



W
—
28
(9133)

DOCUMENTO DE TRABAJO
9133
METODOLOGIA ESTADISTICA DEL ANALISIS
ESTRUCTURAL

ANGEL VEGAS PEREZ

METODOLOGIA ESTADISTICA
DEL
ANALISIS ESTRUCTURAL

Dr. Angel Vegas Pérez
Catedrático jubilado de la
Universidad Complutense de Madrid

Han colaborado en la elaboración de este documento de trabajo los
Profesores Dra. M^a. del Carmen Moral Moral, Dr. Jesús Vegas Asensio
y Dr. José Manuel Vegas Montaner.-

SUMARIO

- I.- *ESTRUCTURA.*
- II.- *ESTRUCTURAS ESTOCASTICAS.*
- III.- *ESTIMACION DE LA ESTRUCTURA.*
- IV.- *IDENTIFICACION ESTRUCTURAL*
- V.- *APLICACIONES A LOS MODELOS BINOMIAL, LINEAL UNIECUACIONAL Y LINEAL MULTIECUACIONAL.*

I.- ESTRUCTURA.

1.- Estructura: Concepto de la forma en que aparecen dispuestas las diferentes partes de un todo.

De este concepto se desprende que hablar en términos "estructurales" equivale a hablar de los aspectos "formales" que corresponden al enlace que une los elementos que integran el conjunto total.

Por otra parte, la idea de "estructura" conduce a lo que suele entenderse por "comprensión" de un concepto, como contenido integrado por los objetos definidos o expresados por él, mediante las referencias, notas o atributos que los caracterizan.

La "comprensión" difiere, por lo tanto, de la suma de las notas de los objetos integrados estructuralmente, así como de los propios objetos en cuanto términos de referencia de dichas notas.

2.- Conjunto estructurado: Una consecuencia interesante de la idea de "comprensión", en el concepto de estructura, es la definición de "conjunto estructurado".

En efecto, la teoría matemática de "conjuntos" se basa fundamentalmente en la idea de "conjunto definido" cuando se dispone de un criterio que permite, inequívocamente, asegurar si un objeto cualquiera, pertenece o no al conjunto. Es evidente que caben dos formas de definición. En efecto, cabe la alusión "concreta" o "nominal" a cada uno de los elementos que lo integran, en cuyo caso se trata de la definición por "extensión", o bien se define a través de las propiedades "estructurales", tratándose entonces de la definición por "comprensión".

Al conjunto definido por "comprensión" lo llamamos conjunto estructurado. Simbólicamente podemos representarlo de la forma $\{ X \}$ en la que X simboliza los elementos que

lo integran, Θ representa a constantes y, por ello, tienen un significado paramétrico y \mathcal{F} es el conjunto de operaciones entre X y Θ , por lo que tiene un significado de naturaleza algorítmica. Θ , por lo tanto, tiene una significación paramétrica, ya que representa a lo que permanece constante frente a la multiplicidad de X .

Supongamos, por ejemplo, que se trata del conjunto de puntos alineados, es decir, que pertenecen a una recta. Sabemos que la ecuación cartesiana de la recta es $Ax+By+C=0$. Todos los puntos de coordenadas, x, y , constituyen el conjunto X y los parámetros $(A B C)$ que satisfacen la ecuación a través de las operaciones de suma y producto, integran la componente estructural Θ .

3.- Sistema. Entendemos por tal un conjunto estructurado en el que la "estructura" consiste en las relaciones que ligán los componentes del mismo. Se trata de una "Organización" en la que todos los elementos están "interconectados".

4.- Imagen del mundo. Se trata del conjunto de ideas vigentes, debidamente trabadas, y que, por ello, constituye una descripción ideal del ámbito en el que pueden presentarse los acontecimientos.

Hablar, por lo tanto, de la obtención de la imagen del mundo, es tanto como plantear en términos "estructurales" los acontecimientos históricos y, por ello, supone la culminación de toda pretensión científica de subrayar lo que hay de formalmente "estable o permanente" en los acontecimientos que integran, en su más amplio sentido, el devenir físico, económico, sociológico, ...

5.- Estructuralismo. Se trata de una teoría común a diferentes ciencias que concibe a cualquier objeto de estudio como un todo cuyos miembros se determinan entre sí, tanto en su naturaleza como en sus funciones, en virtud de leyes generales.

Un ejemplo de "estructuralismo" se ofrece al considerar el plano como conjunto de puntos o como conjunto de rectas, es decir, "puntual" o "reglado". En la concepción "puntual" la recta puede definirse como lugar de los puntos, definido por dos de ellos. En la concepción "reglada" el punto viene definido como el lugar de rectas definido por dos de ellas.

Ambas concepciones del plano, "puntual" y "reglado", responden a la misma expresión estructural: "recta definida por dos de sus puntos" "punto definido por la intersección de dos rectas". Esto equivale a decir que existe un "isomorfismo" entre la recta como conjunto de puntos y punto como haz de rectas. Se trata, pues, de un ejemplo de estructuralismo.

6.- Fenomenología. Entendemos por fenómeno lo que puede ser percibido por los sentidos o por la conciencia.

Se trata de un concepto de carácter "potencial" que incluye la posibilidad de suceder, de acontecer.

Es evidente que si el suceso es lo que acontece, lleva en sí la posibilidad de su percepción y, por ello, lo eventual tiene categoría "fenoménica".

Todo esto conduce a considerar en el "fenómeno", juntamente con su categoría potencial, su categoría eventual, de acontecer.

Evidentemente, puede hablarse de "conjunto estructurado" de fenómenos, o fenómeno estructurado, como la expresión de la propiedad formal que integra el conjunto.

