

## LIBROS

### LA TEORIA PROBABILISTICA EN ECONOMETRIA (\*)

#### P R E F A C I O

Este estudio se considera como una contribución a la econometría. Representa un intento de ofrecer una fundamentación teórica al análisis de las interrelaciones entre las variables económicas. Está basado en la teoría moderna de la probabilidad e inferencia estadística. Dedicaremos unas palabras a justificar tal estudio.

El método de la investigación econométrica intenta, esencialmente, unir la teoría económica y las mediciones reales, empleando la teoría y la técnica de la inferencia estadística como un puente. Pero el puente, por su parte, nunca ha estado completamente construido. Así, el procedimiento común ha sido, primero, construir una teoría económica que envuelva relaciones funcionales exactas y entonces comparar estas teorías con algunas mediciones reales, y, por último, "juzgar" cuando la correspondencia es "buena" o "mala". El instrumento de la inferencia estadística se ha introducido en algún grado para servir de base a tales juicios, así, por ejemplo, el cálculo de unos pocos errores típicos y coeficiente de correlación múltiples. La aplicación de estos sencillos "estadísticos" se ha considerado legítima mientras, al mismo tiempo, la adopción de modelos definidos de probabilidad ha sido considerada como un crimen en la investigación econométrica, una violación de la naturaleza de los datos económicos. Es decir, se ha considerado legítimo emplear algunos de los instrumentos desarrollados por la teoría estadística sin aceptar la verdadera fundamentación sobre lo que se ha construido la teoría estadística. Porque *ningún instrumento desarrollado por la*

---

(\*) La traducción ha sido realizada por Gonzalo Arnáiz Villando.

*teoría estadística tiene significado* —excepto, quizá, para propósitos descriptivos— *sin referirlo a algún esquema estocástico.*

La repugnancia entre los economistas a aceptar modelos probabilísticos como base para la investigación económica tiene como origen, al parecer, un concepto muy estrecho de lo que es probabilidad y variable aleatoria. Creían que los esquemas probabilísticos se podrían aplicar a fenómenos tales como extracciones de lotería o, a lo más, a aquellas series de observaciones en donde cada observación podía considerarse como una extracción independiente de la misma "población". Desde este punto de vista se ha argumentado diciendo, por ejemplo, que muchas series económicas no se ajustaban al modelo probabilístico "porque las observaciones sucesivas no son independientes". Pero no es necesario que las observaciones sean independientes y que tengan la misma ley de probabilidad unidimensional. Es suficiente suponer que el *conjunto total* de las  $n$  observaciones puede considerarse como una observación de  $n$  variables (o punto muestra) con una ley conjunta  $n$  dimensional de probabilidad; la existencia de la misma puede ser meramente hipotética. En este caso podemos contrastar hipótesis considerando esta ley de probabilidad conjunta y obtener inferencias por medio de un punto muestra (en  $n$  dimensiones). La moderna teoría estadística ha hecho considerables progresos resolviendo tales problemas de inferencia estadística.

Así, si consideramos la investigación económica actual —aunque sea realizada por individuos opuestos al empleo de los esquemas probabilísticos— encontramos que reposa en última instancia sobre alguna noción, quizá muy vaga de probabilidad y variable aleatoria. Ya que cada vez que apliquemos una teoría a los hechos no creemos —y no lo esperamos— obtener una coincidencia exacta. Ciertas discrepancias se clasifican como "admisibles", otras como "prácticamente imposibles" bajo los supuestos de la teoría. Y el *principio* de tal clasificación es, en sí mismo, un esquema teórico, a saber, un principio en el cual la expresión vaga "prácticamente imposible" o "casi seguro" se reemplaza por la "probabilidad es casi cero" o la "probabilidad es casi uno".

Esto no es nada más que una forma conveniente de expresar opiniones sobre los fenómenos reales. Pero el concepto de probabilidad tiene la ventaja de ser "analítico": podemos deducir nuevas

afirmaciones por las reglas de la lógica. Así, partiendo de un modelo puramente formal de probabilidad, que envuelve ciertas probabilidades, las cuales, por su parte, pueden no tener contrapartida lógica en el mundo real, podemos deducir afirmaciones como la siguiente: "la probabilidad de A es casi igual a 1". Substituyendo algún fenómeno real por A, y transformando la afirmación "una probabilidad próxima a 1" en "estamos casi seguros que A ocurrirá", hemos realizado una afirmación acerca de un fenómeno real, la verdad de la cual habrá de contrastarse.

La clase de afirmaciones científicas que pueden expresarse en términos probabilísticos es enorme. En efecto, esta clase contiene todas las leyes que pueden formularse. Porque tales leyes no dicen más ni menos que esto: Es casi 1 la probabilidad de que un cierto suceso ocurra.

Así aparece una doble justificación para nuestro intento de dar una fundamentación probabilística más rigurosa a los problemas de investigación económica: Primero, si queremos aplicar la inferencia estadística a la contrastación de las hipótesis de la teoría económica, llevará como consecuencia el que la formulación de teorías económicas representen hipótesis estadísticas, es decir, afirmaciones —quizá algunas poco precisas— sobre ciertas distribuciones de probabilidad. La creencia de que podemos hacer uso de la inferencia estadística sin esta ligazón puede basarse solamente en la falta de precisión al formular los problemas. Segundo, como hemos indicado anteriormente, no hay falta de generalidad al elegir tal teoría. Trataremos de demostrar que es provechoso y útil.

Los principios generales de inferencia estadística introducidos en este estudio están basados en la teoría de contrastación de hipótesis estadística de Neyman-Pearson.

El capítulo I contiene una discusión general de la relación entre los modelos abstractos y la realidad económica.

El capítulo II se refiere al problema de establecer "relaciones constantes en el campo de la economía y al grado de invarianza de las relaciones económicas respecto ciertos cambios en la estructura".

En el capítulo III discutimos la naturaleza de los modelos estocásticos y su aplicabilidad a los datos económicos.

En el capítulo IV se demuestra que un sistema hipotético de relaciones económicas puede transformarse en una afirmación sobre

la ley de distribución conjunta de las variables económicas que envuelve, y que, por lo tanto, un sistema de este tipo puede considerarse como una hipótesis estadística en el sentido de Neyman-Pearson. Una breve exposición de la teoría de Neyman-Pearson, de la contrastación de hipótesis estadística y estimación, se da al comienzo del capítulo.

El capítulo V se refiere al problema de estimación siguiente: Dado un sistema de ecuaciones estocásticas, que comprende cierto número de parámetros, tales que el sistema actualmente se satisface por datos económicos cuando se elige un cierto conjunto de valores de los parámetros, ¿se satisface el sistema para otros valores de los parámetros? Si esto sucediese no se podría obtener una estimación única de los parámetros a partir de los datos (esto es, en el caso de relaciones lineales, el problema muy conocido de la multicolinealidad). Se dan en el mismo reglas matemáticas para investigar tales situaciones.

El capítulo VI contiene una corta descripción de los problemas de predicciones. Se presentan algunos ejemplos para aclarar algunos puntos esenciales.

---

La idea de comenzar este estudio se me ocurrió durante mis trabajos como adjunto del profesor Ragnar Frisch en el Instituto de Economía de Oslo. El lector encontrará en él muchas ideas de Frisch, e indirectamente su influencia puede encontrarse en la formulación de problemas y en los métodos adoptados. Estoy vivamente agradecido a su guía y constante ánimo, así como su paciente enseñanza e interés por mi trabajo.

El análisis aquí presentado se realizó con detalle durante mis períodos de estudio de los Estados Unidos, fué publicado en mimeografía en Harvard en 1941. Doy las más sinceras gracias al profesor Abraham Wald, de la Universidad de Columbia, por las numerosas sugerencias y por la ayuda en muchos puntos durante la preparación del manuscrito. De su conocimiento único, de la moderna teoría estadística y las matemáticas generales, he obtenido mucho provecho. Muchas de las secciones estadísticas de este estudio se han formulado, y otras se han reformulado después de haberlas dis-

cutido con él. El lector puede consultar, en relación con el presente análisis, un estudio del profesor Wald y Dr. H. B. Mann "On the Statistical Treatment of Linear Stochastic Difference Equations". *Econometría*, vol. II, julio-octubre 1943, pág. 173-220. En este artículo se encuentra un tratamiento estadístico más explícito de los problemas que en el estudio presente he mencionado solamente en términos generales.

Debo expresar mi reconocimiento y deuda con el profesor Jacob Marschak, director de investigación de la Cowles Commission, por sus conversaciones estimulantes a este respecto. Quiero, además, expresar mi gratitud al profesor Joseph A. Schumpeter y Edwin B. Wilson, de la Universidad de Harvard, por la lectura de parte del manuscrito original y por críticas que he recogido en la presente formulación. Igualmente estoy en deuda con Mr. Leonid Hurwicz, de la Cowles Commission, y con Miss Edith Elbogen, del National Bureau of Economic Research, por la lectura del manuscrito y sus comentarios valiosos.

Por otra parte, el autor se hace responsable de cualquier error o falta.

Nueva York, junio de 1944.

*Trygve Haavelmo*

## CAPITULO I

## MODELOS ABSTRACTOS Y REALIDAD

1.º—*Introducción*

Los modelos teóricos son instrumentos necesarios en nuestro intento de entender y explicar sucesos de la vida real. En realidad, una sencilla descripción y clasificación del fenómeno real probablemente no será posible o realizable sin contemplar la realidad a través del funcionamiento de algún esquema concebido a priori.

Con estos modelos teóricos sacamos conclusiones del siguiente tipo: "si A es cierto, entonces B es cierto". Podremos también decidir cuándo una afirmación o eslabón de la teoría es exacta o falsa, es decir, cuándo viola o no los requisitos de consistencia íntima de nuestro modelo. En tanto permanezcamos en el mundo de la abstracción y simplificación no hay un límite a lo que se quiera demostrar o rechazar; o, como ha dicho Pareto, "No hay proposición que no se pueda considerar verdadera bajo ciertas condiciones a determinar (1)." Nuestra vigilancia contra especulaciones fútiles consiste en requerir que los resultados de nuestras consideraciones teóricas pueden compararse con algunos fenómenos del mundo real. Esto, por otra parte, no indica que todo resultado teórico, por ejemplo los de la matemática pura, deban tener una aplicación práctica inmediata. Una gran parte de la investigación en teoría pura consiste en deducir afirmaciones rigurosas, las cuales no siempre han de tener una aplicación directa sobre los hechos. Estos, sin embargo, pueden ayudar a consolidar y desarrollar las técnicas e instrumentos del análisis y así aumentar nuestro poder para atacar los problemas de la realidad.

Cuando las afirmaciones deducidas de un modelo teórico se aplican a la realidad, el problema de "cierto" o "falso" es más ambiguo. Los hechos corrientemente diferirán, en muchos aspectos, con cualquier afirmación apriorística y exacta, deducida de un modelo teó-

---

(1) *Manual de Economía Política*, segunda edición, página 9.

rico. En otras palabras, los modelos exactos son falsos en relación con los hechos considerados. ¿Cómo podremos emplear modelos que implican afirmaciones falsas? Es corriente responder a esta pregunta diciendo que, como los modelos abstractos nunca se corresponden exactamente con los hechos, debemos de contentarnos con que las discrepancias no sean muy grandes, o sea que haya "bastante buen acuerdo", etc. Pero posteriormente veremos que tal punto de vista es insostenible, ya que entonces es evidente que tendríamos que tener una regla para decidir de antemano cuándo nuestras afirmaciones a priori son ciertas. Esto es, tales reglas formarían parte de nuestros modelos. Nuestros modelos así desarrollados darían lugar a afirmaciones imprecisas que, cuando se aplican a los hechos, serían verdaderas o falsas.

Sin embargo, sea cual fuere la teoría, no puede ser cierta respecto a un cierto conjunto de hechos si implica una afirmación falsa sobre los mismos. Nos encontraríamos que es prácticamente imposible mantener cualquier teoría que implicase una afirmación no trivial respecto a ciertos hechos a causa de que más pronto o más tarde los hechos contradecirían cualquiera de tales afirmaciones. Por lo tanto, no sólo tendremos que conformarnos con afirmaciones mucho menos precisas que las que corrientemente se dan en un modelo exacto, sino que tendríamos que adoptar un tipo especial de modelo, a saber, modelos que permitan afirmaciones que no lleven en sí nada implícito, sino que tengan una cierta probabilidad de ser ciertas. Esto llevaría a una formulación probabilística de las teorías que se intentan aplicar.

Expresiones como éstas, "la teoría es casi cierta", no tienen sentido si no se especifican de alguna forma análoga a como lo hemos hecho nosotros. Por lo tanto, cuando decimos que una teoría "exacta" es "casi cierta" queremos indicar que esta teoría, aunque sea falsa, en la práctica puede reemplazar otro modelo, el cual, primeramente daría lugar a afirmaciones más generales, y, segundo, permitiría que estas afirmaciones amplias fuesen falsas "en raras ocasiones".

De esta forma, el problema de cuándo un modelo teórico es "casi cierto" es realmente el mismo problema que saber cuándo algún otro modelo más general es cierto respecto a los hechos, o al menos no los contradice. Es con modelos de este último tipo con los

que vamos a tener relación cuando tratemos de contrastar las teorías respecto los hechos. Como ya se ha dicho, veremos que esto conduce a adoptar una formulación probabilística de las teorías a aplicar.

Estas direcciones se aplican, más o menos, a todos los tipos de la teoría económica, bien sean formulados cuantitativamente o no. Pero nosotros no dirigiremos nuestra teoría hacia este amplio campo. A continuación nos ocuparemos de una clase muy importante de teorías económicas, exactamente aquéllas en las que el modelo teórico consiste en un sistema de ecuaciones (ordinarias o funcionales) entre ciertas variables económicas. Se pueden hacer unas pocas observaciones en cuanto al sentido común de este tipo de teoría económica.

En términos amplios, podremos clasificar tales relaciones económicas cuantitativas en tres grupos:

- I. Definiciones que son identidades.
- II. Relaciones técnicas.
- III. Relaciones que describen acciones económicas.

Un ejemplo del primer grupo se puede dar por la siguiente relación: Gasto total = precio multiplicado por la cantidad comprada. Salida total = salida por trabajador multiplicado por el número de los mismos, y "relaciones análogas de contabilidad". Al segundo grupo pertenecen las funciones de producción técnica y otras restricciones naturales o institucionales, las que corrientemente se consideran como datos en la planificación económica. En el tercer grupo encontramos la amplia clase de las relaciones que describen el comportamiento de los individuos o unidades colectivas en su actividad económica, sus decisiones de producir y consumir.

En tales relaciones aparecen dos clases de cantidades, a saber: las variables bajo investigación y los parámetros introducidos en el proceso de análisis (los términos variables y parámetros son relativos al problema particular que se trata, no pueden definirse en ningún sentido absoluto). En las relaciones del tipo I, los parámetros se dan por definición, mientras en las relaciones del tipo II los parámetros están a *nuestra disposición* para adaptar tales relaciones hipotéticas a un conjunto de variables económicas. Desde el punto de vista de la teoría económica esta distinción se aplica especialmente a las relaciones del tipo III. Se aplican quizá menos



a las del tipo II; la elección de la forma en las relaciones técnicas puede considerarse como tareas de otras ciencias.

Consideremos en particular las relaciones del tipo III. Con certeza sabemos que las decisiones de consumir, invertir, etc., dependen de un gran número de factores, muchos de los cuales no pueden expresarse en forma cuantitativa. ¿Para qué entonces vamos a intentar asociar tal comportamiento con solamente un limitado conjunto de fenómenos medibles, los cuales no pueden dar más que un cuadro incompleto del "ambiente" o "atmósfera" en la cual las decisiones y planificaciones económicas tienen lugar? Antes de nada diremos que las "explicaciones" de esta clase se realizan solamente para fenómenos que tienen una naturaleza cuantitativa, tales como precios, valores y volúmenes físicos. Y cuando las decisiones económicas son del tipo "más" o "menos", "mayor" o "menor", éstas pueden tener consecuencias para algunos otros fenómenos medibles. Así, si un hombre decide gastar más de su renta (fija) en una cierta mercancía, deberá gastar menos en otras cosas. Si un fabricante quiere aumentar su producción necesitará pedir más factores de producción. Si su beneficio aumenta, tendrá consecuencia para su política de ahorro-gasto; y así análogamente. Evidentemente será muy artificial suponer que estas cantidades en sí mismas no tendrán influencia en las decisiones tomadas, y que tales influencias no forman un sistema. Es esto, pues, solamente un paso natural, para intentar una descripción aproximada de tales influencias por medio de unos ciertos parámetros de comportamiento.

Por lo pronto, éste es un tipo de "explicación". Otros tipos pueden elegirse. Pero cualquiera que sean las "explicaciones" no hay que olvidar que todas son invenciones artificiales en una investigación para entender la vida real; no son verdades ocultas que haya que "descubrir".

## 2.º—*Definiciones cuantitativas exactas de las variables económicas*

Esta frase ha llegado a ser una especie de tópico entre los economistas modernos, pero a veces parece existe cierta confusión en cuanto a su verdadero significado. Su interpretación sencilla y racional sería la de que al presentarse los hechos más importantes que

queremos estudiar en la vida económica real en forma de medidas numéricas tendremos que elegir nuestros modelos de aquel campo de la lógica que trata de los números, es decir, de las matemáticas. Pero los conceptos de las matemáticas alcanzan su significado cuantitativo de una forma implícita mediante el sistema de operaciones lógicas que imponemos. En la matemática pura no existe realmente un problema tal como el de definición cuantitativa de un concepto "per se" sin referencias a ciertas operaciones.

Por lo tanto, cuando los economistas hablan acerca de los problemas de definiciones cuantitativas de variables económicas, piensan en algo que tenga relación con los fenómenos económicos reales. Mas, precisamente, quieren dar "reglas exactas para medir ciertos fenómenos de la vida real, tratar de "conocer exactamente qué elementos de la vida real se corresponden con los de la teoría". Cuando consideramos un modelo teórico que lleva consigo ciertas variables y ciertas relaciones matemáticas, es corriente preguntar acerca del comportamiento de esta y la otra variable. Pero esta pregunta no tiene sentido dentro de un modelo teórico. Y si la cuestión se aplica a la realidad no tienen una respuesta precisa. La respuesta que podemos dar es, en el caso mejor, una tentativa de descripción que consistirá en palabras que hemos tratado de asociar, más o menos vagamente, con ciertos fenómenos reales.

Una cosa es construir un modelo teórico y otra dar reglas para elegir los hechos a los cuales el modelo teórico se puede aplicar. Esto es, una cosa es elegir el modelo teórico del campo de las matemáticas y otra cosa es clasificar y medir objetos de la vida real. Por esta última razón necesitaremos una buena disposición entre nuestros compañeros de investigación para ponernos de acuerdo, con fines prácticos, sobre cuestiones de definición. No son posibles —"hablando estrictamente"— ambigüedades en la clasificación y medidas de los fenómenos reales. No solamente es nuestra técnica de medición de fenómenos físicos imprecisa, sino que en muchos casos no somos capaces de dar una definición sin ambigüedad del método de medición a emplear, somos capaces de dar reglas precisas para la elección de cosas que hay que medir en relación con una cierta teoría. Así, por ejemplo, consideremos la al parecer sencilla cuestión de medir el consumo total de una mercancía durante un período de tiempo dado en un país. Inmediatamente se presen-

tan dificultades derivadas del hecho que la noción de "mercancía", "consumo", etc., no son términos precisos; pueden discutirse respecto su contenido o medida cuantitativa. Y esto se aplica a todas las cantidades que representan medidas prácticas de objetos reales.

3.º—*Variables «observables», «reales» y «teóricas»;  
una distinción importante*

Aunque nuestro conocimiento real de los hechos económicos está basado en clasificaciones poco precisas y medidas aproximadas, a menudo tenemos la impresión de que "lo podríamos hacer algo mejor", que, en muchos casos, *sería* posible dar descripciones y reglas de medida de tal manera que si dos o más observadores independientes aplicasen esas reglas a un grupo dado de objetos, prácticamente obtendría las mismas cantidades. A menudo, cuando operamos con nociones como la renta nacional, producción de ciertas mercancías, importaciones, exportaciones, etc., creemos que estas cosas tienen un significado cuantitativo definido y que posiblemente pueden medirse más exactamente, pero —por razones financieras o falta de tiempo— no somos capaces de realizar el recuento y medida en la forma que realmente debía hacerse. Corrientemente creemos también que estos problemas de medidas son algo diferente de los consistentes en investigar "explicaciones". Cuando hablamos de ciertos hechos que han de ser explicados, pensamos, en muchos casos, en medidas más correctas y controladas de los hechos que las que se suelen dar por la estadística económica corriente. De la experiencia en varios campos hemos adquirido conocimientos empíricos de las fuentes de error y el grado de precisión relacionados con los tipos corrientes de la técnica estadística de observación. Por lo menos, tal como está la situación ahora en el campo de las estadísticas económicas, podremos casi siempre estar seguros que podrían ser mejores si pudiéramos encontrar el tiempo y dinero necesarios. Cuando hablamos de los "verdaderos" valores de ciertos fenómenos observables, comparándolos con alguna información estadística observada, la distinción se debe a que pensamos en forma parecida a la que acabamos de describir en términos un tanto vagos.

En la teoría pura introducimos variables (o funciones del tiempo) las cuales, por construcción, satisfacen ciertas condiciones de

consistencia interior de los modelos teóricos. A estas variables teóricas corrientemente se las da nombres que indican con qué medidas reales "verdaderas" esperamos que las variables teóricas se identifiquen. Pero las variables teóricas no se definen como idénticas a algunas variables verdaderas. Para que el proceso de medición sea correcto es esencial aplicarlo a cada variable *separadamente*. Imponer alguna relación funcional entre las variables significa ir mucho más allá. Podemos expresar la diferencia diciendo que las "verdaderas" variables (o funciones del tiempo); representan en nuestro ideal, las medidas precisas de la realidad "como ocurre de hecho", mientras que las variables definidas en una teoría son las verdaderas medidas que haríamos si la realidad estuviera de acuerdo con nuestro modelo teórico.

La distinción entre estos tres tipos de variables, aunque algo vaga, es de una gran importancia para la comprensión de sus relaciones con la teoría pura. Trataremos de explicar esto de forma diferente, que quizá sea más clara.

Uno de los hechos más característicos de la moderna teoría económica es el empleo extensivo de símbolos, fórmulas, ecuaciones y otras nociones matemáticas. Libros modernos y artículos de Economía están "llenos de matemáticas". Muchos economistas consideran "la Economía matemática" como una rama separada de la Economía. Este problema sugiere ver qué diferencia existe entre "Economía matemática" y "matemáticas". ¿Un sistema de ecuaciones se hace menos matemático y más económico sólo por llamar a  $x$  "consumo" e  $y$  "precio", etc.? Evidentemente hay muchos ejemplos en que no se va mucho más allá en cuanto se refiere a significado económico. Difícilmente entonces merecen el rango de contribuciones a la Economía. Lo que hace a un trabajo Economía matemática, no sólo matemáticas, yo pienso que es esto: Cuando establecemos o sentamos un sistema de relaciones teóricas y empleamos nombres económicos para las variables teóricas incluidas, tenemos en el pensamiento algunos *experimentos* reales, o un diseño de un *experimento*, que podríamos por lo menos pensar que estábamos realizando, para medir aquellas cantidades que en la vida real económica suponemos obedecen a las leyes impuestas sobre sus nombres teóricos. Por ejemplo, en la teoría de la elección introducimos la noción de superficie de indiferencia, para demostrar cómo un individuo, a precios dados, distribuirá su renta fija entre

varias mercancías. Esto parece "económico", pero así sólo es un esquema matemático; solamente es económico si añadimos un *diseño de experimentos* que pudiese indicar, primero, qué fenómeno real se identifica con los precios teóricos, cantidades y rentas; segundo, lo que entendemos por un "individuo", y, tercero, cómo podríamos ordenar los individuos que actualmente hacen esta elección.

Hay numerosas indicaciones de que los economistas casi siempre tienen algunos de estos diseños de experimentos ideales en el fondo de su pensamiento cuando construyen tales modelos teóricos. Por ejemplo, difícilmente existirá un economista que se sienta satisfecho al identificar series corrientes de renta nacional, consumo, etc., con las variables de esos mismos nombres en sus teorías. O, recíprocamente, encontraría corrientemente muy complicado o quizá hasta falta de interés el intentar construir modelos de forma que las observaciones que le gustase identificar con las correspondientes variables teóricas, correspondieran a aquellas dadas por las estadísticas económicas existentes. En la descripción verbal de su modelo, "en términos económicos", el economista sugiere generalmente, explícita e implícitamente, algunos tipos de experimentos o de medidas controladas destinadas a obtener las variables reales para las cuales cree que su modelo vendría bien. Esto es, él piensa en algunas variables "verdaderas" que le gustaría medir. Los datos que actualmente obtiene, están primero casi siempre perturbados por algunos errores sencillos en las medidas, esto es, por ciertos "hechos" extras que él no intenta "explicar" por medio de su teoría; en segundo lugar, y esto todavía es más importante, el economista es generalmente un observador más bien pasivo respecto a los fenómenos económicos importantes; generalmente no controla la recogida de estadísticas económicas. No está en situación de reforzar las prescripciones de sus propios diseños de experimentos ideales.

Podríamos, quizá, caracterizar también la diferencia entre las variables "verdaderas" y las "observables", del siguiente modo: Las variables "verdaderas" son variables tales que si su comportamiento contradice una teoría, ésta tendría que rechazarse como falsa; mientras que las variables "observables", cuando contradigan la teoría, dejan la posibilidad de haber podido ensayar la teoría sobre hechos para los cuales esa no estaba pensada, la confusión está provocada

por el uso de los mismos nombres para cantidades que son diferentes.

Para contrastar una teoría respecto a los hechos, o para usarla en predicciones, tanto han de corregirse las observaciones estadísticas disponibles, como adaptarse la teoría, de forma que convierte a los hechos que estamos considerando en "las variables verdaderas" necesarias para la teoría, como ya hemos dicho. Para emplear un ejemplo mecánico, supongamos que queremos comprobar la ley de los cuerpos cayendo en el vacío, y que nuestras medidas para tal fin consistiesen en una serie de observaciones con una piedra, por ejemplo, dejada caer desde una altura. Para emplear tales datos por lo menos deberíamos calcular el efecto extra de la resistencia del aire y extraer dicho elemento de los datos. O lo que sería lo mismo, tendríamos que desarrollar la teoría simple de los cuerpos cayendo en el vacío para deducir la resistencia del aire (y probablemente muchos otros factores). Un físico rechazaría esas medidas como absurdas para tal fin, porque fácilmente puede hacerlas mucho mejores. El economista, por otra parte, tiene a menudo que contentarse con medidas brutas y sesgadas. Corrientemente, su tarea es obtener las medidas que necesita de datos que ya están recogidos con algún otro fin; o tiene resultados que, por así decirlo, la naturaleza ha producido en toda su complejidad; su tarea será construir modelos que expliquen lo que se ha observado.

La conclusión práctica del anterior estudio es un consejo que los economistas difícilmente dejan de dar, pero que realmente pocos siguen. Es decir, que se debe estudiar con todo detenimiento las series reales consideradas y las condiciones bajo las que se obtienen, antes de identificarla con las variables de un modelo teórico particular. (Estudiaremos estos problemas en el capítulo II.)

#### 4.º—Modelos teóricos, hipótesis y hechos

Sean  $x'_1, x'_2, \dots, x'_n$ ,  $n$  variables reales, y  $(x'_1, x'_2, \dots, x'_n)$  o, abreviadamente  $(x')$  cualquier conjunto particular de valores de estas variables. Cualquiera de tales conjuntos puede representarse como un punto en un espacio enedimensional cartesiano. Sea  $S$  el conjunto de tales puntos y "A" un sistema de reglas u operaciones que definen un subconjunto  $S_2$  de  $S$  ( $S_2$  puede, por ejemplo, ser una

superficie enedimensional). La regla "A" adscribe a cada punto ( $x'$ ) una propiedad, es decir, la propiedad de pertenecer o de no pertenecer a  $S_a$ . Si admitimos que las  $n$  variables  $x'$  varían con la condición de que ( $x'$ ) pertenezca a  $S_a$ , se forma un modelo teórico para las variables  $x'$ .

De forma análoga, consideremos  $n$  funciones del tiempo  $x'_1(t)$ ,  $x'_2(t)$ , ...,  $x'_n(t)$ . Sea  $F$  el conjunto de todos los posibles sistemas de  $n$  funciones del tiempo, y sea "B" un sistema de reglas u operaciones que definen una subclase  $F_b$  de  $F$ . Cualquier sistema de  $n$  funciones del tiempo tendrá entonces la propiedad de pertenecer o no pertenecer siempre a  $F_b$ . El sistema de reglas  $B$  define un modelo respecto a la serie de tiempo de  $n$  componentes.

Así, un modelo teórico puede considerarse como una sencilla restricción en las variaciones conjuntas de un sistema de cantidades variables (o más generalmente objetos), los cuales, por otra parte, pueden tener cualquier valor o propiedad. De forma aún más general, las restricciones impuestas pueden no excluir en absoluto cualquier valor de las cantidades consideradas, sino sencillamente dar diferentes ponderaciones (o probabilidades) a los varios conjuntos posibles de valores de las cantidades variables. El modelo en cuestión se caracteriza corrientemente por el hecho de que define ciertos conjuntos restringidos del conjunto de todos los valores posibles de las cantidades, de forma que estos subconjuntos tienen casi todo el peso.

Un modelo teórico de este tipo no tiene interés práctico. Y esta situación no cambia, como ya hemos dicho anteriormente, cuando se introducen solamente "nombres económicos" para las cantidades u objetos variables que figuran en él. El modelo tiene interés económico solamente después de que un sistema dado de cantidades u objetos de la vida económica se han elegido o descrito de forma que puedan ser identificados con los comprendidos en el modelo. Esto es, el modelo tendrá un interés económico solamente cuando esté asociado con un diseño de experimentos que describe —e indica cómo se mide— un sistema de variables "verdaderas" (u objetos)  $x_1, x_2, \dots, x_n$  que pueden identificarse con las variables correspondientes de la teoría.

Como una consecuencia de tal identificación, todas las afirmaciones correctas que se puedan hacer dentro del modelo, con respec-

to a las variables teóricas u objetos que en él figuran, se hacen automáticamente también respecto a las variables "verdaderas" reales. El modelo, por lo tanto, se convierte en una hipótesis a priori sobre los fenómenos reales, estableciendo que todo sistema de valores que podemos observar de las variables verdaderas será uno que pertenezca al sistema de valores admisible dentro del modelo. La idea es que la Naturaleza tiene una forma de seleccionar sistemas de valores de las "verdaderas" variables, tal que estos sistemas se comportan como si la selección se hiciese por la regla que define nuestro modelo teórico. Hipótesis como la anterior son las únicas implicaciones conjuntas —y las únicas contrastables, en lo que se refiere a la observación— de la teoría y diseño de experimentos. Parece natural, pues, adoptar el convenio de que una teoría se llame verdadera o falsa, según que la hipótesis sea verdadera o falsa, cuando se contrasta respecto a los datos elegidos como variables verdaderas. Entonces podríamos hablar de intercambiabilidad entre la contrastación de hipótesis y contrastación de teorías.

Si un cierto conjunto de valores de las variables se *excluye del modelo*, entonces que cualquier sistema de valores observados caiga en este conjunto excluido, será suficiente para rechazar la hipótesis (y por lo tanto la teoría) por falsa respecto a las verdaderas variables consideradas. Pero como ya hemos dicho, el modelo debe ser (y creemos que en la práctica así es) de tal forma que no excluya cualquier sistema de valores de las variables, sino que solamente dé diferentes ponderaciones o probabilidades a los distintos sistemas de valores. Estas ponderaciones necesitan una interpretación práctica con objeto que el modelo exprese una hipótesis de comportamiento respecto a las variables "verdaderas" correspondientes. De acuerdo con la experiencia se ha encontrado que corrientemente es fructífero interpretar tales ponderaciones como una medida de la "frecuencia real de ocurrencia". Si las ponderaciones totales adscritas a todos los conjuntos de valores posibles del sistema es finita, podremos entonces decir que el comportamiento del conjunto de sistema de valores que tienen una ponderación casi igual a cero de acuerdo con el modelo, es una hipótesis que dice que la naturaleza tiene una forma de seleccionar sistemas de valores conjuntos de las correspondientes variables "verdaderas" que hace que sea "prácticamente imposible" que un sistema de valores observados pueda



caer dentro de tal conjunto. Para *contrastar* la teoría respecto a algunas otras teorías alternativas, podremos agregar que consideraremos la hipótesis contrastada falsa cuando observemos un cierto número de tales sistemas de valores "casi imposibles". Esto es, ante el riesgo de cometer un error, preferimos adoptar otras hipótesis bajo las cuales las observaciones no son de un tipo casi imposibles.

Si hemos encontrado que una cierta hipótesis, y, por lo tanto, el modelo, es aceptable sobre la base de un cierto número de observaciones, podremos decidir emplear la teoría para la predicción. Si después encontramos que no son muy útiles estas predicciones, nos inclinaremos a dudar de la validez de las hipótesis adoptadas (y, por lo tanto, de la utilidad de la teoría). Podríamos contrastarla otra vez, siempre sobre la base de un conjunto más extenso de observaciones.

Se ha encontrado que es útil introducir un cálculo especial para tratar tales tipos de hipótesis. Este es el cálculo de probabilidades. Posteriormente estudiaremos las razones de sentido común que hay para emplear este cálculo en la deducción de hipótesis sobre los fenómenos económicos.

Supongamos ahora que tenemos un conjunto de observaciones que confirman las afirmaciones que son permisibles dentro del modelo. Entonces estas afirmaciones se convierten en hechos a la luz de nuestro modelo teórico, o en otras palabras, nuestro modelo es aceptable en lo que se refiere a las observaciones conocidas. ¿Podrá el modelo aplicarse a observaciones futuras? No podemos dar ninguna razón a priori para esta suposición. Podemos solamente decir que, de acuerdo con un amplio registro de experiencias reales, parece que será útil creer en la posibilidad de tales inducciones empíricas.

---

A la luz del análisis anterior, podemos ahora clasificar de forma amplia los principales problemas con los que se enfrenta la investigación científica cuantitativa.

1.º *Tentativa de construcción de modelos.*—Parece casi imposible tratar de describir exactamente cómo un científico marcha hacia la construcción de un modelo. Es un proceso de creación, un arte que opera con nociones racionalizadas de algún fenómeno real y con el mecanismo que las produce. La idea de tales modelos

descansa en la creencia ya respaldada por una gran cantidad de experiencias en muchos campos, de la existencia de ciertos elementos de invarianza en una relación entre fenómenos reales, siempre que tengamos éxito al reunir los que deban ser.

2.º *Contrastación de teorías*, que es el problema de decidir sobre la base de los datos, cuando se mantiene y emplea cierta teoría o se deshecha y cambia por otra.

3.º *El problema de la estimación*, que en un sentido amplio es el problema de dividir, tomando como base los datos, todas las posibles teorías a priori acerca de ciertas variables en dos grupos: unos conteniendo las teorías admisibles, y otros que contienen aquellas que han de rechazarse.

4.º *El problema de las predicciones*.

Los problemas 2, 3 y 4 están estrechamente ligados a una formulación probabilística de las hipótesis, y se ha producido mucha confusión por los intentos de tratarlos de otra forma. En una formulación probabilística pueden definirse precisamente, y una gran parte de la confusión existente en la investigación económica puede aclararse. Estos problemas se tratan en los capítulos IV, V y VI.

Muchos economistas, no obstante, consideran los problemas 2-4 como detalles. Su tarea principal tiene un sentido más fundamental: el problema de saber si podremos tener alguna esperanza de construir modelos racionales que contribuyan en algo a nuestro conocimiento de la vida económica real. En el próximo capítulo trataremos de aclarar algunos de los principales argumentos de este estudio.

## CAPITULO II

### EL GRADO DE PERMANENCIA DE LAS LEYES ECONOMICAS

Si comparamos el desarrollo histórico de las distintas ramas de las ciencias cuantitativas, encontramos una semejanza estricta en las huellas que han seguido. Su origen es la curiosidad del hombre por las explicaciones de "hechos curiosos". Las observaciones de tales hechos son más o menos accidentales y en cualquier caso de un carácter pasivo. Sobre las bases de tales—quizá muy vagos—reco-

nocimientos de hechos, hay personas que construyen desarrollos primitivos, corrientemente de tipo metafísico. Entonces algunos empiricistas más fríos van más allá. Quieren “conocer los hechos”, observan, miden y clasifican, y mientras realizan esto creen en la posibilidad de establecer un cierto orden, un cierto sistema, en el comportamiento del fenómeno real. Intentan construir sistemas de relaciones que copien la realidad tal como ellos la ven, o sea desde el punto de vista de un observador cuidadoso pero pasivo. Cuando siguen coleccionando más y más observaciones, ven que su “copia” de la realidad necesita “reparaciones”. Y, sucesivamente, sus esquemas crecen en un laberinto de supuestos extra y casos especiales; el aparato total se hace cada vez más difícil de manejar. Algún trabajo clasificador se impone, y se ha encontrado que la llave de tal clasificación es un *razonamiento a priori*, que conduce a la introducción de unos principios y relaciones muy generales —y corrientemente muy sencillos—, de los cuales la clase total de las cosas aparentemente muy distintas pueden deducirse. En las ciencias naturales el último paso ha suministrado instrumentos mucho más potentes de análisis que la *enumeración empírica de los casos*.

Podríamos inclinarnos a decir que la posibilidad de tales construcciones hipotéticas y deducciones fructíferas depende de dos factores distintos, a saber: por un lado, el hecho de que son leyes de la naturaleza, y por el otro lado, la eficiencia de nuestros instrumentos analíticos. Sin embargo, por una inspección detenida, vemos que tal distinción es dudosa. Indudablemente, podemos tratar de describir tal cosa como una ley de la naturaleza sin referirnos a ciertos *principios de análisis*. Y la frase “en las ciencias naturales tenemos leyes estables”, indica no mucho más ni menos que esto. Las ciencias naturales han elegido formas muy útiles de observar la realidad física. Sin embargo, una frase tal como “en la vida económica no hay leyes constantes” es no solamente muy pesimista, sino que parece insensata. En cualquier caso, no puede contrastarse. Pero podemos discutir cuándo las relaciones que se deducen de nuestro *esquema actual* de la teoría económica son tales que se aplican a los hechos de la vida económica real. Podemos discutir los problemas que surgen cuando intentamos hacer comparaciones entre la realidad y nuestra construcción actual de la teoría económica. Trataremos de encontrar una explicación racio-

nal al hecho de que relativamente pocos intentos de establecer "leyes" económicas han tenido éxito. Yo creo que considerables esfuerzos pueden gastarse en aclarar estos problemas restringidos.

En lo que sigue nos proponemos tratar algunos de los problemas fundamentales que se presentan cuando se juzga el grado de persistencia en el tiempo de las relaciones entre las variables económicas. Por motivos de simplicidad operaremos aquí con la noción de relaciones "exactas" más que "estocásticas". Podemos hacer esto porque los principales puntos que se discuten no están ligados a un tipo particular de relaciones que esperemos establecer. Los problemas que se van a discutir están más directamente relacionados con el problema general de si podemos o no esperar encontrar elementos de invarianza en la vida económica sobre la que se establecen leyes permanentes.

### 5.º—¿Qué entendemos por «Relaciones Constantes»?

Cuando empleamos los términos "relaciones constantes" o "relaciones cambiables inestables" evidentemente nos referiremos al comportamiento de algún fenómeno económico real, *comparado* con algún comportamiento que esperamos de las consideraciones teóricas. La noción de constancia o permanencia de las relaciones *no es*, por lo tanto, una cuestión teórica. Es una propiedad, de los fenómenos reales *al mirarlos desde el punto de vista de una teoría particular*. Mas precisamente, sean  $x'_1, x'_2, \dots, x'_n$   $n$  variables teóricas restringidas por una ecuación

$$f(x'_1, x'_2, \dots, x'_n; \alpha_1, \alpha_2, \dots, \alpha_k) = s' \quad [5.1]$$

en donde las  $\alpha_i$  son constantes y la  $s'$  es una desviación que posee ciertas propiedades específicas. [5.1] no se convierte en una teoría económica sólo por emplear terminología económica en los nombres de las variables que envuelven, [5.1] se convierte en una teoría económica cuando se asocia con una regla de medición de  $n$  variables económicas reales  $x_1, x_2, \dots, x_n$ , que se han de comparar con  $x'_1, x'_2, \dots, x'_n$  respectivamente. La característica esencial de tal regla de medida es que no impone *a priori* las restricciones [5.1] sobre las

variables que han de medirse. Si hiciésemos eso caeríamos en el mundo de la teoría abstracta, a causa de que una de las variables podría deducirse de las medidas de las otras  $n - 1$  variables y las propiedades asignadas a  $s'$ . La regla de medida es esencialmente un procedimiento técnico de medir cada variable separadamente. Es un diseño de experimentos reales para obtener las variables verdaderas como se describió en la sección 3.

Todos los conjuntos de  $n$  valores de las variables teóricas  $x'$  en [5.1] tienen una propiedad común: la de satisfacer la ecuación. Estamos interesados en saber cuándo o no las "verdaderas" variables  $x_1, x_2, \dots, x_n$  tienen la misma propiedad. Sea  $(x_1, x_2, \dots, x_n)$  uno cualquiera de los resultados obtenidos por nuestro diseño de experimentos, y sea  $s$  una variable definida implícitamente por

$$f(x_1, x_2, \dots, x_n; \alpha_1 \dots \alpha_k) = s \quad [5.2]$$

en donde  $f$  es la misma que en [5.1]. Si entonces  $s$  tiene la misma propiedad que  $s'$  en [5.1], cualquiera que sea el sistema de valores experimentales observados  $x_1, x_2, \dots, x_n'$ ; diremos que las variables "verdaderas" observadas  $x_1$  siguen una ley constante.

Por lo tanto, dada una relación teórica, un diseño de experimentos y un conjunto de observaciones, el problema de la constancia o invariancia de una relación económica se convierte en estos dos:

- 1) ¿Hemos observado lo que realmente queríamos observar, es decir, puede el conjunto de observaciones dadas considerarse como un resultado obtenido siguiendo nuestro diseño de experimentos ideales?
- 2) ¿Tienen las variables "verdaderas" reales las mismas propiedades que las variables teóricas?

Un diseño de experimentos (lo que los físicos llaman "experimento crucial") es un apéndice esencial de cualquier teoría cuantitativa. Y corrientemente pensamos en tales experimentos cuando construimos las teorías aunque —desgraciadamente— muchos economistas no describen sus diseños de experimentos explícitamente. Si lo hiciesen, verían que los experimentos que han pensado podrían agruparse en dos clases diferentes: 1), experimentos que podríamos hacer para ver si ciertos fenómenos económicos reales —cuando artificialmente se aíslan de "otras influencias"— pueden cumplir

ciertas hipótesis, y 2), la corriente de experimentos que la Naturaleza está constantemente realizando en su enorme laboratorio, y en los cuales nosotros intervenimos como observadores pasivos. En ambos casos el objeto de la teoría es el mismo, a saber: dominar los hechos de la vida real. Pero nuestra teoría es ligeramente diferente en ambos casos.

