr(+ | \text{Enfermo}) = 0,98.$$

Otro modo de razonar, propio de la estadística clásica, es considerar —puesto que el suceso $+$ se ha observado— el cociente o razón de verosimilitudes, también llamado factor de Bayes

$$\frac{\Pr(+ | \text{Enfermo})}{\Pr(+ | \text{Sano})} = \frac{0,98}{0,05} = 19,6;$$

Si este cociente es significativamente mayor que 1, como ocurre en nuestro ejemplo, se diagnostica que la persona está enferma.

La solución, es decir el diagnóstico probabilístico correcto resulta de aplicar el teorema de Bayes, para lo cual hace falta un dato adicional que es la probabilidad a priori de que una determinada persona del grupo de riesgo tenga la enfermedad, es decir $\Pr(\text{Enfermo})$. En medicina a esta cantidad se la conoce con el nombre de *prevalencia* de la enfermedad.

De estudios epidemiológicos se sabe que la *prevalencia* de la enfermedad en el grupo al cual pertenecía esa persona es aproximadamente $\Pr(\text{Enfermo}) = 0,01$.

La probabilidad de que la persona esté tuberculosa condicionada a que el test ha dado positivo se calcula por el teorema de Bayes como sigue:

$$\Pr(\text{Enfermo} | +) = \frac{\Pr(\text{Enfermo}) \Pr(+ | \text{Enfermo})}{\Pr(\text{Enfermo}) \Pr(+ | \text{Enfermo}) + \Pr(\text{Sano}) \Pr(+ | \text{Sano})} = 0,1653.$$

La conclusión, que aparentemente resulta chocante, es que del colectivo de personas del grupo de riesgo solamente el 16,53 % de las personas que diesen positivo estarían realmente enfermas: la mayoría de las que diesen positivo (el 83,47%) resultarían ser falsos positivos.

La tabla siguiente da las probabilidades a posteriori de padecer o no tuberculosis condicionadas a que el resultado del test haya sido positivo o negativo.

	Enfermo	Sano
Positivo (+)	0,1653	0,8347
Negativo (–)	0,0002	0,9998

REDUCCIÓN SECUENCIAL DE LA INCERTIDUMBRE

Una de las ventajas de medir la incertidumbre mediante una medida de probabilidad es la de poder disponer de todas las herramientas y resultados del cálculo de probabilidades, con lo que, desde la perspectiva bayesiana, la estadística no es sino una aplicación del cálculo de probabilidades.

Predicción de resultados electorales

Los problemas de predicción son desde el punto de vista práctico uno de los problemas más interesantes e importantes de la estadística. La predicción del comportamiento de magnitudes futuras que todavía no se han observado no solo es importante *per se*, sino que además sirve para validar el modelo estadístico que se emplea para predecir.

Desde el punto de vista de la inferencia bayesiana, la predicción se realiza calculando la densidad a posteriori de la magnitud futura condicionada a los datos ya observados.

Así, si estamos interesados en predecir la magnitud futura $T(x_{n+1}, x_{n+2}, \dots, x_{n+m})$ habiendo observado $\mathbf{x} = (x_1, x_2, \dots, x_n)$, la solución viene dada por la fórmula siguiente:

$$f(t|\mathbf{x}) = \int_{\Omega} f(t|\theta)p(\theta|\mathbf{x}) d\mu(\theta),$$

donde $f(t|\theta)$ representa la densidad del estadístico $T(x_{n+1}, x_{n+2}, \dots, x_{n+m})$ condicionada al parámetro θ .

La densidad $f(t|\mathbf{x})$, llamada densidad predictiva, contiene toda la incertidumbre referente a la magnitud futura $T(x_{n+1}, x_{n+2}, \dots, x_{n+m})$.

La predicción de resultados electorales es una de las aplicaciones más llamativas del cálculo de probabilidades. En el ejemplo que sigue consideramos una versión simplificada que consiste en escrutar los votos de una urna cuando hay varios candidatos y se tiene en cuenta además la posibilidad de que haya votos en blanco y nulos.

No estamos interesados en predecir el porcentaje de votos obtenidos por cada uno de los candidatos, sino el problema mucho más interesante de predecir secuencialmente la probabilidad de que cada uno de ellos obtenga una cierta mayoría, a medida que se van escrutando los votos de una urna.

Este problema no es posible abordarlo desde el punto de vista de la estadística clásica pues no se trata de un simple problema de estimación secuencial, ya que al ser el problema discreto —las posibles configuraciones de la urna son finitas—, el espacio paramétrico también lo es y la inferencia clásica únicamente considera espacios paramétricos continuos. Además el cálculo de las probabilidades predictivas implica *necesariamente* la especificación de unas probabilidades a priori, que son ajenas a la estadística clásica.