En este sentido, podemos hablar de una relación estructural entre los elementos que integran las dos categorías, potencial y eventual, estableciendo una correspondencia expresada en

los siguientes términos: Definida una categoría potencial, caben tres posibilidades para la categoría eventual, según que conste de varios sucesos, de uno sólo o de ninguno. En el primer caso, el fenómeno se denomina aleatorio, en el segundo, cierto o determinista, y, en el tercero, imposible.

En el caso especial en el que la indeterminación que comporta la aleatoriedad pudiera expresarse numéricamente mediante un coeficiente de indeterminación, al fenómeno se le denomina estocástico y a dicho coeficiente, probabilidad.

En general, se podrá hablar de estructuras aleatorias o estocásticas, por una parte, y estructuras determinadas o imposibles, por otra.

Supongamos, por ejemplo, que se trata de un fenómeno cuya componente potencial consiste en "color blanco" y la eventual en "color blanco y color negro". Según lo anteriormente dicho, se trata de un fenómeno aleatorio.

Es el caso en que los sucesos "color blanco", "color negro" sean el resultado de la obtención de "bola blanca" o "bola negra" en una extracción realizada en una urna en la que existen bolas blancas y negras en la proporción p y q , que, por definición, expresan numéricamente las probabilidades de obtener bola blanca o bola negra. Se trata, por tanto, de un fenómeno estocástico.

En el caso en que la categoría "eventual" esté integrada por los resultados de un experimento, es decir, provocados según la voluntad del observador con arreglo a un esquema perfectamente conocido, el fenómeno es experimental.

Por el contrario, si el fenómeno es independiente de la voluntad del observador, se trata de un fenómeno observacional. Según la categoría de los fenómenos, cabe hablar de estructuras experimentales y estructuras observables y, por tanto, de leyes experimentales y leyes observables.

II.- ESTRUCTURAS ESTOCASTICAS.

1.- Como dijimos anteriormente, una estructura aleatoria cuya indeterminación puede expresarse en términos de probabilidad recibe la calificación de estocástica.

La "estocasticidad", por lo tanto, está fundamentalmente relacionada con la teoría de la probabilidad, de la que recordaremos algunas cuestiones fundamentales.

2.- Algebra de Sucesos.- Representamos por (S_i) el conjunto de sucesos posibles de un fenómeno.

Definimos como suma de "n" sucesos al suceso $S_{\Sigma} = S_1 \cup S_2 \dots \cup S_n = \bigcup_{i=1}^n S_i$ que se presenta o acaece cuando se presenta cualquiera de los "n" S_i , y como producto $S_{\pi} = S_i \cap S_j$, cuando se presentan ambos sucesos S_i, S_j .

Una relación importante entre dos sucesos S_i, S_j es la llamada condicionante que representamos de la forma $S_k = S_j/S_i$, y supone que el suceso S_j exige, para su realización, que previamente se realice el S_i .

3.- Probabilidad.- Entenderemos por probabilidad del suceso S, la función P(S) del mismo que cumple las siguientes propiedades:

- I $0 \leq P(S) \leq 1$
- II $P(\bigcup_{i=1}^n S_i) \leq P(S_1) + P(S_2) + \dots + P(S_n)$
- III $P(S_i \cap S_j) = P(S_i) P(S_j/S_i) = P(S_j) P(S_i/S_j)$
- IV $\lim P(S_n) = P(\lim S_n)$

En el caso en que se cumpla $P(S_i) = P(S_j/S_i)$ diremos que se trata de dos sucesos estocásticamente independientes. Consecuencia inmediata es que la expresión de la probabilidad del producto $S_i \cap S_j$ adopta la forma $P(S_i \cap S_j) = P(S_i) P(S_j)$.

4.- Variable estocástica.- Si el conjunto de los sucesos puede representarse como conjunto de puntos en un espacio cartesiano, tendremos la variable estocástica o variante. Según nos enseña la matemática de la probabilidad, la distribución de las probabilidades correspondiente al caso unidimensional, se representa por la función:

$$P_r(\xi \leq x) = F(x), \quad P(a < \xi \leq b) = F(b) - F(a)$$

Si la ξ variante fuera continua, cabe hablar de la función de densidad $f(x) = F'(x)$, es decir, $F(x) = \int_{-\infty}^x f(x) dx$.

Si se tratase del caso multivariante, la función de distribución de probabilidad vendría representada de la forma $P(\xi_1 < x_1, \dots, \xi_n < x_n) = F(x_1, x_2, \dots, x_n)$; las llamadas distribuciones marginales, tendrán la forma $F(x_1) = F(x_1, \infty, \dots, \infty)$; $F(x_n) = F(\infty, \dots, \infty, x_n)$.

En el caso continuo tendremos:

$$f(x_1, \dots, x_n) = \frac{\partial^n F(x_1, x_2, \dots, x_n)}{\partial x_1 \partial x_2 \dots \partial x_n}$$

$$F(x_1, \dots, x_2, \dots, x_n) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} f(x_1, x_2, \dots, x_n) dx_1 \dots dx_n$$

En el caso que las $\xi_1, \xi_2, \dots, \xi_n$ fueran estocásticamente independientes, tendremos $f(x_1, \dots, x_n) = f_1(x_1) \dots f_n(x_n)$, es decir, que la función de densidad conjunta es el producto de las funciones de densidad marginales.

5.- Información estocástica. - Dados dos sucesos S_i, S_j nos interesa encontrar la medida de la influencia que S_i ejerce sobre la posible presencia del S_j . Para ello partimos de la definida relación condicionante S_j/S_i expresada en términos de probabilidad, es decir. $P(S_j/S_i)$.