En el primer caso podemos hacer que la concordancia o discordancia entre la teoría y los hechos dependa de dos cosas: los hechos que elegimos para considerar, así como nuestra teoría acerca de ellos. Como ha dicho Bertrand Russell: "El procedimiento actual de la ciencia consiste en una combinación de observaciones, hipótesis, experimentos y teorías."

En el segundo caso, nosotros tratamos de ajustar nuestras teorías a la realidad tal como aparece ante nuestra vista. ¿Cómo se comporta un diseño de experimentos en este caso? Así: tratamos de elegir una teoría y un diseño de experimentos conectado con ella, de tal forma que los datos resultantes *puedan* ser aquellos que nosotros obtenemos por observación pasiva de la realidad. Y en tanto que consigamos hacerlo nos haremos dueños de la realidad por convenio pasivo.

Ahora bien, si examinamos teorías económicas corrientes, vemos que un gran número de ellas, en particular las más profundas, requieren experimentos del tipo primeramente mencionado. Por otra parte, la clase de los datos económicos que actualmente tenemos pertenecen más bien al segundo tipo. En Economía empleamos un vocabulario relativamente pequeño para describir una variedad enorme de fenómenos (y muchas veces los economistas emplean diversos nombres para el mismo fenómeno). El resultado es que muchas cosas distintas aparecen bajo el mismo nombre, y que, por lo tanto, tenemos el peligro de considerarlas como idénticas. De esta forma, las teorías, corrientemente se comparan con datos que no pueden considerarse como observaciones obtenidas siguiendo el diseño de experimentos que hemos ideado cuando construimos la teoría. Por otra parte, cuando una teoría no muestra acuerdo con los hechos, podremos siempre decir que no tenemos la clase de datos que debíamos tener, pero esto es una frase hueca, a menos que podamos decir al mismo tiempo cuál es la clase de datos que necesitamos y la forma de obtenerlos, por lo menos en cuestión de principio. Si

toda teoría fuese acompañada por una cuidadosa descripción del diseño de experimentos, se aclararía mucha de la confusión existente sobre la constancia y el cambio de las leyes económicas.

La descripción del problema de estabilidad o permanencia de las relaciones económicas es muy amplia. Pueden darse unas respuestas preliminares a las críticas muy superficiales sobre la posibilidad de desarrollar la Economía como una ciencia. Pero no se puede responder a los más profundos problemas de detalle con los que nos enfrentamos cuando tratamos realmente de investigar porque la Economía, hasta ahora, no nos ha llevado hasta unas leyes universales y exactas como las que se obtienen en las ciencias naturales.

Consideremos una vez más el argumento general "No hay leyes constantes que describan fenómenos de la vida económica". Anteriormente dijimos que este argumento no tenía sentido. Justificaremos esta afirmación un poco más adelante. No es posible dar una respuesta precisa a este argumento porque en sí mismo no representa un problema preciso, pero tratemos de comprender lo que significa el mismo. Supongamos, primero, que consideramos la "clase de todos los diseños de experimentos", los resultados de los cuales "podrían interesarnos como economistas". Aquí, desde luego, tropezamos inmediatamente con una dificultad, porque probablemente no es posible definir tal clase. No conocemos todos los experimentos en que pudiéramos estar interesados. Por otra parte, consideremos la clase de todas las teorías económicas posibles (del tipo que estamos discutiendo aquí). Para cada diseño de experimentos hay una sucesión definida de medidas reales. Considérese, para cada una de estas medidas, la subclase de teorías con las cuales están de acuerdo. Para una sucesión de medidas tenemos una sucesión de tales subclases de teorías, ahora si estas clases de teorías no tienen ninguna propiedad en común no trivial podríamos decir que las medidas obtenidas por el diseño de experimentos utilizado no sigue ninguna ley. Pero, ¿realmente dice algo esta afirmación? Evidentemente muy poco. Porque es una afirmación acerca de clases de cosas que están completamente indefinidas. Por mucho que inventemos y por mucho que nos equivoquemos, nunca podremos sentar una conclusión como "En la vida económica no hay leyes constantes".

Estudiaremos un problema mucho más restringido, el de has-

ta donde las "leyes hipotéticas" de la teoría económica en su grado presente se aplican a datos como los que conseguimos por medio de observaciones pasivas. Por observaciones pasivas entendemos resultados observables de lo que los individuos, Empresas, etc., *hacen realmente* en el curso de los acontecimientos no lo que ellos *podrían* hacer, o lo que ellos piensan que harán, bajo otras circunstancias especiales. Sería superficial considerar este problema meramente como una cuestión de si nuestra teoría económica presente es *buena o mala*; o que es una exposición provechosa del problema. Tenemos que empezar analizando lo que realmente estamos tratando de conseguir con la teoría económica. Tenemos que comparar sus diseños de experimentos idealizados con aquellos que podrían requerirse para reproducir los fenómenos de la vida económica real que observamos de forma pasiva.

En tal discusión pronto descubrimos que tenemos que tratar con una maraña de cuestiones diferentes. Tratemos de pasar revista a las más importantes:

a) ¿La mayoría de las teorías que construimos sobre "economía racional" son de tal forma que para ellas *no* son experimentos adecuados los datos históricos y las observaciones pasivas? Esta cuestión está relacionada con lo siguiente:

b) ¿Tratamos de construir teorías que describen lo que individuos, Empresas, etc. *hacen* realmente en el curso de los acontecimientos, o construimos teorías que describen modelos de alternativas en un momento dado? Si es este el caso, ¿qué valor tienen tales modelos de alternativas sobre una serie de decisiones y acciones llevadas a cabo ya?

c) ¿Por qué no nos limitamos sólo a las teorías que son comprobables directamente? O, ¿por qué estamos interesados en relaciones para las cuales la Naturaleza no puede proporcionar experimentos?

d) Corrientemente nuestras teorías son tales que creemos que ciertas series directamente observables *podrían* dar resultados experimentales adecuados para una comprobación, con tal que *otras cosas* no cambien. ¿Qué valor pueden tener tales teorías sobre la realidad, si sencillamente despreciamos la influencia de esas "otras cosas"? Esto vuelve a estar relacionado con el problema siguiente:

e) Estamos interesados en describir bien lo que *sucede* ac-



*tualmente*, o lo que *sucedería* si pudiésemos conservar inmutables "otras cosas". En el primer caso construimos teorías para las cuales esperamos que la Naturaleza misma cuida que se den las necesarias condiciones *ceteris paribus*, lo hacemos esto cuando sabemos, por ejemplo, que así ha sucedido aproximadamente en el pasado. En el segundo caso tratemos por nuestra cuenta que se cumplan las condiciones *ceteris paribus* por medio de artificios estadísticos tales como eliminar de los datos las influencias que no se han tenido en cuenta en la teoría (por ejemplo, por análisis de la correlación múltiple).

f) De la experiencia con correlaciones de series de tiempo, sabemos que corrientemente es posible establecer relaciones íntimas entre variables económicas para algún período de tiempo especial mientras que estas relaciones se rompen para el próximo período de tiempo. ¿Significa este hecho que no podemos esperar establecer leyes constantes de vida económica?

Esas cuestiones se han tomado más o menos directamente de discusiones corrientes en problemas de investigación económica; están, como puede verse, horriblemente entrelazadas; ninguna de ellas forma un problema analítico preciso. Por esto preguntamos: ¿Pueden cubrirse estos problemas, al menos en parte, por el análisis de una colección de problemas simplificados y disjuntos? Trataremos ahora de hacerlo así estudiando tres grupos diferentes de problemas a los que abreviadamente podríamos llamar:

I.—Reversibilidad de relaciones económicas.

II.—La cuestión de simplicidad en la formulación de leyes económicas.

III.—La autonomía de una relación económica.

#### 6.—*Reversibilidad de relaciones económicas*

En el campo de las investigaciones económicas, la aplicación a las relaciones de la teoría pura de las series de tiempo o registros históricos ha llegado algunas veces a ser tabú. Muchos economistas no entrenados suficientemente en la teoría estadística, al parecer han sido asustados por trabajos críticos tales como, por ejemplo, el de G. U. Yule. Han llegado a pensar que hay algo inherente en las series de tiempo económicas que *como tales* hacen a sus datos

inaprovechables para las aplicaciones de teoría económica pura. El argumento general es análogo a éste. En la teoría económica, operamos con esquemas hipotéticos de decisiones, que los individuos, Empresas, etc., pueden tomar como respuestas a ciertas condiciones alternativas dadas (por ejemplo, adaptación de la cantidad consumida a un cambio dado del precio). Pero las series económicas de tiempo que muestran resultados reales de decisiones tomadas realmente son solamente descripciones históricas de un viaje en una dirección realizado a través de una serie de circunstancias que están continuamente cambiando, de forma que no es posible hacer predicciones reales por medio de los modelos de alternativas dados por la teoría económica pura.

Tratando de analizar este problema de forma más precisa, vemos primero que el argumento general anterior no desmiente la posibilidad de que las relaciones deducidas de la teoría económica pueden ser muy persistentes y acuradas cuando se aplican a los datos. El argumento solamente dice que los tipos de datos representados por las series de tiempo no son los que deben resultar de los diseños de experimentos prescritos por la teoría económica. Debemos pensar antes de nada en las dificultades que se derivan del hecho de que las series de observaciones pasivas están influenciadas por un gran número de factores no tenidos en cuenta por la teoría; en otras palabras, dificulta el cumplimiento de la condición. "Otras cosas siguen igual". Pero esto es un problema común a todas las observaciones y medidas prácticas; esto es una cuestión de principio, no un defecto particular de las series económicas de tiempo. Si nosotros no podemos preservar los datos de tales "otras influencias" trataremos de introducir estas influencias en la teoría, con objeto de lograr más concordancia entre la teoría y los hechos. También puede ser que los datos, como dados por las series de tiempo económicas, estén restringidos por un sistema total de relaciones, de tal forma que las series no muestren variación bastante para comprobar cada relación separadamente. Estos problemas los discutiremos con extensión en las dos próximas secciones. Además está el problema de errores de medida. Pero este problema es también general, y no peculiar de las series de tiempo económicas.

Si estas dificultades se eliminan, ¿hay todavía alguna propie-

dad peculiar de las series de tiempo económicas que las hacen no aprovechables para las aplicaciones a relaciones deducidas de la teoría económica pura? Aun haciendo una cuidadosa inspección, es difícil ver que tal propiedad pueda existir, a causa de que si podemos construir cualquier ley general, describiendo lo que los individuos *actualmente hacen* y si tenemos una serie de observaciones de lo que los individuos han hecho en el pasado, entonces, necesariamente, la ley teórica deberá ajustarse a esta serie de observaciones. Surge aquí un problema. Aunque yo pienso que se trata de una confusión entre dos clases diferentes de relaciones que aparecen en la teoría económica: 1), las que intentan describir lo que los individuos hacen en cualquier tiempo, y 2), las que describen un modelo de alternativas en un momento dado antes de que se tome cualquier decisión particular. Relaciones del primer tipo corrientemente se deducen de un sistema de *relaciones* del segundo tipo. Para hacer la discusión de este punto más concreta consideraremos un ejemplo sencillo de la demanda del consumidor para una sola mercancía.

Supongamos que un individuo consume  $n$  mercancías diferentes, e indiquemos por  $x_1, x_2, \dots, x_n$  cantidades de estas  $n$  mercancías. Siendo  $p_1, p_2, \dots, p_n$  sus precios correspondientes. Supongamos que los individuos tienen una renta en moneda constante. De acuerdo con la teoría general de la elección del consumidor, escribiremos:

$$x_i = f_i(p_1, p_2, \dots, p_n) \quad (i = 1, 2, \dots, n) \quad [6.1]$$

en donde las  $f_i$  son funciones de demanda. Supongamos ahora que todos los precios, excepto uno,  $p_1$ , son constantes, y consideremos la correspondiente cantidad  $x_1$  de la mercancía número 1. Podremos escribir:

$$x_1 = f(p_1) \quad [6.2]$$

¿Qué es lo que indica esta función bajo los supuestos hechos anteriormente? Puede indicar dos cosas diferentes:

Una interpretación es que, cuando  $p_1$  tiene un valor particular, es decir  $p'_1$ , el individuo decide comprar una cantidad  $x'_1 = f(p'_1)$  de la mercancía número 1.

Otra interpretación es ésta: Supongamos que el individuo está

en una posición en donde paga el precio  $p^0_1$  y consume una cantidad  $x^0_1$ . Él considera en esta posición los posibles cambios en el consumo de la mercancía número 1 que él podría elegir como respuesta a los varios cambios en el precio de  $p^0$ . Si el precio cambiase de  $p^0_1$  a  $p''_1$  él compraría  $x''_1 = f(p''_1)$ ; y así sucesivamente. Esto es decir él tiene un modelo de alternativas respecto al próximo cambio del precio juzgado desde su posición presente ( $x^0_1, p^0_1$ ). Para indicar que este modelo puede depender de su posición presente escribiremos:

$$x_1 = f^0(p_1) \quad [6.3]$$

en donde  $f^0$  satisface  $x^0_1 = f^0(p^0_1)$ .

Es evidente que estos dos tipos de demanda son de naturaleza diferente, y, además, el primero exige más que el segundo. Para el primero se requiere el supuesto de que existe una relación única entre el consumo y los precios, de acuerdo con la cual el individuo actúa sin considerar la posición que ocupa en el momento que tiene que tomar la decisión. El segundo solamente dice que los individuos tienen un modelo de alternativa respecto al cambio próximo del precio juzgado desde su posición presente ( $x^0_1, p^0_1$ ). Después que él ha tomado una decisión en respuesta al cambio del precio, ya no está en la posición ( $x^0_1, p^0_1$ ), y él puede cambiar su modelo de alternativas a causa de que en la nueva posición puede "ver las cosas de forma diferente".

Si el individuo tiene un modelo de demanda fijo, éste es independiente del punto en donde se encuentra en cualquier momento (es decir, un modelo del tipo [6.2]), entonces un registro histórico de precios y correspondientes cantidades consumidas representarán puntos de este modelo de demanda, y podríamos emplearlos para predecir el consumo para cualquier valor del precio (bajo el supuesto anterior que los otros precios no cambien). Por otra parte, si el modelo de demanda depende de la posición actual del individuo, para cada *posición actual* del individuo sería un modelo bien definido de alternativas, las cuales, si las conocemos, nos permitirán predecir la cantidad que podría ser comprada si el precio cambiase de  $p^0_1$  a  $p'_1$ . Pero tan pronto como la nueva posición ( $p'_1, x'_1$ ) se alcanza nece-

sitamos *otro* modelo  $f'$ , para predecir la cantidad comprada si el precio cambiase de  $p'_1$  a  $p''_1$ . Las dos situaciones se representan gráficamente en las figuras 1 y 2.

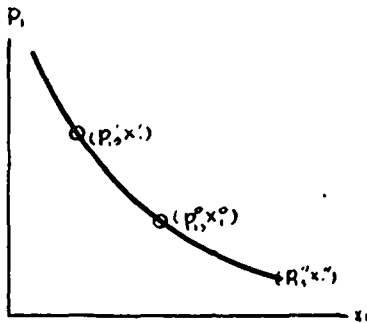


Fig. 1.ª modelo de demanda reversible.

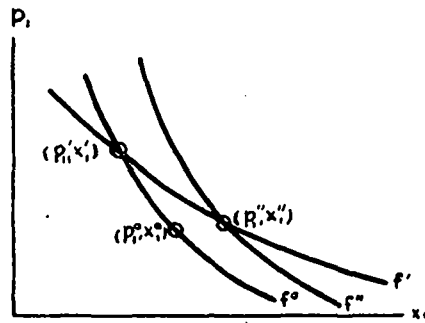


Fig. 2.ª Modelo de demanda irreversible.

En la figura segunda, un registro histórico de las posiciones reales  $(p_1^0, x_1^0)$   $(p_1', x_1')$ , etc. pueden no formar puntos de cualquier curva fijada de demanda. Y si queremos ajustar alguna curva a estos puntos de posiciones reales, tal curva no deberá emplearse para predecir el efecto del próximo cambio del precio. Para encontrar el modelo de un individuo en un momento dado, tendremos que preguntarle, y que él nos responda, lo que haría si el precio cambiase alternativamente a ciertas cantidades.

Podemos considerar la figura primera como un esquema estático, mientras la figura segunda representa uno dinámico, a causa de que en la figura primera la sucesión de cambios en los precios es irrelevante mientras en la segunda es esencial. Sin embargo, no subrayamos aquí tanto la sucesión de los cambios del precio-cantidad en el tiempo como el hecho real de llevar a cabo una decisión planeada que conduce al individuo a un nuevo "medio", por así decirlo, donde él se comporta de forma distinta a la que pensó que se comportaría antes de llegar a ese punto.

Si un esquema igual que el de la figura segunda está más cerca de la realidad que la figura primera, entonces, naturalmente, un intento de emplear el esquema de la figura primera fallará. Por otra parte, si la teoría opera con modelos acotados iguales

a los de la figura segunda, entonces los registros históricos de las combinaciones de los precios-cantidades reales no son los datos importantes de la teoría.

Un esquema irreversible igual que el de la figura segunda corrientemente puede reducirse a uno reversible introduciendo más variables. Podemos, por ejemplo, suponer que los modelos de demanda de la figura segunda cambian de una forma regular con las posiciones iniciales con las que está asociada la teoría. Sean las variables  $\bar{X}_1, \bar{P}_1$ , el precio y cantidad que representan posiciones reales del individuo, y sea  $(x_1, p_1)$  cualquier punto del modelo de demanda que pasa por  $(\bar{x}_1, \bar{p}_1)$ . Puede suceder que el comportamiento individual se pueda describir por una relación del tipo

$$x_1 = F(p_1, \bar{X}_1, \bar{p}_1) \quad [6.4]$$

en donde F es tal que

$$\bar{x}_1 = F(\bar{p}_1, X_1, p_1) \quad [6.4']$$

Esta función deberá entonces ser compatible con la serie de tiempo de los precios y cantidades realmente consumidos. Más específicamente, cada par de puntos sucesivos representan posiciones reales que satisfarán [6.4], es decir, si  $(\bar{x}_1, \bar{p}'_1)$  y  $(x'_1, p'_1)$  son dos puntos sucesivos, tendremos

$$x'_1 = F(\bar{p}'_1, \bar{x}_1, \bar{p}'_1) \quad [6.5]$$

Podríamos entonces determinar los parámetros de F por medio de la serie real de tiempo, por [6.4] podríamos calcular el modelo de demanda para cualquier punto inicial dado  $(\bar{x}_1, \bar{p}_1)$ .

Este esquema es posible que sea demasiado sencillo. En general tendremos probablemente que introducir como variable no solamente la posición instantánea del individuo, sino también la sucesión total de las posiciones pasadas, así como las longitudes de los intervalos de tiempo existentes entre los cambios de precios y la situación, podría, por otra parte, ser aún más complicada cuando todos los otros precios también varíen. Esto se excluye de nuestra discusión anterior. ¿Cuándo no es posible ajustar de esta forma se-

ries históricas a los esquemas de relaciones reversibles? Es una pregunta que no puede contestarse *a priori*. Así hemos tratado de ponerlo de manifiesto.

Además de las dificultades del tipo anteriormente discutido, las cuales parecen —en cuanto a cuestión de principio— muy sencillas y claras, yo no veo que las series de tiempo económicas tengan ninguna otra propiedad “mística” que las haga incompatibles con la teoría económica.

### 7.—*La cuestión de sencillez en la formulación de Leyes Económicas*

Indiquemos por  $y$  una variable económica. Los valores observados de la cual pueden considerarse como resultados de decisiones económicas planeadas, tomadas por individuos, Empresas, etc. (por ejemplo y puede ser el consumo anual de una cierta mercancía dentro de un grupo de individuos o el tanto anual que ellos abonan de su renta, etc., o puede ser la tasa de producción de una industria monopolizada, o las importaciones mensuales de un cierto material, acero, etc., etc.). Comencemos por suponer que la variable  $y$  está influida por un número de factores causales. Este punto de vista está profundamente enraizado en nuestro modo de razonar sobre las cosas que observamos en la realidad.

No necesitamos captar las nociones de causa y efecto en cualquier sentido metafísico. Lo que indicamos es solamente que los individuos, Empresas, etc., están limitados en el proceso de adaptación en sus planes y decisiones por un conjunto de condiciones que son los *datos*. Dentro de los límites dados por estas condiciones, el proceso de adaptación consiste en elegir lo que se designa como la “mejor” decisión, ya sea en un sentido ya sea en otro. Suponemos que los individuos tienen un sistema de modelos de preferencia que determinan las “mejores decisiones” correspondientes a cualquier conjunto dado de condiciones límites. Tendremos, por lo tanto, el siguiente esquema:

Condiciones dadas.	Sistema de modelos	“Mejor decisión”.
Variables independientes.	de preferencia.	Variables dependientes.

[7.1]

Si el sistema de modelos de preferencia establece una correspondencia entre el conjunto de condiciones dadas y el de "mejor decisión", de tal forma que para cada conjunto de condiciones hay una, y sólo una, mejor decisión, podremos saltarnos la parte central de [7.1] y decir que las decisiones de los individuos, Empresas o grupos están determinados por el sistema de las condiciones límites de elección dadas (variables independientes).

En una cuestión de principio, quizá podrían aparecer algunas dificultades lógicas deducidas del hecho de operar con tal relación causal unívoca. En efecto, los economistas modernos han dado mucha importancia a la necesidad de operar con relaciones del tipo de dependencia mutua en lugar de relaciones del tipo causa-efecto. Sin embargo, yo creo que ambas clases de relaciones tienen lugar en la teoría económica, y, por lo tanto, no es necesario oponer unas a otras a causa de que un sistema de relaciones del tipo de dependencia mutua para la economía como un *total* puede construirse de sistemas abiertos de relaciones causales dentro de los sectores varios de la Economía. Los factores causales (o las variables independientes) para una sección de la Economía pueden, a su vez, ser variables dependientes en otras secciones, mientras que en ésta las variables dependientes de la primera sección entran como independientes. El hecho esencial es que mientras para la Economía como un total todo depende de otro todo, para cada individuo, Empresa o grupo, hay ciertos factores a los que los individuos, Empresas o grupos consideran como datos. La noción de factores causales es de un carácter relativo más que absoluto.

Nos permitiremos, por lo tanto, aceptar el punto de vista de que las decisiones que producen el ahorro, el consumo, etc., están influenciadas por un número definido de factores causales relacionados  $x_1, x_2, \dots$ . Nuestra esperanza en teoría e investigación económica es que pueda ser posible establecer relaciones constantes y relativamente sencillas entre las variables dependientes (del tipo descrito anteriormente) y un número relativamente pequeño de variables independientes  $x$ . En otras palabras, esperamos que cada variable  $y$  sea explicada por un número relativamente pequeño de factores explicativos, las variaciones de los cuales son prácticamente decisivos para determinar las variaciones de  $y$ . (El problema de la simplicidad de la forma de una relación es corrientemente mu-



cho menos importante que el del número de variables que lleva consigo a causa de que si sabemos que hay una relación funcional, corrientemente es posible aproximarla desarrollando la función en serie.)

Cuando, o no, pueden establecerse tales relaciones sencillas deben decidirse por ensayos reales. A priori nunca puede demostrarse que sea posible o imposible. Pero podemos hacer algo que nos dé alguna indicación sobre nuestras razones para ser optimistas o pesimistas; trataremos de explicar cómo debe ser la situación real para que no tengamos esperanza de poder establecer relaciones sencillas y estables.

Antes de nada, es necesario definir lo que entendemos por "influencia" de un factor económico. Esta expresión tal como se ha empleado en la literatura económica, parece que tiene diferentes significados. Distinguiremos entre dos nociones diferentes de "influencia", a las que llamaremos respectivamente influencia *potencial* e influencia de "facto". Definiremos primero estos dos conceptos desde un punto de vista formal.

Sea  $y'$  una variable teórica definida como una función de  $n$  variables "causales" independientes  $x_1, x_2, \dots, x_n$

$$y' = f(x_1, x_2, \dots, x_n) \quad [7.2]$$

en donde  $f$  está definido dentro de un cierto dominio de las variables  $x_i$ . La influencia potencial del factor  $x_1$  sobre  $y'$  la definiremos como  $\Delta_1$ , y  $y$  vendrá dado por

$$\Delta_1 y' = f[x_1, x_2, \dots, (x_1 + \Delta x_1), \dots, x_n] - f(x_1, x_2, \dots, x_n) \quad [7.3]$$

en donde  $\Delta x_1$  es una magnitud positiva tal que  $x_1 + \Delta x_1$  está dentro del dominio de definición de  $f$ . Es evidente que la cantidad  $\Delta_1 y$  en general dependerá de las variables  $x$  así como del valor  $\Delta x_1$ . Y por otra parte, lo que entendemos por un grande o pequeño  $\Delta x_1$  dependerá de las unidades de medidas de las variables  $x_i$ . Para comparar el tamaño de la influencia de cada variable  $x_i$  tendremos que elegir para cada punto  $(x_1, x_2, \dots, x_n)$  un desplazamiento  $(\Delta x_1, \Delta x_2, \dots, \Delta x_n)$  que se considere de igual tamaño de acuerdo con alguna *norma de juicio*. (Por ejemplo, una norma particular puede consistir en definir los incrementos  $\Delta x$  en cualquier punto del espacio de

las variables  $x$  como constante. e iguel a un porcentaje de las  $x_1, x_2, \dots, x_n$  respectivamente.) Para un sistema dado de incrementos  $\Delta x_1, \dots, \Delta x_n$ , las influencias potenciales son, evidentemente, propiedades formales de la función  $f$ .

Definamos ahora la noción de influencia de *facto* de  $x_1$  sobre  $y'$ . En contraste con la influencia potencial, la influencia de *facto* se refiere a un conjunto de valores de  $y'$  correspondiente a un conjunto de sistema de valores de las variables  $x_1, x_2, \dots, x_n$  elegidos de acuerdo con algún principio externo. Sea

$$\begin{array}{l}
 y'_1, x_{11}, x_{21}, \dots, x_{n1} \\
 y'_2, x_{12}, x_{22}, \dots, x_{n2} \\
 \dots\dots\dots \\
 y'_N, x_{1N}, x_{2N}, \dots, x_{nN}
 \end{array}
 \tag{7.4}$$

un conjunto de  $N$  sistemas de tales valores. Por influencia de *facto* de  $x_1$  sobre  $y'$  dentro de este conjunto de valores entendemos, a grandes rasgos, las partes de  $y'_1, y'_2, \dots, y'_n$  que pueden adscribirse a las variaciones de  $x_1$ . Esto puede definirse cualitativamente de varias formas. Una puede ser la siguiente: Reemplazamos la variable  $x_1$  en [7.2] por una constante,  $c_1$  que se determina de tal forma que

$$\begin{aligned}
 Q_i = \sum_j^N [f(x_{1j}, x_{2j}, \dots, x_{1j}, \dots, x_{nj}) - \\
 - f(x_{1j}, x_{2j}, \dots, c_1, \dots, x_{nj})]^2
 \end{aligned}
 \tag{7.5}$$

sea mínimo respecto a  $c_1$ .

Suponiendo que tal mínimo existe, la influencia de *facto* sobre  $y'$  de la variable  $x_1$  en el sistema [7.4] podría entonces, por ejemplo, definirse como: Constante  $\sqrt{Q_i^{(mín)}}$ .

De las definiciones anteriores es evidente que la influencia potencial de un factor puede ser grande mientras —al mismo tiempo— la influencia de *facto* del factor en un *conjunto* de datos particulares puede ser cero o muy pequeña. Y recíprocamente, las influencias potenciales pequeñas (pero no idénticas a cero).

Esta distinción es fundamental. Si queremos explicar una cierta variable observada y un sistema de factores causales, no hay, en

general, límite para el número de tales factores que pueden tener influencia potencial sobre  $y$ . Pero la naturaleza puede limitar el número de factores que tienen una influencia *de facto* no despreciable a un número relativamente pequeño. Nuestra esperanza de leyes sencillas en Economía reside en el supuesto de que podamos proceder como si tales limitaciones naturales del número de factores importantes existiese. Discutiremos esto con un poco más detalle.

Supongamos que, fuera de un —quizá infinito— número de factores  $x_1, x_2, \dots$ , con una influencia potencial sobre  $y$ , elegimos un número relativamente pequeño,  $x_1, x_2, \dots, x_n$ , y consideremos una cierta función

$$y^* = U(x_1, x_2, \dots, x_n) \quad [7.6]$$

de estas variables. Supondremos que si todos los otros factores  $x_{n+1}, x_{n+2}, \dots$  (suponiendo que son numerables) *deciden no variar* tendríamos  $y = y^*$  para todo conjunto de valores observados ( $y_1, x_1, x_2, \dots, x_n$ ) ¿Podría el conocimiento de tal relación ayudar a “explicar” los valores reales de  $y$ ? Podría, siempre que la influencia *de facto* de los factores no especificados fuese muy pequeña en comparación de la de los factores especificados  $x_1, x_2, \dots, x_n$ . Este puede ser el caso siempre que [1] los factores no especificados varíen considerablemente. Con tal que su influencia potencial sea muy pequeña [2] las influencias potenciales de los factores no especificados sean considerables, pero al mismo tiempo estos factores no cambien mucho o lo hagan raramente con los factores especificados.

Por otra parte, supongamos que todos los factores  $x_1, x_2, \dots, x_n, x_{n+1}, \dots$ , o por lo menos un gran número de ellos, fuesen del siguiente tipo [1]. Cada factor tiene una influencia potencial considerable sobre  $y$  [2]. Cada  $x$  varía *corrientemente* muy poco, pero *ocasionalmente* tiene alguna variación grande. Cómo hay un gran número de factores  $x$  estamos seguros que *ocasionalmente* ocurrirá alguna gran variación casi siempre en *uno u otro factor*. Elegir un pequeño número de factores  $x$ , suponiendo el resto constante, podría ser una ayuda muy pobre para “explicar” las variaciones actuales observadas en  $y$ , es decir, relaciones del tipo [7.6] tendrían muy poca persistencia en el tiempo si  $y$  fuese sustituido por  $y^*$ , sencillamente a causa de las condiciones *ceteris paribus*  $x_{n+1} = \text{constante}$ ,  $x_{n+2} = \text{constante}$ , no se aproximarían a la realidad. Desde el punto

de vista de "explicar" la realidad, podremos decir que sería prácticamente imposible construir una teoría tal que el diseño asociado a los experimentos pudiese aproximarse a lo que se realiza por la Naturaleza. Desde el punto de vista de comprobar ciertas relaciones simplificadas de la teoría, podríamos decir que, en la situación que acabamos de describir, sería imposible encontrar datos para tal propósito por medio de observaciones pasivas.

¿Cuál es la situación real de lo que sabemos por experiencia en la investigación econométrica? ¿Necesitamos considerar un número enorme de factores para explicar la decisión de producir, consumir, etcétera? Yo creo que nuestra experiencia nos dice todo lo contrario. Cuando tratamos *a priori* de especificar lo que nosotros pensamos que son "factores importantes" nuestra imaginación queda exhausta rápidamente, y cuando tratamos de aplicar nuestra teoría a los datos (por ejemplo, ciertos métodos de regresión) corrientemente nos encontramos con que un gran número de los factores de nuestra primera lista no tienen una importancia de *facto*.

Frecuentemente nuestra mayor dificultad en la investigación econométrica no reside en establecer relaciones sencillas entre series de observaciones reales, sino más bien en el hecho de que las relaciones observables, sobre ciertos intervalos de tiempo, *son más sencillas* que lo que esperábamos que fueran por la teoría; de ese modo desperdiciamos elementos de una teoría que serían suficientes para explicar "roturas en la estructura" que aparecen posteriormente. Este es el problema de la autonomía de las relaciones económicas, que discutiremos ahora.

### 8.—*La autonomía de una relación económica*

Todo investigador en el campo de la Economía ha pasado probablemente por la siguiente experiencia: Cuando trata de aplicar las relaciones establecidas por la teoría económica a las series reales observadas de las variables que se trata, encuentra frecuentemente que las relaciones teóricas son "innecesariamente complicadas", podríamos actuar con menos variables que las supuestas *a priori*. Pero también ocurre que, cuando tratamos de hacer predicciones con tales relaciones simplificadas para un nuevo conjunto de datos, las relaciones éstas corrientemente no se cumplen, es decir,

que aparecen discontinuidades en la estructura de los datos. Para el nuevo conjunto de datos podemos encontrar también una relación sencilla, pero diferente. Si no apareciesen estas discontinuidades, quedaríamos perplejos ante esta inesperada simplicidad, ya que de nuestras consideraciones teóricas tenemos la sensación de que la vida económica es capaz de producir variaciones de un tipo más general. Algunas veces, por otra parte, la situación puede explicarse directamente por el hecho de haber incluido en nuestra teoría factores que no tienen influencia potencial sobre las variables que hay que explicar. Pero frecuentemente yo creo que la sorpresa es el resultado de confundir dos clases diferentes de *variaciones* de las variables económicas, llamadas variaciones hipotéticas *libres* y variaciones que están restringidas por un sistema de relaciones simultáneas.

Veremos mejor esta diferencia considerando las operaciones racionales en virtud de las cuales se construye un sistema teórico de relaciones. Tales sistemas representan intentos de reconstruir de una forma simplificada el mecanismo que pensamos corresponde al fenómeno que observamos en el mundo real. Para tratar de reconstruir estos mecanismos consideremos *una relación en un tiempo*.

Supongamos, por ejemplo, que estamos considerando  $n$  variables teóricas  $x'_1, x'_2, \dots, x'_n$  que comparamos respectivamente con  $n$  variables observadas  $x_1, x_2, \dots, x_n$ . Imponemos ciertas relaciones a las  $n$  variables teóricas, de tal forma que pensamos que estas variables teóricas están obligadas a mostrar alguna correspondencia con las variables observadas.

Esto establecido, consideremos una relación particular, a saber:  $x'_1 = f(x'_2, \dots, x'_n)$ . Para construir tal relación razonamos de la siguiente forma: si  $x'_2$  es tal y tal,  $x'_3$  tal y tal, etc., *entonces* tendremos un cierto valor de  $x'_1$ . En este proceso no es problema cuando estos "supuestos" ocurren o no. Cuando imponemos más relaciones a las variables, una gran cantidad de estos "supuestos", que eran posibles para las relaciones  $x'_1 = f$  separadamente, pueden ser imposibles a causa de que violan las otras relaciones. Después de haber impuesto un sistema total de relaciones, nos quedará muy poco de toda la variación hipotética con la que comenzamos. Al mismo tiempo, si hacemos una cuidadosa elección de las relaciones teóricas, puede ocurrir que las posibles variaciones concuerden bien con las de las variables observadas.

¿Por qué empezamos con unas variaciones mucho más generales que las que necesitamos finalmente? Supongamos, por ejemplo, que el sistema de relaciones walrasiano del equilibrio general fuese una pintura verdadera de la realidad; ¿qué ganaríamos operando con este sistema general en lugar de con la sencilla afirmación de que cada una de las cantidades que aparecen es igual a una constante? La ganancia es esta: Actuando sobre la relación general concebimos un amplio *conjunto de posibilidades* que pueden corresponder a la realidad que está regida por una relación solamente. El sistema de relaciones simultáneas de una explicación del hecho de que, fuera de este enorme conjunto de posibilidades, solamente una *muy particular será la que realmente aparece*. ¿Pero una vez que esto está establecido, podríamos olvidarnos del proceso total y quedarnos con la descripción simplificada? Es aquí donde el problema de la autonomía de una relación económica se presenta. El significado e importancia de esta noción creo que puede aclararse con la siguiente analogía mecánica.

Si hacemos una serie de pruebas de velocidad con un automóvil rodando en una llanura, por una árida carretera, seremos capaces de establecer una relación funcional muy precisa entre la presión del gas en el cilindro (o la distancia del pedal acelerador al suelo del coche) y la velocidad correspondiente del coche. Y el conocimiento de esta relación puede ser suficiente para conducir el coche a una velocidad especificada. Pero si un individuo que no conoce nada de automóviles necesita saber cómo funcionan, nosotros no le aconsejaríamos que perdiese el tiempo en medir una relación análoga a la anterior. ¿Por qué? Primero, a causa de que tal relación deja el total mecanismo interno del coche en un misterio completo, y, segundo, porque tal relación puede romperse en cualquier momento, en cuanto haya alguna discordancia en cualquier punto del motor del coche. (Comparar esto, por ejemplo, con las muy conocidas relaciones desplazadas entre las curvas de Harvard A-B-C.) Decimos que tal relación tiene muy poca *autonomía* (2) a causa de que su

---

(2) Este término, así como muchas ideas del análisis de esta sección, se han tomado de un artículo mimeografiado de RAGNAR FRISH: *Statistical Versus Theoretical Relations in Economic Macro-Dynamics* (Memoria mimeografiada preparada por la Business Cycle Conference at Cambridge England, July 18-20-1938, discutido por S. Tirbergen, publicada Liga de las Naciones, 1938.)

existencia depende del cumplimiento simultáneo de un gran número de otras relaciones, algunas de las cuales son de naturaleza transitoria. Por otro lado, las leyes generales de la termodinámica, la dinámica, la fricción, etc., son relaciones altamente autónomas, con respecto al mecanismo del automóvil, a causa de estas relaciones describen la función de alguna parte del mecanismo, *independientemente* de lo que sucede en otras partes.

Llevemos esta analogía al mecanismo de la vida económica. La teoría económica se construye en el supuesto de que las decisiones de los individuos de producir y consumir pueden describirse por ciertas relaciones fundamentales de comportamiento, y que, a su vez, hay ciertas restricciones institucionales y técnicas en la libertad de elegir (tales con funciones técnicas de producción, restricciones legales, etc.).

Un sistema particular de tales relaciones define una estructura teórica particular de la Economía; es decir, define un *conjunto teórico* de posibles conjuntos de valores simultáneos o conjuntos de series de tiempo para las variables económicas. Puede ser necesario —y esta es la tarea de la teoría económica— considerar varias *alternativas* de tales sistemas de relaciones, esto es, varias *estructuras* alternativas que pueden corresponder aproximadamente a la realidad económica en cualquier tiempo. A causa de que la “estructura real” puede cambiar, y corrientemente ocurre en varios aspectos.

Para hacer esta idea más precisa, supongamos que es posible definir una *clase  $\Omega$  de estructuras* tal que un miembro u otro de esta clase puede, aproximadamente, describir la realidad económica en cualquier *situación prácticamente concebible*. Supongamos que definimos alguna *medida* no negativa del “tamaño” (o “importancia o credulidad”) de cualquier subclase  $\omega$  de  $\Omega$ , incluyendo al mismo  $\Omega$ , tal que si una subclase contiene totalmente a otras subclases, la medida de la primera es mayor, o por lo menos igual, que la de la segunda, y tal que la medida de  $\Omega$  es positiva. Consideramos ahora una subclase particular (de  $\Omega$ ) conteniendo aquellas estructuras —y solamente aquellas que satisfacen una relación particular  $A$ . Sea  $\omega_A$  esta subclase particular (por ejemplo,  $\omega_A$  puede ser la subclase de todas aquellas estructuras que satisfacen una función de demanda particular). Decimos entonces que la relación “ $A$ ” es *autónoma* respecto a la subclase de estructuras  $\omega_A$ . Y decimos

que "A" tiene un cierto grado de autonomía que será tanto mayor cuanto mayor sea el "tamaño" de  $\omega_A$  cuando se compara con el de  $\Omega$ .

La labor principal de la teoría económica es establecer relaciones tales que se pueda esperar que posean grado de autonomía lo mayor posible.

Cualquier relación que se deduzca combinando dos o más relaciones dentro de un sistema se llama relaciones *confluentes*. Tales relaciones confluentes tienen, por otra parte, generalmente un grado de autonomía más bajo (nunca más alto) que el de las relaciones de las que se ha deducido, y todo lo más como la que lo tenga mayor de las relaciones diferente de las que depende. De un sistema de relaciones, con un cierto grado de autonomía, podemos deducir una infinidad de sistemas de relaciones confluentes. ¿Cómo podríamos distinguir el sistema "original" y un sistema de relaciones confluentes? Este no es un problema de independencia matemática, o, en forma más general, no es un problema de pura lógica, sino un problema de conocimiento *real* sobre el comportamiento del fenómeno y de hacer supuestos reales acerca de ellos. Cuando tratamos de establecer relaciones con alto grado de autonomía, tenemos en cuenta cambios en la estructura económica que podrían perturbar nuestras relaciones, trataríamos de obtener de estas relaciones lo que realmente esperaríamos tenga un gran grado de invarianza respecto a ciertos cambios en la estructura "razonables".

Es obvio que la autonomía de una relación de este tipo es un concepto muy relativo, en el sentido de que cualquier sistema de relaciones hipotéticas entre fenómenos reales puede deducirse de otro sistema más básico, es decir, un sistema con alto grado de autonomía respecto a los cambios estructurales.

La construcción de sistemas de relaciones autónomos es, por lo tanto, una materia de intuición, es un arte.

¿Cuál es la relación entre el grado de autonomía de una relación y su grado observable de constancia o persistencia?

Hay que señalar que constancia o persistencia indica sencillamente invarianza respecto ciertos cambios hipotéticos en la estructura, entonces el grado de constancia y el grado de autonomía podrían ser dos nombres diferentes de la misma propiedad de una relación económica. Pero si consideramos la constancia



de una relación como una propiedad del comportamiento de las *observaciones reales*, hay aquí una diferencia clave entre las dos propiedades a causa de que el grado de autonomía se refiere a una clase de variaciones hipotéticas en la estructura, para las cuales las relaciones podrían ser invariantes, mientras la persistencia depende de las variaciones que ocurran realmente. Por otra parte, si siempre tratáramos de formar relaciones autónomas con respecto a aquellos cambios que en efecto son *más probables que ocurran*, y si lo consiguiéramos entonces, desde luego, habría una conexión íntima entre la persistencia actual y el grado teórico de autonomía. Para poner de manifiesto estas ideas un poco más claramente todavía, estudiaremos una sección puramente formal.