Incluso desde el punto de vista bayesiano el cálculo de las probabilidades predictivas es complejo debido al número enorme de posibles configuraciones de la urna, sobre todo cuando hay muchos votos y varios candidatos, por lo que en estas ocasiones se hace imprescindible recurrir a los métodos de Monte Carlo.

Elecciones de Académicos en la Real Academia de Ciencias

Según los estatutos de la Real Academia de Ciencias, para ser elegido académico en la primera votación es necesario obtener al menos un número de votos favorables igual o superior a los dos tercios de los académicos con derecho a voto. En la segunda votación, si ha lugar, se ha de obtener al menos un número de votos favorables igual o superior a los dos tercios del número de académicos presentes en la votación y, en la tercera votación, si ha lugar, solo es necesario obtener la mayoría absoluta entre los académicos presentes.

En el supuesto que sigue hay 41 académicos con derecho a voto, de los cuales 37 estuvieron presentes en las tres votaciones y hay dos candidatos C_1 y C_2 . Los votos en blanco o nulos se representan por B y los votos emitidos por $V.E.$

El número mínimo de votos favorables necesario en cada una de las tres votaciones es por consiguiente: 28 votos en la primera vuelta, 25 en la segunda y 19 en la tercera.

La tabla siguiente muestra cual fue el desarrollo de la votación, en la que el resultado final aparece en negrita.

$V.E.$	C_1	C_2	B	$V.E.$	C_1	C_2	B
1	0	1	0	20	13	5	2
2	0	2	0	21	14	5	2
3	1	2	0	22	14	6	2
4	1	2	1	23	15	6	2
5	2	2	1	24	16	6	2
6	3	2	1	25	17	6	2
7	3	3	1	26	18	6	2
8	4	3	1	27	18	7	2
9	5	3	1	28	19	7	2
10	6	3	1	29	19	7	3
11	7	3	1	30	20	7	3
12	8	3	1	31	21	7	3
13	8	4	1	32	22	7	3
14	8	5	1	33	23	7	3
15	9	5	1	34	24	7	3
16	10	5	1	35	24	8	3
17	11	5	1	36	25	8	3
18	12	5	1	37	26	8	3
19	13	5	1				

Tabla 1. Desarrollo secuencial de la votación.

En las figuras 4 y 5 se muestra el comportamiento, es decir la variación secuencial, de las probabilidades predictivas de que el primer y segundo candidatos obtengan la plaza de Académico en cada una de las votaciones, respectivamente.

Las predicciones futuras se hacen condicionadas a la información obtenida en cada momento del escrutinio, y reflejan con precisión los cambios cuantitativos que se van produciendo a lo largo del mismo.

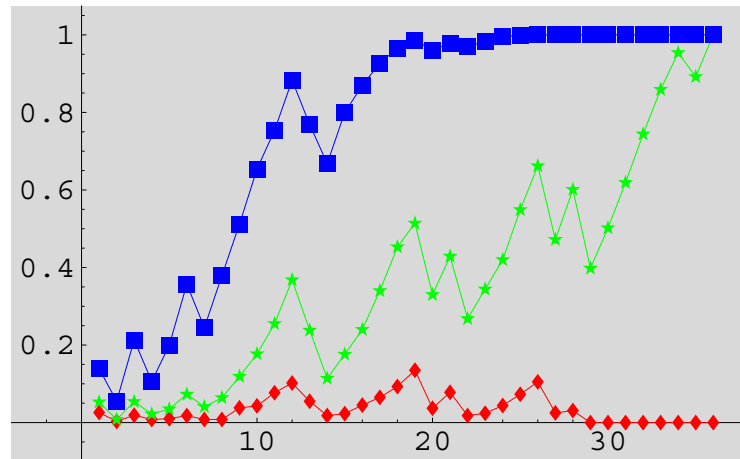


Figura 4. Variación secuencial de las probabilidades de que el primer candidato obtenga la plaza de Académico en primera (♦), segunda (★) y tercera votación (■).

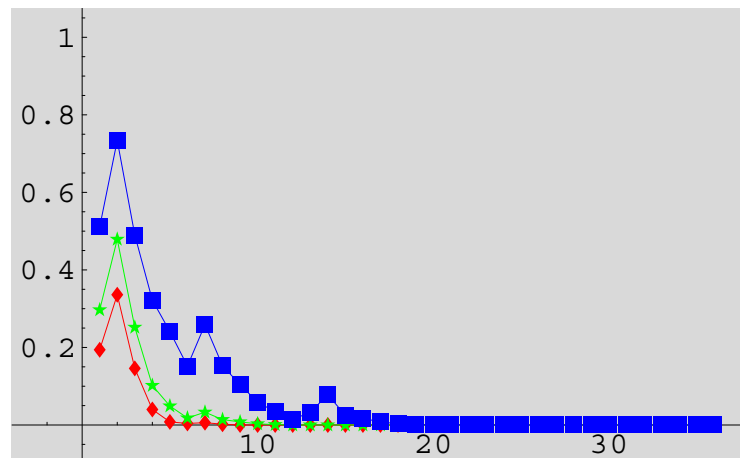


Figura 5. Variación secuencial de las probabilidades de que el segundo candidato obtenga la plaza de Académico en primera (♦), segunda (★) y tercera votación (■).