Definiremos como medida de la información del suceso S_i , respecto al S_j , la expresión

$$I_{ij} = \lg \frac{P(S_j/S_i)}{P(S_j)} .$$

En el caso en que la presencia de S_i no suministre información respecto a la posible realización de S_j , es decir, si $P(S_j)=P(S_j/S_i)$, tendremos $I_{ij} = \lg 1 = 0$.

Por otra parte, es evidente que I_{ij} , toma valores crecientes al crecer $P(S_j/S_i)$, y adquiere su valor máximo cuando $P(S_j/S_i)=1$, lo que supone que $S_i=S_j$, es decir, $I_{ij} = \lg \frac{1}{P(S_i)}$. Este resultado corresponde a la medida de la información que comporta el suceso estocástico S_i de probabilidad $P(S_i)$, cuando se realice.

A partir de estos resultados cabe hablar de la medida de la información que un fenómeno estocástico puede suministrar como consecuencia de su estocasticidad.

Dado el fenómeno con sucesos $S_1 \dots S_n$ y probabilidades $P(S_1) \dots P(S_n)$, las informaciones que pueden suministrar tendrán por medidas: $I_1 = \lg \frac{1}{P(S_1)}$, $I_2 = \lg \frac{1}{P(S_2)}$, ..., $I_n = \lg \frac{1}{P(S_n)}$.

Tomaremos como expresión de la medida de la información del fenómeno, el valor medio:

$$P(S_1) I_1 + P(S_2) I_2 + \dots + P(S_n) I_n = \sum_{i=1}^n P(S_i) \lg \frac{1}{P(S_i)} = \sum_{i=1}^n P(S_i) (\lg 1 - \lg P(S_i)) = - \sum_{i=1}^n P(S_i) \lg P(S_i).$$

Supongamos, por ejemplo, que el fenómeno consista en la obtención de color blanco o negro al extraer al azar una bola de una urna cuya composición consiste en que las bolas blancas y las negras figuran en la proporción p y q , y, por lo tanto, coinciden con la expresión numérica de las probabilidades de obtener bola blanca o bola negra respectivamente.

La medida de la información que puede suministrar este fenómeno, será, por lo tanto
 $I = - (p \lg p + q \lg q)$.

En general, la expresión de la cantidad de información que corresponde a un fenómeno, expresado por la variante \vec{x} .

$$\vec{x} = \begin{pmatrix} x_1 & x_2 & \dots & x_n \\ p_1 & p_2 & \dots & p_n \end{pmatrix}$$

vendrá dado por la expresión $I_{\vec{x}} = - \sum p_i \lg p_i = - H$. H, recibe el nombre de entropía por la similitud que tiene con la medida del fenómeno entrópico, fundamental en importancia en la Termodinámica. Por esta razón la teoría de la información que comporta la aleatoriedad, en la forma en que se ha presentado suele llamarse Teoría entrópica de la información. Una propiedad importante de la función $I(p_1, p_2, \dots, p_n) = - \sum p_i \lg p_i$, consiste en que toma su valor máximo cuando $p_1 = p_2 = \dots = p_n = p$.

En efecto se trata de obtener el máximo de $- \sum p_i \lg p_i$ con la condición $\sum p_i = 1$;

Según sabemos, por el Análisis matemático, dada la partición de la expresión definimos

$$\phi(p_1, p_2, \dots, p_n) = - \sum_i p_i \lg p_i + \lambda \left(\sum_i p_i - 1 \right).$$

Se trata de determinar las p_i que cumpla la condición:

$$\frac{\partial \phi(p_1, \dots, p_n)}{\partial p_i} = 0$$

$$\frac{\partial \phi(p_1, \dots, p_n)}{\partial p_n} = 0$$

como
$$\frac{\partial \phi(p_1, \dots, p_n)}{\partial p_i} = - \left[\lg p_i + K \frac{1}{p_i} \cdot p_i \right] + \lambda p_i =$$

$$= - \lg p_i + K - \lambda = 0$$

luego $P_i = e^{k-\lambda} = C$ ($i = 1, 2 \dots n$) = $\sum P_i = n C = 1$,

en definitiva: $P_1 = P_2 = \dots = P_n = \frac{1}{n}$.

Este resultado indica que la máxima información corresponde al fenómeno de distribución uniforme. El valor de I será entonces , $I = - \sum \frac{1}{n} \lg \frac{1}{n} = - \frac{n}{n} \lg \frac{1}{n} = \lg n$.

Todo lo anteriormente dicho sirve para el caso en el que se tratase de una variante continua cuya función de densidad fuera $f(x)$. La expresión de la cantidad de información

$$I_{\text{cont}} = - \int_{\mathcal{R}} f(x) \lg f(x) dx$$

$$\int_{\mathcal{R}} f(x) dx = 1$$

Se demuestra fácilmente que el máximo de este funcional corresponde al caso en el que $f(x)=k$, para todo x perteneciente al conjunto \mathcal{E} , es decir, se trata de la distribución uniforme.

Estos resultados pueden ser generalizados a fenómenos multidimensionales.

Consideremos, por otra parte, la cantidad de información que corresponde al fenómeno determinista, es decir, aquel en que sólo se presente en un único suceso S_i , es decir,

$$P(S_i) = P_i = 1 ; P(S_j) = P_j = 0 ; i \neq j$$

$$I = - (P_1 \lg P_1 + \dots + P_i \lg P_i + \dots + P_n \lg P_n)$$

cuando $p = 1$, $p \lg p = 0$.

$$\text{Si } p = 0 \quad \lim_{p \rightarrow 0} p \lg p = \lim_{p \rightarrow 0} \frac{\lg p}{\frac{1}{p}} \xrightarrow{K} \frac{\frac{1}{p}}{-\frac{1}{p^2}} = 0.$$

Este resultado indica que la información que corresponde al fenómeno determinista es igual a 0.

$$I = 0$$

El fenómeno dicotómico indeterminista tendrá por nivel de información la expresión $I = \lg 2$.