Supongamos que tenemos un sistema económico, el mecanismo del cual puede caracterizarse por las variaciones de  $n$  cantidades medibles  $x_1, x_2, \dots, x_n$ . Supongamos que la estructura de este mecanismo puede describirse por un sistema de  $m < n$  ecuaciones:

$$f_i(x_1, x_2, \dots, x_n) = 0 \quad (i = 1, 2, \dots, m) \quad [8.1]$$

$(n - m)$  de las variables —podrían ser  $x_{m+1}, x_{m+2}, \dots, x_n$ — se suponen datos dados. El sistema [8.1] podría, por ejemplo, hacer posible expresar cada una de las  $m$  primeras variables unívocamente en función de las restantes  $n - m$ . Tal solución será

$$\begin{aligned} x_1 &= U_1(x_{m+1}, \dots, x_n) \\ x_2 &= U_2(x_{m+1}, \dots, x_n) \\ &\dots\dots\dots \\ x_m &= U_m(x_{m+1}, \dots, x_n) . \end{aligned} \quad [8.2]$$

El sistema [8.2] describiría las covariaciones de las variables tan bien como el sistema original [8.1]. Pero supongamos que hubiese ocurrido un cambio en la estructura del siguiente tipo. Una de las funciones  $f_i$  en [8.1], por ejemplo la  $f_1$ , se reemplaza por otra función  $f'_1$ , mientras todas las demás relaciones de [8.1] permanecen constantes. En general esto podría cambiar totalmente el

sistema [8.2], y si nosotros decimos que no ha cambiado el sistema [8.2] (por ejemplo, a causa de no conocer el sistema original [8.1]). Algunas o todas las relaciones mostrarían falta de constancia respecto a las observaciones que resultarían de la nueva estructura. Por otra parte, las últimas  $m - 1$  ecuaciones en [8.1] podrían —por definición— quedar inalteradas por los cambios estructurales. Puede suceder, como ocurre corrientemente, que una o dos ecuaciones de [8.1] rompan la estructura mientras las restantes permanecen válidas. Entonces cualquier sistema [8.2] correspondiente a un sistema fijo [8.1] podría presentar una pequeña persistencia respecto a las observaciones reales.

En este esquema, las variables  $x_{m+1}, x_{m+2}, \dots, x_n$  eran, por hipótesis, libres, podían moverse de una forma arbitraria. Esto incluía la posibilidad de que, por ejemplo, todas estas variables libres pudiesen moverse como funciones bien definidas del tiempo, por ejemplo:

$$\begin{aligned}
 x_{m+1} &= g_1(t) \\
 x_{m+2} &= g_2(t) \\
 &\dots\dots\dots \\
 x_n &= g_{n-m}(t)
 \end{aligned}
 \tag{8.3}$$

en la medida en que esto sea cierto podremos expresar las variables  $x_1, x_2, \dots, x_m$  como funciones de  $x_{m+1}, x_{m+2}, \dots, x_n$  de formas muy diferentes. Por ejemplo, será posible expresar  $x_1$  como una función de  $x_n$ , es decir

$$x_1 = F(x_n) .
 \tag{8.4}$$

¿Pero puede esta relación emplearse para juzgar el efecto sobre  $x_1$  de varios cambios arbitrarios en  $x_n$ ? Evidentemente no, ya que la existencia real de [8.4] descansa sobre el supuesto de que [8.3] se aplique. La relación [8.4] puede ser altamente inestable para tales cambios arbitrarios, y la persistencia eventual observada por [8.4] en el pasado, cuando [8.3] se aplicaba bien, puede no in-

dicar nada en esta nueva situación. En la próxima situación, el sistema origina [8.1], o incluso el sistema [8.2] puede todavía ser bueno si nosotros lo conocemos. Pero encontrar un sistema tal de relaciones altamente autónomas en el caso actual no es un proceso analítico, es una tarea consistente en hacer hipótesis fructíferas de cómo la realidad es en la actualidad.

Aclararemos estos puntos con dos ejemplos.

Primeramente consideraremos un esquema el cual creo que tiene relación con el problema de deducir curvas de demanda de las series de tiempo.

Sea  $x$  la tasa de consumo por cabeza de una mercancía en un grupo de individuos que tienen igual renta monetaria  $R$ . Sea  $p$  el precio de la mercancía y sea  $P$  el índice de coste de vida. Supongamos que la siguiente función de demanda es la actualmente verdadera:

$$x = a \frac{p}{P} + b \frac{R}{P} + c + \varepsilon \quad [8.5]$$

en donde  $a$ ,  $b$ ,  $c$  son ciertas constantes y  $\varepsilon$  es una variable aleatoria con "varianza bastante pequeña" y tal que los valores esperados de  $x$  son

$$E \left( x \mid \frac{p}{P}, \frac{R}{P} \right) = a \frac{p}{P} + b \frac{R}{P} + c. \quad [8.6]$$

Supongamos que [8.5] es autónoma en el siguiente sentido.

Para cualquier par de valores arbitrarios de  $\frac{p}{P}$  y  $\frac{R}{P}$ , el valor co-

rrespondiente de  $x$  puede estimarse por [8.6]. Supongamos que estamos interesados solamente en variaciones que son pequeñas respecto un cierto nivel de las variables. Entonces podemos aproximar [8.5] por una relación lineal de la siguiente forma: Sean  $p_0$ ,  $R_0$  y  $P_0$  los valores medios de  $p$ ,  $R$  y  $P$  respectivamente. Tendremos:

$$\begin{aligned}
 x &= a \frac{p_0 + (p - p_0)}{P_0 + (P - P_0)} + b \frac{R_0 + (R - R_0)}{P_0 + (P - P_0)} + c + \varepsilon = \\
 &= a \frac{P_0 + (p - P_0)}{P_0} \cdot \frac{1}{1 + \frac{P - P_0}{P_0}} + b \frac{R_0 + (R - R_0)}{P_0} + \\
 &+ \frac{1}{1 + \frac{P - P_0}{P_0}} + c + \varepsilon \simeq a \frac{p_0 + (p - p_0)}{P_0} \left( 1 - \frac{P - P_0}{P_0} \right) + \\
 &+ b \frac{R_0 + (R - R_0)}{P_0} \left( 1 - \frac{P - P_0}{P_0} \right) + c + \varepsilon = \frac{a}{P_0} p - \frac{a p_0}{P_0^2} P + \\
 &+ \frac{a p_0}{P_0} - \frac{a (p - p_0) (P - P_0)}{P_0^2} + \frac{b}{P_0} R - \frac{b R_0}{P_0^2} P + \frac{b R_0}{P_0} - \\
 &\quad - \frac{b (R - R_0) (P - p_0)}{P_0^2} + c + \varepsilon.
 \end{aligned}$$

Si las desviaciones  $(p - P_0)$ ,  $(P - P_0)$  y  $(R - R_0)$  son pequeñas comparadas con  $p_0$ ,  $P_0$  y  $R_0$ , podemos despreciar los términos en que aparecen dos productos de estas desviaciones. Obteniendo entonces

$$x = A p + B R + C P + D + \varepsilon' \tag{8.7}$$

en donde

$$A = \frac{a}{P_0}, \quad B = \frac{b}{P_0}, \quad C = - \left( \frac{a p_0}{P_0^2} + \frac{b R_0}{P_0^2} \right),$$

$$D = \frac{a p_0}{P_0} + \frac{b R_0}{P_0} + c$$

$\varepsilon'$  es un nuevo término residual que también contiene los errores introducidos por las anteriores aproximaciones. Para pequeñas variaciones de las variables  $\varepsilon'$  prácticamente puede no distinguirse de  $\varepsilon$ .

Lo que indicamos ahora es que si los datos  $p$ ,  $P$  y  $R$  que se emplean para deducir funciones de demanda tienden, por una u otra razón, a moverse como funciones regulares del tiempo, entonces puede entre estos datos existir otra relación que tenga exactamente la misma forma que [8.7], pero con diferentes coeficientes, y que puede ajustarse a los datos en general mejor que [8.7]; y si cambiamos esta otra relación por [8.7] tendremos solamente una relación confluyente, y no una aproximación de la función de demanda [8.5].

Para ver esto escribamos [8.5] así:

$$x(t) = a \frac{p(t)}{P(t)} + b \frac{R(t)}{P(t)} + c + \varepsilon(t). \quad [8.5'']$$

Supongamos ahora que las funciones del tiempo  $p(t)$ ,  $P(t)$  y  $R(t)$  —por algún motivo— son tales que satisfacen las relaciones funcionales.

$$\frac{p(t)}{P(t)} = k_1 p(t) + k_2 P(t) + k_0 \varepsilon \quad [8.8]$$

$$\frac{R(t)}{P(t)} = m_1 R(t) + m_2 P(t) + m_0 \quad [8.9]$$

en donde las  $k$  y las  $m$  son ciertas constantes. Una clase amplia de funciones elementales de tiempo satisfacen tales ecuaciones funcionales, y como quiera que este es el caso de las observaciones reales de  $p$ ,  $P$  y  $R$ , una ecuación de la forma [8.7] puede ajustarse a los datos. Pero no podremos emplear la ecuación así obtenida para predecir el efecto de un cambio arbitrario del precio, o de la renta, a causa de que esta ecuación no es en general una aproximación de [8.5] sino un resultado confluyente de [8.5], [8.8] y [8.9].

Y, por lo tanto, puede no aplicarse, por ejemplo, para los cambios en los precios, los cuales violan [8.8] y [8.9] o ambas.

En general, debemos tener mucho cuidado cuando empleemos un conjunto particular de datos en modificar la forma de las relaciones a las cuales se ha llegado en virtud de bases teóricas fuertes. Por ejemplo, en el caso anterior podríamos haber llegado a la conclusión que [8.7] era una forma más correcta de la función de demanda que [8.5], o por lo menos tan buena, mientras realmente, cuando [8.8] y [8.9] se cumplen totalmente, podemos obtener una relación de la forma [8.7] que no es una función de demanda, y que deja de cumplirse tan pronto como  $p(t)$ ,  $P(t)$  y  $R(t)$  tomen otra forma en el tiempo.

Como un ejemplo del problema de la autonomía de una relación económica con respecto a un *cambio en la política económica*, consideremos el modelo económico base de la famosa teoría de la tasa de interés y precio de la mercancía de la teoría wickselliana (con objeto de simplificarla y acortarla, haremos algunos supuestos más restrictivos que los de Wicksell. Nuestro modelo no hace, pues, justicia total a las profundas ideas de Wicksell).

Consideremos una sociedad donde sólo hay tres grupos económicos diferentes: *a*), individuos; *b*), Empresas privadas; *c*), Bancos. Supondremos que 1), todos los individuos dividen su renta en dos partes. Una consistente de gastos + dinero en casa; la otra parte se ahorra, y todos los ahorros se depositan en los Bancos. No hay otro ahorro en la sociedad. 2), toda la producción en la sociedad tiene lugar en las Empresas. Las Empresas son organizaciones impersonales, guiadas en su política de producción por los beneficios esperados solamente. Estas pueden hacer nuevas inversiones por medio de préstamos bancarios solamente. Distribuyen sus beneficios entre los individuos. 3), los precios de bienes y servicios de todas clases varían proporcionalmente a través del tiempo y pueden venir representados por una variable común llamada nivel de precios. 4), los Bancos tienen el poder de expendir o contraer el crédito. Supondremos que solamente hay una tasa monetaria de interés, la cual es la misma para todos los Bancos y para los préstamos y depósitos. (Esto da una descripción aproximada del modelo que estamos discutiendo; difícilmente es posible dar una descripción exhaustiva de un modelo con palabras. La descripción

precisa se da implícitamente a través de las restricciones impuestas al modelo.)

Estamos interesados principalmente en el efecto sobre el precio de ciertos cambios en la política crediticia de los Bancos.

Para ello introduzcamos las siguientes notaciones:

1.  $S(t)$  = ahorro total por unidad de tiempo.
2.  $I(t)$  = inversión total por unidad de tiempo.
3.  $\rho(t)$  = tasa de interés de los Bancos en el tiempo.
4.  $P(t)$  = nivel de precios en el tiempo  $t$ .
5.  $R(t)$  = renta nacional por unidad de tiempo.

Introduciremos ahora un sistema de relaciones fundamentales que describan el mecanismo de nuestro modelo. Consideremos relaciones lineales para mayor sencillez.

Primeramente supondremos que existe una función de oferta del mercado para el ahorro de la siguiente forma:

$$S(t) = a_0 + a_1 \rho(t) + a_2 P(t) + a_3 \dot{P}(t) + a_4 R(t) \quad [8.10]$$

La ecuación dice que la oferta de ahorro (depósito en los Bancos) —aparte una constante— depende de la tasa de interés. La renta total, el nivel de precios y la expectativa respecto al futuro valor real del dinero ahorrado, que está representado por la tasa de cambio en el nivel de precios  $P(t)$ . Es real suponer que  $a_1$  y  $a_4$  son positivos, y  $a_2$  y  $a_3$  negativos.

Consideremos ahora la siguiente función de demanda para los préstamos bancarios:

$$I(t) = b_0 + b_1 \rho(t) + b_2 P(t) + b_3 \dot{P}(t) \quad [8.11]$$

en donde  $b_1$  es negativa y  $b_3$  positiva, mientras el signo de  $b_2$  es incierto, a priori  $b_3$  debe ser positivo a causa de que, cuando el nivel de precios es creciente, las Empresas esperan vender factores de producción en un mercado menos expansivo que en el que últimamente vendían los productos acabados, y el elemento beneficio es un incentivo a la inversión.

Si los Bancos prestan a las Empresas un montante igual a los depósitos, nunca más ni menos, es decir si

$$I(t) = S(t) \quad [8.12]$$

se deduce de [8.10], [8.11] y [8.12] que a cada valor de  $R(t)$ ,  $P(t)$  y  $\dot{P}(t)$  corresponderá a una cierta tasa de interés de equilibrio en el mercado  $\bar{p}(t)$  llamado por Wicksell la tasa normal. Esto es, tendríamos:

$$\bar{p}(t) = \frac{b_0 - a_0}{a_1 - b_1} + \frac{b_2 - a_2}{a_1 - b_1} P(t) + \frac{b_3 - a_3}{a_1 - b_1} \dot{P}(t) - \frac{a_4}{a_1 - b_1} R(t) = A_0 + A_1 P(t) + A_2 \dot{P}(t) + A_3 R(t) \tag{8.13}$$

en donde  $\bar{p}(t)$  es el valor de  $\rho(t)$  que satisface [8.10], [8.11] y [8.12] y donde las A son las notaciones abreviadas para los coeficientes de los términos intermedios.

Si los Bancos quisieran expandir o contraer de forma activa la cantidad de dinero (esto es, si quisieran cambiar la cantidad de dinero que existe fuera de los Bancos) tendrían que fijar una tasa de interés  $\rho(t)$ , que diferiría de  $\bar{p}(t)$  definida por [8.13]. (Hay que hacer notar que  $\bar{p}(t)$  no indica una constante a lo largo del tiempo.) De [8.10] y [8.11] deducimos:

$$I(t) - S(t) = (b_0 - a_0) + (b_1 - a_1) \rho(t) + (b_2 - a_2) P(t) + (b_3 - a_3) \dot{P}(t) - a_4 R(t) \tag{8.14}$$

la cual, para  $\rho(t) = \bar{p}(t)$ , se reduce a

$$0 = (b_0 - a_0) + (b_1 - a_1) \bar{p}(t) + (b_2 - a_2) P(t) + (b_3 - a_3) \dot{P}(t) - a_4 R(t) . \tag{8.15}$$

Restando [8.15] de [8.14] obtenemos

$$I(t) - S(t) = (b_1 - a_1) [\rho(t) - \bar{p}(t)] \tag{8.16}$$

lo cual dice que el montante de "inflación monetaria"  $I(t) - S(t)$



proporcional (negativamente) a la diferencia entre la tasa de interés actual del Banco y la tasa normal definida por [8.13].

Suponiendo que la corriente de "inflación"  $I(t) - S(t)$ , tomada como un barómetro del gasto total, va acompañada por un aumento proporcional en el nivel de precios, tendremos:

$$\dot{P}(t) = k [I(t) - S(t)]. \quad [8.17]$$

Combinando [8.16] y [8.17] obtenemos:

$$\dot{P}(t) = k (b_1 - a_1) [\rho(t) - \bar{\rho}(t)] \quad [8.18]$$

lo cual es una expresión simplificada del teorema fundamental de Wicksell acerca del efecto en el precio de la tasa de interés bancaria cuando difiere de la normal.

Aceptando esta teoría (actualmente no estamos interesados en el análisis de su validez actual ni nada con ella relacionada, ya que lo empleamos solamente como ejemplo), ¿cuál puede ser el grado de autonomía de las tres ecuaciones [8.16], [8.17] y [8.18]?

Consideremos primero la ecuación [8.16]. Su validez en nuestra fundamentación se basa en las dos relaciones fundamentales [8.10] y [8.11]. Suprimiendo estas dos ecuaciones, diremos que no imponemos ninguna restricción a la forma en el tiempo de las funciones  $\rho(t)$ ,  $P(t)$  y  $R(t)$ . Por lo tanto, por hipótesis, cualquiera que sea la forma en el tiempo de estas funciones, la forma en el tiempo de  $I(t)$  y  $S(t)$  —y, por lo tanto, también la forma en el tiempo de  $I(t) - S(t)$ — se deduce de [8.10] y [8.11]. [ [8.16] es solamente otra forma de calcular la diferencia  $I(t) - S(t)$ . ] De [8.13] se deduce que a cada par de funciones en el tiempo  $P(t)$  (siempre que tenga derivada  $\dot{P}(t)$ ) y  $R(t)$  corresponde una función del tiempo  $\bar{\rho}(t)$ , mientras que a cada función del tiempo dada  $\bar{\rho}(t)$  corresponde en general una infinidad de funciones del tiempo  $P(t)$  y  $R(t)$ . La ecuación [8.16] es, por lo tanto —por hipótesis—, autónoma en el siguiente sentido: Para cualesquiera funciones del tiempo arbitrariamente elegidas  $\rho(t)$  y  $\bar{\rho}(t)$ , la inflación del crédito  $I(t) - S(t)$  se calculará a partir de [8.16].

Vemos que esta propiedad de [8.16] —si es cierta— no es una propiedad matemática de la ecuación. No puede funda-

mentarse considerando la ecuación. Se apoya en una hipótesis respecto como la diferencia  $I(t) - S(t)$  se comporta de *facto* para varios cambios en la tasa de interés y la tasa normal. Con otro modelo podríamos obtener una ecuación exactamente de la misma forma, pero sin la misma propiedad de autonomía. Por ejemplo, supongamos que —como una consecuencia de algún modelo, cualquiera que sea, el razonamiento económico particular sobre el que se apoya— todas las funciones anteriores del tiempo tuviesen una tendencia lineal. En particular, supongamos que  $I(t) - S(t) = nt$ ,  $\rho(t) - \bar{\rho}(t) = nt$ . Tendremos entonces:

$$I(t) - S(t) = \frac{m}{n} [\rho(t) - \bar{\rho}(t)] \quad [8.19]$$

el cual es de la forma [8.16]. Pero de [8.19] no podríamos calcular el efecto sobre  $I(t) - S(t)$ , es decir, varios tipos de política de interés, a causa de que cualquier cambio en  $\rho(t)$  que violara la condición  $\rho(t) - \bar{\rho}(t) = nt$ , podría romper el fundamento sobre el que [8.19] reposa. La ecuación [8.19] podría, no obstante, emplearse después de tal violación, pero tendría que deducirse de otro modelo.

La ecuación [8.17] representa, por sí, una relación autónoma respecto a ciertos cambios en la estructura. Es una hipótesis de independencia acerca del nivel de precios, que dice, cualquiera que sea la inflación del crédito  $I(t) - S(t)$ , podremos calcular la correspondiente razón de cambios del nivel de precios. Aquí no podemos conocer si esta propiedad de autonomía es cierta realmente. Es un supuesto, y es tarea de la teoría económica y el investigador justificarlo.

Hemos establecido que [8.16] y [8.17] son, en efecto, relaciones altamente autónomas. ¿Cuál es la situación respecto a la ecuación [8.18]? Evidentemente [8.18] tendrá un menor grado de autonomía que [8.16] y [8.17] separadamente, ya que la clase de funciones de tiempo que satisface [8.18] es —por definición— solamente la clase de funciones que satisfacen [8.16] y [8.17] conjuntamente.

No hemos supuesto relaciones definidas que describan la política crediticia de los Bancos. Meramente hemos descrito el com-

portamiento de individuos o Empresas en respecto a una tasa de interés bancario. Partiendo de ciertos supuestos, tales como el de la buena voluntad del ahorro y el interés, y suponiendo que un aumento de crédito extra en el mercado causa un cambio proporcional en el nivel de precios, hemos obtenido dos relaciones estructurales [8.16] y [8.17]. La variable  $\rho(t)$  se consideró como un parámetro libre. Podría quizá ocurrir que los Bancos, en un cierto período de tiempo por lo menos, decidan seguir una cierta conducta en su política del interés, o que la hagan para asegurar su liquidez particular. Sobre *este período* de tiempo puede suceder que se pudiese añadir una nueva relación a las anteriores, una relación describiendo —temporalmente— la política bancaria. Supongamos, por ejemplo, que los Bancos, en un cierto período de tiempo, proceden como sigue. Cuando comprueban que  $I(t) - S(t)$  tiene un valor positivo proceden a alzar la tasa de interés, con objeto de proteger su liquidez, y, recíprocamente, bajan la tasa de interés cuando  $I(t) - S(t)$  es negativo. Tal política puede describirse por la relación

$$\bar{\rho}(t) = c [I(t) - S(t)] \quad [8.20]$$

en donde  $c$  es una constante positiva. A causa de [8.16] tendremos:

$$\bar{\rho}(t) = c (b_1 - a_1) [\rho(t) - \bar{\rho}(t)]. \quad [8.21]$$

Y combinando [8.18] y [8.21] tendremos:

$$\dot{P}(t) = \frac{k}{c} \dot{\rho}(t). \quad [8.22]$$

La cual, aparentemente, dice que el nivel de precios se mueve en la misma dirección que la tasa de interés. ¿Pero podríamos emplear esta relación para calcular lo que “debía ser” el efecto sobre el nivel de precios de alguna política arbitraria del interés? Evidentemente *no*, ya que [8.22] se aplica solamente cuando

$$R(t), I(t), S(t), P(t)$$

son funciones del tiempo tales que satisfacen simultáneamente [8.13],

[8.16], [8.17] y [8.20]. Por lo tanto [8.22] no se emplea para juzgar el efecto de cambio en la política del interés. Para obtener con este propósito una ecuación, podríamos combinar [8.13] y [8.18]. Lo cual daría una relación de la forma

$$\dot{P}(t) + B P(t) = H_1 \rho(t) + H_2 R(t) + H_0 \quad [8.23]$$

en donde  $B$ ,  $H$ ,  $H_1$ ,  $H_2$  y  $H_3$  son constantes dependientes de las de [8.13] y [8.18]. Aquí —por hipótesis— no hay restricciones en la forma en el tiempo de las funciones  $\rho(t)$ ,  $R(t)$ . Podemos elegir tales funciones arbitrariamente y resolver la ecuación [8.23] para obtener  $P(t)$  como una función explícita de  $\rho(t)$  y  $R(t)$ .

¿Pero cómo hacerlo cuando no sabemos que sea [8.20] la que hay que aplicar en vez de la [8.22]? No hay un método formal en virtud del cual se establezca tal conclusión. En efecto, partiendo de cualquier otro modelo con diferentes supuestos, podríamos alcanzar conclusiones opuestas. Para alcanzar una decisión tenemos que conocer o imaginar —sobre la base de la experiencia general— cuál de las dos relaciones [8.22] u [8.23] debe ser, *en efecto*, la más estable, si una hubiese de emplearse como relación autónoma.

Resumamos la discusión del problema de las relaciones autónomas: En la investigación científica —en el campo de la Economía así como en otros campos— nuestra búsqueda de “explicaciones” consiste en extraer las relaciones más fundamentales entre las que aparecen ante nosotros cuando “contemplamos y miramos”. Cada una de estas relaciones fundamentales la concebimos como invariante respecto a una clase mucho más amplia de variaciones que las particulares que se nos muestran en el curso natural de los hechos. Ahora bien, aunque el fenómeno real que observamos día por día esté realmente regulado por la acción simultánea de un sistema total de leyes fundamentales, vemos solamente un poco de la clase total de variaciones hipotéticas para las cuales cada una de las relaciones fundamentales puede suponerse que se aplica. (Este hecho también se relaciona con el problema más serio de estimar relaciones fundamentales de las observaciones. Este problema se discutirá totalmente en el capítulo V.) Para las variaciones que observamos, si es posible establecer una infinidad de relacio-

nes sencillas combinando dos o más relaciones fundamentales de varias formas. En particular podrá ser posible expresar una variable económica en función de un conjunto de otras variables en una gran variedad de formas. Establecer, por lo tanto, que una variable económica es "alguna función" de un cierto conjunto de otras variables no indica mucho, solamente especifica en qué "medio" la relación se supone que se aplica. Esto es exactamente otro aspecto de la regla general que hemos dado de lado al comienzo de este capítulo. La regla de que toda teoría debe acompañarse por un diseño de experimentos.

### CAPITULO III

#### ESQUEMAS ESTÓCASTICOS COMO UNA BASE PARA LA ECONOMETRIA

Sabemos por experiencia que intentar establecer relaciones exactas funcionales entre variables económicas observables es fútil e inútil. Desde luego sería extraño que esto no ocurriera porque entonces se encontrarían los economistas en una posición más favorable que cualquier otra persona dedicada a la investigación, incluyendo a los astrónomos. Las observaciones reales en cualquier campo que consideremos se desviarán en mayor o menor grado de una relación funcional exacta que pudiéramos tratar de establecer. Por otra parte, como hemos visto, la contrastación de una teoría implica la identificación de sus variables con algunas variables "verdaderas" observables. Si en un caso dado creemos, incluso sin demostrarlo, que tal identificación no es la apropiada, esto no es más que otra forma de decir que la teoría es falsa respecto a las variables verdaderas estudiadas. Para que la contrastación de una teoría tenga alguna significación, tenemos que estar de acuerdo en la identificación de las variables teóricas con las observables, y entonces ver si las observaciones contradicen o no la teoría.

Podríamos, por lo tanto, a priori, decir algo sobre una teoría que creemos podría ser cierta respecto a un sistema de variables observables, es decir, que no debe *excluirse como imposible* cual-

quier sistema de valores de las variables "verdaderas" que ya hayamos observado o que puede esperarse se puedan obtener en el futuro. Pero las teorías que describen meramente un conjunto de valores de las verdaderas variables que concebimos como prácticamente posibles, apenas nos dirían nada del empleo práctico. Tales afirmaciones serían demasiado amplias. Lo que necesitamos son teorías que, sin tener contradicciones lógicas directas, afirmen que las observaciones estarán agrupadas por *regla* general en un subconjunto limitado del conjunto de todas las observaciones posibles, mientras son consistentes con la posibilidad que una observación caiga fuera de este subconjunto alguna vez.

Como se sabe, el esquema de probabilidades y variables aleatorias es, al menos por ahora, el único esquema conveniente para formular tales teorías. Podemos hacer objeciones al empleo de este esquema, pero entre esas objeciones hay por lo menos una que puede eliminarse totalmente, a saber, la objeción de que el esquema de probabilidad y variables aleatorias no es lo suficientemente general para su aplicación a los datos económicos. Sin embargo, como aparentemente esto no está aceptado en general por los economistas, justifica que comencemos nuestro estudio en este capítulo con un buen diseño de la teoría moderna de las variables aleatorias, dando importancia especial a ciertos puntos que parecen relevantes en la Economía.

### 9.—Probabilidad y variables aleatorias

Los más recientes desarrollos de la teoría estadística se basan en lo que se llama la teoría clásica de la probabilidad modernizada. En ella la probabilidad se define como una *función de conjunto* (3)

---

(3) Ver, por ejemplo, STANISLAW SAKS: *Theory of the Integral*. New York, 1937, y NICOLÁS LUZIN: *Les ensembles analytiques*. París, 1930.

Haremos uso frecuente de las siguientes notaciones y definiciones comunes a la teoría de conjuntos:

Si  $A$  es un conjunto de elementos u objetos  $a$ , el símbolo  $a \in A$  se emplea para indicar que  $a$  es un elemento de  $A$  o que pertenece a  $A$ .

Sea  $(A)$  una familia de conjuntos  $A$  y sean  $A_1$  y  $A_2$  los miembros de  $(A)$ .

no negativa y absolutamente aditiva que satisface ciertas propiedades formales (4).

Daremos primero un ejemplo con objeto de aclarar el concepto de probabilidad. Consideremos un dado ordinario con seis caras. Para los propósitos del cálculo de probabilidades, un dado puede

Si todo elemento de  $A_1$  es también un elemento de  $A_2$  diremos que  $A_1$  contiene o cubre a  $A_2$ .

El símbolo  $A_1 + A_2$  (llamado la suma lógica de  $A_1$  y  $A_2$ ) indica el conjunto de todos los elementos  $x$  que pertenece a *uno por lo menos* de los dos conjuntos  $A_1$ ,  $A_2$ .  $A_1 \cdot A_2$  (llamado el producto lógico de  $A_1$  y  $A_2$ ) indican el conjunto de todos aquellos elementos  $a$  que pertenecen a  $A_1$  y  $A_2$  (es decir, su punto común). Estas nociones de sumas y productos pueden extenderse a cualquier sucesión de conjuntos finita o infinita.

Si un producto  $A_1 \cdot A_2$  es vacío,  $A_1$  y  $A_2$  se dice son disjuntos. Si  $A_1$  contiene a  $A_2$ ,  $A_1 - A_2$  se llama diferencia entre  $A_1$  y  $A_2$  e indica el conjunto de elementos que pertenecen a  $A_1$  pero no a  $A_2$ .

Una familia de conjuntos tal que (1) la suma de cualquier sucesión numerable de conjuntos disjuntos miembros de la familia, así como (2) la diferencia  $A_1 - A_2$  de dos cualesquiera miembros de ellos está contenida en la familia, se llama cuerpo de Borel. Y los indicaremos por  $/A/$ .

Supongamos que asociemos con cada miembro  $A$  de  $/A/$  un número finito  $F(A)$ ; entonces  $F(A)$  se llama función de conjunto. (Por ejemplo, si  $A$  es un intervalo en una línea recta, su longitud es una función de conjunto.) La función  $F(A)$  se llama aditiva si para cualquier par de conjuntos disjuntos  $A_1$ ,  $A_2$  en  $\{A\}$  se tiene

$$F(A_1 + A_2) = F(A_1) + F(A_2)$$

$F(A)$  se llama absolutamente aditivo si para una sucesión numerable de subconjuntos *disjuntos* en  $\{A\}$  tenemos

$$F(A_1 + \dots + A_n + \dots) = F(A_1) + \dots + F(A_n) + \dots$$

Por medida de un conjunto  $A$ , perteneciendo al cuerpo  $\{A\}$ , entendemos una función de conjunto absolutamente aditiva, en  $A$  tal que

$$m(A) \geq 0 \quad \text{y} \quad m(\emptyset) = 0$$

cuando  $A$  esté vacío. La longitud áreas y volúmenes son sencillos ejemplos de medidas.

(4) Ver, por ejemplo, J. NEYMAN: *Lectures and Conferences on Mathematical Statistics*. Washington, 1937, págs. 2-18; "L'estimation statistique traitée comme un problème classique de probabilité". *Actualités scientifiques et industrielles*, 739 Conférence internationale de Sciences mathématiques. Paris, 1938, págs. 25-27, y PAUL SEVY: "Théorie de l'addition des variables aleatoires", Paris, 1937; S. S. WILKS: "Statistical Inference 1936-37". Princeton, N. Y., 1937. *Mathematical Statistics* Princeton, 1943.

describirse como un conjunto de seis puntos en una línea recta,  $x = 1, 2, \dots, 6$ . Consideremos ahora todos los puntos de una línea recta de  $-\infty$  a  $+\infty$ ; sobre este conjunto de puntos (es decir, sobre todo el eje real) definimos una *función* medible real no negativa (o un sistema de ponderaciones del siguiente tipo).

1. En el punto  $x = 1$  adscribimos una medida  $P_1$ , en el punto  $x = 2$  adscribimos una medida  $P_2$ , etc. y en el punto  $x = 6$ , por último, adscribimos una medida  $P_6$  tal que

$$P_i \geq 0 \quad i = 1 \dots 6 \quad \text{y tal que} \quad P_1 + P_2 + \dots + P_6 = 1.$$

2. Si  $\omega$  es cualquier subconjunto de puntos (por ejemplo, un intervalo) en el eje  $x$ , la medida  $P(\omega)$  del conjunto  $\omega$  se define como la suma de las medidas  $P_i$  de aquellos puntos cualquiera, entre los seis puntos particulares  $x = 1, 2, \dots, 6$  que pertenezcan al conjunto  $\omega$  (por ejemplo, la medida de un conjunto  $\omega$ , definido por  $1 \leq x < 4$  será  $P_1 + P_2 + P_3$ ).

3. Si  $\omega$  no contiene ningún punto  $x = 1, 2, \dots, 6$ , entonces, para cualquier  $\omega$  en que eso ocurra,  $P(\omega) = 0$  (por ejemplo, si  $\omega$  es el intervalo  $0 \leq x \leq 1/2$ , entonces  $P(\omega) = 0$ ).

$P(\omega) = \theta$ , definido de esta forma, se llama probabilidad de que un punto  $x$  pertenezca al conjunto de puntos  $\omega$ , abreviadamente la probabilidad de  $\omega$ . Se sigue de lo anterior que si  $\omega$  es todo el eje real, entonces  $P(\omega) = 1$ . Si  $\omega$  contiene solamente el punto  $x = 1$ , o  $x = 2$ , ... o  $x = 6$ , entonces  $P(\omega) = P_1$ , o  $P_2$ , o ...  $P_6$  respectivamente.

Consideremos ahora  $n$  dados, numerados  $1, 2, \dots, n$  (o  $n$  tiradas hipotéticas con el mismo dado), teniendo todos el mismo sistema de probabilidad  $P_1, P_2, \dots, P_6$ . Sea  $x_i$  el resultado de una tirada con el dado  $i$ ésimo ( $i = 1, 2, \dots, n$ ), es decir  $x_i = 1$  ó  $2$  ó ...  $6$ , con las probabilidades  $P_1, P_2, \dots, P_6$  respectivamente, todos los otros valores de  $x_i$  tienen la probabilidad cero. Consideremos cualquier sistema posible  $(x_1, x_2, \dots, x_n)$  de valores de las  $n$  variables  $x$ , una para cada dado. Cualquiera de tales sucesiones  $x_1, x_2, \dots, x_n$  pueden representarse por un punto en un espacio euclídeo  $n$ dimensional.

Si definimos la probabilidad de cualquiera de tales puntos como el producto de las probabilidades de cada  $x_i$  separadamente, podremos calcular la probabilidad de punto arbitrario  $(x_1, x_2, \dots, x_n)$ , o,



más generalmente, la probabilidad de cualquier conjunto de puntos arbitrarios en el espacio enedimensional. Es fácil ver que el sistema de todas estas probabilidades satisface unas condiciones exactamente iguales a las (1)-(3) anteriores. La única diferencia es que ahora consideramos puntos de un espacio enedimensional en lugar de puntos sobre una línea recta. Por ejemplo, podríamos calcular la probabilidad de que exactamente  $k$  (no importa cuales), entre las  $n$  variables, tenga el valor 6, es decir, la probabilidad de un punto  $(x_1, x_2, \dots, x_n)$ , teniendo exactamente  $k$  de sus coordenadas iguales

a 6, esta probabilidad es la suma de  $\frac{n!}{k!(n-k)!}$ , productos cada uno de ellos iguales a  $P_6^k (1 - P_6)^{n-k}$ , o

$$\frac{n!}{k!(n-k)!} P_6^k (1 - P_6)^{n-k}$$

Lo cual es, por otra parte, también la probabilidad de una proporción de seis igual a  $\frac{k}{n}$ . De la fórmula [9.a] podemos calcular la probabilidad total de un conjunto de puntos en el espacio  $x$  enedimensional correspondiente al sistema total de valores de  $k$  sencillamente sumando las probabilidades [9.a] para estos valores de  $k$ . De aquí que podamos también calcular, por ejemplo, la probabilidad  $P$ , de

$$P_6 - \varepsilon \leq k/n \leq P_6 + \varepsilon$$

en donde  $\varepsilon$  es cualquier número positivo. Se deduce de la fórmula [9.a], como es bien sabido, que si  $P_6$  es un número finito y si un  $\varepsilon$  positivo se elige, no importa cómo sea de pequeño, entonces  $P$  puede aproximarse a 1 tanto como se quiera, eligiendo  $n$  suficientemente grande.

¿Cuál es la utilidad de tal aparato formal, o, en otras palabras, hay contrapartida en el mundo real?

Antes de nada asignemos un significado práctico a la noción teórica "Una probabilidad próxima a 1". Por esta afirmación—cuando se aplica a fenómenos reales—entendemos "certeza práctica", esto es, cuando decimos en la teoría que la probabilidad

de un suceso es próxima a 1, indica que en las aplicaciones prácticas estamos "casi seguros" que el suceso ocurrirá.

Apliquemos esto al ejemplo anterior del dado. Si la probabilidad de un "seis" es  $P_6$  (no tiene necesariamente que ser  $1/6$ ), entonces el cálculo de probabilidades dice que para un  $n$  suficientemente grande la probabilidad de que una proporción  $k/n$  de seises en  $n$  tiradas independientes esté próxima a  $P_6$  es casi 1. Trasladado al lenguaje práctico esto indica: Si tiramos un dado  $n$  veces cuando  $n$  es un número grande, por ejemplo,  $n = 10.000$ , y obtenemos una proporción  $k_1/n_1$  de seises, prácticamente estamos seguros que en un nuevo gran número de tiradas con ese dado  $n_2 = 10.000$  la proporción  $k_2/n_2$  de seises estará próxima a  $1/5$ . Así, por ejemplo, si hubiéramos obtenido  $k_1/n_1 = 1/5$  para las primeras 10.000 tiradas, y decimos que  $k_2/n_2 = 2/5$  en las segundas 10.000 tiradas, tendremos que pensar en investigar lo que pasa en el dado y procedimiento de lanzamiento, ya que estábamos casi seguros, tomando como base un gran número de experimentos realizados en el pasado, que "algo está mal".

Investigaciones empíricas han puesto de manifiesto que ciertas cosas en el mundo real aparecen solamente raras veces, son "milagros", mientras otras son sucesos corrientes. El cálculo de probabilidades ha llenado totalmente el deseo que había de tener un aparato lógico formal que se refiriese a tales fenómenos de la vida real. La cuestión no reside en cuánta probabilidad de éxito existe; sin embargo —si procedemos *como* si existiese—, somos capaces de hacer afirmaciones acerca de los fenómenos del mundo real que son "correctas para propósitos prácticos".

Los ejemplos anteriores pueden servir para aclarar el comportamiento de la probabilidad y el cálculo de probabilidades. Daremos ahora una definición más general de probabilidad.

Sea  $A$  un conjunto (finito o infinito) de objetos especificados de cualquier clase (por ejemplo, un conjunto de puntos en una cierta región del espacio). Sea  $A_x = A \cdot A_x$  un subconjunto de  $A$  consistente en aquellos elementos de  $A$  que posean una cierta propiedad  $X$  entre un sistema de propiedades  $X$ , tales que la familia de todos los conjuntos correspondiente  $A \cdot A_x$  forma un cuerpo de Borel  $\{A \cdot A_x\}$  y tal que  $A \in \{A \cdot A_x\}$ . Suponemos, además, que hemos definido una medida  $m(A \cdot A_x) \geq 0$  dentro de  $\{A \cdot A_x\}$ , tal que  $m(A) > 0$  y

$m(A \cdot A_x) = 0$  cuando  $A \cdot A_x$  es vacío. El conjunto  $A$  se dice que es probabilizable (Neyman).  $A$  se llama el conjunto fundamental de la probabilidad. Para cualquier elemento  $A \cdot A_x$  de  $\{A \cdot A_x\}$  definimos:

$$P(X|A) = \frac{m(A \cdot A_x)}{m(A)} \quad [9.1]$$

como la probabilidad de un elemento de  $A$  teniendo la propiedad  $X$ . De la definición de cuerpo de Borel y de la definición de la medida  $m(A \cdot A_x)$  se deduce que

$$0 \leq P(X|A) \leq 1 \quad \text{y} \quad P(\bar{X}|A) + P(X|A) = 1$$

en donde  $\bar{X}$  es la propiedad *no*  $X$ .

Cualquier variable real definida como una función unívoca medible de los elementos de un conjunto probabilizado  $A$  se llama *variable aleatoria*. Como un caso particular  $x = x^0 = \text{constante}$  puede tener la probabilidad 1 mientras los otros valores de  $x$  tienen la probabilidad 0. Entonces  $x$  es una constante en el sentido estocástico. Los valores de  $x$  pueden considerarse como propiedades de los elementos de  $A$ .

Una función  $x$  de los elementos del conjunto  $A$  es medible si el subconjunto de  $A$  dado por  $x < c$  es medible en la medida de la probabilidad definida para todo valor finito de  $c$ . Por lo tanto, cualquiera que sean los números reales  $c_1 < c_2$ , la definición de  $A$  y  $x$  determina unívocamente la probabilidad

$$P(c_1 \leq x < c_2 | A) \quad [9.2]$$

y es siempre posible encontrar  $c_1$  y  $c_2$  de tal forma que

$$0 < P(c_1 \leq x < c_2 | A) \leq 1. \quad [9.3]$$

Para cualquier  $c_1$  fijo  $P(c_1 \leq x < c_2 | A)$  es una función de  $c_2$  monótona no decreciente, a lo que se llama la *ley integral de probabilidad de  $x$* .

La definición anterior de probabilidad y de variable aleatoria es prácticamente equivalente a la siguiente definición más directa:

sea  $x$  una variable real; sus valores pueden representarse por puntos de una línea recta de  $-\infty$  a  $+\infty$ . Sea  $\{\omega\}$  un cuerpo de Borel de conjuntos medibles  $\omega$ , en esta línea, tal que, en particular  $\{\omega\}$  contiene el sistema de todos los intervalos  $c_1 \leq x \leq c_2$ , donde  $c_1 < c_2$  son un par de números reales. Sea  $P(\omega)$  una función de conjunto definido sobre  $\{\omega\}$ , tal que  $P(\omega)$  es (1) no negativa, (2) absolutamente aditiva y (3) igual a 1 si  $\omega$  contiene todos los puntos  $x$  de  $-\infty$  a  $+\infty$ . Entonces esto define a  $x$  como una variable aleatoria tal que la probabilidad de que  $(x \in \omega)$  viene dada por  $P(\omega)$ .

Si existe una función medible, en el sentido de Lebesgue, no negativa de  $p(x)$  tal que para todo intervalo  $(c_1, c_2)$

$$P(c_1 \leq x < c_2 | A)$$

está definida, podremos poner

$$P(c_1 \leq x < c_2 | A) = \int_{c_1}^{c_2} p(x) dx \quad [9.4]$$

siendo la integral la de Lebesgue, entonces  $p(x)$  se llama la ley de probabilidad elemental (o función de densidad de probabilidad) de  $x$ .

En estadística corrientemente consideraremos diversos sistemas de variables aleatorias. Hay dos tipos principales de tales sistemas y —aunque realmente no son diferentes desde el punto de vista de la estadística metodológica— la distinción entre ellos aparece cuando comparamos un modelo hipotético con las observaciones.