ESTIMACIÓN PROBABILÍSTICA DE MAGNITUDES CONTINUAS

Los dos ejemplos siguientes muestran cómo se estiman parámetros o magnitudes continuos, desde la perspectiva bayesiana. La solución es, como ya hemos comentado, la densidad a posteriori del parámetro, que en el caso unidimensional puede representarse gráficamente y de la que se puede extraer información precisa sobre el parámetro.

Por razones de espacio no podemos detallar el proceso de obtención de las densidades a posteriori, a partir del modelo estadístico correspondiente. En ambos ejemplos se supone que la información inicial sobre el parámetro es mínima o, de modo equivalente, la incertidumbre inicial es máxima.

Determinación experimental de la *aceleración normal* g_0

El conjunto de 10 medidas independientes, en m/s^2 , del valor en el laboratorio de la aceleración normal de la gravedad g_0 ,

$$D = \{9,78 \ 9,84 \ 9,97 \ 10,02 \ 9,75 \ 9,68 \ 9,84 \ 9,78 \ 9,98 \ 9,73\}$$

se obtuvo en unas prácticas de la asignatura de mecánica realizadas con un péndulo.

Con esos datos, ¿qué podemos afirmar sobre el valor de la magnitud g_0 ?

El modelo estándar para este tipo de datos experimentales es el *modelo normal*, de modo que suponemos que los datos x_1, \dots, x_{10} son observaciones independientes que siguen el modelo probabilístico normal

$$x_i \sim N(g_0, \sigma^2)$$

donde g_0 es el parámetro que queremos estimar y σ^2 es otro parámetro subsidiario, que no interesa estimar en este ejemplo, y que representaría la variabilidad del instrumento de medida, en este caso las observaciones realizadas con el péndulo.

Toda la información sobre g_0 está contenida en la gráfica de la densidad a posteriori de g_0 condicionada a los datos D , que muestra de forma clara la incertidumbre residual sobre g_0 .

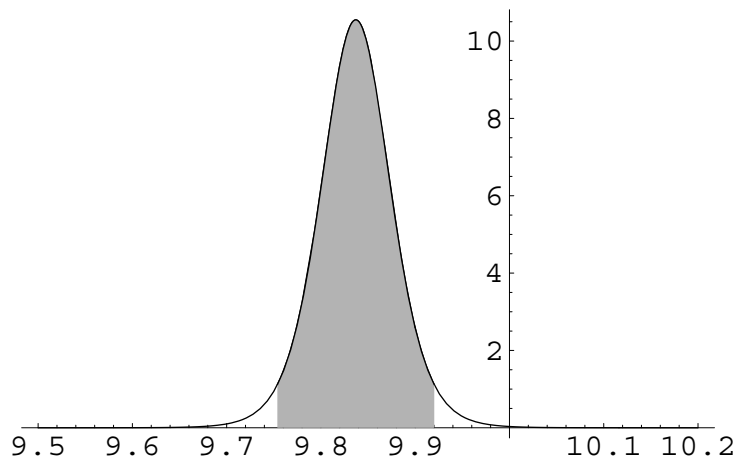


Figura 6. Conclusiones sobre el valor de la aceleración de la gravedad g_0 basadas en el modelo normal y en los resultados experimentales.

De la densidad a posteriori se deduce que estimación puntual de g_0 es la media muestral $\bar{x} = 9,837$, que en este caso además coincide con la moda y mediana de la densidad a posteriori, debido a la simetría de ésta.

De la gráfica se deduce que valores de la aceleración normal menores que 9,7 o mayores que 10 son muy improbables.

De la densidad a posteriori —véase la figura 6— también se pueden obtener afirmaciones probabilísticas del tipo

$$\Pr((9,75 < g_0 < 9,92) | D) = 0,95$$

que son análogas a las que se obtienen con los intervalos de confianza de la estadística clásica.

Es interesante señalar que aunque a g_0 se la considere una *constante física*, de hecho es una magnitud aleatoria que se conoce con mayor o menor precisión según las condiciones experimentales y que además varía según la latitud y la altura sobre el nivel del mar.

Determinación de la proporción θ de individuos seropositivos

Los resultados $D = \{-, -, \dots, -\}$, de los tests de inmuno-deficiencia realizados a 200 estudiantes universitarios elegidos al azar en la Universidad de Valencia, han sido felizmente negativos para todos ellos.

¿Qué podemos afirmar de la proporción θ de individuos seropositivos en el colectivo de todos los estudiantes universitarios de esa universidad?