Es evidente que si tomamos logaritmos en base 2 la cantidad de información será $I = \lg_2 2 = 1$, que recibe el nombre de "bit", unidad de información. Tomaremos para la cuantificación de la información logaritmos en base 2, lo que equivale a establecer como referencia básica la información del fenómeno dicotómico.

$$I = - \sum P_i \lg_2 P_i \quad \text{número de bits}$$

Una aplicación importante de lo expuesto respecto a la máxima información, consiste en aplicar los resultados obtenidos al caso en que se trate de un fenómeno aleatorio del que se desconoce su distribución de probabilidad siendo necesario, por otra parte, disponer de ella.

En este caso podemos adaptar como solución la que corresponde al fenómeno de máxima información, es decir, la distribución uniforme.

Este resultado coincide con el llamado postulado de Bayes-Laplace que dice: "Si se ignoran las probabilidades que corresponden a un fenómeno aleatorio, deben considerarse todas iguales, ya que no parece lógico, al no disponer de conocimientos previos, dar mayor importancia a un suceso que a otro.

Como aplicación de este principio tenemos los siguientes casos:

Si no existe ningún conocimiento la solución es la distribución uniforme.

Si se tratase de un fenómeno continuo de rango $(0, \infty)$, y se conociera el momento de primer orden \underline{m} , la función de densidad sería $f(x) = \frac{1}{m} e^{-\frac{x}{m}} \quad (0, \infty)$.

Si se partiera del conocimiento de m y σ^2 , la solución sería:

$$f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}, \quad (-\infty, +\infty)$$

III.- ESTIMACIÓN DE LA ESTRUCTURA.

1.- Estimación estocástica. - Como hemos dicho anteriormente el conjunto estructurado viene representado por la expresión $f(x, \theta)$ en la que x hace referencia a los elementos u objetos incluidos formalmente en el conjunto, θ representa la propiedad genérica que caracteriza a las x y tiene, por tanto, una significación paramétrica, y por último, la f representa a las operaciones entre x y θ , por lo que tiene carácter algorítmico.

Supongamos, por ejemplo, que se trata del conjunto de puntos que equidistan de un dado, P .

Los puntos x vendrán representados cartesianamente mediante las coordenadas (x, y, z) en un sistema tridimensional ortogonal y cuyo origen coincide con el punto P . La distancia de x al punto P vendrá dada por la relación: $\sqrt{x^2 + y^2 + z^2} = r = \theta$ que, como se sabe, es la ecuación de la esfera de radio r , y, por lo tanto, se trata de la familia de esferas concéntricas.

La estructura vendrá definida por el parámetro θ .

El problema que pretendemos resolver es el de la determinación de θ y, para ello, supondremos que se trata de un fenómeno estocástico, es decir, que θ es una variante cuya función de probabilidad es $\pi(\theta)$ que, en el caso de continuidad, será la función de densidad. El valor estimado vendrá dado por la relación $\hat{\theta} = E(\theta) = \int_{\mathbb{R}} \theta \pi(\theta)$.

a) Supongamos que se trata de la distribución rectangular de rango (a,b) , es decir, $\pi(\theta) = \frac{1}{b-a}$, por lo tanto, el valor medio vendrá dado por la expresión:

$$\hat{\theta} = E(\theta) = \frac{1}{b-a} \int_a^b \theta d\theta = \frac{b^2 - a^2}{2(b-a)} = \frac{a+b}{2}.$$

Este resultado tiene una segunda interpretación. En efecto, si sólo sabemos que $a \leq \theta \leq b$ parece que tiene una justificación, siquiera operativa, estimar θ por el promedio de los extremos. Ahora bien, recordando que el postulado de Bayes-Laplace consiste en adjudicar a la distribución uniforme una preferencia lógica, respecto a cualquier otra distribución, ya que no existiendo en principio ninguna razón para justificar diferencias en las probabilidades correspondientes a los valores del intervalo $a \leq \theta \leq b$, debe ser adoptada la función de probabilidad $\pi(\theta) = \frac{1}{(b-a)}$.

Por último, existe otra justificación en el ámbito definido por la Teoría entrópica de la información. Conforme ya vimos en el capítulo anterior, la máxima información corresponde al fenómeno en el que los posibles sucesos tiene una función de probabilidad constante, por ello, la distribución uniforme conduce a una lógica justificación, en el ámbito que considera la aleatoriedad como fuente de información.

b) En el caso anterior hemos supuesto que no se dispone de ninguna información estadística respecto a los elementos que integran el conjunto. Vamos ahora a suponer que disponemos de dicha información, es decir, que mediante un adecuado método obtenemos un conjunto de

observaciones $x_1^o, x_2^o, \dots, x_n^o$ de los elementos x_1, x_2, \dots, x_n que integran el conjunto estructurado $f(x, \theta)$.

Esta metodología supone que los x están estocásticamente relacionados con θ según una función de probabilidad $P(x_1, x_2, \dots, x_n, \theta) = P(x, \theta)$ que toma el valor $P(x_1^o, x_2^o, \dots, x_n^o, \theta) = l(x^o, \theta)$ cuando se trata de la probabilidad correspondiente a los valores observados y que recibe el nombre de "Verosimilitud". Esta función comporta un profundo sentido crítico respecto de los valores del parámetro estructural.