El primer tipo se refiere a un sistema de diversas variables aleatorias  $x_1, x_2, \dots, x_n$ , asociadas con cada elemento de un conjunto fundamental de probabilidad (por ejemplo, el conjunto fundamental de probabilidad puede ser todas las personas que vivan en los Estados Unidos durante todo el año 1940,  $x_1$  puede ser la renta por persona,  $x_2$  su fortuna privada, etc.). Para cada elemento del conjunto fundamental de probabilidad, el sistema de valores  $x_1, x_2, \dots, x_r$  puede representarse por un punto  $E_r$  de un espacio  $r$ -dimensional  $R_r$ . Si  $\omega$  es cualquier conjunto medible de puntos en  $R_r$ , indicaremos por

$$P(E_r \in \omega | A) \quad [9.5]$$

o, abreviadamente,  $P(\omega)$ , la probabilidad de que un punto arbitrario  $E_r$  pertenezca a  $\omega$ . (A continuación emplearemos la notación abreviada  $P(\omega)$  para aquellos casos en que no existe peligro de confusión en el espacio variable considerado.)  $P(\omega)$ , considerada como una función del conjunto  $\omega$ , se llama la *integral simultánea de la ley de probabilidad* de  $x_1, x_2, \dots, x_n$ , dentro del conjunto fundamental de probabilidad  $A$ .

Se habrá notado que empleamos el mismo símbolo  $P$  para indicar dos cosas diferentes, a saber: 1), un número, y 2), una función. Si el argumento  $\omega$  es un conjunto fijo de puntos  $\omega$ , entonces  $P(\omega)$  indica un número llamado la probabilidad de  $\omega$ . Si  $\omega$  se considera como un argumento variable arbitrario, entonces  $P(\omega)$  indica la función de probabilidad. Emplearemos letras particulares y subíndices, etc., para indicar conjuntos fijos en el espacio-variable en cuestión y que no puedan presentarse confusiones.

Si existe una función medible de Lebesgue no negativa  $p(x_1 \dots x_r)$  tal que para todo  $\omega$  para el que  $P(\omega)$  esté definido, tendremos:

$$P(\omega) = \int \int \dots \int_{(\omega)} p(x_1 \dots x_r) dx_1 \dots dx_r \quad [9.6]$$

entonces  $p(x_1, x_2, \dots, x_r)$  se llama la *ley de probabilidad conjunta elemental* de  $x_1, x_2, \dots, x_r$ .

Sean  $p_1(x_1), p_2(x_2), \dots, p_r(x_r)$  las leyes de probabilidad elemental de la  $r$  variables  $x$  tomadas separadamente (es decir, la distribución marginal de las  $x$ ) dentro de  $A$ . Si

$$p(x_1, x_2, \dots, x_r) = p(x_1) p_2(x_2) \dots p_r(x_r) \quad [9.7]$$

las variables  $x_1, x_2, \dots, x_r$  se dice son estocásticamente *independientes*.

El segundo tipo de sistemas de variables aleatorias se refiere al *muestreo aleatorio*. Consideremos un conjunto fundamental de probabilidad  $A$ ; cada elemento del cual viene caracterizado por los valores de  $r$  variables aleatorias  $x_1, x_2, \dots, x_r$  y supongamos que fijamos una cierta regla por la que tomamos un sistema de  $s$  elementos de  $A$ . Sean  $(x_{11}, x_{21}, \dots, x_{r1})$  el sistema de valores del primer elemento extraído  $(x_{12}, x_{22}, \dots, x_{r2})$  el segundo elemento, y

así sucesivamente. Indiquemos por  $B_i$  el subconjunto de  $A$  correspondiente a todos los posibles sistemas de valores a priori  $(x_{1i}, x_{2i}, \dots, x_{ri})$  cuando el elemento extraído es el  $i$ ésimo ( $i = 1, 2, \dots, s$ );  $B_i$  puede considerarse como el conjunto fundamental de las variables aleatorias  $x_{1i}, x_{2i}, \dots, x_{ri}$ .

El sistema

$$\begin{aligned} & (x_{11}, x_{21}, \dots, x_{r1}) \\ & (x_{12}, x_{22}, \dots, x_{r2}) \\ & \dots\dots\dots \\ & (x_{1s}, x_{2s}, \dots, x_{rs}) \end{aligned} \tag{9.8}$$

se llama *muestra de tamaño s* del conjunto fundamental de probabilidad  $r$ -variante  $A$ , o lo que significa lo mismo,  $s$  muestras de una observación cada una, que dan un sistema de valores  $(x_1, x_2, \dots, x_r)$  para cada conjunto fundamental de probabilidad  $B_i$ . La distribución conjunta de probabilidad de  $(x_{1i}, \dots, x_{ri})$  evidentemente pueden cambiar con  $i$ . El sistema [9.8] puede también considerarse como una muestra de *una* observación extraída de una población  $r$ . s. llamada  $B$ . Cada elemento de  $B$  entonces vendrá caracterizado por un conjunto de valores de la variable aleatoria  $r$ . s. indicada en [9.8], y la distribución de probabilidad asociada a  $B$  será de  $r$ . s. dimensiones. Cada sistema de valores [9.8] puede representarse por un punto  $E$ , en un aspecto euclídeo de  $r$ . s. dimensiones. Al punto  $E$  se le llama punto muestra, o un punto en el espacio muestra de  $r$ . s.-dimensiones.

Por muestra aleatoria corrientemente se entiende un experimento tal que los varios conjuntos  $E_i = (x_{1i}, \dots, x_{ri})$  ( $i = 1, 2, \dots, s$ ) de [9.8] son independientes, es decir tales que existe la ley de probabilidad elemental.

$$p(E) = p_1(E_1) p_2(E_2) \dots p_s(E_s) \tag{9.9}$$

la dependencia o independencia dentro de cada sistema

$$E_i = (x_{1i}, \dots, x_{ri})$$

corrientemente viene "dada por la Naturaleza".

Cuando la ley de probabilidad (integral o elemental) de un

sistema de variables aleatorias es conocida, hay unas reglas matemáticas para deducir las leyes de probabilidad de funciones de estas variables (ver, por ejemplo, S. U. Uspensky. "Mathematical Probability", New York, 1937).

#### 10.—*El comportamiento práctico de las afirmaciones de probabilidad*

Al principio de la sección precedente hemos dado un ejemplo sencillo del comportamiento práctico de las afirmaciones de probabilidad. Vamos ahora a dar una interpretación más general de tales afirmaciones.

Supongamos que conocemos que las  $n$  variables observables  $x_1, x_2, \dots, x_n$  tienen una ley conjunta de probabilidad elemental  $p(x_1, x_2, \dots, x_n)$ . ¿Cuáles son las afirmaciones prácticas que podemos hacer acerca del conjunto de valores  $(x_1, x_2, \dots, x_n)$  que estamos observando? Se ha encontrado que es fructífero en muchos campos de investigación emplear la "frecuencia de ocurrencia" observada de un suceso como contrapartida práctica de la noción puramente teórica de probabilidad. Esto es, si la ley elemental de probabilidad  $p$  implica que la probabilidad de una cierta región o conjunto  $w$ , en el espacio  $x$ ,  $n$ -dimensional es  $P(w)$  diremos que para observaciones repetidas de puntos  $(x_1, x_2, \dots, x_n)$  en el espacio  $x$ , la frecuencia relativa de los puntos que caen en  $w$ , para un gran número de puntos observados, será próxima a  $P(w)$ .

Sin embargo, por lo general no tenemos un interés particular en hacer afirmaciones acerca de un número muy grande de observaciones. Corrientemente estamos interesados en afirmaciones que puedan hacerse con un número relativamente pequeño de puntos observados; o quizá, aún más frecuentemente, estamos interesados en una afirmación *a priori* acerca de una nueva observación solamente. Entonces es de relativamente poco valor práctico conocer cuánto vale  $P(w)$ , esto es, si será 0,4, 0,5 ó 0,6. Ya que entonces no podremos tener una gran confianza en la afirmación de que el próximo punto observado caerá fuera de  $w$ . Con objeto de poder hacer una afirmación útil, la situación debe ser tal que exista un subconjunto  $w$  "interesante" para el cual la probabilidad  $P$  sea

próxima a 1, la interpretación práctica sería que "casi todas" las observaciones caerán en él. Entonces podremos decir que ocurrirá un "milagro" si una observación particular cae fuera de  $w$ . Esto es, estamos casi seguros que esto no ocurrirá. La experiencia ha demostrado que el concepto de distribuciones de probabilidad es un instrumento útil para deducir afirmaciones prácticas.

Anteriormente consideramos la frecuencia de ocurrencia como una contrapartida práctica de la probabilidad. Pero en muchos casos tal interpretación podrá parecer algo artificial, así, por ejemplo, en las series económicas en donde una repetición del "experimento", en el sentido corriente, no sea posible realizarla. Podemos entonces, de forma alternativa, interpretar la "probabilidad" sencillamente como una medida de nuestra *confianza a priori* en la ocurrencia del suceso. También entonces la noción teórica de distribución de probabilidad sirve como un instrumento para deducir afirmaciones que tienen una probabilidad muy alta de ser ciertas, la contrapartida práctica es la de que "estamos casi seguros de que el suceso ocurrirá".

Han tenido lugar muchas discusiones fútiles, al considerar la pregunta de lo que realmente la probabilidad es, los tipos de sucesos para los cuales la probabilidad "existe" y demás. Varios tipos de "fundamentos de la probabilidad" han aparecido, algunos de ellos partiendo de las frecuencias observadas de los sucesos, otros apelando a la idea de credulidad *a priori* o alguna otra noción de la realidad. Otras "fundamentaciones" son de naturaleza puramente formal, sin ninguna referencia a los fenómenos reales. Pero todas ellas tienen una cosa en común, a saber: que al final se *encuentran* basadas en un cierto concepto de probabilidad que es de naturaleza puramente abstracta. Así, todas las "fundamentaciones", cualquiera que sea el sistema de probabilidades que envuelvan, exigen finalmente que satisfaga unos requisitos de consistencia lógica, y para cumplirlos se debe pagar un precio, el cual invariablemente consiste en dar la equivalencia exacta entre las probabilidades teóricas y cualquier fenómeno real que podamos considerar. En este aspecto los esquemas de probabilidad no son diferentes de otros esquemas teóricos. La noción rigurosa de probabilidad y distribución de probabilidad "existen solamente en nuestro pensamiento racional, sirviendo solamente como un instru-



mento para deducir afirmaciones prácticas del tipo descrito anteriormente.

Cuando establecemos que un cierto número de variables observables tienen una cierta ley conjunta de probabilidad podemos considerar esto como la construcción de un *mecanismo* racional capaz de producir (o reproducir) los valores observables de las variables consideradas. Cuando hemos observado un conjunto de valores de  $n$  variables observables ( $x_1, x_2, \dots, x_n$ ) podemos decir, sin ninguna posibilidad de contradicción, que estos  $n$  valores representan un punto muestra sacado de un universo que obedece a alguna desconocida ley de probabilidad enedimensional. Cualquiera que sea la afirmación *a priori* que podamos hacer acerca de los valores de las  $n$  variables observables, podremos deducir esta afirmación de una de las distintas (quizá infinitas) leyes de probabilidad enedimensionales convenientemente elegidas. La clase de todas las leyes de probabilidad enedimensionales puede, por lo tanto, considerarse como una *clasificación a priori de todos los mecanismos convenientes* que pueden reglar el comportamiento de las  $n$  variables consideradas.

Como la asignación de una cierta ley de probabilidad a un sistema de variables observables es un artificio nuestro, inventado con propósitos analíticos, y puesto que algunos resultados observables pueden producirse bajo una gran variedad de esquemas posibles, la cuestión que se presenta es la de cómo ha de elegirse la probabilidad en cualquier caso dado para representar el mecanismo verdadero bajo el cual los datos considerados se están produciendo. Para convertirlo en un problema racional de inferencia estadística hemos de establecer un axioma, postulando que todo conjunto de variables observables está asociado con una ley de probabilidad "verdadera", pero desconocida. Puesto que el conocimiento de esta ley verdadera de probabilidad puede permitir responder a cualquier pregunta que posiblemente podría contestarse de antemano respecto a los valores tomados por las variables observadas, el problema total de la inferencia cuantitativa puede entonces considerarse en cada caso como un problema de falta de información acerca de alguna ley de probabilidad desconocida.

## 11.—*Variables aleatorias y distribuciones de probabilidad en relación con los datos económicos.*

La experiencia nos ha enseñado mucho acerca del tipo de fenómenos reales a los cuales los esquemas de la teoría de la probabilidad se aplican con mayor éxito (mostraremos más adelante que el campo de aplicación de los esquemas probabilísticos es mucho más amplio que el indicado en esta sección). Estos fenómenos se agrupan bajo el nombre de "experimentos aleatorios". No se puede dar una respuesta precisa a lo que es un experimento aleatorio a causa de que no es un concepto abstracto, sino solamente un nombre aplicado a ciertos fenómenos reales. No obstante, indicaremos algunas de las propiedades principales que se asignan a tales experimentos. Primero, la noción de experimento aleatorio implica corrientemente alguna posibilidad hipotética o real de "repetir el experimento" aproximadamente bajo "las mismas condiciones". Segundo, se supone que tales repeticiones pueden dar resultados variables. Y tercero, las inferencias que sacamos de los experimentos aleatorios son del tipo. ¿Con cuánta frecuencia debe un resultado aparecer?

¿Puede esta descripción aplicarse a los datos económicos?

Aquí yo creo que es útil—aunque no siempre posible—hacer una distinción entre dos clases diferentes de experimentos, a saber: por un lado, aquellos que planeamos y realizamos nosotros como trabajos de investigación para analizar ciertos hechos que se presentan; por otro lado los experimentos de los que ya he hablado, y que son producto de la *Naturaleza* y en los cuales los hechos nos aparecen. Para destacar esta distinción de forma más clara consideremos un ejemplo.

Supongamos que tratamos de explicar el tamaño y variación del consumo de una cierta mercancía, A, en una sociedad o grupo consistente en N individuos o familias. Lo que corrientemente entendemos por "explicación" en tales casos es la elección de ciertos factores medibles, la variación de los cuales—por hipótesis o experiencia—esperamos que tenga influencia *de la misma forma* en el comportamiento de cada individuo, familia, etc. Supongamos que hemos especificado un cierto número de tales factores, en el caso

presente, por ejemplo, precio de la mercancía A, precio de otras mercancías, renta individual (o de la familia), edad de los individuos, etc. Designemos los  $n$  factores, especificados por  $x_1, x_2, \dots, x_n$ , y sea el consumo real de la mercancía A para un individuo dado  $y$ . Por el momento rechazamos los errores de observación debidos a la falta de precisión en la definición de lo que  $y$  y las variables  $x$  representan, así como la imprecisión en las medidas. En otras palabras, estamos tratando de variables verdaderas, como se describieron en la sección 3.<sup>2</sup>

Supongamos que para cada individuo podamos explicar el consumo de A por una ecuación:

$$y^* = f(x_1, x_2, \dots, x_n) \quad [11.1]$$

en donde  $y^*$  se obtiene para cada individuo poniendo en la derecha de [11.1] aquellos valores de los factores influyentes  $x$  que figuran en él. Sin embargo, si nosotros hacemos esto para cada individuo encontraremos—no importa cuál sea la función,  $f$ —que nuestra explicación es incompleta. De forma más específica encontraremos que dos individuos, o el mismo individuo en dos períodos de tiempo diferentes, pueden enfrentarse con exactamente el mismo conjunto especificado de factores influyentes  $x$  (y de aquí tener la misma  $y^*$  por [11.1]), y, sin embargo, los dos individuos tener cantidades diferentes de  $y$ , ninguna de las cuales es igual a  $y^*$ . Podríamos tratar de eliminar tales discrepancias introduciendo más “factores explicativos”  $x$ . Pero corrientemente agotaremos el número de factores que pueden considerarse como comunes a todos los individuos y que al mismo tiempo no tienen una influencia apreciable sobre  $y$ . Las discrepancias  $y - y^*$  para cada individuo pueden depender de una gran variedad de factores. Estos factores pueden ser diferentes de un individuo a otro y pueden variar con el tiempo en cada individuo.

De una forma puramente formal podemos reemplazar  $y^*$  por  $y$  en [11.1] y además sumarle una desviación  $s$  que se hace cargo de la diferencia entre  $y$  y  $y^*$ ; es decir:

$$y = f(x_1, x_2, \dots, x_n) + s \quad [11.2]$$

Supongamos, por ejemplo, que sabemos o suponemos que para cada

conjunto de valores de las variables  $x, s$  ( $y$ , por lo tanto,  $y$ ) es una variable aleatoria, teniendo una cierta distribución de probabilidad con media cero. ¿Cuál es el comportamiento de tal esquema?

Elijamos un subgrupo de individuos del grupo total de los  $N$ , tal que para cada miembro de este subgrupo los factores  $x$  sean idénticamente los mismos. Cuando, no obstante esto, las cantidades  $y$  para los miembros de este subgrupo son diferentes, nos indica que las decisiones de los individuos, aun después de haber fijado los valores de  $x_1, x_2, \dots, x_n$ , están expuestas a la incertidumbre. Los individuos no actúan todos igual. Cuando suponemos que  $s$  tiene para cada conjunto de valores fijados de  $x$  una cierta distribución de probabilidad aceptamos los parámetros (o algunas propiedades más generales) de esta distribución, como ciertas características adicionales del modelo teórico. Estos parámetros (o propiedades) describen la estructura del modelo, lo mismo que lo hace la influencia sistemática de  $x_1, x_2, \dots, x_n$  sobre  $y$ . Tales elementos aleatorios no son, pues, meros aditamentos superficiales "para propósitos estadísticos".

Cuando describimos  $s$  como una variable aleatoria con cierta distribución de probabilidad para un conjunto fijo de valores de las variables  $x$  estamos pensando en una clase de poblaciones infinitas hipotéticas, cada una de las cuales queda completamente descrita por el esquema [11.1] y por las características de la distribución de  $s$ . El número total de individuos  $N$  actualmente presente puede considerarse como una muestra mixta consistente en submuestras sacadas de los miembros de la clase hipotética de las poblaciones. No hay dificultad lógica en considerar "la población total como una muestra" de la clase de poblaciones con que estamos tratando, que *no* consisten en una infinidad de individuos diferentes, sino en una infinidad de *decisiones* posibles que pueden tomarse respecto al valor  $y$ . Y todas las decisiones tomadas por todos los individuos que estuvieron presentes durante un año pueden considerarse como una muestra, todas las decisiones tomadas quizá por los *mismos* individuos durante otro año pueden considerarse como otra muestra y así sucesivamente. Desde este punto de vista podemos considerar el número total de posibles observaciones (el número total de decisiones de consumir  $A$  por todos los individuos) como resultado de un procedimiento de muestreo que

la *Naturaleza* lleva a cabo, y en el que hacemos el papel de observadores pasivos.

Es con intención por lo que hemos empleado como aclaración un ejemplo de comportamiento económico individual en lugar de una relación media del mercado. Pues parece racional introducir los supuestos acerca de los elementos estocásticos de nuestra teoría económica en las "leyes" de comportamiento individual para un solo individuo, empresa, etc., como una característica de su comportamiento y entonces deducir las relaciones medias del mercado o relaciones para la sociedad total de estas leyes individuales. Será posible, por ejemplo, en muchos casos demostrar que, siempre bajo supuestos muy débiles acerca de las distribuciones de los elementos estocásticos en estas relaciones individuales, la media o relación total para el mercado o la sociedad total estará caracterizada por ciertas variables estocásticas compuestas (por ejemplo, sumas de los términos de error individuales), las cuales, por la ley de los grandes números, se distribuirán aproximadamente normalmente.

Una investigación activa podría producirse con otro tipo de experimento aleatorio. Así, en el ejemplo anterior podríamos elegir por algún proceso aleatorio un subgrupo de todos los individuos actualmente presentes y medir la  $y$  y las  $x$ . De este subgrupo podemos sacar inferencias sobre el comportamiento del grupo total. Pero la relación entre tal subgrupo y el grupo total que *podríamos* haber observado es diferente de la que existe entre este grupo total de individuos (o decisiones) presente y la clase hipotética de la población infinita de la cual ese grupo total presente se supone sacado, ya que la primera conexión es esencialmente dependiente de la elección del procedimiento de muestreo aleatorio que haya de emplearse. Eligiendo otro procedimiento tendremos otra conexión, y podríamos eliminar gradualmente todos los posibles errores de muestreo aumentando el tamaño de la muestra hasta que, por último, obtuviéramos una *descripción verdadera* de la muestra de todos los individuos presentes. Pero la incertidumbre en la correspondencia entre esta muestra de todos los individuos y la clase hipotética de la población infinita permanece. Un problema es la construcción de modelos hipotéticos de probabilidades de los cuales sea posible por extracciones aleatorias reproducir muestras del tipo dado por la "Naturaleza". Otro problema es el de hacer

mediciones exactas de estas muestras. La primera tarea pertenece esencialmente a la teoría conómica. La segunda es una técnica de observación estadística y teoría "clásica" de muestras. Por otra parte, después que el esquema estocástico ha sido elegido no hay una diferencia esencial entre los problemas de inferencia estadística que presentan.

12.—*El método de dividir las variables observadas en "parte sistemática" y "perturbación".*

Las variables económicas observables no satisfacen relaciones exactas (excepto quizá algunas identidades triviales). Por lo tanto, si funcionamos con un esquema teórico tendremos—con objeto de aplicarlo—que añadir algunos elementos estocásticos para tapan la brecha entre la teoría y los hechos. Una forma muy discutida de proceder es la de adoptar el convenio de que las variables observables están formadas de dos partes: una *sistemática*, la cual, por supuesto, satisface la relación exacta considerada, y un error o perturbación de naturaleza estocástica (5).

Sean  $x'_1, x'_2, \dots, x'_n$   $n$  variables teóricas que satisfacen una relación exacta funcional. Y  $x_1, x_2, \dots, x_n$  los correspondientes valores observables considerados. Escribimos:  $x_i = x'_i + x''_i$   $i = 1, 2, \dots, n$ , donde las variables  $x''_i$  son ciertas variables estocásticas. Con objeto de que nuestra relación entre las variables  $x'$  puedan también expresar algo acerca de las variables observables  $x$  tendremos que hacer ciertos supuestos adicionales acerca de la distribución de las variables  $x''$ . Entonces nuestra relación exacta entre las variables  $x'$  será una relación estocástica en las variables  $x$  y  $x''$  por la sustitución de  $x'$  por  $x - x''$ .

Es importante decir, sin embargo, que tal división de las variables necesariamente es de una naturaleza relativa, dependiendo del sistema particular de ecuaciones teóricas con que estemos relacionados.

(5) Este esquema es, por ejemplo, la base para el método de Frisch del "Análisis de la Confluencia". Ver "Ragnar Frisch Statistical Confluence Analysis by Means of Complete Regression Systems Oslo", 1934. Ver también T. KOOPMANS: "Linear Regression Analysis of Economic Time". Series De Erven F. Bohn, N. V., 1937.

Esto puede ponerse de manifiesto por medio de un ejemplo teórico.

Consideremos a este objeto tres dados ordinarios, uno negro, otro rojo y otro blanco, y realicemos la siguiente serie de experimentos. Primero lanzamos los tres dados. Obtenemos como resultado tres números  $x_n$ , para el negro;  $x_r$ , para el rojo, y  $x_b$ , para el blanco. Sea la suma de estos números  $X = x_n + x_r + x_b$ . Después hagamos que el dado negro permanezca en la posición en que quedó en la primera tirada, y volvemos a lanzar el blanco y el rojo. El resultado de este experimento será:  $Y_n = x_n$ ,  $Y_r$ ,  $Y_b$ , y sea  $Y = y_n + y_r + y_b$ . Ahora, por último, dejamos al negro y al rojo en la posición que tienen y tiremos el blanco una vez más. El resultado de este experimento será:  $z_n (= Y_n = X_n)$ ,  $Z_r (= Y_n)$ ,  $Z_b$  y  $Z = Z_n + Z_r + Z_b$ . Suponiendo que repetimos este experimento  $N$  veces obtendremos las series

$$\begin{aligned} X_1, Y_1, Z_1 & \text{ (1.º experimento)} \\ X_2, Y_2, Z_2 & \text{ (2.º experimento)} \\ \dots\dots\dots & \\ X_{1r}, Y_n, Z_n & \text{ (3.º experimento)} \end{aligned} \tag{12.1}$$

De la fundamentación de estos experimentos es evidente que las tres series,  $X$ ,  $Y$ ,  $Z$ , están correlacionadas, a causa de que tienen componentes comunes, indudablemente para cualquier  $X_i$ ,  $Y_i$ ,  $Z_i$  (resultado del experimento  $i$ ésimo) tendremos:

$$\begin{aligned} X_i &= x_{ni} + x_{ri} + x_{bi} \\ Y_i &= x_{ni} + y_{ri} + y_{bi} \\ Z_i &= x_{ni} + y_{ri} + z_{bi} \end{aligned} \tag{12.2}$$

Supongamos que ahora queremos estudiar la interdependencia entre las tres variables  $X$ ,  $Y$ ,  $Z$ , separando como "perturbación" los factores que no son "causas comunes". De [12.2] deducimos:

$$\begin{aligned} Y_i - (y_{ri} + y_{bi}) &= X_i - (x_{ri} + x_{bi}) \\ Z_i - (y_{ri} + z_{bi}) &= X_i - (x_{ri} + x_{bi}) \\ Z_{i1} - (z_{bi}) &= Y_i - (y_{bi}) \end{aligned} \tag{12.3}$$

en donde las expresiones entre paréntesis indican perturbaciones.

La composición de las perturbaciones depende claramente de la *relación* que estamos investigando. Despreciar esto puede quitar eficacia a la teoría.

Esta esquemática exposición tiene yo creo alguna importancia en muchos problemas importantes de economía. Así, por ejemplo, si  $X$ ,  $Y$ ,  $Z$  representan resultados de decisiones tomadas en alguna planificación económica. Entonces el esquema anterior puede considerarse de la siguiente forma: primero,  $X$  se determina por algunas consideraciones, las cuales no investigamos en este problema. Una vez que esta decisión se ha tomado, la decisión de  $Y$  no es libre, ya que está influida por  $X$ , aunque también hay otros factores que determinan  $Y$  que no tienen nada que ver con  $X$ , que son  $Y_r$  e  $Y_b$ . Estos factores, sin embargo, actúan como perturbaciones en  $Y$  respecto la causa  $X$ . Son ellos a su vez parte sistemática, "causas" respecto la decisión de  $Z$ , tomada después de haber elegido  $Y$ .

Consideremos un ejemplo de la economía dinámica. La interrelación entre inversión y beneficio. Por  $V(t)$  indicamos la actividad inversionista observada por año en el tiempo  $t$ , y sea  $Z(t)$  el beneficio observado. Supongamos que no hay error al observar estas cantidades. Hagamos la siguiente hipótesis: actividad de la inversión en  $t$  depende del beneficio obtenido en algún tiempo previo,  $t - \theta$ , mientras el beneficio en  $t$  depende de la inversión en  $t$ . Por  $\varepsilon_1(t)$  y  $\varepsilon_2(t)$  indicamos ciertas desviaciones aleatorias; podremos expresar la hipótesis por

$$V(t) = f[Z(t - \theta)] + \varepsilon_1(t) \quad [12.4]$$

$$Z(t) = g[V(t)] + \varepsilon_2(t) \quad [12.5]$$

en donde  $\theta$  es positiva y  $f$  y  $g$  son ciertas funciones. Puede suceder que en [12.4] tengamos que admitir una perturbación considerable,  $\varepsilon_1(t)$ , en  $V(t)$ , comparada con la parte de  $V(t)$ —exactamente  $V(t) - \varepsilon_1(t)$ —, que está "explicada" por  $Z(t - \theta)$ . Pero esto *no* indica que solamente *esta* parte de  $V(t)$  tenga influencia sobre  $z(t)$  a través de [12.5]; es decir, que nosotros podamos reemplazar  $V(t)$  por  $V(t) - \varepsilon_1(t)$  en [12.5]. Exactamente, la inversión *actual*—es decir,  $V(t)$ —tiene una relación más directa sobre el beneficio  $z(t)$  que la "parte sistemática" construida hipotéticamente por nosotros.



La aparición de tales situaciones tiene consecuencias muy importantes en el problema de *enlazar conclusiones* sacadas de relaciones diferentes, y que veremos en la próxima sección.

### 13.—Ecuaciones estocásticas hacia ecuaciones exactas.

La afirmación “un conjunto de variables que satisface una cierta ecuación” tiene un comportamiento diferente según se aplique a un esquema matemático abstracto o a variables observadas en la vida real.

En un esquema matemático abstracto la afirmación indica lo siguiente. Sean  $x'_1, x'^2 \dots x'_n$  variables reales. Cada conjunto de valores de estas  $n$  variables puede representarse por un punto en un espacio cartesiano  $n$ -dimensional. Indiquemos por  $S$  el conjunto de *todos* los puntos de este espacio y sea  $A$  una regla por la cual se elige un cierto subconjunto de puntos  $S_A$  de  $S$ . Esto es, se *excluyen* todos los puntos de  $S$  que no pertenecen a  $S_A$ . Entonces si existe una función  $f$  que no es idénticamente cero, pero tal que

$$f(x'_1, x'_2, \dots, x'_n) = 0 \quad [13.1]$$

para *todos* los puntos pertenecientes a  $S_A$ , diremos que las variables  $x'_1, x'^2, \dots, x'_n$  (las variaciones de las cuales están limitadas por la regla  $A$ ) tienen la propiedad de satisfacer la ecuación  $f = 0$ . Aquí el conjunto total  $S_A$  viene dado por definición a través de una operación lógica  $A$ , y podríamos comprobar cuándo la afirmación [13.1] es cierta o falsa.

Afirmaciones análogas acerca de las variables observadas en la vida real son de un carácter mucho más hipotético. Cuando hacemos afirmaciones del tipo [13.1] acerca del conjunto de variables observables, es decir,  $x_1 \dots x_n$ , suponemos que la Naturaleza tiene una regla para elegir tales puntos observados ( $x_1, \dots, x_n$ ) en el espacio  $x$  de tal forma que ninguno de estos puntos contradiga la hipótesis [13.1] cuando las variables  $x'$  se reemplazan por las variables  $x$ . Entonces decimos que [13.1] es una *ley de la Naturaleza*. Tratamos de establecer tales leyes contrastando la verdad de [13.1] respecto de observaciones pasadas. Pero aunque satisfagan [13.1] no podemos saber si la próxima observación la cumplirá. Corrien-

temente, sin embargo, creemos que esto será así porque tenemos un enorme registro de casos empíricos que nos dicen que tales inducciones empíricas han sido muy fructíferas. Al mismo tiempo también sabemos que en una investigación empírica es útil reemplazar la expresión "un conjunto de variables satisfaciendo cierta ecuación" por la expresión "satisfaciendo aproximadamente" tal ecuación. Esto indica que si ponemos los puntos observados ( $x_1, x^2, \dots, x_n$ ) en el lado izquierdo de [13.1] obtendremos en el lado derecho una cierta variable  $s$ .

Si—como hemos discutido anteriormente—queremos que una expresión tal como "aproximadamente satisface" no tenga un significado trivial haremos de cambiar la hipótesis [13.1] de tal forma que exprese *qué clase de aproximación* suponemos. Una forma de realizar esto consiste en cambiar la hipótesis [13.1] en

$$f(x_1, x_2, \dots, x_n) = s \quad [13.2]$$

y asignar a  $s$  ciertas propiedades generales que no contradigan los datos. Estamos particularmente interesados en los esquemas en que se adscriben a  $s$  ciertas propiedades generales de una variable aleatoria, primeramente a causa de que tenemos un gran número de casos empíricos demostrando que tales esquemas se han aplicado con éxito a fenómenos observados, y segundo, porque la teoría de tales esquemas se ha desarrollado más que cualquier otro esquema de aproximación. Y encontramos una justificación para aplicarlos a los fenómenos económicos en el hecho de que corrientemente solamente queremos —y nos interesan— efectos medios o totales de muchas decisiones individuales, que parcialmente están guiadas por factores comunes e individuales (véase sección II).

En el caso de que se suponga que  $s$  es una variable aleatoria, diremos que la variable  $x$  satisface la ecuación estocástica [13.2]. Esta, por otra parte, es un tipo muy particular de ecuación estocástica. Ahora no "vituperaremos" a un elemento particular de nuestro esquema, por el hecho de que las variables observadas ( $x_1, \dots, x_n$ ) no satisfaga [13.1] exactamente. Podemos operar con otros esquemas especificando con más detalle *dónde* está el elemento estocástico. En general, podríamos dar la siguiente definición. Si  $x_1, x_2, \dots, x_n$  es un conjunto de variables observadas, y si  $\varepsilon_1, \dots, \varepsilon_m$  son  $M$  varia-

bles aleatorias, y si existiese una función no idénticamente cero, tal que para todas las observaciones

$$F(x_1, x_2, \dots, x_n; \varepsilon_1, \varepsilon_2, \dots, \varepsilon_m) = 0 \quad [13.3]$$

entonces  $x_1, x_2, \dots, x_n$  se dice satisfacen una ecuación estocástica. Así, una ecuación estocástica en  $n$  variables es una ecuación exacta en  $m + n$  variables.

Supongamos que nuestra observación material consiste en  $N \geq n$  puntos en el espacio enedimensional de las variables  $x$ , y que adscribimos a la distribución de probabilidad conjunta de  $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_m$ , ciertas propiedades *a priori*. Insertamos sucesivamente los  $N$  puntos observados para las variables  $x$  en [13.3] y para cada punto observado elegimos un conjunto de valores de las  $\varepsilon$  de tal forma que [13.3] se cumpla. Así tenemos una muestra de  $N$  puntos en el espacio cartesiano enedimensional de las  $\varepsilon$ . Por otra parte, asignando *a priori* ciertas propiedades a la distribución de probabilidad de la  $\varepsilon$ , y excluyendo la posibilidad de obtener ciertas muestras de las  $\varepsilon$ , que *ahora* serán improbables (en uno u otro sentido, esto se discutirá posteriormente), habremos construido un conjunto probabilístico límite del subconjunto de muestras admisible de las  $\varepsilon$ . Sea este conjunto de puntos admisibles para los  $\varepsilon$   $Q$ . Diremos que si los  $N$  puntos observados en el espacio  $x$  son tales que —bajo la condición [13.3]— es posible elegir una muestra de  $N$  conjuntos de  $\varepsilon$  que pertenecen al conjunto  $Q$ , no podremos rechazar la hipótesis de que las  $n$  variables  $x_1, x_2, \dots, x_n$  satisfacen la ecuación estocástica [13.3].

De un esquema estocástico de la forma [13.3] podemos deducir ciertas ecuaciones exactas que no contengan las variables aleatorias  $\varepsilon$ , dando a una o más de las variables  $x$  una *nueva interpretación*. Hay dos tipos importantes y diferentes de tales ecuaciones exactas. El primer tipo podría llamarse “Si no hubiese errores en las ecuaciones”, y el segundo, “Valor esperado de las ecuaciones”.

El primer tipo se obtiene asignando a las variables aleatorias en [13.3] ciertos valores constantes. En muchos casos podríamos formular la ecuación estocástica de tal forma que fuese posible que estos valores constantes de las  $\varepsilon$  fuesen cero. Entonces exigiríamos que

$$F(x_1, x_2, \dots, x_n; 0, 0, \dots, 0) = 0 \quad [13.4]$$

imponemos una condición a las variables que en muchos casos quedarán violadas por las observaciones. Por lo tanto, si se emplea [13.4], una o más de las variables  $x$  deben querer decir —no lo que son realmente— sino lo que *serían* si “no hubiese errores”. Esta clase de ecuaciones exactas simplificadas, por lo tanto, representa una corrección hipotética de los puntos individualmente observados en el espacio  $x$ .

El segundo tipo de ecuaciones “exactas”, por otra parte, representa relaciones medias en un grupo de observaciones. En este caso *no se simplifica* el esquema original, sino que nos limitamos a estudiar ciertas propiedades estocásticas límites del esquema. Podemos ilustrar la diferencia con un ejemplo sencillo.

Consideremos un grupo de familias de igual tamaño y composición. Sea  $r$  la renta de la familia y  $x$  el gasto familiar durante un cierto período de tiempo. Supondremos todos los precios constantes y los mismos para todas las familias durante este período. Entre todas las familias con la *misma* renta, el montante del gasto  $x$  variará de una familia a otra a causa de una gran cantidad de factores poco importantes. Supongamos que los hábitos de gasto de una población infinita de tales familias pudiera describirse por medio de la siguiente ecuación estocástica:

$$\log_e x = k \log_e r + k_0 + \varepsilon \quad (k \text{ y } k_0 = \text{constante}) \quad [13.5]$$

en donde  $\varepsilon$  es una variable aleatoria normalmente distribuída con medio cero y varianza  $= \sigma^2$ . De este esquema estocástico podemos deducir las dos ecuaciones exactas siguientes:

Primero.—Imaginemos que podemos, de algún modo, eliminar las fuerzas que causan las discrepancias. En esta población hipotética todas las familias con la misma  $r$  actuarán igual, y tendremos:

$$\log_e x = k \log_e r + k_0 \quad [13.6]$$

Segundo.—Imaginemos que los “errores”  $\varepsilon$  permanecen en el esquema, pero consideremos solamente consumo esperado o medio para aquellas familias que tienen la misma renta  $r$ . Esto es:

$$E(x|r) = \bar{x}(r) = e^{k_0} \sigma^k \int_{-\infty}^{\infty} e^x \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}} dx = e^{k_0 + \frac{1}{2}\sigma^2} r^k \quad [13.7]$$

en donde  $E(x|r)$  significa valor esperado de  $x$  dado  $r$ .

Por lo tanto, lo que la media familiar indica en el esquema [13.5] *no* es que necesariamente todas las familias actúen igual.

Es de mucha importancia tener en cuenta la diferencia entre estos dos tipos de relaciones cuando tratamos de realizar *operaciones algebraicas* dentro de los sistemas de ecuaciones estocásticas. Por ejemplo, del esquema teórico [13.6] podemos deducir

$$x \quad [13.8]$$

Pero de

$$E(\log_e x|r) = k \log_e r + k_0$$

no podemos deducir que

$$E(x|r) = e^{k_0} r^k$$

Por lo tanto, cuando realizamos tales operaciones, debemos tener en cuenta que estamos empleando el esquema hipotético "si no hubiese error", y no el esquema "valor esperado". Se presentan confusiones en este punto corrientemente por el hábito de omitir la operación simbólica  $E$  (o un trazo sobre  $X$ , etc.) en ecuaciones tales como [13.7]. Particularmente aparecen confusiones cuando tenemos un *sistema* de ecuaciones algebraicas y aplicamos procesos de eliminación algebraicos a las ecuaciones "valores esperados". El error corriente reside en identificar los valores esperados de una variable en *una* ecuación con los valores esperados de esa misma variable en *otra* ecuación. Esto puede dar lugar a resultados sin sentido. El siguiente es uno de ellos.

Sean  $x_1, x_2, x_3$ , tres variables observadas, definidas por

$$x_1 = \varepsilon_1 + \varepsilon_2, \quad x_2 = k_1 \varepsilon_1, \quad x_3 = k_2 \varepsilon_2$$

en donde  $\varepsilon_1$  y  $\varepsilon_2$  son dos variables aleatorias independientes con medio cero. Tendremos:

$$x'_1 = E(x_1 | x_2) = \frac{1}{k^2} x^2 \quad x''_1 = E(x_1 | x_3) = \frac{1}{k_2} x_3$$

Ahora, si identificamos las dos variables  $x'_1$  y  $x''_1$ , designando ambas por  $x_1$ , tendríamos  $x_2 = \frac{k_1}{k_2} x_3$ , lo que *no es cierto*.

## CAPITULO IV

### LA CONTRASTACION DE HIPOTESIS

Los estadísticos han argüido corrientemente, con mucha razón, que los economistas no presentaban sus teorías de forma tal que éstas fuesen hipótesis estadísticas bien definidas, y que, por lo tanto, los estadísticos no “entendían el lenguaje” de los economistas. Los economistas, por otra parte, no son los únicos que han sido vituperados. Indudablemente, la teoría estadística como un todo, estuvo hasta hace muy poco, en un estado de gran confusión. Esta situación está desapareciendo rapidísimamente gracias a un cambio muy fructífero de dirección aparecido a partir del trabajo fundamental de S. Neyman y E. S. Pearson (6), en el que introducen unos principios muy generales —y en sí mismos sencillos— de contrastación de hipótesis estadísticas, y estimación, que han abierto el camino para una corriente de trabajos de alta calidad, los cuales gradualmente están elevando la teoría estadística a un nivel científico real. Los trabajos, aparte detalles técnicos sobre los principios generales de Neyman-Pearson, están solamente en sus principios. Y problemas técnicos muy difíciles pueden probablemente aparecer. La importancia fundamental de los principios de Neyman y Pearson reposa en el hecho de que estos principios especi-

(6) Ver, especialmente, “Statistical Research Memoirs”, vol. 1, 1936; vol. II, 1938. London. Otras referencias se dan en lo que sigue del texto.

fican claramente la *clase de problemas* que caen dentro del campo de la teoría e inferencia estadística. Así ha sido posible a los economistas ver ahora exactamente *cómo* han de formular sus teorías si quieren la asistencia de un estadístico. Es de gran importancia que el economista conozca estos principios de formulación aunque no sea un estadístico experto, ya que por lo menos podrá formular inteligentemente los problemas estadísticos.

Daremos a continuación una breve descripción de los principios básicos de la teoría de Neyman-Pearson para la contrastación de hipótesis estadísticas y estimación, y, además, emplearemos estos principios para la *formulación estadística de hipótesis construidas en la teoría económica*. Esta será la fundamentación de unas discusiones surgidas en relación con el problema de "comprobación" estadística de relaciones económicas.

#### 14.—Una descripción de la teoría de Neyman-Pearson de contrastación de hipótesis y estimación

Indiquemos por  $x_1, x_2, \dots, x_n$   $n$  variables aleatorias definidas dentro de un conjunto fundamental de probabilidad  $A$ , y sea

$$P(E_n \in w_n | A)$$

o, abreviadamente,  $P(w)$  una ley integral conjunta de probabilidad.

Cualquier tentativa de afirmación  $H$ , referente a la ley integral de probabilidad  $P(w)$  de las variables  $x_1, x_2, \dots, x_n$  (o referente a su ley elemental de probabilidad  $p(x_1, x_2, \dots, x_n)$  que se supone existen) se llama una *hipótesis estadística*. Mas preciso sea  $\Omega_n$ , o abreviadamente  $\Omega$ , el conjunto o clase de *todas las posibles* leyes integrales de probabilidad  $c$ -covariantes, y sea  $\omega$  cualquier subconjunto especificado de  $\Omega$  ( $\omega$  puede, por ejemplo, ser el conjunto de todas las distribuciones normales  $c$ -covariantes, o el conjunto de todas las distribuciones  $c$ -covariantes continuas, o cualquier otro subconjunto de  $\Omega$ ). Una afirmación de la forma

$$P(w) \in \omega \quad [14.1]$$

(léase la ley integral de probabilidad de  $x_1, x_2, \dots, x_n$  pertenece a la clase  $\omega$ ) se llama hipótesis estadística.