El estimador intuitivo de θ , que es el estimador de máxima verosimilitud, sería 0, que no es admisible pues no tendría sentido afirmar que en la población universitaria no hay ningún seropositivo, por el hecho de no haber observado ninguno en la muestra.

Como en el ejemplo anterior, toda la información sobre θ está contenida en la densidad a posteriori de θ condicionada a los datos D , véase la figura 7.

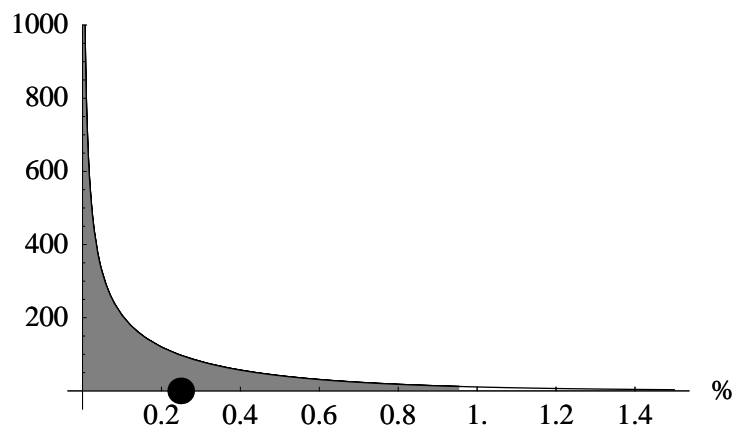


Figura 7. Conclusiones sobre la proporción θ de seropositivos entre los estudiantes basadas en el modelo binomial y en los resultados experimentales.

La densidad a posteriori en este caso es muy asimétrica pues, como cabe esperar, refleja la asimetría de los datos: *todos* los resultados fueron negativos.

La estimación puntual basada en la media de la distribución a posteriori (señalada con un punto) es 0,25%, mientras que la basada en la mediana es 0,11%, valores muy pequeños pero no nulos.

La afirmación probabilística

$$\Pr(0\% < \theta < 0,9546\% | D) = 0,95$$

mide la incertidumbre residual sobre θ , después de analizar los datos, con una probabilidad alta, en este caso de 0,95, con lo que podríamos concluir que la probabilidad de que el porcentaje de infectados en la población de estudiantes sea menor que el 1% es muy alta.

PREDICCIÓN PROBABILÍSTICA

El ejemplo siguiente, un problema de predicción relativamente sencillo, se resuelve calculando las distribuciones predictivas correspondientes a tres instantes o intervalos de tiempo futuro. La solución del problema pone de manifiesto la ventaja de trabajar en términos probabilísticos, calculando las respectivas distribuciones predictivas, en lugar de dar, como es habitual, un estimador puntual o numérico de la predicción (generalmente el valor esperado) que no solamente es insuficiente —como lo es toda estimación puntual— sino que puede ser ridículo.

En cambio, la predicción probabilística muestra *toda* la información relevante al problema de predicción.

Previsión del número de accidentes de carretera

En un tramo recién inaugurado de una carretera comarcal no se ha registrado ningún accidente durante cinco semanas consecutivas, es decir, se ha observado el vector $D = \{0, 0, 0, 0, 0\}$.

¿Qué podemos afirmar, es decir predecir, sobre el número de accidentes que se va a producir la semana próxima, o en las cinco o diez semanas próximas, suponiendo que no hay variaciones sustanciales en el tráfico como, p. ej., las debidas a puentes, fiestas nacionales o autonómicas, etc.?

Un modelo probabilístico adecuado para describir este tipo de situaciones es el de Poisson homogéneo en el tiempo. Así que supondremos que el número de accidentes semanales x_1, \dots, x_5 , y el de futuros accidentes $x_6, \dots, x_{15}, \dots$ siguen un modelo de Poisson de parámetro λ , es decir

$$\Pr(x_i | \lambda) = \frac{\lambda^{x_i} e^{-\lambda}}{x_i!} \quad i = 1, \dots, 5, 6, \dots, 15, \dots$$

Como en el ejemplo anterior, el estimador intuitivo, que es el estimador de máxima verosimilitud, del parámetro λ es 0, que es inadmisibles: sería obviamente imprudente afirmar, esto es predecir, que en ese tramo nunca más habrá accidentes.

No detallamos la expresión, ni la gráfica, de la densidad a posteriori de λ , obtenida bajo la hipótesis de mínima información, o máxima incertidumbre, sobre el parámetro, pero mostramos en cambio, en la tabla 2, las tres distribuciones predictivas del número de accidentes correspondientes a los tres futuros períodos de tiempo especificado, a saber: una semana, cinco semanas y diez semanas.

Nº. acc.	1 semana	5 semanas	10 semanas
0	0.912871	0.707107	0.577350
1	0.076073	0.176777	0.192450
2	0.009509	0.066291	0.096225
3	0.001321	0.027621	0.053458
4	0.000193	0.012084	0.031184
5	0.000029	0.005438	0.018710
6	0.000004	0.002492	0.011434
7	0.000001	0.001157	0.007078

Tabla 2. Probabilidades predictivas del número de accidentes que ocurrirán en los tres períodos de tiempo especificados.