2.- Inferencia bayesiana.

Si consideramos los dos fenómenos estocásticos x y θ , correspondientes a los puntos x y al parámetro θ , tendremos: $P(x \cap \theta) = P(x) \pi(\theta/x) = \pi(\theta) P(x/\theta)$,

$$\text{luego } \pi(\theta/x) = \frac{P(x/\theta)}{P(x)} = \pi(\theta)$$

Si nos referimos a los valores observados x^o , tendremos: $\pi(\theta/x^o) = \frac{l(x^o, \theta)}{P(x^o)} \pi(\theta)$.

Este resultado conduce a la importante relación entre el "a priori" y el "a posteriori" respecto a la información suministrada por los valores observados x^o . Esta relación constituye el elemento básico de lo que suele llamarse "inferencia bayesiana" en la "teoría del conocimiento estadístico".

En efecto, la teoría clásica de la "inferencia estadística" en lo que se refiere al "contraste de las hipótesis" formulada y sistematizada de forma magistral por Neyman-Pearson, se basa en la definición exhaustiva de los dos posibles errores que pueden cometerse al admitir o rechazar una hipótesis. Se entiende por error de primera especie el que se comete al rechazar la hipótesis cuando debiera ser admitida, y error de segunda especie al que se comete al admitir la hipótesis, cuando debiera ser rechazada.

En un congruente planteamiento lógico los dos errores, atendiendo a sus consecuencias, son esencialmente distintos, ya que si la hipótesis surge de la pretensión de completar la "imagen del mundo", es más lamentable incurrir en error rechazando lo que convenía admitir dentro de las exigencias de una inducción, consistente en sus fundamentos lógicos, que admitirla aunque dicha admisión fuera errónea. Por ello, conviene expresar, en términos de probabilidad, la categoría correspondiente a dichos errores llamando nivel de significación a la probabilidad de cometer un error al rechazar la hipótesis, error que hemos llamado de primera especie, y potencia a la probabilidad de rechazar la hipótesis cuando es falsa.

Si llamamos error de segunda especie al que se comete al admitir la hipótesis siendo falsa, su probabilidad será la complementaria a la potencia.

Estas probabilidades de los errores dependerán de la información que suministran los valores observados x^o , a través de la verosimilitud, $l(x^o, \theta)$.

Llamaremos región crítica al subespacio observacional, que se caracteriza por que, si las observaciones obtenidas pertenecen a él, la hipótesis debe ser rechazada. A la probabilidad que corresponde a esta región se le llama tamaño de la región.

Dada una región crítica de tamaño α diremos que es prepotente cuando tiene la máxima potencia, es decir, cuando es mínima la probabilidad de un error de segunda especie. Suele llamarse a la región crítica prepotente región crítica propiamente dicha.

El problema fundamental que resuelve la inferencia, desde el punto de vista de Neyman-Pearson, según lo dicho anteriormente, es disuasorio en sentido estricto, ya que la información estadística de los sucesos observados es utilizada para admitir o rechazar la hipótesis sobre los valores estructurales.

Por el contrario, el planteamiento bayesiano trata de utilizar la información suministrada por las observaciones para mejorar la estimación. En efecto, si suponemos conocida la función de probabilidad $\pi(\theta)$, la estimación del parámetro estructural, será, como sabemos, $\hat{\theta} = \int_{\mathcal{R}} \theta \pi(\theta) d\theta$.

Supongamos que x_0 son las observaciones de los elementos x . La expresión de la función de probabilidad, condicionada por las x_0 , será $\pi(\theta/x_0)$ y, por lo tanto, la estimación del parámetro será: $\hat{\theta}_{x_0} = \int_{\mathcal{R}} \theta \pi(\theta/x_0) d\theta$

Como hemos visto anteriormente $\pi(\theta/x_0) = \frac{l(x_0, \theta)}{P(x_0)} \pi(\theta)$
 Teniendo en cuenta que $\int_{\mathcal{R}} \pi(\theta/x_0) d\theta = 1$

tendremos, $\frac{1}{P(x_0)} \int_{\mathcal{R}} l(x_0, \theta) \pi(\theta) d\theta = 1$

luego $P(x_0) = \int_{\mathcal{R}} l(x_0, \theta) \pi(\theta) d\theta$

En definitiva $\pi(\theta/x_0) = \frac{l(x_0, \theta) \pi(\theta)}{\int_{\mathcal{R}} l(x_0, \theta) \pi(\theta) d\theta}$

La estimación $\hat{\theta}$ será entonces $\hat{\theta}_{x_0} = \frac{\int_{\mathcal{R}} \theta l(x_0, \theta) \pi(\theta) d\theta}{\int_{\mathcal{R}} l(x_0, \theta) \pi(\theta) d\theta}$

De forma análoga, si disponemos de los valores observados en n agrupaciones, $x^{(1)}, \dots, x^{(n)}$, tendremos la sucesión de probabilidades del parámetro estructural:

$$\pi(\theta/x_0^{(1)}) = k_1 l(x_0^{(1)}, \theta) \pi(\theta)$$

$$\pi(\theta/x_0^{(1)}, x_0^{(2)}) = k_2 l(x_0^{(2)}, \theta) \pi(\theta/x_0^{(1)}) = k_1 k_2 l(x_0^{(1)}, \theta) l(x_0^{(2)}, \theta) \pi(\theta)$$

$$\pi(\theta/x_0^{(1)}, x_0^{(2)}, \dots, x_0^{(n)}) = k_1 k_2 k_3 \dots k_n l(x_0^{(1)}, \theta) l(x_0^{(2)}, \theta) \dots l(x_0^{(n)}, \theta) \pi(\theta)$$

Este resultado nos indica la ley de incorporación de las nuevas observaciones para conseguir una estimación más apropiada de la probabilidad del parámetro estructural.