La afirmación [14.1] puede ser falsa, y entonces la alternativa es que

$$P(w) \in (\Omega - \omega) \tag{14.2}$$

Anteriormente, la única cosa que se había supuesto es que conocíamos con certeza que  $P(w) \in \Omega$ , lo cual es trivial. Corrientemente, sin embargo, sabemos —o por lo menos *suponemos* que conocemos— más que esto. Designamos por  $\Omega^0$  un subconjunto de  $\Omega$ , y sea  $\omega^0$  cualquier subconjunto de  $\Omega^0$ . Si, por otra parte, *sabemos o suponemos* que la afirmación

$$P(w) \in \omega^0 \tag{14.4}$$

puede ser falsa, entonces  $\Omega^0$  se llama *el conjunto de la hipótesis admisible a priori* respecto a la ley de probabilidad  $P(w)$  (por ejemplo,  $\Omega^0$  podría ser el conjunto de las leyes enedimensionales de probabilidad para las cuales existe la ley elemental de probabilidad, y  $\omega^0$  podría ser el conjunto de todas las leyes elementales de probabilidad que son simétricas respecto la media). En los problemas de contrastación de hipótesis estadísticas la especificación del conjunto de hipótesis admisibles *a priori* es, como veremos, de importancia fundamental.

Una hipótesis estadística se llama *simple* si especifica completamente la ley de probabilidad  $P(w)$ . Por ejemplo, la afirmación

$$P(w) = \int \int_w \int \frac{1}{(\sigma \sqrt{2\pi})^n} e^{-\frac{1}{2\sigma^2} \sum_1^n (x_i - \bar{x}_i)} dx_1, \dots, dx_n \tag{14.5}$$

en donde  $\bar{x}_i$ , ( $i = 1, 2, \dots, n$ ) y  $\sigma$  son parámetros constantes numéricos especificados, es una hipótesis simple. Cualquier hipótesis que no es simple se llama compuesta. Por ejemplo, si el valor de el parámetro  $\sigma$  o alguna de las medias  $\bar{x}_i$ , o todas, no están unívocamente especificadas, entonces [14.5] es una hipótesis compuesta.

Un conjunto de  $\Omega^0$  de hipótesis admisibles se llama *paramétrico* si todas las leyes de probabilidad  $P(w)$  pertenecientes a  $\Omega^0$  vienen dadas por expresiones analíticas, las cuales difieren unas de



otras solamente por los valores numéricos de un número finito de parámetros. Por ejemplo, todas las leyes de probabilidad [14.5] tales que  $\bar{x}_i > 0$ ,  $i = 1, 2, \dots, n$ , forma un conjunto paramétrico. Un conjunto que no es paramétrico se llama *no paramétrico*. Si  $\Omega^0$  es paramétrico, entonces  $\omega^0$  debe ser paramétrico. Pero si  $\Omega^0$  no es paramétrico,  $\omega^0$  puede ser o no paramétrico en [14.4].

Un *contraste* de una hipótesis estadística es una regla de rechace o no rechace de la hipótesis, tomando como base un *punto muestra* dado. Sea  $x_1, x_2, \dots, x_n$   $n$  variables aleatorias, y sea  $\Omega^0$  el conjunto de todas *las hipótesis admisibles a priori* acerca de su ley simultánea de probabilidad  $P(w)$ . Para cualquier miembro particular del conjunto  $\Omega^0$  y para cualquier subconjunto particular  $w$  de puntos en el espacio muestra  $R_n$  podemos calcular la probabilidad que un punto muestra  $E$  caiga en  $w$ . Si  $w$  queda fijado, la probabilidad de que  $E$  caiga en  $w (= w^0)$  variará generalmente de acuerdo con el elemento de  $\Omega^0$  que se emplea para calcularlo. Esto es, una parte "improbable" del espacio muestra respecto una ley de probabilidad de  $\Omega^0$  puede ser más probable para otra ley de  $\Omega^0$ . Y este hecho forma, por otra parte, la base para contrastar cualquier hipótesis particular dentro de  $\Omega^0$  respecto las otras admisibles *a priori*.

Muchas discusiones se han entablado sobre este punto en la antigua literatura, en particular porque un razonamiento sacado de un punto muestra de la verdadera población puede llevar consigo la noción de "probabilidad inversa". Corrientemente encontramos expresiones tales como "la distribución más probable" de la que una muestra dada ha sido obtenida. Tal afirmación, por otra parte, lleva consigo una cierta distribución de probabilidad de la hipótesis dentro de  $\Omega^0$ . En muchos casos, no obstante, tal modelo puede no tener sentido, ya que cuando sacamos una muestra la tomamos de un elemento *fijo pero desconocido* de  $\Omega$ . La probabilidad de que cualquier elemento de  $\Omega^0$  se el verdadero, es decir de donde sacamos la muestra, es, por lo tanto, 0 ó 1, con independencia del punto muestra observado.

Por otra parte, si establecemos una regla en virtud de la cual se rechaza o no la hipótesis, y si la decisión que se hace depende únicamente de la localización del punto muestra, podremos hablar

de probabilidad de que nuestra decisión sea verdadera o falsa, porque la decisión —por ser función del punto muestra  $E$ — es una variable aleatoria.

Visto esto, formularemos de forma más precisa lo que es un contraste de hipótesis estadística. Sea  $\Omega^0$  el conjunto de todas las hipótesis admisibles a priori como leyes de probabilidad  $P(w)$  de las variables  $x_1, x_2, \dots, x_n$ , y sea  $P(w) \in \omega^0$ , donde  $\omega^0$  es un subconjunto de  $\Omega^0$  la hipótesis  $H_0$  que hay que contrastar. Esto significa: tenemos seguridad de que, cualquiera que sea el punto muestra observado, la verdadera distribución de probabilidad de la variable aleatoria  $n$ -dimensional es un miembro (fijo, pero desconocido) del conjunto  $\Omega^0$ , y nuestra hipótesis es que  $P(w)$  pertenece al conjunto de distribuciones más restringido  $\omega^0$  dentro  $\Omega^0$ . La clase  $\omega^0$  puede contener solamente un solo miembro (hipótesis simple) o diversos miembros (hipótesis compuesta). En el último caso todos los miembros de  $\omega^0$  se consideran como equivalentes, y no tenemos interés en hacer distinciones entre ellos.

Sea  $W_0$  un conjunto de puntos en el espacio muestral  $n$ -dimensional  $R_n$ , tal que cuando un punto muestra cae en  $W_0$ , es decir  $E \in W_0$ , rechazamos la hipótesis  $H_0$ , en caso contrario no;  $W_0$  se llama la *región crítica* (o más generalmente, conjunto crítico de puntos) para contrastar la hipótesis  $H_0$ , es decir  $P(w) \in \omega^0$  respecto a las alternativas  $P(w) \in (\Omega^0 - \omega^0)$ . Una región crítica y un contraste son evidentemente dos cosas iguales con nombres diferentes.

En casos particulares, un contraste de hipótesis  $H_0$  puede ser decisivo, a saber: en los casos donde existe un subconjunto del espacio muestra que tiene probabilidad = 1 respecto  $H_0$ , pero probabilidad = 0 respecto cualquier otro miembro de  $\Omega^0$ . Entonces, por medio de un solo punto, podemos decidir —con probabilidad = 1 de ser cierto— cuando  $H_0$  es verdadera o falsa, rechazando  $H_0$ , sólo si  $E \in (R_n - \bar{W})$ . También se supone que el conjunto  $\Omega^0$  de hipótesis  $H$  puede dividirse en un sistema de  $k$  subconjuntos disjuntos  $\Omega^0_1, \Omega^0_2, \dots, \Omega^0_k$ , correspondientes unívocamente con  $k$  conjuntos no solapados  $W_1, W_2, \dots, W_k$  del espacio muestra, tal que

$$P(w_i | H \in \Omega^0_j)$$

sea  $= 0$ ,  $i \neq j$  y  $1$  cuando  $i = j$ . ( $i, j = 1, 2, \dots, k$ ). Entonces un solo punto muestra podría restringir el conjunto de hipótesis admisibles *a priori* a uno de estos subconjuntos  $\Omega^0$ . Tales casos, aunque importantes, son triviales desde el punto de vista de la teoría estadística. Podríamos, por lo tanto, suponer que el conjunto  $\Omega^0$  se ha reducido de antemano de forma que cualquier subconjunto  $W$  del espacio muestra tiene una probabilidad  $1$  de ser un miembro de  $\Omega^0$  y tiene, además, una probabilidad positiva con respecto a todos los otros miembros de  $\Omega^0$ . La aplicación de un contraste tal como se define anteriormente siempre llevará algún peligro de decisiones erróneas.

Ahora bien, si la región de rechace  $W_0$  fuera el espacio muestra total  $R_n$  (o el espacio total menos una parte que tiene una probabilidad cero respecto a cualquier miembro de  $\Omega^0$ ) entonces siempre (o casi siempre) podríamos rechazar  $H_0$ . Esto, evidentemente, no es lo que queremos, ya que cuando deseamos contrastar  $H_0$  esto lleva consigo el que la creamos correcta; y en este caso, el contraste daría lugar constantemente a decisiones falsas. Por otra parte, si

$$P(R_n - W_0 | H_0) \text{ y } P(W_0 | H_0)^2$$

son ambas positivas, corrientemente tendremos dos *clases de riesgo* de tomar decisiones falsas por medio del contraste.

Supongamos primero que la hipótesis es cierta y que al mismo tiempo el punto muestra cae en  $W_0$  (lo cual —por hipótesis— es posible), entonces rechazamos  $H_0$  y cometemos un error. Que se llama error de primera clase.

Supongamos ahora que la hipótesis es falsa (es decir, cierta una de las hipótesis alternativas), y al mismo tiempo el punto no cae en  $W_0$ . Entonces aceptamos  $H_0$  y cometemos un error. Al que se llama error de segunda clase.

Para cualquier tamaño dado de muestra podemos hacer la probabilidad de uno u otro error tan pequeña como queramos por una elección apropiada de  $W_0$ , pero no es posible hacerlo con los dos errores al mismo tiempo. Debemos, por lo tanto, hacer un convenio que dependerá de la clase de error que queramos reducir, y ésta siempre depende de las consecuencias que tenga una decisión falsa en cada caso particular.

El problema total de contrastar hipótesis estadísticas y de la estimación consiste en deducir "las mejores regiones críticas"  $W_0$  tomando como base ciertos *parámetros del riesgo*, los cuales, a su vez, vienen dados por consideraciones externas y se consideran como datos en la teoría estadística. Expondremos brevemente la teoría de Neyman-Pearson para la solución de este problema. El principio fundamental de esta teoría reside en la distinción entre las dos clases de errores anteriormente descritos. Una distinción sugerida por sí misma, al reconocer el sencillo hecho de que el desear contrastar una hipótesis lleva consigo el hecho de que *pueda ser falsa*, y, por lo tanto, es necesario especificar en qué sentido *puede ser falsa*. El reconocimiento y formulación de tal principio elemental —aparentemente casi trivial— se encuentra frecuentemente incluido entre las grandes retalizaciones de la investigación científica.

Consideremos primeramente el caso en que  $\omega^0$  consiste de una sola distribución de probabilidad  $P_0$ , y supongamos que el conjunto  $\Omega^0 - \omega^0$  solamente contiene otro solo elemento:  $P_1 \neq P_0$ . Queremos contrastar, sobre la base de un punto muestra  $E$ , la hipótesis  $H_0$  de que la verdadera distribución de probabilidad es  $P_0$ , la única alternativa es  $P_1$ : sea  $W_0$  una región crítica tal que la probabilidad  $P(W_0 | P_0)$  sea exactamente igual a  $\alpha$  (por ejemplo,  $\alpha = 0,05$ ),  $\alpha$  recibe el nombre de *nivel de significación* o también tamaño de la región crítica  $W_0$  y es un parámetro de riesgo elegido *a priori*. Hecho esto, si elegimos  $W_0$  como región crítica para el rechace de la hipótesis  $H_0$ , la probabilidad de que rechazemos la hipótesis cuando es cierta (es decir, probabilidad del error de primera clase) es exactamente  $\alpha$ . Pero en general hay diferentes regiones  $W_0$  del mismo tamaño  $\alpha$ . Ahora bien, si la hipótesis *no* es cierta, es decir, si la verdadera distribución es  $P_1$ , queremos, por otra parte, tener una probabilidad la mayor posible de rechazar la hipótesis  $H_0$ , es decir, queremos que la probabilidad  $P(W_0 | P_1)$  de que  $E$  caiga en  $W_0$  cuando  $P_1$  es cierta sea lo mayor posible. Esta probabilidad  $P(W_0 | P_1)$  se llama la *potencia del contraste*  $W_0$  respecto la alternativa  $P_1$ . Sea  $W_0^*$  la región de tamaño  $\alpha$  para la que esta potencia es máxima. Evidentemente debemos emplear la región  $W_0^*$  como nuestra región crítica en lugar de cualquier otra región de

tamaño  $\alpha$ . A  $W^*_0$  se la llama entonces la mejor región crítica para contrastar  $P = P_0$  respecto a la alternativa  $P = P_1$ .

Supongamos que ampliamos el conjunto  $\Omega^0 - \omega^0$  para que comprenda un sistema total de distribuciones de probabilidad alternativas. Entonces si  $W^*_0$  es al mismo tiempo la mejor región crítica para la contrastación de  $P = P_0$  respecto a todo el elemento del conjunto de hipótesis alternativas,  $W^*_0$  recibe el nombre de *contraste uniformemente más potente*. En unos pocos casos importantes puede demostrarse que tales regiones existen. Pero esto se aplica solamente en ciertos tipos de contrastación de hipótesis respecto ciertos conjuntos restringidos de alternativas. Y si un contraste así no existe tendremos que elegir alguna región crítica de tamaño  $\alpha$  la cual sea lo "más potente posible" respecto al conjunto de hipótesis alternativas en cuestión. La elección del "mejor" contraste será entonces algo más subjetivo. Puede suceder que tomemos en consideración un cierto sistema de *ponderaciones en importancia* para los errores de segunda clase para los diversos elementos del conjunto de hipótesis alternativas. Por ejemplo, si la hipótesis a contrastar es que un cierto parámetro  $\theta$ , en una distribución de probabilidad (la forma de la cual es conocida) es igual a un valor especificado  $\theta^0$ , las alternativas posibles de todos los valores de  $\theta$  que van de  $-\infty$  a  $+\infty$  puede ser que, por alguna razón, podamos considerar más importante considerar las alternativas  $\theta > \theta^0$  que las alternativas  $\theta < \theta^0$ . El problema de introducir tal función de ponderación ha sido tratado por A. Wald (7).

Anteriormente hemos supuesto que la hipótesis que va a ser contrastada era simple, pero la idea general es extenderla a las hipótesis compuestas aunque las dificultades técnicas para deducir regiones críticas del tipo aquí discutido son más serias. El problema de determinar regiones  $W_0$  que tengan el mismo tamaño para *todo miembro* del conjunto  $\omega^0$  que ha de ser contrastado, presenta siempre problemas matemáticos complicados, y algunas veces *no* existe tal región (8).

(7) A. WOLD: "Contribution to the theory of Statistical Estimation and Testing Hypothesis". *Annals of Mathematical Statistics*, vol. 10, diciembre 1939, págs. 299-329.

(8) Ver, por ejemplo, W. FELLER: "Note on Regions Similar to the Sample Spaces". *Statistical Research Memoirs*, vol. II, London, 1938, págs. 107-125

Cualquiera que sean los principios por los cuales elegimos una "mejor" región crítica de tamaño  $\alpha$ , lo esencial es que un contraste se desarrolle siempre respecto a un conjunto dado de alternativas posibles  $\Omega^0$ . Si sobre la base de algún principio general puede encontrarse un mejor contraste o región  $W^1_0$  para contrastar una hipótesis dada  $P \in \omega^0$  respecto al conjunto  $\Omega^0$  de las hipótesis admisibles *a priori*, y si nos fijamos en otro conjunto de hipótesis admisibles *a priori*  $\Omega^1$ , también conteniendo  $\omega^0$ , el mismo principio general corrientemente no dará otra "mejor" región crítica  $W''_0$ . En otras palabras, si un contraste se ha desarrollado tomando como base un conjunto dado de hipótesis alternativas *a priori*  $\Omega^0$ , el contraste será, en general, válido solamente para este conjunto  $\Omega^0$ . Extendiendo el conjunto de hipótesis admisibles de forma que incluyamos nuevas alternativas *sin cambiar* la región crítica, siempre podríamos encontrar alternativas tales, que cualquiera que sea la región crítica elegida, su potencia respecto a alguna de las nuevas hipótesis alternativas será muy pobre. Es esta una formulación más precisa de frases como ¿para qué sirve contrastar la significación del coeficiente de correlación cuando quizá el supuesto de regresión lineal es falso? Este es exactamente el tipo de argumento que hemos discutido anteriormente. Generalmente cuando contrastamos el significado del coeficiente de regresión, el conjunto de hipótesis alternativas  $\Omega^0$  es sólo el sistema de ecuaciones de regresión *de la misma forma* con coeficientes de regresión que son distintos de cero.  $\Omega^0$  no debe incluir otra forma de regresión aunque éstas podrían muy bien aparecer.

En general, si una región crítica  $W_0$  para una hipótesis dada  $H_0$  se ha desarrollado tomando como base un conjunto  $\Omega^0$  de hipótesis admisibles *a priori*, y si la verdadera hipótesis —en lugar de pertenecer a  $\Omega^0$  como se ha supuesto— pertenece a  $\Omega - \Omega^0$  (es decir, al conjunto complementario de  $\Omega^0$ ) habremos perdido el control de los errores originariamente adscritos al contraste. Y puede, por otra parte, ocurrir que la potencia del contraste respecto a estas hipótesis "fuera del esquema" sea bastante bueno, es decir, cuando una de estas nuevas alternativas es cierta en vez de la hipótesis contrastada, la probabilidad de que el punto muestra caiga en  $W_0$  puede ser alta. Pero esta probabilidad puede también ser

pequeña, incluso más pequeña que  $\alpha$ , lo cual indica que podríamos tener menor probabilidad de rechazar la hipótesis contrastada cuando es falsa que cuando es correcta.

El requerimiento de la especificación del conjunto de hipótesis admisibles *a priori antes* de construir un contraste fuerza a establecer explícitamente que lo suponemos conocido y fuera de duda y que deseamos contrastarla.

---

El problema de estimación es el problema de obtener inferencias de un punto muestra, así como la ley de probabilidad del conjunto fundamental de probabilidad de la cual la muestra se ha obtenido. El problema de estimación está muy relacionado con el problema de contrastar hipótesis estadísticas; en efecto, la estimación puede considerarse como una forma particular de contrastar hipótesis.

Sean  $x_1, x_2, \dots, x_n$   $n$  variables con la ley de probabilidad (desconocida)  $P(w)$ . Si sabemos que  $P(w)$  pertenece a una clase paramétrica de distribuciones  $\Omega^0$ , es decir,  $P(w)$  es conocida excepto para los valores de un cierto número finito de parámetros. Podremos expresar esto así:

$$P(w) = P(w | \theta^1, \theta_2, \dots, \theta_k)$$

o abreviadamente  $P(w | \theta)$ , en donde la función  $P$  es conocida. Una muestra  $E$  se obtiene de uno de los miembros de  $\Omega^0$ , aunque no sabemos de cual. El problema es sacar inferencias de  $E$  respecto los valores correspondientes de los parámetros  $\theta$ . Sean los verdaderos, pero desconocidos valores  $\theta^0_1, \theta^0_2, \dots, \theta^0_k$ . Cualquier sistema de valores de los parámetros  $\theta$  pueden representarse por un punto  $\theta$  en el espacio paramétrico, es decir, en un espacio euclídeo  $k$  dimensional, donde los ejes representan los  $k$  parámetros  $\theta$ . El problema de estimación consiste en definir una función la cual *asocia* cada punto  $E$  en el espacio muestra con un *conjunto* de puntos  $\theta$  en el espacio *paramétrico*. Si esta función es tal que a cada punto  $E$  en el espacio muestra corresponde uno, y solamente un punto  $\theta$ , en el paramétrico, hablaremos de *estimación por pun-*

tos. Si a cada punto  $E$  en el espacio muestra adscribe la fórmula de estimación una *región*  $I(E)$  [o más generalmente un conjunto de puntos  $I(E)$ ] en el espacio paramétrico, hablaremos de intervalo (o conjunto) de estimación. En el caso particular de estimación por puntos  $I(E)$  contiene solamente un punto  $\theta$  para cada  $E$ .

El intervalo (o conjunto)  $I(E)$  es, evidentemente, un conjunto aleatorio, ya que es una función del punto muestra  $E$ . Podríamos, por tanto, hablar de la probabilidad  $\beta$ , de que el conjunto  $I(E)$  cubra el verdadero punto paramétrico  $\theta^0$ , y podríamos elegir el valor de  $\beta$  de acuerdo con el montante del riesgo que deseamos considerar así  $\beta = 0,95$ . Como no conocemos el verdadero punto paramétrico  $\theta^0$ ,  $\beta$  será independiente de  $\theta^0$ , es decir, cualquiera que sea el verdadero punto paramétrico  $\theta^0$  de la distribución de la cual obtenemos la muestra, la probabilidad  $P(\theta^0 \in I | \theta^0)$  (9) podrá ser la misma.  $\beta$  se llama el coeficiente de confianza para la estimación de  $\theta^0$ , y el  $I(E)$  correspondiente se llama un *intervalo de confianza* (o más generalmente un conjunto de confianza) para el verdadero punto paramétrico.

Consideremos el conjunto de todos los puntos paramétricos admisibles *a priori* correspondientes a  $\Omega^0$ . Este conjunto de puntos paramétricos puede considerarse como el conjunto de todas las hipótesis simples contenidas en  $\Omega^0$ , es decir, todas las hipótesis  $\theta = \theta^0$  en donde  $\theta^0$  puede ser cualquier punto entre el conjunto de puntos paramétricos admisibles *a priori* (ahora consideraremos  $\theta^0$  como un punto variable). Supongamos que para toda hipótesis simple  $\theta = \theta^0$ , en el conjunto admisible *a priori*  $\Omega^0$ , construimos por algún medio una *mejor región crítica*  $W_0(\theta^0)$  de tamaño  $\alpha$ , como hemos descrito anteriormente,  $W_0(\theta^0)$  es la región (o conjunto) de rechace de

$$\theta = \theta^0. \quad R_n - W_0(\theta^0)$$

es, por lo tanto, la región de no rechace o, brevemente, la región de aceptación de  $\theta = \theta^0$ , y su tamaño es  $1 - \alpha$ , sea  $1 - \alpha = \beta = \alpha$  el coeficiente de confianza para estimar los puntos paramétricos por

---

(9) Cuando empleamos la notación  $\theta^0 \in I$  recordamos que  $\theta^0$  es el elemento constante, mientras  $I$  el variable.



medio de un punto muestra. Sea  $E_1$  cualquier punto muestra arbitrariamente fijado. Puesto que suponemos que la verdadera hipótesis está contenida en  $\Omega^0$  es razonable exigir que en nuestro sistema de regiones de aceptación  $R_n - W_0(\theta^0)$  puede haber *por lo menos una* región tal que  $E_1$  pertenece a él. En general  $E_1$  será un elemento del sistema total de regiones de aceptación. Consideremos todas las regiones de aceptación de tamaño  $\beta$  de las cuales  $E_1$  es un elemento. A cada región de aceptación  $R_n - W_0(\theta^0)$  corresponde un punto en el espacio paramétrico llamado el punto  $\theta^0$ , representando la hipótesis  $\theta^0$  para la cual  $W_0(\theta^0)$  es una región de rechazo. Al sistema de todas las regiones de aceptación de las cuales  $E_1$  es un miembro, corresponde, por lo tanto, un conjunto de puntos paramétricos  $I(E_1)$ . Como  $E_1$  era arbitrario y, por lo tanto, puede ser cualquier punto  $E$  del espacio muestra, define una función  $I(E)$  para todo  $E$ . Esta  $I(E)$  evidentemente tiene las propiedades de un conjunto de confianza para estimar los parámetros  $\theta$  por medio de un punto muestra  $E_1$  a causa de que cualquiera que sea el verdadero punto paramétrico la probabilidad de que el punto muestra  $E$  caiga en su región correspondiente de aceptación es  $1 - \alpha = \beta$  constante, y *cuando*  $E$  cae en la región de aceptación para  $\theta - \theta^0$  entonces también  $\theta^0 \in I(E)$ . La probabilidad de que  $I(E)$  abarque el verdadero punto paramétrico, cualquiera que sea, es igual a  $\beta$ .

El problema de estimación puede formularse más generalmente. Sean  $x_1, x_2, \dots, x_n$   $n$  variables aleatorias con distribución de probabilidad  $P(w)$  de la que sólo se conoce que pertenece a un cierto conjunto admisible *a priori*  $\Omega^0$  de funciones de distribución.  $\Omega^0$  puede considerarse como el conjunto de todas las hipótesis admisibles *a priori*. Para cada una de estas hipótesis simples construyamos una cierta región de aceptación  $U$ , de tamaño  $\beta$ , y sea  $(U)$  la familia de todas las tales regiones de aceptación correspondientes al conjunto  $\Omega^0$ . Se da un punto muestra  $E_1$ . Sea  $[U(E_1)]$  la familia de todas aquellas regiones de aceptación de las cuales  $E_1$  es un miembro, y sea  $I(E_1)$  el conjunto de todas las hipótesis simples (contenidas en  $\Omega^0$ ) que corresponden al sistema de regiones  $[U(E_1)]$ . Como  $E_1$  puede ser cualquier punto  $E$ , corresponderá un  $I(E)$  a todo  $E$ .  $I(E)$ , definido de esta forma, es un conjunto

de confianza con el coeficiente de confianza  $\beta$ , es decir, la probabilidad de que  $I(E)$  contenga el miembro verdadero de  $\Omega^0$ , es igual a  $\beta$ .

15.—*Propiedades generales del problema de contrastación de relaciones económicas*

La teoría de Neyman-Pearson de contrastación de hipótesis estadística es puramente abstracta, al igual que cualquier otro esquema teórico. Esta cuestión es de interés por lo siguiente: ¿Este esquema representa un instrumento útil en relación con el problema de contrastar teorías económicas? ¿Puede ayudar a entender mejor la naturaleza de estos problemas e investigar soluciones prácticas de ellas? Yo creo que estas preguntas pueden responderse casi todas afirmativamente. La discusión siguiente se hace para mantener este punto de vista.

Intentaremos dar una formulación general axiomática al problema de contrastar relaciones económicas, empleando los principios de la teoría de Neyman-Pearson.

A.—*Datos importantes en la investigación económica*

El objeto de la investigación económica son las variaciones y covariaciones entre grupos de fenómenos de la vida real. Sea  $K_1, K_2, \dots, K_n$  uno de tales grupos.  $K_1$  puede ser, por ejemplo, un cierto tipo de bienes de consumo,  $K_2$  puede indicar el fenómeno llamado "precio" de  $K_1$ , etc. Cada  $K$  es exactamente el nombre de una cierta categoría de fenómenos reales concebidos de forma más o menos equivalente, y distintos de los de otras categorías. Muchas clases de variaciones y desviaciones pueden aparecer dentro de cada categoría. Estamos interesados solamente en variaciones que se muestran por medio de un cierto número de características medibles de cada  $K$ . Sean éstas  $n$  características numerables  $x_1, x_2, \dots, x_n$ , y sea

$$(x_1 t_1, x_2 t_1, \dots, x_n t_1)$$

el conjunto de valores observados conjuntamente por las  $n$   $K$ acs

indica "observación en el tiempo  $t_i$ " o simplemente observación número  $i$  ( $t_1, t_2, \dots$ , etc., necesariamente no tienen que ser equidistantes).

Sea

$$\begin{aligned}
 & (x_1 t_1, x_2 t_1, \dots, x_n t_1) \\
 & (x_1 t_2, x_2 t_2, \dots, x_n t_2) \\
 & \dots \\
 & (x_1 t_N, x_2 t_N, \dots, x_n t_N)
 \end{aligned}
 \quad [15.1]$$

un sistema de  $N$  observaciones conjuntas. Cada columna en [15.1] representa una serie de medidas de la "misma variable", por ejemplo, series de tiempo.

#### B.—*Supuestos fundamentales acerca de la naturaleza de los datos económicos*

*Los  $nN$  valores*

$$(x_{1t}, x_{2t}, \dots, x_{nt}) \quad t = t_1, t_2, \dots, t_N$$

en el sistema [15.1] de  $N$  conjunto de valores, puede considerarse como un punto muestra  $E$  en el espacio muestra  $nN$  dimensional de  $nN$  variables aleatorias

$$(x_{1t}, x_{2t}, \dots, x_{nt}) \quad t = t_1, t_2, \dots, t_N,$$

con cierta ley integral de probabilidad  $P(w)$ . ( $w$  indica un conjunto de puntos arbitrarios en el espacio muestra de  $nN$  dimensiones). Este supuesto indica lo siguiente: Consideramos la situación antes de que la muestra [15.1] se obtuviere, es decir, consideremos el sistema [15.1] como  $nN$  células vacías. Y consideremos el conjunto total de sistemas alternativos, con  $nN$  elementos, que *a priori* puedan caer en las  $nN$  células. El supuesto anterior supone —por hipótesis o por ser verdad— que antes de que la muestra se obtuviese hay un conjunto de sistemas satisfaciendo los requisitos del conjunto fundamental de probabilidad tal como se define en la sección 9. Este supuesto es extremadamente general, como puede

verse de la definición de variable aleatoria de la misma sección.

Indudablemente es difícil concebir un caso que pueda contradecir este supuesto. Para el propósito de la contrastación de hipótesis no es siempre necesario suponer que la muestra pueda repetirse. Hacemos afirmaciones hipotéticas antes de obtener la muestra y tenemos sólo que ocuparnos cuando la muestra acepta o rechaza la hipótesis. El anterior supuesto cubre también, como un caso particular, la situación en donde, para ciertas células de [15.1], hay exactamente un sistema fijo de números que deben ocupar estas células, es decir, el caso en que —para algunas de las células de [15.1]— ciertos valores fijos de las  $x$  correspondientes tengan una probabilidad = 1 (es decir, son estocásticamente constantes). Esto es de importancia en muchos problemas económicos en donde algunas de las variables se consideran como dadas automáticamente.

### C.—*La transformación de un esquema estocástico teórico*

Dos son las clases de esquemas abstractos que aparecen en la teoría económica, a saber: un tipo que introducimos como materia de ejercicio en razonamiento lógico o como un modelo de una economía idealizada (es decir, esquemas para los cuales una comparación con la realidad no tiene sentido), y otros tipos que —aunque abstractos— creemos tienen alguna aplicación en los fenómenos económicos reales. Para nuestro estudio solamente el último es importante.

Cuando construimos esquemas de este último tipo, pensamos siempre en algún fenómeno real cercano y tratamos de incluirlo en el esquema —siempre de forma simplificada—, comprobamos que tales esquemas nunca pueden dar una pintura exacta de la realidad. Podemos admitir ciertas discrepancias. En el Capítulo III hemos discutido cómo se puede emplear un esquema estocástico con este objeto a causa de la definición tan general de variables aleatorias, los esquemas estocásticos representan casos extremadamente generales de modelos teóricos. Podremos, por lo tanto, suponer que el problema de contrastar relaciones económicas consiste en confrontar ciertos modelos estocásticos especificados con un conjunto de datos [15.1].

Sea

$$\begin{array}{l}
 x'_1 t_1, x'_2 t_1, \dots, x'_n t_1 \\
 x'_1 t_2, x'_2 t_2, \dots, x'_n t_2 \\
 \dots\dots\dots \\
 x'_1 t_N, x'_2 t_N, \dots, x'_n t_N
 \end{array} \quad [15.2]$$

un sistema de variables aleatorias teóricas que han de compararse con las variables observadas correspondientes de [15.1].

Además, sea

$$\begin{array}{l}
 \varepsilon_1 t_1, \varepsilon_2 t_1, \dots, \varepsilon_m t_1 \\
 \varepsilon_1 t_2, \varepsilon_2 t_2, \dots, \varepsilon_m t_2 \\
 \dots\dots\dots \\
 \varepsilon_1 t_N, \varepsilon_2 t_N, \dots, \varepsilon_m t_N
 \end{array} \quad [15.3]$$

otro sistema de  $nN$  variables aleatorias introducidas en el esquema teórico como *parámetros aleatorios auxiliares*, teniendo ciertas propiedades especificadas su distribución conjunta. (Las  $\varepsilon$  pueden también introducirse como contrapartida de algún fenómeno real. Ver Sección II.)

Por último, sea

$$x_1, x_2, \dots, x_k \quad [15.4]$$

un conjunto de constantes.

Impongamos un sistema de restricciones

$$\begin{aligned}
 & f t_i [x'_1 t_i, x'_1 t_{i-1}, \dots, x'_1 t_1; x_2 t_i, \dots, x'_2 t_1; \dots; \\
 & x'_n t_i, x'_n t_{i-1}, \dots, x'_n t_1; (X_0); x_1, x_2, x_k; \varepsilon_1 t_i \dots \\
 & \dots \varepsilon_m t_i] = 0 \quad i = 1, \dots, N
 \end{aligned} \quad [15.5]$$

a las cantidades [15.2] — [15.4]  $f t_i$  es una función especificada para cada valor de  $i$ ,  $i = 1, 2, \dots, N$  (en particular, todas las  $f$  pueden ser las mismas e independientes de  $t$ ; entonces solamente los argumentos de la función cambiarían).  $(X_0)$  es un símbolo abreviado de un *conjunto de condiciones iniciales*, es decir, los valores de  $x_{jt}$  ( $j = 1, 2, \dots, n$ ) para  $t = t_0, t_{-1}, t_{-2}$ . Tales cantidades pueden o no entrar en [15.5]. Si entran supondremos entonces que son *constantes* que tienen valores *conocidos*.

Para cada punto del tiempo  $t = t_1, t_2, \dots, t_n$  [15.5] es una relación estocástica que define implícitamente una de las variables, por ejemplo  $x'_1 t_1$ , como una función de

1. Los valores previos de la primera variable.
2. Los valores previos y simultáneos de las otras variables.
3.  $m$  variables aleatorias  $\epsilon$ .

Supongamos que [15.5] debe contrastarse en nuestra teoría económica; las variables aleatorias  $\epsilon$  tienen una distribución con ciertas propiedades prescritas. La principal tarea de una teoría económica es hacer una elección fructífera de la forma de  $f$ .

En esta formulación general, [15.5], con los supuestos respecto  $\epsilon$ , puede representar una teoría *estática* o *dinámica*. Supongamos, como anteriormente, que cada ecuación [15.5] puede resolverse para  $x'_1 t_i, i = 1, 2, \dots, N$ . La teoría es *estática* si: 1), entran solamente variables para el mismo período de tiempo en cada una de las ecuaciones [15.5], y, al mismo tiempo, 2), las  $n - 1$  variables aleatorias  $x'_2 t_i, \dots, x'_n t_i$  y las  $m$  variables

$$\epsilon_1 t_i, \epsilon_2 t_i, \dots, \epsilon_m t_i \quad (i = 1, 2, \dots, N)$$

se suponen estocásticamente independientes de los valores *previos* de las variables  $x'$  y  $\epsilon'$ . Por otra parte, la teoría es *dinámica* en el sentido de que "lo que sucede en el tiempo  $t$  depende de lo que haya sucedido antes".

Por otra parte, [15.5] es una afirmación vacía acerca de las variables [15.2] en el caso de que *solamente* conozcamos algo acerca de las variables aleatorias  $\epsilon$ , además de [15.5] para que —cualquiera que sean las variables  $x'$ — podamos definir las variables  $\epsilon$  de forma que [15.5] se verifique. Debemos hacer alguna afirmación adicional respecto las propiedades de la ley condicional de probabilidad conjunta de todas las variables para valores dados de las  $(n - 1) N$  variables "independientes", las cuales se suponen son

$$x'_{2t_1}, x'_{3t_1}, \dots, x'_{nt_1} \quad (t = t_1, t_2, \dots, t_N).$$

Quando ocurre esto se deduce de [15.5] que la ley de probabilidad conjunta de todas las variables  $x'$  de [15.2] no pueden tener cualquier distribución, deben pertenecer a una clase *restringida* (mayor o menor) de leyes de probabilidad.

Por ejemplo, supongamos que [15.5] fuera de la forma

$$x'_1 t_i - x_1 x'_2 t_i - \varepsilon_1 t_i = 0 \quad [15.5']$$

y supongamos que las variables  $\varepsilon$  se distribuyen independientemente de las variables  $x'_2 t_i$ . Sea

$$p_1 (\varepsilon_1 t_i, \dots, \varepsilon_1 t_N)$$

la ley de probabilidad conjunta de las  $N$  variables  $\varepsilon$ . Entonces se deduce que para valores dados de las variables  $x'_2 t_i$  las variables  $x'_1 t_i$  tienen la ley de probabilidad conjunta elemental

$$p_1 [(x'_1 t_1 - x_1 x'_2 t_1), \dots, (x'_1 t_N - x_1 x'_2 t_N)].$$

Y de aquí, cualquiera que sea la ley de probabilidad elemental de  $p_1$ , es decir  $N$  de todas las variables  $x_{2i}$ , la ley de probabilidad conjunta  $p_2$  de las  $2N$  variables  $x'$  tendrán la forma

$$p_3 = p_1 - p_2.$$

Así [15.5], junto con cualquier supuesto adicional referente a la distribución y propiedades de las  $\varepsilon$ , implicaría que la ley de probabilidad  $nN$  dimensional de las  $nN$  variables aleatorias  $x'$  pertenecen a una subclase restringida  $w$  de las clases de todas las posibles leyes de probabilidad  $nN$  dimensionales. Al mismo tiempo, es evidente que *toda nuestra teoría considera lo más posible las observaciones de las variables  $x'$  que estamos considerando* (las ecuaciones [15.5], por otra parte, dicen mucho más acerca de las variables  $x'$  y  $\varepsilon$  cuando se consideran conjuntamente, pero no es posible observar las  $\varepsilon$  —por hipótesis—). Si añadimos un nuevo sistema de  $nN$  ecuaciones,  $x = x'$ , es decir, si identificamos cada variable teórica  $x'$  en el sistema [15.2] con la correspondiente variable observada en [15.1], nuestra teoría se convierte en una hipótesis estadística, a saber: la hipótesis  $p(w) \in \omega$ . Formularemos esto con algo más de detalle.

D.—*Formulación de [15.5] como una hipótesis estadística respecto a la ley de probabilidad de las variables observadas*

Indiquemos por  $\in$  un punto en el espacio muestra  $mN$  dimensional de la variable  $\varepsilon$  en [15.3].  $D(\in \varepsilon v)$ , en donde  $v$  es el argumento de la función de conjunto  $D$ , indica la integral de la ley de

probabilidad de las  $mN$  variables  $\epsilon$  dados los valores de las  $(n - 1)N$  variables

$$x'_{2t}, x'_{3t}, \dots, x_{nt} \quad (t = t_1, t_2, \dots, t_N)$$

(las "variables independientes"). Esta distribución está a nuestra disposición para formular la teoría. Por lo tanto, pertenece —por hipótesis— a un cierto conjunto  $S$  de leyes de probabilidad  $mN$  dimensionales. En el caso en que se halla especificado totalmente la distribución  $D$  de las variables en nuestra teoría,  $S$  contiene solamente un elemento.

Consideremos el caso general en que los valores de los parámetros  $\alpha$ , en [15.5], no vienen dados por la teoría, sin que estén a nuestra disposición, es decir, estamos preparados a aceptar valores cualesquiera de  $\alpha$ . Entonces  $S$ , y las restricciones [15.5], definen una cierta clase,  $\omega$ , de leyes de probabilidad para  $x'$ . Esta clase se podría imaginar que se obtiene por el proceso siguiente:

Consideremos un solo miembro  $D$  del sistema  $S$  y todas las distribuciones conjuntas posibles de las variables  $x'$  sujetas a las restricciones [15.5] para un sistema fijo arbitrario de valores de  $\alpha$ . Repetido este proceso para 1), todos los sistemas de valores posibles de los parámetros, y 2), para todo miembro del sistema  $S$ . Todas las leyes de probabilidad conjunta de las variables  $x'$  obtenidas de esta forma forman la clase  $\omega$ .

Estamos interesados en conocer cuándo  $P(\omega)$ , es decir, la ley de probabilidad de las  $mN$  variables observadas  $x$ , pertenecen a  $\omega$ . La hipótesis que ha de contrastarse es, por lo tanto,

$$P(\omega) \in \omega; \text{ alternativas admisibles } P(\omega) \in (\Omega - \omega).$$

Esta formulación del problema de contrastar relaciones económicas es muy general. Con objeto de desarrollar contrastes no triviales es necesario imponer restricciones supletorias a los conjuntos  $\Omega$  y  $\omega$  (en particular para restringir el conjunto  $S$  de leyes de probabilidad condicional de las variables aleatorias  $\epsilon$ ). Mencionaremos algunos tipos importantes de restricciones del conjunto  $\Omega$  y  $\omega$ .

1.—Restricciones en las variables aleatorias  $\epsilon$  exigiéndolas que sigan ciertas leyes de probabilidad sencillas, o restricción en el



sistema  $S$  convirtiéndolo en cierta familia paramétrica de distribuciones, o en una distribución especificada.

2.—Restricción del conjunto de hipótesis admisible *a priori* del conjunto  $\Omega^0$  y del conjunto  $\omega$  definidos anteriormente, es decir, los conjuntos de todas las distribuciones de probabilidad que son compatibles con [15.5] para por *lo menos un* sistema de valores de los parámetros  $\alpha$ , y entonces la restricción del conjunto de las leyes de probabilidad debe contrastarse respecto un conjunto particular  $\omega^0$  de este  $\Omega^0$ , correspondiente a un sistema fijo de valores de los parámetros  $\alpha$  (ejemplo de contraste de significación). Esto indica que estamos seguros —o que lo aceptamos sin contraste— que la teoría [15.5] es correcta cualquiera que sea la forma de las funciones  $f$ .

3.—Restricciones impuestas a las variables  $x'$  por algunas *otras* relaciones basadas en la teoría económica *además* de las [15.5]. Éste es un caso muy corriente cuando consideramos *sistemas* de relaciones económicas, y *debe* tenerse en cuenta cuando se formula el conjunto  $\omega^2$  anterior.

Una pregunta interesante e importante es la siguiente: ¿Un contraste de la hipótesis [15.6] es también un contraste de la corrección en la forma de las  $f$  en [15.5]?