Las *predicciones puntuales*, basadas en la media de las distribuciones predictivas del número de accidentes para la semana siguiente, para el conjunto de las cinco semanas siguientes y para el conjunto de las diez semanas siguientes son $t_1 = 0,1$, $t_5 = 0,5$ y $t_{10} = 1,0$, respectivamente. Las modas y las medianas de las tres distribuciones predictivas tampoco son muy útiles como estimadores puntuales pues son todas igual a 0.

Estas estimaciones puntuales, además de ser ridículas, dan muy poca información sobre el número de accidentes que, *se predice*, va a ocurrir. Sin embargo, la tabla de las probabilidades predictivas muestra claramente lo que puede ocurrir en cada uno de los tres casos. En todos ellos, las probabilidades predictivas del número de accidentes son decrecientes con el número de estos, y la tasa de decrecimiento es menor cuanto más largo sea el período para el que se realiza la predicción, de acuerdo con la intuición.

En particular, de la tabla 2 resultan los intervalos de credibilidad predictivos:

- 0 ó 1 accidentes en la semana siguiente, con probabilidad 0,988944.
- 0, 1 ó 2 accidentes en las cinco semanas siguientes, con probabilidad 0,950175.
- 0, 1, 2, 3 ó 4 accidentes en las diez semanas siguientes, con probabilidad 0,950667.

Si llegara a observarse un número superior de accidentes en los períodos especificados, habría fundadas sospechas para suponer que o bien ha ocurrido un suceso raro o bien ha habido un aumento sustancial en la siniestralidad debido a posibles factores como, p. ej., un aumento en la densidad de tráfico.

LA TEORÍA DE LA UTILIDAD

Hemos visto que la manera de cuantificar o medir la incertidumbre es expresándola mediante una distribución de probabilidad, que se va transformando mediante la aplicación del teorema de Bayes, a medida que se va incorporando nueva información.

Sin embargo, nuestras decisiones no solo entrañan incertidumbre sino que tienen unas consecuencias que también tenemos que valorar, y que son el otro ingrediente de los problemas de toma de decisiones individuales.

La *medida* de las consecuencias de nuestros actos, aunque se trate de consecuencias numéricas, es el objetivo de la *Teoría de la utilidad*.

El origen de la teoría de la utilidad se debe a Daniel Bernoulli, quien en su trabajo de (1738) expone una crítica del concepto de *esperanza matemática* que debería sustituirse por el de *esperanza moral* o, en terminología moderna, *utilidad esperada*.

En ese artículo aparece por vez primera un problema, que ahora se conoce como *la paradoja de San Petersburgo* por haberse publicado en la revista de la Academia de Ciencias de esa ciudad, y que consiste en determinar el precio justo de un juego cuya esperanza matemática es infinita, a pesar de que solo se puede ganar una cantidad finita.

El problema se puede formular de la siguiente manera: ¿cuánto se estaría dispuesto a pagar por jugar a una lotería L que ofrece premios por valor de $1000 \times 2^{i-1}$ pesetas, con probabilidades respectivas de $1/2^i$, que podemos representar por

$$L = \begin{pmatrix} 1/2 & 1/4 & 1/8 & 1/16 & \dots \\ 1000 & 2000 & 4000 & 8000 & \dots \end{pmatrix}$$

y cuya esperanza matemática es infinita ya que $E[L] = 500 + 500 + \dots = \infty$.

La experiencia demuestra que la mayoría de las personas no estarían dispuestos a pagar una cantidad demasiado elevada, p. ej. superior a 4 o 5 mil pesetas, por jugar al juego que da origen a la paradoja de San Petersburgo.

Sin embargo, en la teoría elemental del cálculo de probabilidades se dice, con demasiada frecuencia, que la esperanza es el precio justo que se debe pagar por jugar a un juego de azar. El argumento que se esgrime es que, *a la larga*, es decir, si se jugase muchas veces y en virtud de la ley fuerte de los grandes números, el valor medio de todos estos juegos se aproximaría a la esperanza, lo que no parece razonable cuando la esperanza es infinita. Incluso si la esperanza fuese finita el argumento anterior no tedaría en cuenta las fluctuaciones aleatorias ni la posibilidad de ruina de los jugadores, (véase Feller, 1968, cap. 3).

La solución que propone Bernoulli es sustituir el valor monetario por un *índice de utilidad* u , de modo que la nueva esperanza sea finita

$$E[u(L)] = \sum_{i=1}^{\infty} u(1000 \times 2^{i-1}) \times 2^{-i} < \infty,$$

con lo que el *precio justo* p_0 , que se pagaría por jugar, sería la solución de la ecuación

$$u(p_0) = E[u(L)].$$

La idea de Bernoulli quedó olvidada durante más de dos siglos pero resurgió en uno de los libros más influyentes del siglo XX, *Theory of Games and Economic Behavior* del matemático John von Neumann y el economista Oskar Morgenstern (1944).