Ahora bien, en todas las fórmulas aparece la probabilidad $\pi(\theta)$, es decir, la que corresponde al caso en que no se dispone de ninguna información observacional y, por lo tanto, de carácter totalmente "apriorístico".

Para proceder a la determinación de $\pi(\theta)$ tendremos que utilizar el postulado de Bayes-Laplace, suponiendo que $\pi(\theta)$ es constante, es decir, que corresponde a la distribución uniforme. Otra forma de dar solución a la determinación de $\pi(\theta)$ consiste en proceder por simulación, es decir, aprovechando la información que suministra una experimentación, en un modelo isomorfo.

Supongamos que se trata de estimar la estructura de un fenómeno dicotómico, es decir, que sus posibles sucesos pueden tener el carácter a o el carácter b con probabilidades p y q, siendo $p + q = 1$. Evidentemente, el parámetro estructural será $p = \theta$. Si de n sucesos obtenemos r con el carácter a y, por tanto, n-r con el carácter b, la función de verosimilitud será: $l(x_0, \theta) = \theta^r (1 - \theta)^{n-r}$ y, por tanto,

$$\pi(\theta/x_0) = \frac{\theta^r (1 - \theta)^{n-r} \pi(\theta)}{\int_0^1 \theta^r (1 - \theta)^{n-r} \pi(\theta) d\theta}$$

Si suponemos $\pi(\theta) = k$,

$$\pi(\theta/x_0) = \frac{\theta^r (1 - \theta)^{n-r}}{\int_0^1 \theta^r (1 - \theta)^{n-r} d\theta} = \frac{(n+1)}{r(n-r)} \theta^r (1 - \theta)^{n-r},$$

por tanto

$$\hat{\theta} = \frac{(n+1)}{r(n-r)} \int_0^1 \theta^{r+1} (1 - \theta)^{n-r} d\theta = \frac{(n+1)}{r(n-r)} \frac{(r+1)(n-r)}{(n+2)} = \frac{r+1}{n+2}$$

3.- Nitidez y borrosidad.- En todo lo que llevamos diciendo el problema de la estimación del parámetro estructural depende fundamentalmente de la calidad de las observaciones apreciadas por la contemplación de los elementos observados.

Entenderemos por información nítida aquella que es suministrada por elementos que no ofrecen duda respecto a la pertenencia al conjunto definido por el parámetro estructural. Por el contrario, cuando la observación del elemento en cuestión no ofrece la certeza absoluta de pertenencia, pero tampoco de no pertenencia al conjunto cuya estructura es objeto de estudio y determinación, se dice que es borrosa.

De acuerdo con esto, llamaremos elemento nítido a aquel que pertenece totalmente al conjunto y, por el contrario, elemento borroso al que pertenece parcialmente.

Con el propósito de poder dar un tratamiento cuantitativo a la borrosidad, designaremos por α al coeficiente de pertenencia. El parámetro α estará definido en el intervalo $[0, 1]$. El valor 0 corresponde a la no pertenencia y el 1 a la pertenencia absoluta. Cualquier otro valor del intervalo corresponde a una cierta borrosidad.

Si, por ejemplo, α fuese igual a $1/2$, se diría que el valor observado corresponde a un elemento que pertenece al conjunto en un cincuenta por ciento.

Evidentemente, los α_i dependen de θ , $\alpha_i = f_i(\theta)$.

La probabilidad que corresponde al elemento x_i con un coeficiente de pertenencia α_i , puede considerarse proporcional a α_i , es decir, $P(x_i) = C \alpha_i$.

Si x_1, x_2, \dots, x_n son los elementos del conjunto $\sum_1^n P(x_r) = C (\alpha_1 + \alpha_2 + \dots + \alpha_n)$

luego
$$P(x_i) = \frac{\alpha_i}{\alpha_1 + \alpha_2 + \dots + \alpha_n}$$

La función de verosimilitud correspondiente será

$$\begin{aligned}
 l(x_1, x_2, \dots, x_n, \theta) &= P(x_1, \dots, x_n) = P(x_1)P(x_2)\dots P(x_n) = \\
 &= \frac{\alpha_1 \alpha_2 \dots \alpha_n}{(\alpha_1 + \alpha_2 + \dots + \alpha_n)^n} = \frac{f_1(\theta) f_2(\theta) \dots f_n(\theta)}{(f_1(\theta) + f_2(\theta) + \dots + f_n(\theta))^n}
 \end{aligned}$$

A partir de este resultado, puede obtenerse la expresión

$$\pi(\theta/x_0) = c l(x, \theta) \pi(\theta)$$

IV.- IDENTIFICACION ESTRUCTURAL.

1.- Equivalencia estructural.- Como hemos visto anteriormente, la metodología correspondiente a la estimación de la estructura se basa esencialmente en que los elementos \underline{x} son observables y están relacionados estocásticamente con el parámetro estructural según una función de probabilidad $P_{\theta}(x)$ de la que se deduce la "verosimilitud" $l(x_0, \theta)$ correspondiente a una observación concreta x_0 .

Basándonos en el profundo sentido crítico que la función de verosimilitud tiene respecto a los posibles valores del parámetro estructural θ definiremos como estructuras observacionalmente equivalentes aquellas para las cuales las verosimilitudes correspondientes son iguales.

Representaremos esta definición de la forma $\theta' \sim \theta'' \rightarrow l(x_0, \theta') = l(x_0, \theta'')$.