Antes de nada preguntemos ¿qué es un sistema correcto de funciones  $f$ ? Demos una respuesta precisa a esta pregunta: *Cualquier* sistema de funciones  $f$  tal que

$$[P(\omega)] \in \omega(f_{11} \dots f_{1N})$$

en donde  $\omega(f_{11} \dots f_{1N})$  o, abreviadamente,  $\omega(f)$  indica el conjunto  $\omega$  (o  $\omega^0$ ), correspondiente a este sistema de las  $f$ , es un sistema correcto de funciones  $f$ . Habrá, en general, una infinidad de teorías “correctas” [15.5]. En particular, habrá varios sistemas diferentes de  $f$ , las cuales —junto con varios supuestos acerca de las propiedades de la distribución de las  $\varepsilon$ — *darán todas idénticamente* el mismo conjunto de leyes de probabilidad  $\omega$ , es decir, *son indistinguibles* desde el punto de vista de las observaciones. Esto, por otra parte, no indica que todas las formas “correctas” de la teoría sean igualmente buenas o interesantes; para los propósitos de la predicción de la “bondad” de una relación estocástica, si es co-

recta, en general se juzgará a partir de las propiedades de las variables aleatorias  $\varepsilon$  que contiene. Generalmente queremos que los errores sean pequeños en uno u otro sentido.

Sea  $\omega(f^0, S)$  un conjunto de leyes de probabilidad de las variables  $x'$  definida para un sistema particular  $f^0$  de funciones en [15.5] y un conjunto  $S$  de distribuciones de  $\varepsilon$ . Entonces si un contraste  $W_0$  de la hipótesis

$$P(\varepsilon) \in \omega(f^0, S)$$

debe tener una *alta potencia* con respecto de toda alternativa no contenida en  $\omega(f^0, S)$  el contraste  $W_0$  podrá, por su parte, tener también una alta potencia de detección, en particular de decisiones falsas sobre las formas de las  $f^0$ .

Si queremos, no obstante, contrastar una hipótesis [15.6], siendo las alternativas *casi todas* (es decir, el conjunto de hipótesis admisibles *a priori* es  $\Omega$ ) entonces no importa el contraste elegido; habrá siempre dentro de estas "casi todas" alternativas algunas para las cuales la potencia del contraste es muy pobre, en el caso de que una de estas alternativas fuese cierta tendríamos sólo una posibilidad muy pequeña de rechazar la hipótesis comprobada. En todos los casos prácticos, es por lo tanto necesario que seamos capaces de restringir, de antemano, el conjunto de hipótesis admisibles  $\Omega^0$  tanto como sea posible, teniendo al mismo tiempo fuertes motivos para creer que la hipótesis verdadera no está fuera de  $\Omega^0$ .

---

No hemos entrado en detalles técnicos acerca de la construcción de contrastes, la teoría de los cuales se ha descrito brevemente en la sección 14. Nuestro propósito ha sido demostrar cómo debe un economista formular problemas de contrastación para que los pueda resolver un estadístico. Para dar un ejemplo más concreto, en la próxima sección consideremos un ejemplo sencillo, pero importante, de economía estadística: el problema de contrastar la tendencia de una serie de tiempo suponiendo que las variables aleatorias adicionales son de un tipo muy sencillo.

16.—Ejemplo de contrastación de hipótesis. Un ejemplo sencillo del ajuste de una tendencia

Sea  $y_t$  una serie de tiempo observable, en donde  $t = 1, 2, \dots, N$  indica  $N$  puntos equidistantes en el tiempo. Supongamos que *conocemos, o creemos sin contraste*, que el siguiente modelo (en donde  $E$  indica "valor esperado de") es cierto:

$$y_t = k t + b + \varepsilon_t \quad (t = 1, 2, \dots, N) \quad [16.1]$$

$$E(y_t | t) = k t + b \quad (t = 1, 2, \dots, N) \quad [16.2]$$

$$E(\varepsilon_t) = 0 \quad E(\varepsilon_t^2) = \sigma^2 \quad (\text{independiente de } t) \quad [16.2']$$

$$\rho(y_t | t) = \frac{1}{\sigma \sqrt{2\pi}} \exp \left[ -\frac{1}{2\sigma^2} (y_t - k t - b)^2 \right] \quad [16.3]$$

$\sigma$  se supone numéricamente conocido.

Consideremos  $N$  poblaciones (o universos) correspondiente a  $N$  valores fijados  $1, 2, \dots, N$  de  $t$ . Para cada  $t$ ,  $y_x$  se distribuye normalmente con media  $(k t + b)$  y varianza  $\sigma^2$ . Para cada valor de  $t$  suponemos que obtenemos exactamente un valor de  $y_t$  de tal forma que estas extracciones sean estocásticamente independientes. La distribución en el muestreo de estas  $N$  extracciones es, por lo tanto:

$$\rho(y_1, y_2, \dots, y_N) = \frac{1}{(\sigma \sqrt{2\pi})^N} \exp \left[ -\frac{1}{2\sigma^2} \sum (y_t - k t - b)^2 \right]$$

( $\Sigma$  significa  $\sum_{t=1}^N$  en toda esta sección)

Todas estas cosas se suponen conocidas; el único elemento desconocido son los valores de  $k$  y  $b$ , es decir, sabemos que es posible elegir  $b$  y  $k$  tal que la serie observable  $y_t$  satisfaga nuestro modelo.

Por el método de los mínimos cuadrados (o por el de la máxima verosimilitud aplicado a [16.4] obtenemos la siguiente fórmula de estimación para el parámetro  $k$ :

$$\text{Estimación de } k = k = \frac{\sum (t - \bar{t}) (y_i - \bar{y})}{\sum (t - \bar{t})^2} \quad [16.5]$$

en donde  $\bar{t}$  e  $\bar{y}$  indican medias aritméticas observadas de  $t$  e  $y$  respectivamente.  $\hat{k}$  es, por otra parte, una variable aleatoria en muestras repetidas (de  $N$  extracciones las  $t$  son las mismas siempre). Empleando [16.1] tendremos:

$$\hat{k} = \frac{\sum (t - \bar{t}) (k t + b + \varepsilon_i - k \bar{t} - b - \bar{\varepsilon})}{\sum (t - \bar{t})^2} \quad [16.6]$$

de esta forma  $E(\hat{k}) = k$ , es decir, será una estimación inesgada. Queremos contrastar la hipótesis de que  $k = 0$ . ¿Cuál es el conjunto de hipótesis admisibles *a priori*, o sea el conjunto  $\Omega$ ? Es éste. El sistema de todas las distribuciones de probabilidad [16.4] obtenidas dejando que  $k$  y  $b$  recorran (independientemente) todos los valores de  $-\infty$  a  $+\infty$  y *no otros*. La hipótesis que hay que contrastar es que  $k = 0$  frente a estar  $b$  comprendido entre  $-\infty + \infty$ , es decir, el conjunto  $\omega^0$  es el sistema de todas las distribuciones de probabilidad obtenidas de [16.4], haciendo  $k = 0$  y dejando que  $b$  tome sucesivamente todos los valores de  $-\infty$  a  $+\infty$ . Tenemos que contrastar, por lo tanto, una hipótesis compuesta.

Para contrastar  $k = 0$  debemos elegir una región crítica de rechazo  $W_0$  en el espacio muestra  $N$  dimensional de las variables y tal que la probabilidad de que un punto muestra caiga en  $W_0$ , no importa cuál sea el valor de  $b$ , es igual a  $\alpha$  (es decir, 0,05) cuando la hipótesis  $k = 0$  es cierta; y además, la región  $W_0$  deberá ser tal que la probabilidad de que un punto muestra caiga en ella cuando la hipótesis  $k = 0$  es *falsa* sea lo mayor posible, e independiente del valor de  $b$ .

Con este propósito consideremos la distribución del estimador  $k$ . De [16.6] se ve que  $k$  es una función lineal de  $N$  variables  $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_N$  lineales independientemente distribuidas y normales, las  $t$  son un conjunto de constantes; por hipótesis,  $\hat{k}$  se distribuye normalmente con

$$\text{media} = k \quad \text{varianza} = \frac{\sigma^2}{\Sigma (t - \bar{t})^2} \quad [16.7]$$

La distribución de  $\hat{k}$  es independiente de  $b$ , y tendremos:

$$\rho(\hat{k}) = \frac{\sqrt{\Sigma (t - \bar{t})^2}}{\sigma \sqrt{2\pi}} \exp \left[ -\frac{\Sigma (t - \bar{t})^2}{2\sigma^2} (\hat{k} - k)^2 \right] \quad [16.8]$$

y, de acuerdo con nuestra hipótesis a contrastar,  $k = 0$ , tendremos:

$$\rho_0(\hat{k}) = \frac{\sqrt{\Sigma (t - \bar{t})^2}}{\sigma \sqrt{2\pi}} \exp \left[ -\frac{\Sigma (t - \bar{t})^2}{2\sigma^2} \hat{k}^2 \right]. \quad [16.8']$$

Consideremos las dos siguientes *colas* iguales de la distribución:

$$\hat{k} < -K = \frac{-c\sigma}{\sqrt{\Sigma (t - \bar{t})^2}} \quad \text{y} \quad \hat{k} > +K = \frac{+c\sigma}{\sqrt{\Sigma (t - \bar{t})^2}} \quad [16.9]$$

en donde  $c$  es una constante positiva determinada, esto es:

$$1 - \int_{-K}^{+K} \rho_0(\hat{k}) d\hat{k} = \alpha. \quad [16.10]$$

Los dos intervalos [16.9] definen una cierta región de rechazo  $W_0$  en el espacio muestra de las variables  $y$  y a causa de que  $k$  es, por [16.5], una función unívoca de las  $y$ . La probabilidad —cuando la hipótesis  $k = 0$  es cierta— de que  $\hat{k}$  puede caer en cualquiera de los dos intervalos [16.9] es la misma que la probabilidad de que el punto muestra caiga en  $W_0$ , y esta probabilidad es exactamente igual a  $\alpha$ . Por otra parte, ¿cuáles son las propiedades de esta región crítica si la hipótesis es falsa, es decir, si  $k \neq 0$ ? Se ha demostrado que la región de rechazo  $W_0$  correspondiente a las dos colas [16.9] tiene las siguientes propiedades (10):

(10) Ver NEYMAN: "Lectures and Conferences on Mathematical Statistics", Washington, 1937, pág. 20.

Cuando la hipótesis  $k = 0$  es falsa, es decir, cuando  $k \neq 0$ , la probabilidad de que el punto muestra caiga en  $W_0$  (es decir, la potencia del contraste) es  $\alpha > 0$ , lo cual indica que el contraste es inesgado. Y que para cualquier otra región crítica inesgada de tamaño  $\alpha$  la potencia es menor.

Si rechazamos la hipótesis  $k = 0$  cuando  $\hat{k}$  cae en uno de los intervalos [16.9] tendremos el *mejor contraste inesgado* de la hipótesis  $k = 0$  correspondiente al nivel de significación  $\alpha$ .

La probabilidad de que  $\hat{k}$  pueda caer en cualquiera de los intervalos [16.9] cuando  $k \neq 0$ , es decir, la potencia del contraste, puede calcularse como una función de  $k$  directamente de [16.8]. Esta *función de potencia* —se la designa por  $\beta(k)$ — es sencillamente:

$$\beta(k) = 1 - \int_{-K}^K \frac{\sqrt{\Sigma(t - \bar{t})^2}}{\sigma \sqrt{2\pi}} \exp\left(-\frac{\Sigma(t - \bar{t})^2}{2\sigma^2} (\hat{k} - k)^2\right) d\hat{k} \tag{16.11}$$

en donde  $K$  viene dada por [16.9].

Como ejemplo tenemos

$$N = 9, \quad \sigma = 1, \quad \alpha = 0,05, \quad c = 1,96$$

(de las tablas de la distribución normal). Tendremos entonces:

$$\Sigma(t - \bar{t})^2 = 60.$$

Si introducimos estos valores numéricos y cambiamos la variable de integración por la transformación

$$K = (1/\sigma) \sqrt{\Sigma(t - \bar{t})^2} (\hat{k} - k)$$

[16.11] se convierte en:

$$\beta(k) = 1 - \frac{1}{\sqrt{2\pi}} \int_{-1,96 - \sqrt{60}k}^{1,96 - \sqrt{60}k} \exp\left(-\frac{1}{2}K^2\right) dK \tag{16.11'}$$

Valores de  $\beta(k)$  para valores diferentes de  $k$  se obtienen directamente de las tablas de la distribución normal.

En la tabla 1 se dan los resultados para unos pocos valores de  $k$ . La curva de la figura 3 representa la función continua de la potencia  $\beta(k)$ .

TABLA 1

Si el verdadero valor de $k$ es	La probabilidad de que rechacemos $k = 0$ por el contraste [16.9] (es decir, la potencia del contraste)
$k$	$\beta(k)$
0	0,05
$\pm 0,1$	0,12
$\pm 0,2$	0,34
$\pm 0,3$	0,64
$\pm 0,4$	0,87
$\pm 0,5$	0,97

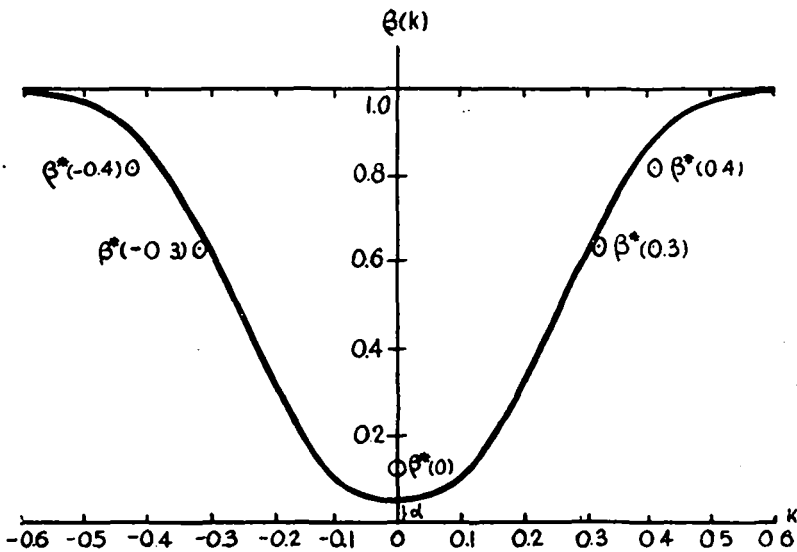


Figura 3.—Función de potencia del contraste [16.9]

(N = 9,  $\sigma = 1$ ,  $\alpha = 0,05$ ,  $c = 1,96$ )

En el eje horizontal van valores dados de  $k$  ( $k = 0$  es la hipótesis contrastada; otros valores de  $k$  representan hipótesis alternativas).

En el eje vertical van valores de  $\beta(k)$  representando la probabilidad de que  $\hat{k}$  caiga en la región de rechace [16.9] para la hipótesis  $k = 0$ , cuando  $k$  es el verdadero valor de estos parámetros.  $\alpha = 0,05$  representa el nivel de significación.

Los puntos señalados por un círculo (©) indican la potencia del mismo contraste [16.9] cuando las  $\epsilon$  son independientes y vienen definidas por [16.12].

Este gráfico indica para  $k \neq 0$  la probabilidad de rechazar —correctamente— la hipótesis  $k = 0$  cuando es falsa. Cuanto más nos alejemos de  $k = 0$  mayor será la probabilidad de rechazar  $k = 0$ ;  $\beta(k)$  es la probabilidad de no cometer un error de la segunda clase, considerado como función de  $k$ .

Consideremos un ejemplo mostrando lo que sucede si la hipótesis alternativa que realmente es cierta no se incluye en el conjunto de hipótesis admisibles *a priori*  $\Omega^0$  que fué la base del contraste anterior.

Una de las restricciones impuestas anteriormente fué la de que las  $\epsilon$ , en las  $N$  observaciones, eran estocásticamente independientes. Esto se considera como un hecho conocido y no como una hipótesis que pudiese ser cierta o falsa. Supongamos que esto no estuviese justificado. Así, por ejemplo, sin *nuestro conocimiento*, y mientras procedíamos como si nuestro esquema original fuese correcto, las series de las  $\epsilon$  eran del siguiente tipo:

Sean

$$\xi_0, \xi_1, \dots, \xi_N,$$

$N + 1$  variables distribuidas independientemente con medio cero y varianza  $\sigma^2 =$  que la de las anteriores  $\epsilon$ . Consideremos una nueva serie de  $\epsilon$  dadas por las fórmulas

$$\epsilon_t = \frac{1}{\sqrt{2}} (\xi_{t-1} + \xi_t) \quad (t = 1, 2, \dots, N). \quad [16.12]$$

Cada una de estas nuevas  $\epsilon$  tomadas separadamente tiene como



medio cero, y la misma varianza  $\sigma^2$ . Pero las  $\varepsilon_t$  y  $\varepsilon_{t+1}$  están positivamente correlacionadas (coeficiente de correlación 1/2).

Supongamos que procedemos como si estuviéramos tratando con la serie  $\varepsilon$  original, en lugar de [16.12]. Por [16.6]  $\hat{k}$  es una función lineal de variables normalmente e independientemente distribuidas, es decir:

$$\hat{k} = k + \frac{\sum (t - \bar{t}) (\xi_{t-1} + \xi_t)}{\sqrt{2} \sum (t - \bar{t})^2} \quad [16.13]$$

y, por lo tanto,  $\hat{k}$  ahora también se distribuye normalmente con media =  $k$ , y la varianza de  $\hat{k}$  es ahora la de la función lineal

$$\frac{\sum (t - \bar{t}) (\xi_{t-1} + \xi_t)}{\sqrt{2} \sum (t - \bar{t})^2} \quad [16.14]$$

la cual da

$$\begin{aligned} \sigma^2_{\hat{k}} &= \frac{1}{[\sum (t - \bar{t})^2]^2} [\sum (t - \bar{t})^2 \sigma^2 + \\ &+ \sum_1^{N-1} (t - \bar{t}) (t + 1 - \bar{t}) \sigma^2]. \end{aligned} \quad [16.15]$$

Tomando como en el ejemplo previo

$$N = 9, \quad \sigma = 1 \quad c = 1,96$$

obtenemos

$$\sigma^2_{\hat{k}} = \frac{60^2}{1} (60 + 40) = \frac{36}{1} \quad [16.16]$$

y por lo tanto, por analogía con [16.11], tendremos:

$$\beta^*(\hat{k}) = 1 - \frac{1}{\sqrt{2\pi}} \int_{-1,52 - 6k}^{1,52 - 6k} \exp\left(-\frac{1}{2} K^2\right) dK \quad [16.17]$$

$\beta^*(k)$  es la probabilidad de que  $\hat{k}$  —calculado por [16.5] aunque las  $\varepsilon$  están definidas por [16.12]— caiga en cualquiera de los dos intervalos definidos por [16.9]; esta probabilidad se considera como una función del parámetro  $k$ . Como ejercicio, los valores de  $\beta^*(k)$  para

$$k = 0 \quad k = \pm 0,3 \quad \text{y} \quad k = \pm 0,4$$

se han señalado en la figura 3 (puntos dentro de un círculo). Estos valores obtenidos de [16.17] son

$$\beta^*(0) = 0,13, \quad \beta^*(\pm 0,3) = 0,61 \quad \text{y} \quad \beta^*(\pm 0,4) = 0,81.$$

¿Qué es lo que indican estos resultados? Indican que el contraste [16.9] para una hipótesis alternativa (dependiente de  $\varepsilon$ ) no incluida en  $\Omega^0$  puede —incorrectamente— rechazar la hipótesis contrastada (es decir,  $k = 0$ ) cuando es cierta más frecuentemente que lo supuesto (13 % en lugar del 5 %). Habíamos construido el contraste de forma que rechazaba la hipótesis  $k = 0$  —cuando era cierta— en solamente el 5 % de los casos en que el contraste se aplica. Pero esto ahora no ocurre. El motivo es fácil de encontrar. Con objeto de hacer  $\alpha = 0,05$  en nuestro primer ejemplo (con errores independientes), fijábamos un valor  $c$  tal que la integral en [16.10] fuese igual a 0,95. En el caso presente, la integral sobre el mismo recorrido (dado por [16.9]) es menor que 0,95 a causa de que la varianza de  $k$  ahora es mayor (1/36 en lugar de 1/60) uno menos, esta integral es, por lo tanto, mayor que  $\alpha = 0,05$ .

También vimos que rechazábamos la hipótesis falsa  $k = 0$  en el 87 % de aquellos casos en que  $k = \pm 0,4$ , mientras de hecho ahora es solamente en el 81 % de los casos a causa de que cuando construimos el contraste para  $k = 0$  no tomamos en consideración la posibilidad de que las  $\varepsilon$  pudiesen ser dependientes.

La hipótesis  $k = 0$ , así como la hipótesis alternativa acerca de  $k$  en la última construcción, no indica la misma cosa que en el primer ejemplo con errores independientes. En particular, la hipótesis contrastada (es decir,  $k = 0$ ) no es para la que hemos construido el contraste, a causa de que ahora incluye la posibilidad de que los errores sean dependientes. En otras palabras: Aunque la hipótesis  $k = 0$  puede ser cierta, hay algo aquí que es falso en este

caso —respecto la hipótesis contrastada en el caso de errores independientes— ya que el coeficiente de correlación entre las  $\varepsilon$  hay que tomarlo en consideración. Es de interés hacer notar que el contraste anterior indica esto bastante bien al rechazar la hipótesis  $k = 0$  un 13 % en lugar del 5 % en los casos en que  $k = 0$  es cierta. Este resultado, sin embargo, no es general. Lo opuesto puede ocurrir en otros casos.

Por otra parte, en el caso anterior, el fallo puede no ser muy grande a causa de que *sucede* que la potencia del contraste es bastante buena también para la hipótesis fuera de  $\Omega^0$  que habíamos considerado. Y en muchos casos importantes esto puede suceder, es decir, siempre que desarrollemos un contraste solamente respecto a una clase muy restringida de hipótesis admisibles *a priori*  $\Omega$ , este contraste podría ser bueno respecto a una clase de alternativas más amplia.

El ejemplo anterior aclara, yo creo, un método muy útil de proceder para contrastar relaciones económicas. Definiremos primeramente un cierto conjunto de esquemas admisibles *a priori*,  $\Omega^0$ , conteniendo las que consideramos como las más importantes alternativas, y son al mismo tiempo de tal forma que pueden experimentarse sin dificultades técnicas prohibitivas. Entonces, si llegamos a tener sospechas —por una u otra razón— sobre si  $\Omega^0$  es *completa*, podríamos estudiar la potencia del contraste para ciertos esquemas exteriores no contenidos en  $\Omega^0$ . Por ejemplo, puede suceder que una cierta hipótesis exterior a  $\Omega^0$ , es decir, una hipótesis rechazada *a priori*, si a pesar de todo fuese la verdadera, podría tener consecuencias importantes para nuestra decisión. Para ver en *qué riesgo estamos* incurriendo empleando un contraste que rechaza la posibilidad de que esta hipótesis sea cierta, calcularemos la potencia de la prueba para esta hipótesis exterior.

Por otra parte, cualquiera que sea el contraste desarrollado sobre la base de un cierto conjunto  $\Omega^0$ , de hipótesis admisibles *a priori*, será siempre posible encontrar hipótesis fuera de  $\Omega^0$ , tal que la potencia del contraste respecto a estas hipótesis sea muy pobre, o por lo menos esto es lo que queremos que ocurra con un contraste que sea bueno *dentro* de  $\Omega^0$ . Para tener alguna posibilidad de alcanzar conclusiones no triviales debemos suponer que tenemos

un cierto conocimiento *a priori*, o deseamos tener un cierto riesgo con objeto de restringir  $\Omega^0$ ; el riesgo total que lleva el restringir  $\Omega^0$  es tal que no puede calcularse en términos de probabilidad. La elección de un conjunto admisible *a priori*  $\Omega^0$  es, indudablemente, objeto de conocimiento general e intuición.

La discusión anterior da, yo lo creo, una interpretación clara a la frase "supongamos que el total fundamento formal de la teoría es falso, ¿qué significado tiene contrastar coeficientes, etc.? De hecho, esta cuestión está, hablando estrictamente, justificada, siempre que tratemos de explicar la realidad por un modelo teórico. Pero si esta actitud se lleva al extremo nunca seremos capaces de realizar nada en el camino de explicar fenómenos reales.

#### 17.—*El significado de la frase "Fórmulas teóricas por contemplación de los datos"*

Todos los modelos de teoría económica, por abstractos que sean, probablemente se deducen de la consideración de algún fenómeno real. Los "datos", en el sentido amplio del conocimiento empírico, estarán siempre en alguna extensión influyendo nuestra formulación de teorías acerca de ellos.

Si tratamos de dar solamente descripciones simplificadas y condensadas de *casos empíricos*, no hay riesgo al elegir una teoría que "se ajuste bien". El riesgo se presenta si generalizamos en el siguiente sentido. Especificamos una *clase empírica* de fenómenos (por ejemplo, la clase de todos los valores correspondientes de precios y cantidades vendidas de una cierta mercancía). Conocemos empíricamente un cierto número de miembros de estas clases. Especificamos una *clase teórica* (por ejemplo, una relación estocástica: precio-cantidad) que abarca en particular los miembros conocidos de la clase empírica. *Esperamos* que la clase teórica *cubrirá* todos los miembros de la clase empírica. La construcción de tal clase teórica es, sin duda, el problema de la ciencia inductiva. Lleva consigo riesgos de fallos que están fuera de nuestro control. Una discusión general sobre "verdad" o "mentira" en relación con tales procesos empíricos inductivos cae en la metafísica.

Pero la frase "fórmulas teóricas por contemplación de datos"

tiene, entre los que trabajan en investigaciones económicas, un estricto significado, el cual consideraremos mientras se aclara. El argumento corriente es el siguiente: Supongamos que tenemos un cierto número de observaciones o valores simultáneos de un sistema de variables económicas y que hay *algunas* relaciones entre las variables sin especificar la forma de las mismas. Eliminamos un gran número de relaciones y encontramos una que se "ajusta a los datos" (en un sentido u otro). Ahora bien, si encontramos la forma de una relación que se "ajuste bien", ¿es esto, en sí mismo, una comprobación de la bondad de esta relación como una teoría?, ¿no es esta fórmula una afirmación trivial respecto los hechos?

Muchas discusiones se han suscitado por este motivo en relación con el problema de contrastar teorías de la vida de los negocios. Por ejemplo, una gran cantidad de modelos simplificados dinámicos implican que cada una de las variables que se presentan satisfaga (aparte de un término de error) alguna ecuación en diferencias lineales de un cierto orden, con coeficientes constantes. Es evidente que podríamos alcanzar este mismo resultado partiendo de modelos fundamentalmente diferentes, es decir, podríamos construir un gran número de modelos que fuesen muy distintos, así como sus supuestos básicos o el tipo de mecanismo económico que describen. Ahora bien, si las series observadas satisfacen algunos ciclos bastante regulares, tales ecuaciones en diferencias pueden corrientemente ajustarse a estas series muy bien por una elección adecuada de los coeficientes. Y si aceptamos esto como una comprobación de que la serie real satisface tal ecuación en diferencias finitas, podríamos decir que la teoría "correcta" podría pertenecer a la clase de modelos que se relacionan con tales ecuaciones en diferencias. Pero no podemos, de este ajuste, deducir la teoría "correcta" de las clases de modelos admisibles. Si elegimos un modelo particular, el hecho de que la ecuación correspondiente en cada variable pueda ajustarse a los datos no da una garantía de que el modelo sea correcto. Se arguye generalmente que tales ajustes buenos de ecuaciones "finales" no indican mucho desde el punto de vista de comprobación de teorías.

Este argumento no cubre totalmente el punto real de fricción. En efecto, si pudiésemos establecer que las variables observadas satisfacen muy acusadamente un cierto sistema de ecuaciones en

diferencias lineales (por ejemplo), tendríamos una restricción muy útil y fuerte sobre la clase de los modelos admisibles *a priori*. En general, cuando podemos establecer que ciertos datos satisfacen relaciones, añadimos algo más a nuestro conocimiento: una restricción en la clase de las hipótesis admisibles. La *dificultad real* reside en decidir cuándo una relación dada es o no compatible con los datos; y la cosa importante que hay que analizar es la realidad del contraste por el que se hace la decisión, puesto que tratamos de relaciones estocásticas y variables aleatorias, no relaciones exactas.

Desde este punto de vista, no hay objeción justificada cuando se consideran varias teorías para encontrar una que se "ajuste a los datos". Pero se pueden hacer objeciones respecto ciertos "métodos de contrastar el ajuste". Examinaremos esto con más detalle.

Consideremos un sistema de variables aleatorias observadas, como en [15.1], y una relación a contrastar, tal como la [15.5]. La teoría define una clase  $\omega^0$  de leyes de probabilidad, y queremos contrastar  $P(\omega) \in \omega^0$ . Hemos visto que, con objeto de desarrollar un contraste de esta hipótesis, tenemos que definir un conjunto,  $\Omega^0$ , de hipótesis admisibles *a priori*. Sea  $(\omega^0)$  un sistema de conjuntos diferentes  $\omega^0$ , correspondiente a distintas relaciones a contrastar y tal que cada  $\omega^0$  esté contenido en  $\Omega^0$ . Para *cualquier* conjunto  $\omega^0$  podríamos contrastar la hipótesis  $P(\omega) \in \omega^0$ ; el conjunto de hipótesis admisibles *a priori* será siempre el mismo  $\Omega^0$ . Es, evidentemente irrelevante, cómo debemos elegir la hipótesis que ha de ser contrastada *dentro* de  $\Omega^0$ . En particular, la hipótesis puede ser una que se sugiera por la inspección de los datos. Esto es perfectamente legítimo, tanto más cuanto el *conjunto*  $\Omega^0$  de *alternativas es y permanece fijo a priori*. Entonces podremos calcular la potencia del contraste empleado y veremos qué riesgo tenemos si aceptamos la hipótesis contrastada. Lo que no se puede permitir es que  $\Omega^0$  sea *una función del punto muestra*. A causa de que el contraste no controla los dos tipos de error. Si  $\Omega^0$  se fija sobre la base de un punto muestra, y se desarrolla un contraste respecto este conjunto de hipótesis admisibles, no tendremos nunca idea de cuándo la verdadera hipótesis está contenida en  $\Omega^0$  o no. Tendríamos la situación inconcebible de que el método de contrastación variaría aleatoriamente de prueba a prueba.

La cosa esencial es, por lo tanto, no la forma en que elijamos la hipótesis de contrastar. Lo esencial es que *sepamos o creamos* cuál es la clase de la hipótesis admisible *a priori*, y que la potencia de nuestro contraste de rechazar la hipótesis contrastada sea "realmente diferente" cuando la alternativa sea cierta.

## CAPITULO V

### PROBLEMAS DE ESTIMACION

En la sección 14 describimos el problema y el principio general de la estimación estadística. Problemas más específicos de estimación se presentan en varios campos de aplicaciones. A continuación discutiremos un problema que es particularmente importante en la investigación económica: el de estimar parámetros en un sistema de ecuaciones estocásticas.

Un procedimiento peligroso —pero corrientemente empleado— en este campo es ajustar cada ecuación separadamente sin considerar el hecho de que las variables que existen se suponen, generalmente, que satisfacen simultáneamente, un cierto número de *otras* relaciones estocásticas. Si esto es cierto será una casualidad que no creemos una inconsistencia interior en el sistema total. Así, por ejemplo, el supuesto que alguna de las variables en una ecuación permanece constante en muestras repetidas —a causa de su inclusión en otra ecuación del sistema— es imposible. Aclaremos esto por un ejemplo posterior (ver sección 21).

Aunque tal inconsistencia no se produzca, el procedimiento de ajustar cada ecuación separadamente, corrientemente no dará las estimaciones más eficientes de los parámetros. Porque una información adicional acerca de los parámetros contenidos en una ecuación puede obtenerse del hecho de que, simultáneamente, las variables satisfacen otra ecuación. Y que esto es importante lo podemos deducir del hecho de que algunos de los parámetros a estimar pueden, en efecto, ser *arbitrarios* respecto al *sistema* de ecuaciones. Este es el *lado estadístico* del problema de las relaciones autónomas que se discutió en la sección 8, y que puede describirse con las siguientes palabras:

Supongamos que un cierto conjunto de variables económicas satisfice un sistema de ecuaciones (estáticas o dinámicas), cada una de las cuales esperamos que tenga un cierto grado de autonomía; estamos interesados en medir los parámetros constantes que aparecen (por ejemplo, ciertas elasticidades). De este sistema de ecuaciones, por operaciones algebraicas, podemos deducir una infinidad de sistemas confluentes. Supongamos que, en particular, es posible deducir una infinidad de nuevos sistemas que tienen exactamente la misma forma que el sistema original, pero con diferentes valores de los coeficientes (corrientemente esto indica que el número de parámetros del sistema de ecuaciones puede reducirse, como se explicará en la sección 19). Entonces, si no conocemos nada acerca de los valores de los parámetros en el sistema de ecuaciones originales, es evidente que no es posible obtener una estimación única para cualquier número de observaciones de la variable. Y si decimos que hemos obtenido alguna estimación que parece ser única en este caso puede deberse solamente a la aplicación de fórmulas de estimación que dan lugar a resultados espúreos o sesgados. Así, el problema de deducir curvas de demanda y oferta del mismo conjunto de datos de precio y cantidad es un ejemplo clásico de este tipo de problema.

Esta cuestión (en el caso de relaciones lineales, conocida como el problema de multicolinealidad) es de gran importancia en la investigación económica, a causa de que tales investigaciones se han realizado basándose en observaciones pasivas de los hechos, en lugar de en datos obtenidos en experimentos racionalmente planeados (ver capítulo II). Y esto indica que podemos obtener solamente tales datos como resultado del sistema económico que en efecto es, y no como el que debía ser bajo aquellas variaciones hipotéticas irrestrictas con las cuales operamos en teoría económica, y que son las que nos interesan para los propósitos de Economía política. Una considerable clarificación de este punto se ha alcanzado en los últimos años después del primer trabajo de Frisch (11).

---

(11) R. FRISCH: "Correlation and Scatter in Statistical Variables", *Nordic Statistical Journal*, vol. 1, 1929, págs. 36-102; "Statistical Correlation and the Theory of Cluster Types" (junto con B. D. MUDGETT), *Journal of American Statistical Association*, vol. 26, December 1931, págs. 375-392; *Pitfalls in the Statistical Construction of Demand and Supply Curves* ("Veröffentlichungen



A continuación veremos que la investigación de este problema de coeficientes indeterminados, así como otros problemas de estimación en relación con los sistemas de ecuaciones económicas, se convierten en la misma cosa: *el estudio de las propiedades de la distribución de probabilidad conjunta de las variables aleatorias (observables) en un sistema de ecuaciones estocásticas.*

18.—*Formulación general del problema de estimar parámetros en sistemas de relaciones económicas* (12).

Discutiremos una clase general de sistemas estáticos dinámicos.

A.—*Sistemas estáticos.*

Indiquemos por

$$\xi_{1j}, \xi_{2j}, \dots, \xi_{mj}, \dots, \xi_{nj} \quad (j = 1, 2, \dots, N).$$

$N$  medidas verdaderas de  $n$  variables económicas. El subíndice  $j$  indica observación número  $j$ . Las medidas de estas variables pueden (y corrientemente ocurre) estar sometidas a *errores de medida*. Las variables correspondientes realmente observadas serán  $x_{ij}$  definidas por

$$x_{ij} = G_{ij}(\xi_{ij}, \eta_{ij}) \quad (i = 1, 2, \dots, n; j = 1, 2, \dots, N) \quad [18.1]$$

o cuando se resuelve para  $\eta_{ij}$

$$\eta_{ij} = g_{ij}(x_{ij}, \xi_{ij}) \quad [18.1']$$

---

der Frankfurter Gesellschaft für Konjunkturforschung, *Neue Folge*, Heft 5), Leipzig, 1933, 39 págs.; *Statistical Confluence Analysis by Means of Complete Regression Systems*, Publication no. 5 from the Institute of Economics, Oslo, 1934; "Statistical versus Theoretical Relations in Economic Macro-Dynamics" (Memorandum mimeografiado preparado por la Conferencia de Business Cycle de Cambridge, England, July 18-20, 1938, to discuss J. Tinbergen's Publications of 1938 for the League of Nations). Ver también J. MARSCHAK: "Economic Interdependence and Statistical Analysis", in *Studies in Mathematical Economics and Econometrics*, in Memory of HENRY SCHULTZ, Chicago, 1942, págs. 135-150.

(12) Por motivos de sencillez nos limitamos en esta y en la sección siguiente al caso en que se supo existen leyes fundamentales de probabilidad. Sin embargo, no habrá ninguna dificultad para reformular la teoría sobre la base de la ley integral de probabilidad.

en donde las  $\eta_{ij}$  son variables aleatorias que caracterizan los errores de medidas y  $G_{ij}$  (así como  $g_{ij}$ ) son ciertas funciones *conocidas*. Introducimos estas funciones  $G_{ij}$  por la siguiente razón:

Si escribimos

$$x_{ij} = \xi_{ij} + \text{error} \tag{18.2}$$

la distribución de los errores en general dependerá de  $\xi_{ij}$ . Si esto fuese el caso, supondremos que es posible escribir la parte de error como una función *conocida* de  $\xi_{ij}$  y una nueva variable  $\eta_{ij}$ , la cual es estocásticamente independiente de  $\xi_{ij}$  (y también de  $\xi_{hk}$  cuando  $hk \neq ij$ ). Estas transformaciones están expresadas por las funciones  $G_{ij}$  en [18.1]. Esto da lugar:

*Supuesto 1.º* Las  $nN$  variables aleatorias

$$\eta_{ij} (i = 1, 2, \dots, n; j = 1, \dots, N)$$

tienen una ley conjunta de probabilidad (13) *elemental*

$$\rho_1 (\eta_{11} \dots \eta_{nN}; \gamma_1 \dots \gamma_q) \tag{18.3}$$

*conocida*, salvo —quizá— los valores de las  $q$  parámetros  $\gamma_1 \dots \gamma_q$ , e independiente de las variables  $\xi_{ij}$  y  $\varepsilon$  definidas anteriormente.

*Supuesto 2.º* Las  $(n - m)$   $N$  cantidades

$$\xi_{m+1 \cdot j} \dots \xi_{n \cdot j} (j = 1 \dots N; m < n)$$

se consideran *como constantes en muestras repetidas*. El significado económico de ellas es que estas variables son parámetros autónomos fijados por fuerzas externas en el sector económico bajo consideración.

*Supuesto 3.º* Las  $mN$  cantidades

$$\xi_{1j} \dots \xi_{mj} (j = 1, 2, \dots, N)$$

son variables aleatorias (“variables dependientes”) en muestras repetidas, y se sabe que satisfacen  $m$  ecuaciones estocásticas

$$f_i (\xi_{1j} \dots \xi_{mj}; \xi_{m+1 \cdot j} \dots \xi_{nj}; \alpha_1 \dots \alpha_k; \varepsilon_{1j} \dots \varepsilon_{hj}) = 0 \tag{18.4}$$

$$h \geq m; i = 1, 2, \dots, m; j = 1, 2, \dots, N$$

---

(13) Este supuesto puede y es deseable reemplazarlo por el supuesto de que las  $\xi$  autónomas son variables aleatorias. Esto puede causar solamente pequeños cambios en la subsiguiente formulación.

en donde

$$\alpha_1, \alpha_2, \dots, \alpha_k$$

son  $k$  constantes desconocidas y donde

$$\varepsilon_{1j}, \dots, \varepsilon_{hj} \quad (j = 1, \dots, N)$$

son  $hN$  variables aleatorias. Aquí  $\alpha_1 \dots \alpha_k$  indican todas las constantes desconocidas en el sistema de ecuaciones [18.4].

Estos pueden estar unos pocos presentes en cada una de las  $m$  ecuaciones; de forma análoga ocurre con las  $h, \varepsilon$ .

[18.4] representa una teoría económica cuando se imponen ciertas restricciones en la distribución de  $\varepsilon$  que caracterizan el modelo estocástico.

*Supuesto 4.º* Las  $hN$  variables

$$\varepsilon_{1j}, \dots, \varepsilon_{hj} \quad (j = 1, \dots, N)$$

tienen una ley elemental de probabilidad

$$\rho_2(\varepsilon_1, \dots, \varepsilon_{hN}; \beta_1, \beta_2, \dots, \beta_r) \quad [18.5]$$

(es decir, la distribución condicional de las  $hN\varepsilon$  cuando se dan las  $\xi$  autónomas) las cuales se conocen excepto —quizá— en los valores de  $r$  parámetros  $\beta_1 \dots \beta_r$ . Introduciendo un número considerable de  $\beta_1, \beta_2$  puede extenderse a una amplia clase de distribuciones.

*El problema es estimar los valores de*

$$\alpha_1, \alpha_2, \dots, \alpha_k$$

sobre la base de un punto muestra

$$(x_{11}, x_{21}, \dots, x_{n1}, \dots, x_{n1}, \dots, x_{2N})$$

en el espacio muestra  $nN$  dimensional de las variables observadas  $x$ . Puede ser o no necesario estimar los parámetros  $\gamma$  y  $\beta$  de [18.3] y [18.5]. Veremos ahora que este problema es un problema de estimación estadística igual al ya descrito en la sección 14.

De las  $mN$  ecuaciones [18.4] podremos (bajo ciertas condicio-

nes de resolubilidad) expresar  $mN$  de estas  $hN\epsilon$ , como funciones de las  $mN$  variables  $\xi_{11} \dots \xi_{mN}$  y las  $(h - m)N$  restantes  $\epsilon$ . Estas funciones en general llevarán parámetros  $\alpha$  y  $(n - m)N\xi$  autónomas. Introduciendo estas expresiones para las  $mN\epsilon$  en [18.5], multiplicando por el jacobiano de la transformación e integrando sobre las  $(h - m)N$  restantes  $\epsilon$  de  $-\infty$  a  $+\infty$  obtenemos la ley de probabilidad de las  $mN$  variables ("dependientes")  $\xi_{11} \dots \xi_{mN}$ . Esta ley de probabilidad será

$$\rho_3 [\xi_{11}, \dots, \xi_{mN} | \alpha_1, \alpha_2, \dots, \alpha_k; \beta_1, \dots, \beta_r; \xi_{m+1,1}, \dots, \xi_{nN}] \quad [18.6]$$

Esta es la distribución *condicional* de las  $mN$  variables  $\xi_{11} \dots \xi_{mN}$  para valores *dados* de las  $\xi$  autónomas. Por supuesto, esta distribución es independiente de las variables  $n$  definidas por [18.1]. Por lo tanto, la distribución conjunta de

$$(\xi_{11}, \dots, \xi_{mN}) \text{ y } (\eta_{11}, \dots, \eta_{nN})$$

será igual a

$$\rho_1 \cdot \rho_2 \quad [18.7]$$

Introduciendo la transformación [18.1'] en [18.7] e integrando el resultado respecto las  $mN$  variables aleatorias de  $-\infty$  a  $+\infty$ .  $\xi_{11} \dots \xi_{mN}$  tenemos la ley elemental de probabilidad conjunta de las  $nN$  variables  $x_{ij}$  (variables observables). Esta ley de probabilidad será:

$$\psi [x_{11} \dots x_{nN} | \xi_{m+1,1}, \dots, \xi_{nN}; \alpha_1 \dots \alpha_k; \beta_1 \dots \beta_r; \gamma_1 \dots \gamma_q] \quad [18.8]$$

Podremos decir ahora: Nuestra teoría económica, así como las variables observables  $x$  se consideran *como indistinguibles tomando como base el que las variables  $x$  tienen la ley de probabilidad conjunta [18.8] en donde  $\psi$  es una función conocida.* Y el problema de estimar los parámetros desconocidos se reduce a un problema ordinario de estimación.