Daniel Bernoulli (1700–1782) □



John Von Neumann (1903–1957)

Función de utilidad: aversión al riesgo

¿Por qué personas distintas pagarían precios distintos cuando se enfrentan a la misma situación de riesgo, como en el ejemplo de la paradoja de San Petersburgo, donde las probabilidades son objetivas y fácilmente comprensibles?

La respuesta es que distintas personas pagarían distintas cantidades, dependiendo de su situación económica y de su actitud ante el riesgo, que son propias de cada individuo.

Consideremos un ejemplo sencillo de determinar el valor, o precio justo, de una lotería simple como la siguiente

$$L' = \begin{pmatrix} 1/2 & 1/2 \\ 0 & 100.000 \end{pmatrix}$$

que se puede interpretarse como un juego donde con probabilidad 1/2 el jugador se queda como está y con probabilidad 1/2 su patrimonio aumenta en 100.000 ptas.

¿Cuánto se estaría dispuesto a pagar por jugar a esa lotería?

La esperanza matemática de la lotería L' es igual a 50.000 ptas. pero, además de lo anteriormente señalado, la mayoría de las personas estaría dispuesta a pagar menos que el valor de la esperanza. Esto nos lleva a preguntarnos sobre la utilidad del dinero, no en el sentido estricto y puramente crematístico de que el dinero sea útil, sino en el de *valorarlo o medirlo* a través de un índice o función de utilidad que, de algún modo, refleje la afición o aversión de las personas hacia situaciones que entrañan riesgo, que tenga en cuenta además su posición económica o patrimonio.

Por situaciones que entrañan riesgo entendemos aquellas en que las probabilidades de los sucesos inciertos están especificadas, por lo que se identifica el riesgo con la incertidumbre probabilística.

La utilidad del dinero

Los comentarios precedentes plantean, entre otras, las siguiente preguntas:

- ¿Por qué la esperanza *monetaria* no es un buen criterio para tomar decisiones?

La respuesta es que la *utilidad* del dinero, aunque es una función creciente del dinero, *no* es una función lineal del dinero para la mayoría de las personas.

- ¿Por qué se suele pagar menos que el valor de la esperanza de una lotería?

La respuesta es porque la mayoría de las personas suele tener *aversión al riesgo*.

De hecho, la aversión a la riesgo se define por la propiedad anteriormente señalada de que el precio que pagaría una persona por jugar a cualquier lotería es siempre menor que la esperanza matemática de dicha lotería.

Esta definición equivale, desde el punto de vista matemático a que la función de utilidad de las personas que tienen aversión al riesgo sea una función cóncava. La figura 8 muestra la forma de una posible función de utilidad cóncava como las que a veces se utilizan en la práctica. Por el contrario, las personas que muestran *afición al riesgo* —aquellas que por jugar a una lotería pagarían una cantidad superior a la esperanza matemática de ésta— tendrían una función de utilidad convexa.

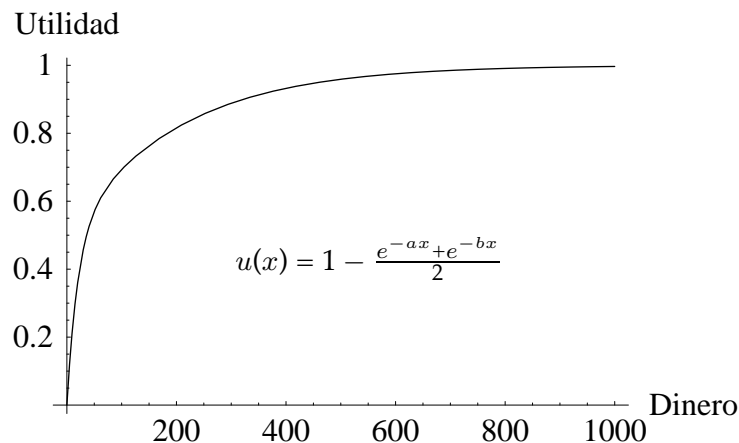


Figura 8. Forma característica de una función de utilidad que representa aversión al riesgo decreciente.

¿Cómo se calcularía el precio justo p que hay que pagar por jugar a una lotería cuyo premio es una variable aleatoria L de distribución conocida, supuesto que el patrimonio de la persona es x_0 y su función de utilidad es $u(x)$?

La solución es igualar la esperanza de utilidad de la lotería a la utilidad de la situación sin riesgo $u(x_0)$, es decir

$$u(x_0) = Eu(x_0 + L - p)$$

y resolver la ecuación en la variable p .