Esta definición nos permite considerar el espacio estructural Θ como conjunto de estructuras integradas en clases de equivalencia.

2.- Identificación estructural.- Cuando la clase de equivalencia consta de un sólo elemento diremos que la estructura está identificada. En este caso, la equivalencia es una igualdad.

La información observacional es entonces suficiente para la estimación de la estructura.

Estos resultados nos conducen a considerar a la identificación estructural como problema primario y fundamental.

3.- Función identificada.- Una función de la estructura $f(\theta)$ se dice identificada cuando toma iguales valores para estructuras equivalentes. Es decir, $\theta' \sim \theta'' \rightarrow f(\theta') = f(\theta'')$.

En el caso en que esta relación fuese recíproca, es decir, $f(\theta') = f(\theta'') \rightarrow \theta' \sim \theta''$, entonces la función recibe el nombre de identificante y se representará por $I(\theta)$.

La relación entre el concepto genérico de función identificada y el específico de identificante es de singular importancia en la solución del problema de la identificación de la estructura y viene expresada por el siguiente teorema:

Si $f(\theta)$ es función del identifiante, es función identificada.

En efecto, si $f(\theta) = g(I(\theta))$, para $\theta' \sim \theta''$, $g(I(\theta')) = g(I(\theta''))$ ya que, $I(\theta') = I(\theta'')$, luego $f(\theta') = f(\theta'')$. Recíprocamente, si $\theta' \sim \theta'' \rightarrow f(\theta') = f(\theta'')$ existirá una función del identifiante $g(I)$, tal que $f(\theta) = g(I(\theta))$. Si no existiera tal función se podría dar el caso en que $\theta' \sim \theta''$ no llevara consigo que $f(\theta') = f(\theta'')$, en contra del supuesto.

En definitiva, podremos afirmar que la condición necesaria y suficiente para que una función de la estructura sea identificada es que pueda expresarse como función del identifiante.

De todo lo que llevamos diciendo aparecen las siguientes conclusiones de profundo significado, tanto en el orden teórico como en el práctico:

1º.- La función de verosimilitud es identificada y, por ello, función de cualquier identificante.

En efecto, al ser función del parámetro estructural θ , por una parte, y al tomar valores iguales para estructuras equivalentes, es función identificada $\theta' \sim \theta'' \iff l(x, \theta') = l(x, \theta'')$ por lo tanto, $l(x, \theta) = g(I(\theta))$.

2º.- Toda estructura identificada es función del identificante.

En efecto, al ser la estructura identificada, se cumple

$$\theta' \sim \theta'' \rightarrow \theta' = \theta''$$

Este resultado nos indica cómo se puede obtener el subconjunto $\bar{\theta}$ de un modelo θ identificable a partir del conocimiento del identificante.

$$\bar{\theta} = g(I(\theta))$$

3º.- Teorema de Kadane.

Supongamos el espacio estructural θ identificable, cuyo identificante es $i = I(\theta)$.

Sea $\bar{\theta}$ un elemento estructural correspondiente al subconjunto $\bar{\theta}$, el valor correspondiente del identificante será $i = I(\bar{\theta})$. La probabilidad correspondiente $\bar{\theta}$ será:

$$\pi(\bar{\theta}) = \pi(i) \pi(\bar{\theta}/i)$$

y, por lo tanto,

$$\pi(\bar{\theta}/x_0) = \frac{l(x_0, \bar{\theta})}{P(x_0)} \pi(\bar{\theta})$$

Al ser la función de verosimilitud identificada $l(x_0, \theta) = f(i)$ en la que $f(i) = l^*(x_0, i)$. Evidentemente la interpretación de la función $f(i)$ es la de la función de verosimilitud cuando el parámetro estructural es sustituido por el correspondiente valor del identificador $l^*(x_0, i) = P(x_0/i)$.

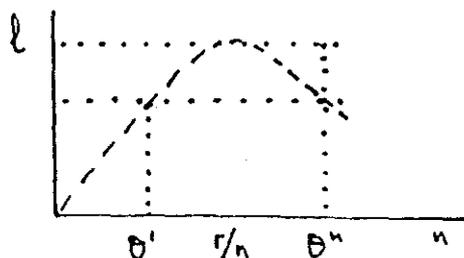
En resumen, tendremos:

$$\begin{aligned} \pi(\bar{\theta}/x_0) &= \frac{P(x_0/i)}{P(x_0)} \pi(i) \pi(\bar{\theta}/i) = \\ &= \frac{P(x_0) \pi(i/x_0)}{P(x_0)} \pi(\bar{\theta}/i) = \pi(\bar{\theta}/i) \pi(i/x_0) \end{aligned}$$

Este resultado constituye en esencia el Teorema de KADANE, que podemos enunciar de la siguiente forma: "La probabilidad "a posteriori" de un parámetro $\bar{\theta}$ perteneciente al conjunto definido por el identificador, es igual al producto de la probabilidad de $\bar{\theta}$ condicionada por el identificador, multiplicada por la probabilidad del identificador condicionada por los valores observados.

V.- APLICACIONES A LOS MODELOS BINOMIAL, LINEAL UNIECUACIONAL Y LINEAL MULTIECUACIONAL.