Si, en particular, todas las variables se observan *sin error de medida*, nuestra teoría económica se podrá expresar por [18.6].

Si todas las  $\xi$  autónomas se miden sin error, podremos —en lugar de [18.8]— poner (14):

$$\psi_1 (x_{11} \dots x_{mN} | \xi_{m+1,1}, \dots, \xi_{nN}; \alpha_1 \dots \alpha_k; \beta_1 \dots \beta_r; \gamma'_1 \dots \gamma'_q) \quad [18.8']$$

(14)  $\gamma'_i$  indica los parámetros en una distribución  $mN$  dimensional en lugar de una distribución  $nN$  dimensional.

es decir, una distribución con solamente  $m$   $N$  en lugar de  $n$   $N$  variables aleatorias.

En [18.8] las  $(n - m)$   $N$   $\xi$  autónomas son parámetros conocidos, los cuales pueden o no, necesariamente, estimarse con objeto de estimar las  $\alpha$ .

Evidentemente, no se puede dar una descripción más completa de las interconexiones entre un cierto número de variables que la dada por la ley de probabilidad conjunta. Si, por lo tanto, *dos formulaciones* diferentes de una teoría económica dan la misma ley de probabilidad conjunta de las variables aleatorias observables, no podremos distinguirlas tomando como base las observaciones (pero las teorías pueden no ser equivalentes en ciertos aspectos).

La ley de probabilidad conjunta de todas las variables cubre también el caso particular en que el conjunto de variables aleatorias pueden dividirse en subgrupos de variables independientemente distribuidas con parámetros diferentes a estimar *apareciendo en la distribución de cada subgrupo*. En todos los otros casos la ley de probabilidad conjunta de todas las variables contiene más información que la obtenida de la ley de probabilidad de subgrupos de variables. Es, por lo tanto, evidente que la *ley de distribución conjunta* de todas las variables aleatorias observables en un sistema económico es la única base general para estimar los parámetros desconocidos del sistema.

### B.—Sistemas dinámicos.

Consideremos el siguiente tipo general de sistemas económicos dinámicos (haciendo supuestos análogos a los anteriores).

Sean

$$\xi_1(t_1), \xi_2(t_1) \dots \xi_m(t_1) \dots \xi_n(t_1)$$

$n$  series de tiempo definidas para los puntos en el tiempo

$$t_N, t_{N-1} \dots t_1, t_0, t_{-1}, t_{-2} \dots \quad [18.9]$$

Por el momento no consideraremos los problemas de errores de medidas.

Las  $(n - m)$  series

$$\xi_{m+1}(t_1) \dots \xi_n(t_1)$$

se supone que son variables autónomas que permanecen fijas en muestras repetidas.

Para cada punto en el tiempo [18.9] las cantidades

$$\xi_1(t_i) \dots \xi_m(t_i)$$

son variables aleatorias definidas implícitamente por un sistema de relaciones dinámicas del tipo

$$\begin{aligned} \xi_j(t_i) = F^{(j)} t_i [ & \xi_1(t_i), \xi_1(t_{i-1}) \dots; \xi_2(t_i), \xi_2(t_{i-1}) \dots; \\ & \xi_j(t_{i-1}), \xi_j(t_{i-2}) \dots; \xi_m(t_i), \xi_m(t_{i-1}) \dots; \alpha_1, \alpha_2, \dots, \alpha_k; \\ & \varepsilon_1 t_1 \dots \varepsilon_h t_i ] \quad (i = 1, 2, \dots, N; j = 1, 2, \dots, m; h > m) \end{aligned}$$

[18.10]

o dicho con palabras: Cada una de las variables “dependientes”

$$\xi_1(t_i) \dots \xi_m(t_i)$$

es una función de: 1), los valores previos de esta misma variable; 2), los valores simultáneos y previos de las otras  $n - 1$  variables. Las  $m$  funciones  $F^{(j)} t_i$  pueden ser diferentes para cada punto del tiempo, pero tiene *forma conocida*.

El sistema [18.10] comprende, además,  $k$  constantes desconocidas

$$\alpha_1, \alpha_2, \dots, \alpha_k$$

algunas de las cuales pueden faltar en alguna ecuación particular. Para cada punto del tiempo  $t_i$  el sistema comprende también  $h$  variables aleatorias  $\varepsilon$  que tienen ciertas propiedades conocidas en su distribución. Referiremos todas estas  $h$  variables  $\varepsilon$  al mismo punto del tiempo que la variable que se encuentra a la izquierda de [18.10], aunque los sucesos reales de los cuales *se deducen* pueden tener lugar en diferentes puntos del tiempo. Esto es sencillamente una transformación de variables en la ley de distribución de probabilidad conjunta de las  $\varepsilon$ . Si ocurriese que hubiese una relación funcional entre las  $\varepsilon$  en dos puntos del tiempo diferentes (por ejemplo,  $\varepsilon_s t_i = \varepsilon_r t_j$ ) la dimensión de la distribución conjunta de las  $h N \varepsilon$  podrían reducirse.

En el conjunto [18.10] de  $m N$  ecuaciones podremos (bajo cier-

tas condiciones de resolubilidad) expresar las  $mN$  variables aleatorias

$$\xi_1(t_i) \dots \xi_m(t_i) \quad (i = 1, 2, \dots, N)$$

como funciones de: 1), condiciones iniciales, es decir, todos (o algunos de) los valores de

$$\xi_1(t_i) \dots \xi_n(t_i) \quad \text{para } i = 0, -1, -2, \dots;$$

2), los valores de las variables autónomas

$$\xi_{m+1}(t_i) \dots \xi_n(t_i) \quad \text{para } i = 1, 2, \dots, N;$$

3), las  $Nh$  variables aleatorias

$$\varepsilon_1 t_i \dots \varepsilon_h t_i \quad (i = 1, 2, \dots, N).$$

Supondremos que las condiciones iniciales son *dadas* y constantes en muestras repetidas. Abreviadamente designaremos el conjunto total de condiciones iniciales por  $(\xi^0)$ .

Sea

$$\rho'_2 [\varepsilon_1 t_1, \dots, \varepsilon_h t_N \mid \xi_{m+1}(t_1) \dots \xi_n(t_N); (\xi^0); \beta_1 \dots \beta_r] \quad [18.11]$$

la probabilidad elemental conjunta de las  $hN$  variables  $\varepsilon$  para condiciones *iniciales dadas* de las  $\xi$  y valores dados de las  $\xi$  *autónomas* ( $\rho'_2$  puede o no puede depender de estas cantidades). Las  $\beta$  son parámetros que pueden o no pueden conocerse.

Puesto que las  $mN$  variables aleatorias

$$\xi_1(t_i) \dots \xi_m(t_i) \quad (i = 1, \dots, N)$$

pueden expresarse como funciones de las variables aleatorias  $\varepsilon$ . Podremos deducir la distribución conjunta de las  $mN$  variables aleatorias

$$\xi_1(t_i) \dots \xi_m(t_i) \quad (i = 1, \dots, N)$$

de la misma forma que se discutió en A. Esta distribución de probabilidad será:

$$\rho'_2 [\xi_2(t_1) \dots \xi_m(t_N) \mid \xi_{m+1}(t_1) \dots \xi_n(t_N); (\xi^0); \alpha_1 \dots \alpha_k; \beta_1 \dots \beta_r]. \quad [18.12]$$

Si las medidas de las  $\xi$  (pero no las de las condiciones iniciales) están sujetas a error, tendremos un problema adicional exactamente igual al que se discutió para los sistemas estáticos.

El problema de estimar los parámetros en un sistema dinámico de la forma [18.10] se reduce al problema de estimar los parámetros de una  $mN$  ( $-0nN$ ) ley elemental de probabilidad por medio de un punto muestra asociado con esta ley de probabilidad (15).

Esta forma de condensar las afirmaciones implicadas en un sistema de relaciones estocásticas puede extenderse a una clase más general de esquemas estocásticos, y este procedimiento es no sólo conveniente, sino yo creo que necesario, si queremos estar seguros de que los supuestos varios, hechos acerca de las propiedades de la distribución de las variables aleatorias que se trata, no dan lugar a contradicciones interiores análogas a las que mencionamos en la introducción de este capítulo.

---

Estamos ahora en condiciones de formular de forma precisa los dos problemas fundamentales de estimación en investigaciones económicas: 1), el problema de relaciones confluentes, y 2), el problema de las "mejores estimaciones".

I.—*El problema de relaciones confluentes* (el problema de parámetros arbitrarios).

Si dos sistemas de ecuaciones estocásticas dan la misma distribución conjunta de probabilidad de las variables aleatorias observables, son *indistinguibles* (tomando como base las observaciones). En particular, los sistemas son tales que difieren solamente en los valores de los parámetros (desconocidos). El problema de los coefi-

---

(15) Para las fórmulas de estimación y límites de confianza, etc. en el caso de un sistema de ecuaciones en diferencias estocásticas, Lineal, ver un artículo de H. B. MENN y A. WALD "On the Statistical Treatment of linear Stochastic Difference Equations". *Econometrica*, vol. II, julio-octubre 1943, páginas 173-220 en particular la parte II pág. 192-216.



cientes arbitrarios está, por lo tanto, incluido en el siguiente problema matemático general:

Sea

$$\rho(x_1, x_2, \dots, x_s | \theta_1 \dots \theta_K; z_1 \dots z_r)$$

una función de  $s$  variables independientes  $x_1 \dots x_s$  que comprende  $k$  parámetros desconocidos  $\theta$  y  $r$  parámetros conocidos  $z$ . Sea  $\theta^0_1 \dots \theta^0_K$  o abreviadamente,  $\theta^0$ , un punto en el espacio paramétrico  $K$  dimensional de las  $\theta$ . ¿Existirá o no existirá, por lo menos, un punto paramétrico  $\theta'$  ( $\neq \theta^0$ ) tal que

$$\begin{aligned} \rho(x_1 \dots x_s | \theta^0_1 \dots \theta^0_K; z_1 \dots z_r) &\equiv \\ \equiv \rho(x_1 \dots x_s | \theta'_1 \dots \theta'_K; z_1 \dots z_r) & \end{aligned} \quad [18.13]$$

para *todos* los valores de las variables  $x$ ? La respuesta a esta pregunta depende de una de las siguientes cosas: 1), la forma de la función  $\rho$ ; 2), el punto paramétrico  $\theta^0$ , y 3), los valores de los parámetros conocidos  $z$ .

Si [18.13] se cumple, y si  $\theta^0$  es el punto paramétrico “verdadero” entonces no importa cuántas observaciones tengamos de las variables  $x$ , no hay una estimación *única* para  $\theta^0$  a causa de que no podemos distinguir entre  $\theta^0$  y  $\theta'$  (el problema muy conocido de la “multicolinealidad”, éste, por otra parte, incluido en esta formulación como un caso muy especial del problema de los parámetros arbitrarios) (16).

## II.—El problema de la “mejor estimación”.

Sea

$$y = \rho(x_1 \dots x_s | \theta_1 \dots \theta_K; z_1 \dots z_r) \quad [18.14]$$

una familia paramétrica de leyes de probabilidad conjuntas de  $s$  variables aleatorias  $x_1 \dots x_s$  que comprende  $k$  parámetros desconocidos  $\theta_1, \theta_2, \dots, \theta_K$ . Si, para valores dados de los parámetros conoci-

(16) Ver la discusión de MANN y WALD del problema relacionado con las ecuaciones “reducidas” o las ecuaciones “originales” obra citada, pág. 200-202.

dos  $z$ , hay una correspondencia univoca entre los puntos paramétricos  $\theta$  y los miembros de la familia de  $k$  parámetros [18.14], y si  $\theta^0$  es el verdadero punto paramétrico, ¿cuál es la mejor estimación de  $\theta^0$  obtenida de un punto muestra  $(x_1 \dots x_s)$ ?

El problema II —en cuestión de principio— es un problema de estimación estadística, y no hay necesidad de una discusión separada del problema estadístico.

Lo mismo podría, por otra parte, decirse acerca del problema I. Este es un problema de matemáticas puro. Este problema, sin embargo, es de una significación particular en el campo de la econometría, e importante en la construcción de modelos económicos, y, además, este problema matemático particular, no parece que ha atraído el interés de las matemáticas; nosotros, por lo tanto, desarrollaremos algunos instrumentos de análisis con este objetivo particular.

19.—*Reducibilidad de una función respecto su número de parámetros*

Sea

$$y = f(x_1, x_2, \dots, x_s; \theta_1, \theta_2, \dots, \theta_k) \tag{19.1}$$

una función real de  $s$  variables independientes (no *relacionadas funcionalmente*)  $x_1 \dots x_s$  que comprende  $k$  parámetros  $\theta_1, \theta_2, \dots, \theta_k$  (por ejemplo,  $f$  puede ser la función [18.14] para valores fijos de las  $z$ ). Sea  $\theta^0$  un punto en el espacio paramétrico  $k$  dimensional de las  $\theta$ . Designemos por  $S(\theta^0)$  el conjunto de todos los puntos  $(y, x_1, \dots, x_s)$  en el espacio de las variables de  $(S + 1)$  dimensiones definidos por [19.1] cuando  $\theta = \theta^0$ . Por  $\theta'$  indicamos otro punto paramétrico  $\neq \theta^0$  y sea  $S(\theta')$  el correspondiente conjunto de puntos  $(y, x_1, x_2, \dots, x_s)$ . Si existe por lo menos un punto paramétrico  $\theta' \neq \theta^0$  tal que

$$S(\theta^0) = S(\theta') \tag{19.2}$$

o. —lo que es lo mismo— tal que

$$f(x_1 \dots x_s; \theta^0_1 \dots \theta^0_k) = f(x_1 \dots x_s; \theta'_1 \dots \theta'_k) \tag{19.3}$$

para *todos* los valores de las variables  $x$  diremos que el punto pa-

ramétrico  $\theta$  tiene (una cierta cantidad de) *arbitrariedad respecto al conjunto*  $S(\theta^0)$ .

Podemos distinguir entre los casos siguientes:

A) Existe un entorno finito del punto  $\theta^0$  tal que dentro de este entorno no hay un punto  $\theta' \neq \theta^0$  que satisfaga [19.3], mientras fuera, en la frontera de este entorno, hay uno o más puntos  $\theta'$  que satisfacen [19.3].

Ejemplo 1:

$$y = \theta^2_1 x_1.$$

en donde si  $\theta_1 = \theta^0_1 > 0$ , no existe un punto  $\theta'_1 \neq \theta^0_1$  en el recorrido  $\theta_1 > -\theta^0_1$  que satisfaga [19.3], mientras en el recorrido  $\theta_1 \leq -\theta^0_1$  hay exactamente un punto  $\theta'_1$  que satisface [19.3],  $\theta'_1 = -\theta^0_1$ .

Ejemplo 2:

$$y = \theta_1 \text{ sen } (\theta_2 + \theta_3 x_1)$$

en donde  $\theta^0$  es un punto paramétrico; aquí no hay puntos paramétricos en la vecindad inmediata de  $\theta^0$  satisfaciendo [19.3], pero hay una infinidad de puntos paramétricos aislados  $\theta'$  satisfaciendo [19.3] que son

$$\theta'_1 = \theta^0_1; \theta'_2 = \theta^0_2 + 2\pi n; \theta'_3 = \theta^0_3; \quad n = 1, 2, 3, \dots$$

Ejemplo 3:

$$y = \nu(\theta_1) x_1; \nu(\theta_1) = (|\theta_1| + \theta_1 + 1) - 2(|\theta_1 - 1| + (\theta_1 - 1)).$$

Supongamos que  $\theta^0_1 = 2$ , entonces  $\nu(\theta^0_1) = 1$ . Ahora, si  $\theta_1 > 2$  entonces  $\nu(\theta_1) < 1$  y ningún punto  $\theta'_1 > 2$  satisfará [19.3]. Si  $2 > \theta_1 > 0$ , entonces  $\nu(\theta_1) > 1$ ; por lo tanto, ningún punto  $\theta'$  tal que  $2 > \theta'_1 > 0$  satisface [19.3]. Pero si  $\theta_1 \leq 0$  entonces  $\nu(\theta_1) \equiv 1$ , de aquí todos los puntos  $\theta'_1 \leq 0$  satisfarán [19.3].

B). Si se elige un entorno de  $\theta^0$ , no importa lo pequeño que sea, donde hay puntos  $\theta' \neq \theta^0$  que satisfacen [19.3].

Ejemplo:

$$y = (\theta_1 + \theta_2) x_1 + \theta_3 x_2$$

Vamos ahora a deducir ciertas condiciones generales bajo las cuales (A) o (B) ocurrirán.

Con este propósito consideremos la función  $f$  en [19.1] como una función de  $S + K$  variables independientes

$$x_1, x_2, \dots, x_s, \theta_1 \dots \theta_K$$

Supondremos durante el resto de esta sección que:

1)  $f$  está definida sobre un cierto dominio  $D$  del espacio  $x$ , ese dimensional, y sobre una cierta región simplemente conexa  $D_\theta$ , del espacio paramétrico  $K$  dimensional, y es una función uniforme para todo punto  $x \in D_x$  y para todo punto  $\theta \in D_\theta$ .

2) Para todo punto  $x \in D_x$  y para todo punto interior  $\theta$  de  $D_\theta$   $f$  tiene derivadas parciales de primer orden

$$\frac{\partial f}{\partial \theta_i} \quad (i = 1, \dots, K)$$

continuas (es decir, continuas en las  $\theta$ ).

*Definición:* La función

$$f(x_1, x_2, \dots, x_s; \theta_1 \dots \theta_K)$$

se dice que es  $v$  — veces reducible ( $K \geq v > 0$ ) en el punto paramétrico  $\theta^0$ , siendo  $\theta^0$  un punto interior de  $D_\theta$  si existen  $K - v$  funciones

$$u_1(\theta_1, \theta_2, \dots, \theta_K), u_2(\theta_1 \dots \theta_K) \dots u_{K-v}(\theta_1 \dots \theta_K)$$

no dependientes del punto  $x$  y una función

$$\bar{f}(x_1 \dots x_s; u_1 \dots u_{K-v})$$

que tiene las siguientes propiedades:

a)  $f(x_1 \dots x_s; \theta_1 \dots \theta_K) \equiv \bar{f}(x_1 \dots x_s; u_1 \dots u_{K-v})$  para todo punto  $x \in D_x$  y para todo punto  $\theta$  de un entorno arbitrariamente pequeño de  $\theta^0$ .

b)  $\partial u_i / \partial \theta_j$  ( $i = 1, 2, \dots, K - v; j = 1, 2, \dots, K$ ) existen y son

continuas para todo punto  $\theta$  dentro de un entorno arbitrariamente pequeño finito de  $\theta^0$ .

c) La matriz jacobiana

$$\frac{\delta(u_1 \dots u_{K-v})}{\delta(\theta_1 \dots \theta_K)}$$

es de rango  $K-v$  en  $\theta = \theta^0$ .

Si una función  $f$  tiene estas propiedades en un punto paramétrico  $\theta^0$ , entonces, evidentemente, existen infinitos puntos  $\theta^0$  en el entorno de  $\theta^0$ , tales que [19.3] se satisface para  $u_1 \dots u_{K-v}$  fijos;  $v$  parámetros  $\theta$  pueden elegirse arbitrariamente en un cierto entorno de  $\theta^0$  sin cambiar los valores de  $f$ , cualquiera que sea  $x \in D_x$ .

Teorema 1.—Si una función

$$f(x_1, x_2, \dots, x_n; \theta_1 \dots \theta_K)$$

es  $v$  — veces reducible en el punto paramétrico  $\theta^0$ , existe un sistema de funciones

$$\lambda_{ij}(\theta_1 \dots \theta_K) \quad (i = 1, 2, \dots, K; j = 1, 2, \dots, v)$$

que son independientes del punto  $x$  y continuas en el entorno de  $\theta$ , y tales que

$$\left\| \begin{array}{cccc} \lambda_{11} & \lambda_{21} & \dots & \lambda_{K1} \\ \lambda_{21} & \lambda_{22} & \dots & \lambda_{K2} \\ \dots & \dots & \dots & \dots \\ \lambda_{1v} & \lambda_{2v} & \dots & \lambda_{Kv} \end{array} \right\| \quad [19.4]$$

es de rango  $v$  para  $\theta = \theta^0$ , y

$$\lambda_{1j} \frac{\delta f}{\delta \theta_1} + \lambda_{2j} \frac{\delta f}{\delta \theta_2} + \dots + \lambda_{Kj} \frac{\delta f}{\delta \theta_K} \equiv 0 \quad (j = 1, 2, \dots, v) \quad [19.5]$$

para todos los puntos  $x \in D_x$  y para todos los puntos  $\theta$  de un entorno arbitrariamente pequeño pero finito de  $\theta^0$ .

Demostración. Como la matriz jacobiana

$$\frac{\delta(u_1 \dots u_{K-v})}{\delta(\theta_1 \dots \theta_K)}$$

es de rango  $K - v$  en  $\theta = \theta^0$ , contiene por lo menos un determinante  $(K - v)$  que no es cero en  $\theta = \theta^0$ . Como el número de las  $\theta$  es arbitrario, podremos, sin falta de generalidad, suponer que

$$\frac{\partial(u_1 \dots u_{K-v})}{\partial(\theta_1 \dots \theta_{K-v})}$$

es de rango  $K - v$  para  $\theta = \theta^0$ , entonces el sistema

$$\begin{aligned} u_1 & (\theta_1, \theta_2, \dots, \theta_K) = u_1 \\ u_2 & (\theta_1, \theta_2, \dots, \theta_K) = u_2 \\ & \dots \dots \dots \dots \dots \dots \dots \\ u_{K-v} & (\theta_1, \theta_2, \dots, \theta_K) = u_{K-v} \end{aligned} \tag{19.6}$$

tiene como solución

$$\theta_i = \psi_i(u_1, u_2, \dots, u_{K-v}; \theta_{K-v+1} \dots \theta_s) \quad (i = 1 \dots K - v) \tag{19.7}$$

que es única en un cierto entorno de  $\theta = \theta^0$  y tal que

$$\frac{\partial \psi_i}{\partial u_j} \quad (i = 1, 2, \dots, K - v; j = 1, 2, \dots, K - v)$$

y

$$\frac{\partial \psi_i}{\partial \theta_k} \quad (i = 1, \dots, K - v; k = K - v + 1, \dots, K)$$

existe y son funciones continuas de

$$u_1, u_2, \dots, u_{K-v}; \theta_{K-v+1} \dots \theta_K$$

en este entorno (esto se deduce de la teoría clásica de los determinantes funcionales).

De aquí tendremos

$$\begin{aligned} & f(x_1, x_2, \dots, x_n; \theta_1 \dots \theta_K) \equiv \\ \equiv & f(x_1 \dots x_n; \psi_1 \dots \psi_{K-v}, \theta_{K-v+1} \dots \theta_K) \equiv \\ \equiv & \bar{f}(x_1, x_2, \dots, x_n; u_1, u_2, \dots, u_{K-v}) . \end{aligned} \tag{19.8}$$

Por definición,  $f$  tiene derivadas continuas parciales

$$\frac{\partial f}{\partial \theta_i} (i = 1, 2, \dots, K);$$

por lo tanto,

$$\frac{\partial f}{\partial \psi_i} (i = 1, 2, \dots, K - v)$$

existe y son continuas en el entorno de  $\theta^0$ . Como también

$$\frac{\partial \psi_i}{\partial u_j} (i = 1, 2, \dots, K - v; j = 1, 2, \dots, K - v)$$

existen y son continuas, como se demostró anteriormente,  $f$  tiene derivadas parciales continuas

$$\frac{\partial f}{\partial u_j} (j = 1, 2, \dots, K - v)$$

en un cierto entorno finito de  $\theta^0$ . Pero cuando  $f \equiv \bar{f}$ , también

$$\frac{\partial \bar{f}}{\partial u_j} (j = 1, 2, \dots, K - v)$$

existen y son continuas en el mismo entorno. Tendremos, por lo tanto,

$$\frac{\partial f}{\partial \theta_1} \equiv \frac{\partial \bar{f}}{\partial u_1} \cdot \frac{\partial u_1}{\partial \theta_1} + \dots + \frac{\partial \bar{f}}{\partial u_{K-v}} \cdot \frac{\partial u_{K-v}}{\partial \theta_1}$$


---


$$\frac{\partial f}{\partial \theta_K} \equiv \frac{\partial \bar{f}}{\partial u_1} \cdot \frac{\partial u_1}{\partial \theta_K} + \dots + \frac{\partial \bar{f}}{\partial u_{K-v}} \cdot \frac{\partial u_{K-v}}{\partial \theta_K}$$

[19.9]

[19.9] puede considerarse como una transformación lineal *singular* de las  $K - v$  variables

$$-\frac{\partial \bar{f}}{\partial u_j} \quad (j = 1, 2, \dots, K - v)$$

en las  $K$  variables

$$\frac{\partial f}{\partial \theta_i} \quad (i = 1, 2, \dots, K)$$

la matriz de la transformación será —por definición— de rango  $K - v$  para  $\theta = \theta^0$ , y tendrá elementos continuos en la vecindad de  $\theta^0$ . Pero entonces el teorema 1 se deduce inmediatamente de la teoría de la dependencia lineal.

El recíproco del teorema 1 puede demostrarse, bajo ciertas restricciones adicionales sobre las  $\lambda$ .

Vamos ahora a demostrar un teorema que dé una condición suficiente para la *no* existencia de una relación del tipo [19.3] en la vecindad de un punto paramétrico  $\theta^0$ .

**Teorema 2.** Si las funciones

$$\frac{\partial f}{\partial \theta_1} \dots \dots \dots \frac{\partial f}{\partial \theta_K}$$

son linealmente independientes en el punto  $\theta^0$ , es decir, en el punto paramétrico  $\theta^0$

$$\sum_{i=1}^K \lambda^0_i \frac{\partial f}{\partial \theta_i} \neq 0 \tag{19.10}$$

cualquiera que sea el sistema de constantes  $\lambda^0_1 \dots \lambda^0_K$ , no todas iguales a cero, existe un entorno finito del punto paramétrico  $\theta^0$  tal que, en este entorno, no hay puntos paramétricos  $\theta \neq \theta^0$  para los cuales [19.3] se satisfaga.

**Demostración.** Primeramente es fácil ver que la independencia lineal de la función

$$\frac{\partial f}{\partial \theta_i} \quad \text{en } \theta = \theta^0$$



lleva consigo que el conjunto  $S(\theta^0)$  definido anteriormente contenga por lo menos  $K$  puntos diferentes

$$(y^{(j)}, x^{(j)}_1, \dots, x^{(j)}_s) \quad (j = 1, 2, \dots, K)$$

tales que si

$$y^{(j)} = f^{(j)} \quad (j = 1, 2, \dots, K) \tag{19.11}$$

es el sistema de ecuaciones obtenidas insertando sucesivamente estos  $K$  puntos en [19.1]. El jacobismo

$$\begin{vmatrix} \frac{\partial f^{(1)}}{\partial \theta_1} & \frac{\partial f^{(1)}}{\partial \theta_2} & \dots & \frac{\partial f^{(1)}}{\partial \theta_K} \\ \frac{\partial f^{(2)}}{\partial \theta_1} & \frac{\partial f^{(2)}}{\partial \theta_2} & \dots & \frac{\partial f^{(2)}}{\partial \theta_K} \\ \dots & \dots & \dots & \dots \\ \frac{\partial f^{(K)}}{\partial \theta_1} & \frac{\partial f^{(K)}}{\partial \theta_2} & \dots & \frac{\partial f^{(K)}}{\partial \theta_K} \end{vmatrix} \neq 0. \quad \text{Para } \theta = \theta^0 \tag{19.12}$$

Puesto que también, por definición,  $\delta f / \delta \theta$  son funciones continuas del parámetro  $\theta$  y las  $x$ , en [19.11], son constantes, se deduce de la teoría de los determinantes funcionales que el sistema [19.11] puede resolverse para  $\theta_1, \theta_2, \dots, \theta_K$ , y la solución es única y, por lo tanto, igual a  $\theta^0_1, \theta^0_2, \dots, \theta^0_K$ . Dentro de un entorno finito del punto paramétrico  $\theta^0$  no hay otros puntos paramétricos  $\neq \theta^0$  satisfaciendo [19.11]. Esto prueba el teorema 2.

De aquí se deduce inmediatamente.

Teorema 3. Si

$$\frac{\delta f}{\delta \theta_1} \quad \dots \quad \frac{\delta f}{\delta \theta_k}$$

son linealmente independientes para todo punto  $\theta^0$  en el interior de  $D_\theta$ , entonces [19.3] es, a lo más, satisfecho para puntos paramétricos entre los cuales hay una distancia finita mayor que un cierto  $\varepsilon$  positivo.

Así, en los casos más prácticos, la cuestión de coeficientes arbitrarios puede responderse investigando cuándo o no las derivadas parciales de  $f$  respecto las  $\theta$  son linealmente dependientes. Pero para esto necesitamos, por lo menos en los casos más complicados, una regla que decida, en un número finito de pasos, cuándo tales dependencias lineales existen o no. En la próxima sección daremos tal regla.

20.—*El criterio de Gram para dependencias de funciones lineales. Extendido a funciones de varias variables*

Sea

$$\begin{aligned} U_1 &= U_1(v_1, v_2, \dots, v_n) \\ U_2 &= U_2(v_1, v_2, \dots, v_n) \\ &\dots\dots\dots \\ U_m &= U_m(v_1, v_2, \dots, v_n) \end{aligned} \tag{20.1}$$

$m$  funciones reales de  $n > m$  variables reales independientes  $v_1, v_2, \dots, v_n$ .

Supuesto:  $U_1, U_2, \dots, U_m$  son funciones continuas de las  $n$  variables  $v$  sobre un cierto dominio cerrado  $W$  en el espacio  $v$  definido por

$$a_i \leq v_i \leq \bar{a}_i \quad (i = 1, 2, \dots, n) \tag{20.2}$$

en donde

$$a_i, \bar{a}_i \quad (i = 1, 2, \dots, n)$$

son  $2n$  números reales.

Consideremos la expresión

$$s = c_1 U_1 + c_2 U_2 + \dots + c_m U_m \tag{20.3}$$

en donde las  $c$  no dependen de las  $v$  ni de las  $a$  de [20.2]. Si se puede encontrar un sistema de  $c$  no todas cero, tal que

$$S \equiv 0 \tag{20.4}$$

para *todos* los valores de  $v_1, v_2, \dots, v_n$  en el dominio definido por

[20.2], las  $m$  funciones [20.1] se dice que son linealmente dependientes en  $W$ .

Consideremos la integral

$$S = \int \int \dots \int_{(w)} s^2 dv_1 dv_2 \dots dv_n \quad [20.5]$$

tendremos

$$S \cong 0 \quad [20.6]$$

$S$  es cero cuando y solamente cuando [20.4] es cierto. Por lo tanto, si un conjunto de  $c$  no todas cero existe para el cual [20.4] se satisface, debe —al mismo tiempo— ser el conjunto de las  $c$  tal que haga  $S$  *mínimo e igual a cero*. Y recíprocamente, si hay un conjunto de  $c$  no todas cero tales que  $S = 0$ , entonces [20.4] se cumple. El problema de dependencia lineal se reduce, por lo tanto, a estudiar el mínimo de  $S$  respecto a  $c$ .

Puesto que no todas las  $c$  serán cero, podremos suponer que

$$\sum_{i=1}^m c^2_i = 1. \quad [20.7]$$

Tendremos que encontrar el mínimo de  $S$  bajo la condición [20.7] o, lo que es lo mismo, encontrar el mínimo, sin restricciones de

$$S' = S - \lambda \sum_{i=1}^m c^2_i \quad [20.8]$$

en donde  $\lambda$  es un multiplicador de Lagrange. Introduzcamos la siguiente notación

$$M_{ij} = \int \int \dots \int_{(w)} u_i u_j (dv_1 \dots dv_n) \quad [20.9]$$

$$(i = 1, 2, \dots, m; j = 1, 2, \dots, m)$$

un sistema de  $c$  que hagan mínimo [20.8] deben satisfacer el siguiente sistema de ecuaciones lineales

$$\begin{aligned} (M_{11} - \lambda) c_1 + M_{12} c_2 + \dots + M_{1m} c_m &= 0 \\ \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots & \\ M_{m1} c_1 + M_{m2} c_2 + \dots + (M_{mm} - \lambda) c_m &= 0 \end{aligned} \quad [20.10]$$

[20.10] tiene una solución de  $c$  no todas cero cuando, y solamente cuando, el determinante formado por los coeficientes de las  $c$  es igual a cero, es decir

$$\begin{vmatrix} (M_{11} - \lambda) & M_{12} & \dots & M_{1m} \\ M_{21} & (M_{22} - \lambda) & \dots & M_{2m} \\ \dots & \dots & \dots & \dots \\ M_{m1} & M_{m2} & \dots & (M_{mm} - \lambda) \end{vmatrix} = 0 \quad [20.11]$$

Ahora  $S$  es una forma cuadrática definida (semi) positiva. Por lo tanto todas las raíces de [20.11] son no negativas.  $S$  tiene, por lo tanto, un mínimo  $= 0$  para otros valores de las  $c$  distintos de cero solamente si la raíz mínima  $\lambda$  en [20.11] es igual a cero. Una condición necesaria y suficiente para la dependencia lineal de las  $m$  funciones [20.1] es, por lo tanto, que

$$|M_{ij}| = 0 \quad [20.12]$$

en donde  $|M_{ij}|$  es el determinante [20.11] para  $\lambda = 0$ .

### 21.—Un ejemplo del problema de estimación

Consideremos un esquema sencillo de oferta-demanda, que incluye ciertos elementos aleatorios, y un impuesto de venta autónomo.

Indiquemos por  $\xi^{(D)}_{1t}$  la cantidad demandada en el punto del tiempo  $t$ ,  $\xi^{(S)}_{1t}$  la cantidad ofrecida,  $\xi_{2t}$  el precio por unidad vendida y  $\xi_{3t}$  un impuesto de venta por unidad. Consideremos estas variables en  $N$  puntos equidistantes del tiempo  $t = 1, 2, \dots, N$ . Supondremos que sabemos que estas variables satisfacen el sistema de ecuaciones aleatorias

$$\xi^{(D)}_{1t} = \alpha_1 \xi_{2t} + \varepsilon_{1t} \quad (t = 1, 2, \dots, N) \quad [21.1]$$

es una curva lineal de demanda con desviaciones aleatorias  $\varepsilon_{1t}$ ;

$$\xi^{(S)}_{1t} = \alpha_2 (\xi_{2t} - \xi_{3t}) + \varepsilon_{2t} \quad (t = 1, 2, \dots, N) \quad [21.2]$$

$$\xi^{(D)}_{1t} = \xi^{(S)}_{1t} = \xi_{1t} = \text{cantidad vendida en } t. \quad [21.3]$$

$$\varepsilon_{11}, \varepsilon_{12}, \dots, \varepsilon_{1N}; \varepsilon_{21} \dots \varepsilon_{2N}$$

$$\varepsilon_{1t}, t = 1, 2, \dots, N$$

$$\varepsilon_{2t}, t = 1, 2, \dots, N$$

observamos

$$x_{1t} = \xi_{1t} + \eta_{1t} \quad (t = 1, 2, \dots, N) \quad [21.4]$$

mientras el precio  $\xi_{2t}$  y el impuesto  $\xi_{3t}$  se observa sin error, es decir:

$$x_{2t} = \xi_{2t} \quad x_{3t} = \xi_{3t} \quad (t = 1, 2, \dots, N) \quad [21.5]$$

Supondremos que las  $N$  variables aleatorias  $\eta_{11}, \eta_{12}, \dots, \eta_{1N}$  se distribuyen independientemente y normalmente con media cero y la misma varianza  $\sigma^2$  y su distribución no depende de las  $\xi$  ni de las  $\varepsilon$ .

Los  $N$  números  $\xi_{31}, \xi_{32}, \dots, \xi_{3N}$  se suponen permanecen fijos en muestras repetidas.

A causa de [21.3],  $\xi_{1t}$  y  $\xi_{2t}$  serán variables aleatorias. De [21.3], [21.1] y [21.2] obtenemos (siempre que  $\alpha_1 \neq \alpha_2$ )

$$\xi_{1t} = \frac{\alpha_1 \alpha_2}{\alpha_2 - \alpha_1} \xi_{3t} + \frac{\alpha_2 \varepsilon_{1t} - \alpha_1 \varepsilon_{2t}}{\alpha_2 - \alpha_1} \quad [21.6]$$

$$\xi_{2t} = \frac{\alpha_2}{\alpha_2 - \alpha_1} \xi_{3t} + \frac{\varepsilon_{1t} - \varepsilon_{2t}}{\alpha_2 - \alpha_1}$$

lo cual demuestra que  $\xi_{1t}$  y  $\xi_{2t}$  son funciones de dos variables aleatorias independientes.

Si pudiéramos hacer experimentos y estudiar separadamente la función de demanda [21.1] y la de oferta [21.2] podríamos razonar de esta forma: 1), para un valor dado de  $\xi_{2t}$ ,  $\xi^{(D)}_{1t}$  es una variable aleatoria con valor esperado igual a  $\alpha_1 \xi_{2t} + \bar{\varepsilon}_1$ ; 2), para valores dados de  $\xi_{2t}$  y  $\xi_{3t}$ ,  $\xi^{(S)}_{1t}$  es una variable aleatoria con valor esperado igual a

$$\alpha_2 (\xi_{2t} - \xi_{3t}) + \bar{\varepsilon}_2$$

Y podríamos ajustar "cada una de las ecuaciones [21.1] y [21.2] separadamente" a los datos respectivos obtenidos por las dos series

de experimentos. Pero en nuestro caso, a causa de la relación del mercado [21.3] no podemos suponer que  $\xi_{2t}$  permanece fija en muestras repetidas. Esto sería sencillamente inconsistente con el supuesto primitivo de que los errores  $\varepsilon_1$  y  $\varepsilon_2$  eran independientes.

Para comprobar claramente todas las implicaciones de nuestro esquema hemos de considerar la *distribución de probabilidad conjunta* de las variables observadas

$$x_{1t} \text{ y } x_{2t} \text{ dadas } x_{3t}, t = 1, \dots, N.$$

Introduciendo [21.4] y [21.5] en [21.1] y [21.2], obtenemos:

$$x_{1t} = \alpha_1 x_{2t} + \varepsilon_{1t} + \eta_{1t} \quad [21.1']$$

$$x_{1t} = \alpha_2 (x_{2t} - x_{3t}) + \varepsilon_{2t} + \eta_{2t} \quad [21.2']$$

o

$$x_{1t} = \frac{\alpha_1 \alpha_2}{\alpha_2 - \alpha_1} x_{3t} + \frac{\alpha_2 \varepsilon_{1t} - \alpha_1 \varepsilon_{2t}}{\alpha_2 - \alpha_1} + \eta_{1t} \quad [21.6']$$

$$x_{2t} = \frac{\alpha_2}{\alpha_2 - \alpha_1} x_{3t} + \frac{\varepsilon_{1t} - \varepsilon_{2t}}{\alpha_2 - \alpha_1}$$

$x_{1t}$  y  $x_{2t}$  se distribuyen conjuntamente normales, porque son funciones lineales de las variables distribuidas normalmente  $\varepsilon_{1t}$ ,  $\varepsilon_{2t}$ ,  $\eta_{1t}$ . Tendremos para cualquier punto fijo  $t$

$$\rho_t(x_{1t}, x_{2t} | x_{3t}) = \frac{1}{2\pi \mu_1 \mu_2 \sqrt{1 - \rho^2}} \exp \left\{ -\frac{1}{2(1 - \rho^2)} \left[ \frac{(x_{1t} - \bar{x}_{1t})^2}{\mu_1^2} + \frac{2\rho(x_{1t} - \bar{x}_{1t})(x_{2t} - \bar{x}_{2t})}{\mu_1 \mu_2} + \frac{(x_{2t} - \bar{x}_{2t})^2}{\mu_2^2} \right] \right\} \quad [21.7]$$

en donde  $\mu_1^2$ ,  $\mu_2^2$  son las varianzas de  $x_{1t}$  y  $x_{2t}$ , respectivamente;  $\bar{x}_{1t}$ ,  $\bar{x}_{2t}$  sus valores medios,  $\rho$  su coeficiente de correlación;

todos estos valores y los de  $x_{3t}$  se obtienen de [21.6']. Obtenemos:

$$\bar{x}_{1t} = \frac{\alpha_1 \alpha_2}{\alpha_2 - \alpha_1} x_{3t} + \frac{\alpha_2 \bar{\varepsilon}_1 - \alpha_1 \bar{\varepsilon}_3}{\alpha_2 - \alpha_1} \quad [21.8]$$

$$\bar{x}_{2t} = \frac{\alpha_2}{\alpha_3 - \alpha_1} x_{3t} + \frac{\bar{\varepsilon}_1 - \bar{\varepsilon}_3}{\alpha_2 - \alpha_1} \quad [21.9]$$

$$\mu^2_1 = E(x_{1t} - \bar{x}_{1t})^2 = \frac{1}{(\alpha_2 - \alpha_1)^2} (\alpha_2^2 \sigma^2_1 + \alpha_1^2 \sigma^2_2) + \sigma^2 \quad [21.10]$$

$$\mu^2_2 = E(x_{2t} - \bar{x}_{2t})^2 = \frac{1}{(\alpha_2 - \alpha_1)^2} (\sigma^2_1 + \sigma^2_2) \quad [21.11]$$

$$\rho = \frac{1}{\mu_1 \mu_2} E(x_{1t} - \bar{x}_{1t})(x_{2t} - \bar{x}_{2t}) = \frac{\alpha_2 \sigma^2_1 + \alpha_1 \sigma^2_2}{(\alpha_2 - \alpha_1)^2 \mu_1 \mu_2} \quad [21.12]$$

Esto demuestra que solamente las medias  $\bar{x}_{1t}$  y  $\bar{x}_{2t}$  dependen de  $t$ , mientras los otros parámetros son independientes de  $t$ .