Como aplicación de este criterio consideremos el problema de determinar el precio justo que se pagaría por jugar a las loterías L de la paradoja de San Petersburgo y L' del ejemplo 2, respectivamente, según sea el patrimonio de cada persona, cuando la función de utilidad es la que muestra la figura 8, a saber

$$u(x) = 1 - \frac{e^{-0,005x} + e^{-0,05x}}{2}$$

donde la variable x está medida en miles de pesetas.

En la tabla 3, el patrimonio x_0 y el precio justo p se especifican en pesetas.

L		L'	
Patrimonio: x_0	Precio justo: p	Patrimonio: x_0	Precio justo: p
0	3.068	0	16.158
10.000	3.134	10.000	17.289
100.000	4.273	100.000	38.352
200.000	4.425	200.000	43.728
400.000	4.427	400.000	43.814
1.000.000	4.427	1.000.000	43.815

Tabla 3. Precio justo que se pagaría por las loterías L y L' , según el patrimonio de una persona con función de utilidad $u(x)$.

Comparación de inversiones

En situaciones inciertas más generales, ¿cómo se toman decisiones individuales que entrañan riesgo o incertidumbre probabilística?

La respuesta —que se conoce como el paradigma de la *maximización de la utilidad esperada*— es valorar las consecuencias asociadas a cada posible decisión y a cada suceso incierto elemental por un índice o función de utilidad; calcular la esperanza de este índice respecto de la probabilidad a posteriori, y tomar aquella decisión que haga máximo el valor de esa esperanza.

Si representamos por D el conjunto de las posibles decisiones del individuo, por $u(d, \theta)$ la utilidad de la consecuencia que se deriva de tomar la decisión $d \in D$ cuando el suceso incierto es $\theta \in \Omega$, y por $p(\theta | \mathbf{x})$ la densidad a posteriori del individuo en el momento de tomar la decisión, entonces la regla de decisión óptima se obtiene resolviendo el problema de optimización

$$\max_{d \in D} \int u(d, \theta) p(\theta | \mathbf{x}) d\mu(\theta).$$

Este problema [de optimización] puede ser muy complejo dependiendo del tamaño de los espacios D y Ω y la complejidad y propiedades de la función de utilidad $u(d, \theta)$ y de la densidad a posteriori $p(\theta | \mathbf{x})$, por lo que no podemos entrar en detalles acerca de su resolución.

Concluimos este trabajo con una aplicación del paradigma anterior al problema de seleccionar una cartera o inversión. Para simplificar la exposición, el ejemplo se centra en la comparación de solamente dos carteras.

Vamos a suponer que la rentabilidad de la cartera i -ésima, R_i , se distribuye según una normal de media μ_i y volatilidad o desviación típica σ_i , es decir $R_i \sim N(\mu_i, \sigma_i)$. Si una persona opta por invertir su patrimonio x_0 en la cartera i -ésima, su situación actual vendría representada por la variable aleatoria $x_0(1 + R_i)$.

La esperanza de utilidad de esta variable es:

$$E[u(x) | R_i, x_0] = \int_{-\infty}^{\infty} u(x) \frac{1}{x_0 \sigma_i \sqrt{2\pi}} \exp \left[-\frac{(x - x_0(1 + \mu_i))^2}{2x_0^2 \sigma_i^2} \right] dx,$$

y se seleccionaría la cartera que hiciese máxima la integral anterior.

En ciertas circunstancias no se hace necesario el cálculo de la esperanza anterior, debido a que hay teoremas matemáticos que garantizan que no es necesario conocer con precisión la función de utilidad de la persona para saber que la esperanza de utilidad de una inversión es mayor que la de otra.

Un ejemplo de tal situación ocurre cuando las dos carteras tiene la misma volatilidad y diferentes rentabilidades medias. En este caso, una de las inversiones domina estocásticamente a la otra y un teorema garantiza que la cartera de mayor rentabilidad media siempre tiene esperanza de utilidad mayor *cualquiera que sea la función de utilidad de la persona* siempre que sea creciente con el dinero.

La figura 9 muestra esta situación, en la que la primera cartera tiene una rentabilidad $R_1 \sim N(8, 1)$ y la segunda $R_2 \sim N(6, 1)$, por lo que la primera *siempre* sería preferible a la segunda.

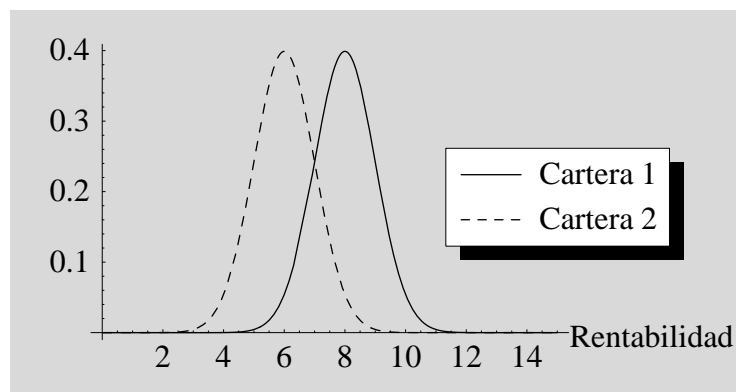


Figura 9. Carteras con distinta rentabilidad y la misma volatilidad.