Puesto que las variables aleatorias  $x_{1t}$ ,  $x_{2t}$  para un valor de  $t$  se distribuyen independientemente de cualquier otro valor de  $t$ , la distribución conjunta de las  $2N$  variables  $x_{11} \dots x_{2N}$  es

$$\rho(x_{11}, \dots, x_{2N}) = \pi \rho_t (t = 1, 2, \dots, N) \quad [21.13]$$

Introduzcamos nuevos parámetros por las transformaciones

$$Q_1 = \frac{\alpha_1 \alpha_2}{\alpha_1 \alpha_2} \quad [21.14]$$

$$Q_2 = \frac{\alpha_2 \bar{\varepsilon}_1 - \alpha_1 \bar{\varepsilon}_2}{\alpha_2 - \alpha_1} \quad [21.15]$$

$$b_1 = \frac{\alpha_2}{\alpha_2 - \alpha_1} \quad [21.16]$$

$$b_2 = \frac{\bar{\varepsilon}_1 - \bar{\varepsilon}_2}{x_2 - x_1} \quad [21.17]$$

Entonces [21.8] y [21.9] se convierten

$$\bar{x}_{1t} = a_1 x_{3t} + a_2 \quad [21'8]$$

$$\bar{x}_{2t} = b_1 x_{3t} + b_2 \quad [21'9]$$

Introduciendo [21.7] y [21.13] y empleando [21.8'] y [21.9'], obtenemos:

$$\rho = C e^Q, \quad [21.18]$$

donde

$$Q = \frac{1}{2(1-\rho^2)} \sum \left( \frac{(x_{1t} - a_1 x_{3t} - a_2)^2}{\mu_1^2} + \frac{2\rho(x_{1t} - a_1 x_{2t} - a_2)(x_{2t} - b_1 x_{3t} - b_2)}{\mu_1 \mu_2} + \frac{(x_{2t} - b_1 x_{3t} - b_2)^2}{\mu_2^2} \right) \quad [21.19]$$

y

$$C = \frac{1}{(2\pi \mu_1 \mu_2)^N (1-\rho^2)^{N/2}} \quad [21.20]$$

La distribución [21.18] está, por lo tanto, caracterizada por siete parámetros desconocidos:

$$a_1, a_2, b_1, b_2, \mu_1, \mu_2, \rho.$$

Si existe una estimación única de estos siete parámetros, el problema de unicidad de los parámetros originales depende solamente de las transformaciones [21.10] — [21.12] y [21.14] — [21.17]. Es fácil ver que estas transformaciones establecen una corresponden-



cia biunívoca entre los nuevos y viejos parámetros sobre el espacio total paramétrico, excepto para un conjunto de medida cero

$$(\alpha_1 = 0 \quad 0 \alpha_2 = 0 \quad 0 \alpha_3 = \alpha_2).$$

Tendremos, por lo tanto, que investigar la unicidad de los parámetros en [21.18]. Esto puede hacerse por medio de los teoremas de la sección (19).

Los derivados parciales de  $\rho$  (en [21.18]) respecto a los parámetros son:

$$\begin{aligned} \frac{\partial \rho}{\partial a_1} &= C e^Q \frac{\partial Q}{\partial a_1} & \frac{\partial \rho}{\partial a_2} &= C e^Q \frac{\partial Q}{\partial a_2} \\ \frac{\partial \rho}{\partial b_1} &= C e^Q \frac{\partial Q}{\partial b_1} & \frac{\partial \rho}{\partial b_2} &= C e^Q \frac{\partial Q}{\partial b_2} \\ \frac{\partial \rho}{\partial \mu_1} &= e^Q \left( \frac{\partial e}{\partial \mu_1} + e \frac{\partial Q}{\partial \mu_1} \right) & \frac{\partial \rho}{\partial \mu_2} &= e^Q \left( \frac{\partial e}{\partial \mu_2} + e \frac{\partial Q}{\partial \mu_2} \right) \\ & & \frac{\partial \rho}{\partial \rho} &= e^Q \left( \frac{\partial e}{\partial \rho} + e \frac{\partial Q}{\partial e} \right) \end{aligned} \quad [21.21]$$

De acuerdo con la sección 19, tenemos interés en saber cuándo estas 7 derivadas parciales son linealmente independientes. Si este fuera el caso, existirían  $7\lambda_1 \lambda_2 \dots \lambda_7$  independientes de las variables  $x$ , no todas cero, y tales que

$$\begin{aligned} & \lambda_1 c \frac{\partial Q}{\partial a_1} + \lambda_2 c \frac{\partial Q}{\partial a_2} + \lambda_3 c \frac{\partial Q}{\partial b_1} + \\ & + \lambda_4 c \frac{\partial Q}{\partial b_2} + \lambda_5 \left( \frac{\partial c}{\partial \mu_1} + c \frac{\partial Q}{\partial \mu_1} \right) + \\ & + \lambda_6 \left( \frac{\partial c}{\partial \mu_2} + c \frac{\partial Q}{\partial \mu_2} \right) + \lambda_7 \left( \frac{\partial c}{\partial \rho} + c \frac{\partial Q}{\partial \rho} \right) \equiv 0 \end{aligned} \quad [21.22]$$

para todos los valores de las variables  $x$ . Como no conocemos los valores verdaderos de los parámetros, estamos interesados en conocer cuándo hay un punto paramétrico cualquiera para el cual [21.22] se cumpla.

De [21.19] y [21.20] vemos que el lado izquierdo de [21.22] será un polinomio de segundo grado en las variables  $x$ , [21.22] puede cumplirse solamente si los coeficientes de términos iguales en el polinomio desaparecen separadamente. Empleando esto comprobamos fácilmente que las 7  $\lambda$  serán *iguales a cero*, cualquiera que sea el verdadero punto paramétrico (excepto para  $\rho = \pm 1$ , que es trivial) siempre que entre el conjunto de  $N$  constantes

$$x_{31}, x_{32}, \dots, x_{3N}$$

haya dos por lo menos que sean diferentes.

Los 7 parámetros de [21.18] pueden, por lo tanto, en general, estimarse.

Consideraremos, en particular, la estimación máxima verosimilitud (17) de los parámetros en [21.18], es decir, los valores de los parámetros obtenidos igualando cada una de las 7 derivadas de [21.21] a cero, y resolviendo este sistema de 7 ecuaciones. Obtengamos las siguientes ecuaciones, que definen la estimación máxima verosimilitud (que indicaremos por  $d_1, d_2$ , etc.):

$$\Sigma (x_{1t} - \hat{a}_1 x_{3t} - \hat{a}_2) = 0 \tag{21.23}$$

$$\Sigma (x_{2t} - \hat{b}_1 x_{3t} - \hat{b}_2) = 0 \tag{21.24}$$

$$\Sigma (x_{1t} - \hat{a}_1 x_{3t} - \hat{a}_2) x_{3t} = 0 \tag{21.25}$$

---

(17) El método de la máxima verosimilitud, empleado comunmente por estadísticos, originariamente estuvo fundamentado sobre la intuición, pero recientemente, se ha demostrada por A. WALD que el método, bajo ciertas condiciones, puede justificarse sobre la base de la teoria moderna de los intervalos de confianza. Ver sus artículos "A New Foundation of the Method of Maximun Likelihood in Statistical Theory" Cowles Commission for Ressearch in Economics, *Report of Sixth Annual Research Conference on Economics and Statistics...* 1940, pág. 33-35 y "Asymtotically Most Powerful Test of Statistical Hypotheses". *Annals of Mathecatical Statistics*. Vol. 12 marzo 1941, páginas 1-19.

$$\Sigma (x_{2t} - \hat{b}_1 x_{3t} - \hat{b}_2) x_{3t} = 0 \quad [21.26]$$

$$N \hat{\mu}_1^2 - \Sigma (x_{1t} - \hat{a}_1 x_{3t} - \hat{a}_2)^2 = 0 \quad [21.27]$$

$$N \hat{\mu}_2^2 - \Sigma (x_{2t} - \hat{b}_1 x_{3t} - \hat{b}_2)^2 = 0 \quad [21.28]$$

$$N \rho = \frac{\Sigma (x_{1t} - \hat{a}_1 x_{3t} - \hat{a}_2) (x_{2t} - \hat{b}_1 x_{3t} - \hat{b}_2)}{\hat{\mu}_1 \hat{\mu}_2} = 0 \quad [21.29]$$

Es fácil comprobar que este sistema tiene, en general, una solución única respecto los siete parámetros  $\hat{a}_1, \hat{a}_2, \dots$ , etc. Por ejemplo, de las primeras cuatro ecuaciones obtenemos:

$$\hat{a}_1 = \frac{m_{13} - m_1 m_3}{m_{33} - m_3^2} \quad [21.30]$$

$$\hat{b}_1 = \frac{m_{23} - m_2 m_3}{m_{33} - m_3^2} \quad [21.31]$$

en donde

$$m_{ij} = \frac{1}{N} \Sigma x_{it} x_{jt} \quad m_i = \frac{1}{N} \Sigma x_{it} \quad [21.32]$$

Estos son los mismos resultados que obtendríamos escribiendo las relaciones "confluentes" [21.6'] en la forma

$$x_{1t} = a_1 x_{3t} + a_1 + \text{error} \quad x_{2t} = b_1 x_{3t} + b_1 + \text{error} \quad [21.6']$$

y ajustando cada una de estas ecuaciones a las dadas por el método de mínimos cuadrados, considerando  $x_{1t}$  y  $x_{2t}$ , respectivamente, como las variables dependientes. Esto podría, por lo tanto, ser un procedimiento correcto. Pero los resultados [21.30] y [21.31] no son los mismos que se obtendrían ajustando las dos ecuaciones *originales* [21.1'] y [21.2'] separadamente, considerando  $x_{1t}$  como variables dependientes en ambas ecuaciones. Por ejemplo, de [21.14], [21.16], [21.30] y [21.31] obtenemos:

$$\hat{Q}_1 = \frac{\hat{a}_1}{\hat{b}_1} = \frac{m_{13} - m_1 m_3}{m_{23} - m_2 m_3} \quad [21.33]$$

mientras si ajustamos [21.1'] directamente por el método de los mínimos cuadrados, obtenemos:

$$\alpha^*_1 = \frac{m_{11} - m_1 m_2}{m_{22} - m_2^2} \quad [21.34]$$

la cual es diferente de [21.33] puesto que [21.34] no dependerá directamente de  $x_{3t}$ , mientras [21.33] se aplica  $\alpha^*_1$  en [21.34] no es sencillamente una estimación de  $\alpha_1$  sino algo más. Es esto: Consideremos las ecuaciones [21.1']. De esta ecuación tendremos:

$$E(x_{1t} | x_{2t}) = \alpha_1 x_{2t} + E(\varepsilon_{1t} | x_{2t})$$

[21.34] hubiera sido una estimación de

$$\alpha_1 \text{ si } E(\varepsilon_{1t} | x_{2t})$$

y fuese independiente de  $x_{2t}$  y  $x_{3t}$ . Pero este no es el caso. Y, por lo tanto, el supuesto sobre la cual la estimación mínima cuadrática [21.34] se basa es que

$$E(x_{1t} | x_{2t}) = \alpha_1 x_{2t} + \text{const.}$$

es sencillamente falso. En efecto, de la distribución conjunta [21.18] de  $x_{1t}$  y  $x_{2t}$  y las transformaciones [21.10] — [21.12] y [21.14] — — [21.17] obtenemos fácilmente que  $E(x_{1t} | x_{2t})$  es función lineal de  $x_{2t}$  y  $x_{3t}$ , es decir:

$$E(x_{1t} | x_{2t}) = \frac{\alpha_2 \sigma_1^2 + \alpha_1 \sigma_2^2}{\sigma_1^2 + \sigma_2^2} x_{2t} + \frac{\alpha_1 \sigma_1^2}{\sigma_1^2 + \sigma_2^2} x_{3t} + \text{const.} \quad [21.35]$$

Si queremos predecir  $x_{1t}$ , dada  $x_{2t}$  y  $x_{3t}$ , esta fórmula [21.35] es la única que debe usarse. Para realizar nuestro propósito, podríamos escribir [21.35] como

$$E(x_{1t} | x_{2t}) = A x_{2t} + B x_{3t} + C$$

y ajustar esta ecuación directamente a los datos por el método de los mínimos cuadrados. Esto da el mismo resultado que si nosotros estimamos primeramente todos los coeficientes en [21.35] por el método de máxima verosimilitud, como se describió anteriormente, y entonces ponemos estas estimaciones en [21.35].

Así vemos que el método de mínimos cuadrados aplicado a las ecuaciones originales [21.1'] y [21.2'] separadamente, nunca da estimaciones correctas para la predicción. Esto demuestra la importancia de estudiar la distribución conjunta de *todas* las variables aleatorias observables en un sistema de relaciones estocásticas.

## CAPITULO VI

### *Problemas de predicción*

Una predicción estadística significa sencillamente una afirmación (probabilística) acerca de la localización de un punto muestra no observado. Si consideramos  $n$  variables aleatorias  $x_1, x_2, \dots, x_n$ , y conocemos su ley conjunta de probabilidad, podemos, al menos en principio, calcular la probabilidad de que un punto muestra  $(x_1, x_2, \dots, x_n)$  caiga en cualquier región o conjunto de puntos del espacio muestra, o podemos tomar un cierto nivel de probabilidad y deducir un sistema de regiones (o conjuntos de puntos) los cuales tienen esa probabilidad. En la práctica nos interesará que la región con un nivel de probabilidad dado sea la "menor" (en uno u otro sentido). Así, si conocemos la ley de probabilidad conjunta de las variables que hay que predecir, el problema de deducir una fórmula de predicción que tenga unas propiedades deseadas será un problema del cálculo de las probabilidades. Y el problema de elegir la mejor fórmula de predicción es un asunto, en gran parte

subjetivo, esto es, un problema del tipo de "juego" que aceptaremos gustosos.

Corrientemente, sin embargo, no conoceremos la ley de probabilidad de las variables que hay que predecir. Entonces el problema de predicción queda íntimamente relacionado con los problemas de contrastación de hipótesis y estimación. Ahora tendremos que hacer inferencias respecto la ley de probabilidad de las variables y predecirla de las muestras ya observadas. Intentaremos dar una formulación más general y rigurosa de estos problemas.

22.—*Formulación general del problema de predicción*

Consideremos  $n$  sucesiones o series de tiempo de variables aleatorias

$$x_{it} \quad (i = 1, 2, \dots, n)$$

observables desde  $t = 1$ . Los valores de cualquiera de las variables anteriores a  $t = 1$  los consideraremos como constantes dadas. Supongamos que podemos observar los valores de cada serie en un cierto punto del tiempo. Sea  $t = s$  este punto del tiempo para la serie  $i$ . Consideremos que el problema es predecir los resultados de las últimas observaciones no hechas.

Tendremos entonces el siguiente modelo de variables aleatorias a considerar:

$$x_{i,t} = x_{i,1}, x_{i,2}, \dots, x_{i,s}, x_{i,s+1}, x_{i,s+2} \dots \quad (i = 1, 2, \dots, n)$$

[22.1]

$x_{1,t}, x_{2,t}, \dots, x_{n,t}$  pueden, por ejemplo, ser  $n$  series de tiempo relacionadas  $t = s$ , indica el último punto del tiempo para el cual tenemos una observación de  $x_{i,t}$ . Queremos predecir el próximo valor en una o más series, o cualquier otro sistema conjunto de valores futuros de las variables no observadas. Consideremos cualquier sistema de  $M$  variables elegidas entre las variables

$$x_{i,s+r} \quad (i = 1, 2, \dots, n; r = 1, 2, 3, \dots)$$

A la vez que las

$$s_1 + s_2 + \dots + s_n = N$$

variables observadas, las variables a predecir formarán un sistema de  $N + M$  variables aleatorias. Para mayor sencillez cambiaremos la notación de estas variables, indicando las  $N$  variables observadas por  $x_1, x_2, \dots, x_N$  y las  $M$  variables a predecir por  $x_{N+1} \dots x_{N+M}$ . Hay una correspondencia biunívoca entre estas variables y las  $N + M$  variables  $x_{i,t}$  consideradas.

El problema de predicción se convierte entonces en el problema de establecer ciertas funciones de las variables observadas  $x_1, x_2, \dots, x_N$  que puedan emplearse para conjeturar el comportamiento de las futuras observaciones de

$$x_{N+1}, x_{N+2}, \dots, x_{N+M} \quad (18).$$

Supondremos que, cualquiera que sea  $s_1, s_2, \dots, s_n$  y cualquiera que sea el conjunto de  $M$  variables futuras consideradas, la ley de probabilidad conjunta de las  $N + M$  variables

$$x_1 \dots x_N, x_{N+1} \dots x_{N+M}$$

existe (pero puede no ser —y corrientemente ocurrirá— conocida). Indicaremos la probabilidad conjunta por

$$\rho(x_1, x_2, \dots, x_N, x_{N+1}, \dots, x_{N+M})$$

o abreviadamente por  $\rho$ . Esta probabilidad corrientemente se describe implícitamente por un sistema de relaciones estocásticas entre las variables consideradas, como se explica en los capítulos IV y V.

Supongamos por un momento que  $\rho$  es conocido. De  $\rho$  podremos calcular la ley de probabilidad elemental condicional de las  $M$  variables  $x_{N+1}, \dots, x_{N+M}$  dadas las  $N$  variables  $x_1, x_2, \dots, x_N$ . Indiquemos esta ley de probabilidad condicional por

$$\rho_2(x_{N+1} \dots x_{N+M} | x_1, x_2, \dots, x_N)$$

---

(18) Ver Harold Hotelling "Problems of Predictions". The American Journal of Sociology Vol. 48 July 1942, págs. 61-67.

o abreviadamente por  $\rho_2$ . Sea  $\rho_1(x_1 \dots x_N)$ , o abreviadamente  $\rho_1$ , la ley de probabilidad conjunta de las  $N$  variables observables. Podremos escribir

$$p = p_1 p_2 \quad [22.2]$$

Indiquemos por  $E$  cualquier sistema de valores particulares —uno para cada— de las variables observadas  $x_1, x_2, \dots, x_N$ ; y análogamente por  $E_2$  designaremos cualquier sistema de valores de las variables futuras  $x_{N+1}, \dots, x_{N+M}$ . Cualquier  $E_1$  puede representarse por un punto en el espacio muestra  $n$ -dimensional  $R_1$  de las variables  $x_1, x_2, \dots, x_N$ ; y análogamente cualquier  $E_2$  puede representarse por un punto en el espacio muestra  $M$ -dimensional  $R_2$  de las variables  $x_{N+1}, \dots, x_{N+M}$ , que han de predecirse. Finalmente, por  $E$  indicaremos un punto en el espacio muestra  $R$  de las  $M + N$  variables.

Ahora, dado cualquier  $E_1$ , en particular, podríamos, de  $\rho_2$ , calcular la probabilidad condicional de que  $E_2$  caiga en un conjunto de puntos del espacio muestra  $R_2$ . Esta probabilidad corrientemente será una función de  $E_1$ . También, para cualquier  $E_1$  dado y para cualquier nivel de probabilidad  $\beta$ , podremos deducir un sistema de puntos o regiones en  $R_2$  tal que la probabilidad de  $E_2$  caiga en cualquier conjunto particular sea  $\beta$ . Esto es, podemos predecir, con probabilidad  $= \beta$  de acertar, que  $E_2$  caerá en un conjunto de puntos particularmente elegidos entre estos conjuntos de puntos. Cualquiera de tales conjuntos de puntos en  $R_2$  lo llamaremos *región de predicción*, y designaremos tal región por  $W_2$ .

En general, sin embargo, no todas las regiones  $W_2$  de probabilidad  $\beta$  son igualmente "interesantes". Generalmente (aunque no siempre) tendremos interés en que la región con probabilidad  $\beta$  sea más estricta en un sentido o en otro. También podríamos estar interesados en predecir cuándo el punto muestra *no* caerá en cierta región. En cualquier caso, la elección del nivel de probabilidad  $\beta$  y la localización de esta región  $W_2$  con probabilidad  $\beta$ , que queremos emplear como fórmula de predicción, dependerá del uso práctico que queramos hacer de ella. Esta elección no es un problema estadístico. Supondremos sencillamente que, cualquiera que sea la probabilidad condicional  $\rho_2$ , el propósito de nuestro intento de predecir será encontrar una, y sólo una, región  $W_2$  de predicción  $E_2$



para todo conjunto de valores de los "predictores"  $x_1, x_2, \dots, x_N$ .

Si, por lo tanto, conocemos  $\rho_2$ , el problema de predicción será meramente un problema de cálculo de probabilidad, y no uno de inferencia estadística a partir de una muestra. En muchos casos prácticos  $\rho_2$  no se conoce, y entonces tendremos que buscar información acerca de  $\rho_2$  de las muestras  $E_1$  de las observaciones previas. La posibilidad de realizar esto se fundamenta en el supuesto básico que formularemos como sigue: *La ley de probabilidad  $\rho$  de las  $M + N$  variables*

$$x_1, x_2, \dots, x_N, x_{N+1}, \dots, x_{N+M}$$

*es de tal tipo que la especificación de  $\rho_1$  lleva consigo la especificación completa de  $\rho$  y, por lo tanto, de  $\rho_2$ .*

Por ejemplo, si  $\rho$  se caracteriza por un cierto número de parámetros desconocidos, entonces todos estos parámetros deben también caracterizar a  $\rho_1$ , con lo cual  $\rho_2$  no contendrá nuevos parámetros distintos de los que aparecen en  $\rho_1$ . Esto no es más que una forma más precisa de establecer que habrá una cierta persistencia en el tipo de mecanismo que produce la serie que ha de predecirse.

Supongamos ahora que la única cosa conocida acerca de  $\rho_1$  es que pertenece a una cierta clase específica  $\Omega_1$  de leyes elementales de probabilidad, y que, por lo tanto,  $\rho_2$  pertenece a una cierta clase correspondiente  $\Omega_2$ . Por  $\rho^*_1$  indicaremos cualquier miembro arbitrario de  $\Omega_1$ . Sea  $W_1(\rho^*_1)$  la región crítica de tamaño  $(1 - \alpha)$  en  $R_1$ , elegida de acuerdo con alguna regla, tal que la hipótesis  $\rho_1 = \rho^*_1$  se rechaza cuando, y solamente cuando,  $E_1$  cae en  $W_1(\rho^*_1)$ . Se establece un sistema tal de regiones críticas en  $R_1$  una para cada  $\rho^*_1$  perteneciente a  $\Omega_1$ . Si  $E_1$  cae fuera de  $W_1(\rho^*_1)$ , entonces  $\rho_1 = \rho^*_1$  no se rechaza. Si el sistema de regiones  $W_1(\rho^*_1)$  no es trivial, cualquier punto muestra  $E_1$  es un punto muestra observado de las  $N$  variables observables, sea  $\omega(E_1)$  el subconjunto de  $\Omega_1$  en cuya región crítica —de tamaño  $(1 - \alpha)$ —  $E_1$  no cae. Como explicábamos en la sección 14, parece razonable estimar la ley de probabilidad desconocida  $\rho_1$  sobre la base de  $E_1$  estableciendo que  $\rho_1 \in \omega(E_1)$ . Ahora bien, hemos supuesto anteriormente que para todo miembro  $\rho^*_1$  de  $\Omega_1$  (o lo que es lo mismo, para todo  $\rho_2$  de  $\Omega_2$ ) y para todo conjunto de valores de  $x_1, x_2, \dots, x_n$ , nuestra elec-

ción de la fórmula de predicción da una, y una sola, región de predicción  $W_2$  de tamaño  $\beta$ . A la subclase  $\omega(E_1)$  corresponderá, por lo tanto, una cierta subclase de tales regiones de predicción. Sea  $K(E_1)$  la suma (lógica) de todos los elementos  $W_2$  de esta subclase. Puede parecer razonable predecir  $E_2$ , tomando como base el punto muestra  $E_1$ , estableciendo que

$$E_2 \text{ cae en } K(E_1) . \tag{22.3}$$

¿Cuál es la probabilidad de que esta afirmación sea cierta? Indiquemos por

$$g [K | \rho_1 \in \omega(E_1) ]$$

o abreviadamente por  $g(K)$  la probabilidad de que  $E_2$  caiga en  $K$  cuando  $\rho_1 \in \omega(E_1)$ . Sea

$$\bar{g} \{K | \rho_1 \in | \Omega_1 - \omega(E_1) \}$$

o abreviadamente  $\bar{g}(K)$ , la probabilidad de que  $E_2$  caiga en  $K$  cuando  $\rho_1$  cae fuera de  $\omega(E_1)$ . La probabilidad  $\rho(E_2 \in K)$  de que [22.3] sea cierta, evidentemente será:

$$P(E_2 \in K) = \alpha g(K) + (1 - \alpha) \bar{g}(K) \tag{22.4}$$

es decir, la probabilidad de que [22.3] sea cierta, es la probabilidad de que  $\omega(E_1)$  cubra  $\rho_1$  por la probabilidad de que  $E_2$  caiga en  $K$  mas la probabilidad de que  $\omega(E_1)$  no cubra  $\rho_1$  por la probabilidad de que  $E_2$  caiga en  $K$ . Ahora bien, las probabilidades  $g(K)$  y  $\bar{g}(K)$  serían, en general, funciones de la verdadera distribución  $\rho_1$ . Pero podemos dar desigualdades para  $\rho(E_2 \in K)$ . Evidentemente,

$$1 \geq g(K) \geq \beta \quad \text{mientras} \quad 0 \leq \bar{g}(K) \leq 1 .$$

Por lo tanto,

$$1 \geq P(E_2 \in K) \geq \alpha \beta . \tag{22.5}$$

(Para  $\Omega_1$  particulares pueden existir limites amplios.)

El procedimiento descrito puede también considerarse de la siguiente forma: Suponemos que para todo miembro  $\rho^*_1$  de  $\Omega_1$  exis-

te una cierta región de predicción  $W^*_2$  que emplearíamos si  $\rho^*_{.1}$  fuese la verdadera distribución. Si  $\rho^*_{.1}$  es la verdadera distribución la probabilidad que  $K(E_1)$  cubra la correspondiente región de predicción, evidentemente es igual a  $\alpha$ . Por lo tanto,  $K(E_1)$  puede considerarse como una región de confianza con coeficiente de confianza  $\alpha$ , para la estimación de la localización de la región "ideal" de predicción  $W_2$  correspondiente a la hipótesis verdadera.

El problema corriente en la práctica es, sin embargo, deducir regiones de predicción para  $E_2$ , la cual, con un nivel de probabilidad dado, sea lo menor posible. Entonces las regiones  $K$ , deducidas como se dijo anteriormente, no tienen necesariamente que ser las mejores regiones. Mas, precisamente, si para un  $\beta$  dado las regiones  $W_2$  fuesen las menores regiones (de acuerdo con alguna medida), y si el conjunto confidencial  $\omega(E_1)$  fuese el menor conjunto confidencial. El problema de cuándo o no la correspondiente  $K(E_1)$ , medida con la misma medida que las regiones  $W_2$ , puede ser la menor región de predicción dependerá de la forma como se define el término "menor" respecto  $\omega(E_1)$ . O expresado en términos simples, la elección de un sistema particular de conjuntos de confianza para la estimación de  $\rho$  dependerá de algún sistema de ponderaciones de los tipos de error que pueden cometerse estableciendo que  $\omega(E_1)$  cubrirá  $\rho_1$ . Si por otra parte el propósito es deducir una región de predicción  $K(E_1)$  una ponderación diferente de los errores de estimación será necesario con objeto de llegar a la ponderación deseada de los errores posibles de predicción.

Vemos, por lo tanto, que, al parecer, el procedimiento lógico bi-étápico de estimar primero la distribución desconocida de las variables que han de predecirse y entonces, empleando esta estimación, deducir una fórmula de predicción para las variables puede no ser eficiente. Discutiremos un método más sencillo y directo para deducir fórmulas de predicción que evite las dificultades discutidas anteriormente.

Indiquemos por  $E_1$  cualquier punto en el espacio muestra  $R_2$  de  $x_{N+1}, \dots, x_{N+M}$  e indiquemos por  $\bar{E}_2$  un punto en  $R_2$  que se va a emplear para predecir  $E_2$ . Consideremos el problema de definir  $\bar{E}_2$  como una función de  $x_1, x_2, \dots, x_n$ , de tal forma que sea alta la probabilidad de que  $\bar{E}_2$  caiga en  $E_2$  (en un sentido o en otro).

Llamaremos a  $\bar{E}_2$  *función de predicción*, si establecemos que  $E_2$  coincidirá con  $\bar{E}_2$  y que esto puede no ocurrir, cometemos un error, la consecuencia del cual dependerá del propósito de la predicción. Empleando una idea de A. Wald (19) podemos asignar un sistema de ponderaciones de los varios errores posibles. Supongamos que este sistema viene definido por una función de ponderación  $Q(E_2, \bar{E}_2)$  tal que

$$Q = 0 \quad \text{si} \quad \bar{E}_2 = E_2$$

y  $Q \geq 0$  (y no idénticamente cero) para todos los puntos  $E_2 \neq \bar{E}_2$ ,  $Q$  puede considerarse como la "pérdida" incurrida si  $E_2 \neq \bar{E}_2$ . El valor esperado  $r$  de esta pérdida, en muestras repetidas, viene dado por

$$r = \int_R Q(E_2, \bar{E}_2) p dE \quad [22.6]$$

la integral se toma sobre el espacio total muestra  $R$  de las  $N + M$  variables

$$x_1, x_2, \dots, x_N; x_{N+1}, \dots, x_{N+M}.$$

Elegimos  $\bar{E}_2$  como una función de  $x_1, x_2, \dots, x_N$  y trataremos de elegir  $E_2(x_1, x_2, \dots, x_N)$  de tal forma que  $r$  (el riesgo) sea lo menor posible.

Supongamos que existe una función de predicción  $\bar{E}_2(x_1, x_2, \dots, x_N)$  dependiente de  $x_1, x_2, \dots, x_N$  solamente, tal que para otra función particular  $r$  pueda ser un mínimo, independientemente, de cual sea la verdadera distribución  $P_1$  (dentro de  $\Omega$ ). Entonces naturalmente elegiríamos esta función como la mejor predicción respecto a la función de ponderación  $Q$ . Podemos llamar a tal función de predicción la mejor uniformemente (dentro de  $\Omega$ ) respecto a la función de ponderación dada.

---

(19) Ver A. WALD: "Contributions to the Theory of Statistical Estimation and Testing Hypotheses", *Annals of Mathematical Statistics*. Vol. 10, december 1939, págs. 299-326.

En unos pocos casos sencillos tales funciones de predicción pueden existir. En general, sin embargo, debemos esperar que no exista una función de predicción uniformemente mejor. Teniendo que introducir algunos principios adicionales con objeto de elegir la función de predicción. Entonces, primeramente, disgregaremos todas aquellas funciones de predicción que haga  $r$  pequeño para todo miembro de  $\Omega_1$ . Si esto no es el caso, llamaremos a la función de predicción considerada función de predicción admisible. Para elegir entre diversas funciones de predicción *admisible*. Para elegir entre diversas funciones admisibles podemos adoptar el siguiente principio introducido por Wald: Para toda función admisible de predicción  $\bar{E}_2$  el riesgo es una función de la verdadera distribución  $p$ . Consideremos la función de predicción  $\bar{E}_2$  entre las admisibles, para la cual el mayor valor de  $r$  es un mínimo (es decir, menor o a lo más igual al mayor valor de  $r$  para cualquier otra  $\bar{E}_2$  admisible). Tal función de predicción, si existe, será la de menor riesgo entre las funciones de predicción admisibles. El problema de deducir tales funciones de predicción está estrechamente unido con el problema semejante de deducir mejores estimaciones (20).

### 23.—*Algunas sugerencias prácticas para la deducción de fórmulas de predicción*

De la discusión llevada a efecto vemos que la elección de una fórmula de predicción no puede, en general, hacerse totalmente sobre bases objetivas. La elección de la función de ponderación  $Q$ , por ejemplo, no es un problema estadístico objetivo. También la elección de una fórmula de predicción, cuando no existe una fórmula de predicción uniformemente mejor, es una materia más o menos subjetiva. La ventaja del procedimiento formal que hemos diseñado es, sin embargo, que describe precisamente dónde y cómo los elementos subjetivos entran en el cuadro y cuáles son sus consecuencias lógicas. El procedimiento descrito da un instrumento más eficiente para formar las funciones de predicción de acuerdo

---

(20) Para una discusión de los problemas de predicción dentro de un modelo de ecuaciones lineales en diferencias estocásticas, ver MANN y WALD, obra citada, págs. 192-202.

con nuestros deseos. Así, por ejemplo, la noción de una función de ponderación  $Q$  es útil en el sentido de que si elegimos una función de predicción más o menos arbitraria (por un método libre), la función de ponderación correspondiente que podría haberse elegido arbitrariamente como la mejor podría ser tal que no fuese aceptable, esto es, podría suceder que la elección arbitraria de la función de predicción no fuese muy buena después de todo.

Una regla práctica, no aceptada quizá con toda generalidad, en relación con diversas sucesiones de tiempo simultáneas, es la siguiente: Si queremos predecir valores futuros para una o más sucesiones, corrientemente es necesario deducir fórmulas de predicción sobre la base de la distribución conjunta de los elementos observables en todas las series. Esto es, hemos de tomar en consideración no solamente la dependencia serial estocástica entre las observaciones sucesivas en la misma secuencia, sino también la interdependencia entre las sucesiones varias consideradas. La situación es análoga a la que se presenta cuando se considera la estimación de parámetros desconocidos, como se discutió en el capítulo V (21).

El procedimiento descrito en la sección precedente, aunque sencillo en principio, llevará consigo problemas matemáticos considerables y álgebra superior. Hay, sin embargo, casos importantes donde son suficientes procedimientos más sencillos. Sugeriremos uno de tales procedimientos que pueden aplicarse con éxito en ciertos casos que frecuentemente se presentan en Econometría y otros trabajos de investigación estadística.

Supongamos que tenemos un caso en donde los siguientes supuestos se cumplen (empleando las notaciones de la sección 22) :

1. La distribución  $p_1$  de  $x_1, x_2, \dots, x_n$  se sabe que pertenece a una familia paramétrica de distribuciones que comprende parámetros desconocidos  $z_1, z_2, \dots, z_k$ , es decir, podremos escribir

$$p_1 = p_1(x_1, \dots, x_n; z_1, \dots, z_k)$$

o abreviadamente  $p_1 [E_1; (z)]$ .

---

(21) Para ulterior discusión de este problema particular ver el artículo del autor "Statistical Implications of a System of Simultaneous Equations" *Econometrica*. Vol. II January 1953, págs. 1-12. Ver también la discusión de H. B. MANN y A. WALD obra citada, págs. 215-216.

2. La distribución  $p$  de todas las  $N + M$  variables consideradas se obtiene sencillamente sustituyendo  $N + M$  por  $N$  en  $p_1$ ;  $N$  y  $M$  son enteros positivos arbitrarios (excepto quizá cuando  $N$  pueda ser mayor que un cierto entero positivo  $N_0$ ),  $p_2$  es, por lo tanto, también conocida, excepto para los valores de los parámetros  $\alpha$ .

3. Se establece que la estimación máxima verosimilitud de las  $\alpha$  deducida de  $p_1 [E_1; (\alpha)]$  para una muestra observada  $E_1$  es inesgada y converge estocásticamente a los verdaderos valores de los parámetros cuando aumenta  $N$ , y estas estimaciones son "buenas" estimaciones también para valores moderados de  $N$ .

Consideremos el "riesgo condicional"  $r$  definido por

$$\bar{r} = \int_n Q(E_2, E_2) \rho_2 [E_2; (\alpha)] dE_2.$$

Para un  $E_1$  fijo, podemos considerar  $r$  como una función de  $E_2$ . Podremos entonces proceder como sigue, para deducir la función de predicción

$$E_2 = E_2(x_1, x_2, \dots, x_N)$$

1.º Encontrar qué punto  $E_2$ , para un conjunto dado de  $\alpha$  y una muestra  $x_1, x_2, \dots, x_n$ , hace  $\bar{r}$  un mínimo (suponiendo que tal mínimo existe). El punto  $E_2$  correspondiente a este mínimo de  $\bar{r}$  sería en general una función de  $\alpha$  y de las variables observadas  $x_1, x_2, \dots, x_N$ . Indicando esta función por  $E_2$  podríamos, por lo tanto, escribir

$$\bar{E}_2 = \bar{E}_2(x_1, x_2, \dots, x_N; \alpha_1, \alpha_2, \dots, \alpha_K)$$

2.º En la función  $\bar{E}_2$  se insertan en lugar de las  $\alpha$  los valores

$$\hat{\alpha}_1, \hat{\alpha}_2, \dots, \hat{\alpha}_K$$

estimados por el método máxima verosimilitud y deducidos de las observaciones  $x_1, x_2, \dots, x_n$  y la distribución  $p_1$ . La fórmula de predicción resultante

$$\bar{E}_2 = \bar{E}_2(x_1, x_2, \dots, x_N; \hat{\alpha}_1, \dots, \hat{\alpha}_K)$$

contiene solamente elementos conocidos y, por lo tanto, está determinada.

Este procedimiento puede demostrarse que da las mismas fórmulas de predicción en ciertos casos ordinarios que los que se establecen como "mejores" tomando como bases la teoría de la estimación estadística. Demos un ejemplo:

Consideremos una sucesión de variables aleatorias definidas por la fórmula de recurrencia

$$x_t = k x_{t-1} + \varepsilon_t \quad (t = 1, 2, \dots) \quad [23.2]$$

en donde  $x_0$  es una constante, mientras  $k$  es una constante desconocida, y donde las  $\varepsilon$  se distribuyen normalmente e independientemente con media cero y varianza  $\sigma$ . Supongamos que hemos observado de  $x_t$  en adelante, incluyendo  $x_N$  y que queremos predecir  $x_{N+1}$  y  $x_{N+2}$ . Supondremos, además, que hemos elegido una función de ponderación del siguiente tipo:

$$Q = a(x_{N+2} - \bar{x}_{N+2})^2 + 2b(x_{N+2} - \bar{x}_{N+2})(x_{N+1} - \bar{x}_{N+1}) + c(x_{N+1} - \bar{x}_{N+1})^2 \quad [23.3]$$

en donde  $\bar{x}_{N+1}$  y  $\bar{x}_{N+2}$  indican los valores predichos de  $x_{N+1}$  y  $x_{N+2}$  y donde  $a > 0$ ,  $b$  y  $c$  son constantes tales que  $ac > b^2$ . (Esto es que la ponderación de un error en la predicción es constante a lo largo de una elipse con centro en  $\bar{x}_{N+1}$ ,  $\bar{x}_{N+2}$ .)

La distribución de  $x_{N+1}$  y  $x_{N+2}$  viene dada por las  $x_t$  anteriores,

y es

$$\varphi_2 = \frac{1}{2\pi\sigma^2} e^{-\frac{1}{2\sigma^2} \gamma} \quad [23.4]$$

en donde

$$\gamma = (x_{N+1} - k x_N)^2 + (x_{N+2} - k x_{N+1})^2. \quad [23.5]$$



La esperanza condicional de  $Q$  es entonces (ver [23.1])

$$r = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma^2} Q \beta e^{-\frac{1}{2\sigma^2} \bar{y}} d d x_{N+1} d x_{N+2}. \quad [23.6]$$

Haciendo mínimo  $\bar{r}$  respecto  $\bar{x}_{N+1}$  y  $\bar{x}_{N+2}$  obtenemos las dos ecuaciones siguientes para  $\bar{x}_{N+1}$  y  $\bar{x}_{N+2}$

$$\begin{aligned} Q \bar{x}_{N+2} + b x_{N+1} &= a k^2 x_N + b k x_N \\ b \bar{x}_{N+2} + c x_{N+1} &= b k^2 x_N + c k x_N \end{aligned} \quad [23.7]$$

lo cual da

$$\begin{aligned} \bar{x}_{N+1} &= k x_N \\ \bar{x}_{N+2} &= k^2 x_N \end{aligned} \quad [23.8]$$

independientemente de los valores de  $a$ ,  $b$  y  $c$ . Esto es, los “mejores” valores predichos respecto la función de ponderación [23.3] son los valores esperados de  $x_{N+1}$  y  $x_{N+2}$ . Pero no conocemos  $k$ . Su estimación máxima verosimilitud es, sin embargo,

$$\hat{k} = \frac{\sum_{t=1}^N x_t x_{t-1}}{\sum_{t=1}^N x_{t-1}^2}. \quad [23.9]$$

Nuestra fórmula de predicción está de acuerdo con el principio adoptado

$$\begin{aligned} \bar{x}_{N+1} &= \hat{k} x_N \\ \bar{x}_{N+2} &= \hat{k}^2 x_N \end{aligned} \quad [23.10]$$

Para juzgar la realidad de la predicción podríamos, por ejemplo, considerar la probabilidad de que

$$(x_{N+1} - \bar{x}_{N+1}) \quad \text{y} \quad (x_{N+2} - \bar{x}_{N+2})$$

estén dentro de ciertas cotas y las variables  $\bar{x}_{N+1}$  y  $\bar{x}_{N+2}$  están definidas por [23.9] y [23.10], ahora de forma sencilla podemos estudiar los valores del riesgo, calculado a partir de [22.6].

## CONCLUSIÓN

El paciente lector, llegado al final de nuestro análisis, puede encontrarse con que la teoría que hemos esbozado, aunque sencilla en cuestiones de principio, en muchos casos llevaría consigo una cantidad enorme de trabajo. Entonces, sarcásticamente diría "necesito mi vida entera para obtener una sencilla elasticidad de la demanda". Y pensaría ¿vale esto la pena? No podríamos, desde un principio y para fines prácticos, hacer uso de los métodos cortos, por ajustes de curvas, o hacer conjeturas favorables combinando nuestras experiencias generales con la inferencia que parece razonable a la vista de los datos particulares que tenemos a mano.

Sería pretencioso, e incluso injustificado, condenar todos los métodos abreviados y los trabajos de conjeturas prácticos, en los que cientos de economistas se apoyan en sus trabajos diarios de administradores o como consejeros de aquellos que regulan nuestra economía. Lo que hemos intentado demostrar es que esta clase de inferencias está basada, implícitamente, y quizá en el subconsciente, en los mismos principios que hemos tratado de describir con más precisión en nuestro análisis. Hemos puesto de manifiesto que los economistas pueden obtener información más útil y real (y también pocos resultados espúrcos), a partir de sus datos, adoptando modelos de probabilidad claramente formulados, y que tal formulación puede ayudar a sugerir qué datos faltan y cómo deben recogerse. Creemos que si la economía tiene que establecerse como una ciencia cuantitativa con prestigio, muchos economistas habrán de revisar sus ideas sobre el nivel de la teoría y técnica estadística y montante de trabajo que se requerirá, incluso para los más modestos proyectos de investigación. Por otra parte, debemos contar el tiempo y trabajos que puede ahorrarse eliminando una buena cantidad de malos planes y cubileteos con cifras. También espero de los estadísticos expertos, una vez que se les pueda convencer

que se tomen interés por los problemas particulares de la estadística relacionados con la Econometría, que sean capaces de suministrar muchas fórmulas tipo y tablas. Uno de los objetivos del análisis precedente ha sido mostrar la clase de lenguaje que creemos debe el economista adoptar con objeto de formular al estadístico claramente el problema. No dudamos que los estadísticos, a su vez, serán capaces de cumplir su objetivo.

En otras ciencias cuantitativas, el descubrimiento de "leyes", aun en campos altamente especializados, ha partido del estudio privado en enormes laboratorios científicos donde trabajan muchos expertos no solamente realizando medidas, sino también comprobando con minuciosa precisión las fórmulas que han de contrastarse y los planes de los experimentos fundamentales a efectuar. ¿Se puede exigir menos a la investigación económica cuando sus resultados han de ser la base de la política económica, de la que puedan depender billones de dólares de la renta nacional y el bienestar económico general de millones de personas?