Consideremos ahora la situación en la que las dos carteras tienen la misma rentabilidad media y distinta volatilidad. Esta situación es la que se ilustra en la figura 10, suponiendo que $R_1 \sim N(5, 5,5)$ y $R_2 \sim N(5, 3)$.

Si la persona tiene aversión al riesgo, entonces, por otro teorema, *siempre* preferirá aquella que tenga mínima volatilidad.

Sin embargo, si la persona tiene afición al riesgo, ocurre lo contrario, es decir, preferirá aquella cartera con mayor volatilidad.

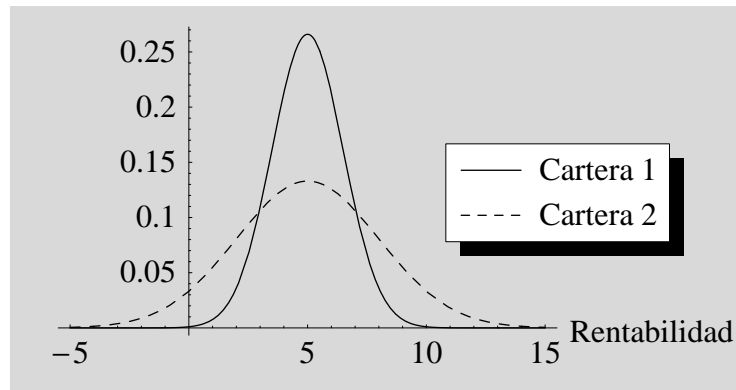


Figura 10. Carteras con la misma rentabilidad y distinta volatilidad.

Si la función de utilidad tiene partes convexas y otras cóncavas no se puede asegurar lo anterior y se hace necesario el cálculo de las esperanzas de utilidad.

Por último, si la rentabilidad media y la volatilidad son distintas, entonces hay que calcular la esperanza de utilidad de cada cartera y elegir aquella que la maximice. Esta situación es la que se ilustra en la figura 11, donde las dos rentabilidades son $R_1 \sim N(5, 5)$ y $R_2 \sim N(10, 20)$, respectivamente.

Así, p. ej., suponiendo que la función de utilidad del inversor $u(x)$ es la misma que antes, y suponiendo que se tiene invertido todo el patrimonio en alguna de las dos carteras representadas en la gráfica, si el patrimonio es mayor que 533.287 ptas., entonces la primera cartera es preferible a la segunda, y viceversa. El resultado tiene su lógica pues si el patrimonio invertido es muy grande, se tienen pérdidas negativas cada vez mayores en valor absoluto que a su vez —por la forma de la función de utilidad en la parte negativa [que no se muestra en la figura 8]— tienen una utilidad negativa cada vez mayor.

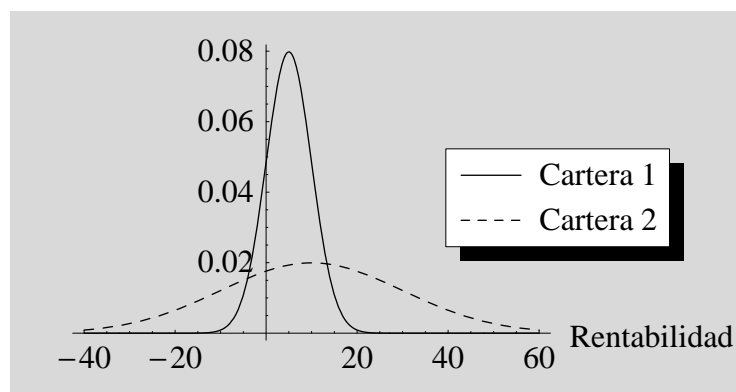


Figura 11. Carteras con distinta rentabilidad y volatilidad.

BIBLIOGRAFÍA

- Acton, F. (1970). *Numerical Methods that Work*. Harper & Row: New York.
- Bernardo, J. M. (2001). Bayesian Statistics. *Encyclopædia of Life Support Systems (EOLSS)*. Paris: UNESCO (en prensa).
- Bernoulli, D. (1738). Specimen theoriae novae de mensura sortis. *Comentarii Academiae Scientiae Imp. Petropolitanae*, V, pp. 175–192.
- Feller, W. (1968). *An Introduction to probability Theory and its Applications*. 3rd Edition. Vol. 1. New York: Wiley.
- Girón, F. J. (2001). Determinismo, Caos, Azar e Incertidumbre. En *Horizontes Culturales: Las fronteras de la Ciencia 1999*. Madrid: Espasa (en prensa).