1.- Modelo binomial.- Como hemos visto anteriormente, al fenómeno dicotómico le corresponde el modelo binomial, siendo su función de verosimilitud $l(x_0, \theta) = \theta^r (1 - \theta)^{n-r}$, cuya representación viene dada por la figura



$$\begin{aligned} l(x_0, \theta') &= l(x_0, \theta'') \\ \updownarrow \\ \theta' &\sim \theta'' \end{aligned}$$

Tomamos $\bar{\theta} = i = \bar{\theta}^r (1 - \bar{\theta})^{n-r}$; según el Teorema de Kadane tendremos:

$$\pi(\theta/x_0) = \pi(\bar{\theta}/i) \pi(i/x_0)$$

Evidentemente,

$$\pi(\bar{\theta}/i) = 1$$

$$\pi(i/x) = \frac{\binom{n+1}{r} \bar{\theta}^r (1 - \bar{\theta})^{n-r}}{\binom{n-r}{r}}$$

luego
$$\hat{\theta} = \frac{\binom{n+1}{r} \int_0^1 \bar{\theta}^{r+1} (1 - \theta)^{n-r} = \frac{r+1}{n+2}}$$

Resultado ya conocido.

2.- Modelo lineal uniecuacional.

Supongamos la relación: $ax_i + b = \varepsilon_i$; $i = (1, 2, \dots, n)$, en la que a y b son parámetros, x_i valores observables de x y ε_i variantes que dan categoría estocástica del modelo.

Vamos, de acuerdo con lo dicho anteriormente, a obtener la función de verosimilitud que corresponde a las variables observables.

Suponemos que la función de probabilidad conjunta correspondiente a las variantes ε_i se representa por $f(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)$. La función de verosimilitud vendrá expresada de la forma:

$$\begin{aligned} L(x_1, x_2, \dots, x_n, \theta) &= \prod (ax_i + b, ax_i + b, \dots, ax_n + b) \left| \frac{\theta(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)}{\theta(x_1, x_2, \dots, x_n)} \right| = \\ &= f(ax_1 + b, ax_2 + b, \dots, ax_n + b) |a|^n \end{aligned}$$

Supongamos el modelo que resulta de multiplicar por la constante k:

$$Kax_i + Kb = K \varepsilon_i = \eta_i; i = (1, \dots, n)$$

La función de probabilidad correspondiente a las η_i cumple la relación

$$P(\eta_1, \eta_2, \dots, \eta_n) = f(\eta_1/k, \dots, \eta_n/k) |1/k|^n,$$

ya que

$$\frac{\theta(\varepsilon_1, \dots, \varepsilon_n)}{\theta(\eta_1, \dots, \eta_n)} = (1/k)^n.$$

La función de verosimilitud será

$$\begin{aligned} L'(x_1, \dots, x_n, \theta') &= \varphi(Kax_1 + Kb, \dots, Kax_n + Kb) |ka|^n = \\ &= f(a x_1 + b, \dots, a x_n + b) |(1/k)|^n |k|^n |a|^n = \\ &= L(x_1, x_2, \dots, x_n, \theta). \end{aligned}$$

Este resultado nos conduce a la conclusión de que dos estructuras θ y θ' son equivalentes observacionalmente, es decir, que el producto de los parámetros estructurales a, b, por una constante $k \neq 0$ conduce a una estructura observacionalmente equivalente.

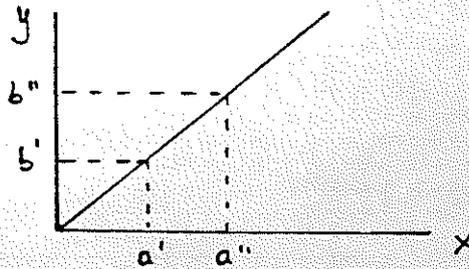
Para resolver el problema de la identificación estructural, pretendemos la determinación de un identificador.

Consideremos dos estructuras $\theta = (a, b)$ y $\theta' = (a', b') \equiv (Ka, Kb)$ siendo $K \neq 0$ que, como hemos visto, son observacionalmente equivalentes, y definiremos la función: $f(\theta) = b/a$. Si $(ab) \sim (a'b') \rightarrow b/a = b'/a'$. Recíprocamente:

$$\begin{aligned} b/a = b'/a' &\rightarrow b'/b = a'/a \rightarrow a' = Ka \rightarrow \theta \sim \theta' \\ &b' = Kb \end{aligned}$$

Estos resultados nos indican que se trata de un identificador $I(\theta) = \frac{b}{a}$.

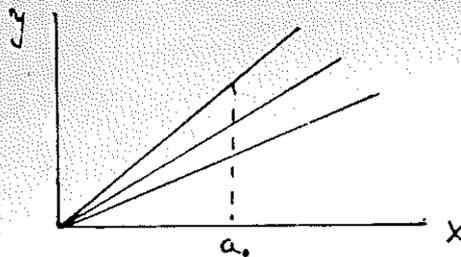
Si representamos cartesianamente Θ (ab) veremos que las rectas que pasan por el origen son el lugar de estructuras equivalentes observacionalmente



$$\frac{b'}{a'} = \frac{b''}{a''} = I(\theta) = b$$

$$y = ix$$

Es claro que los puntos de la recta $x=a_0$, perpendicular al eje x , son función del identificante.



Elegiremos como modelo θ identificado $\tilde{\theta} = i = b/a_0$.

Según el teorema de Kadane

$$\pi(\tilde{\theta}/x_0) = \pi(\tilde{\theta}/i) \pi(i/x_0)$$

$$\pi(\tilde{\theta}/i) = 1$$

$$\pi(i/x_0) = L(x_1, x_2, \dots, x_n, \tilde{\theta}) = f(a_0(x_1+i), \dots, a_0(x_n+i)) |a_0|^n$$

Supongamos que E_1, E_2, \dots, E_n son estocásticamente independientes y normales $N(0, 1)$ es decir, $\frac{1}{\sqrt{2\pi}} e^{-E_i^2/2}$

$$L(x_1, \dots, x_n, \tilde{\theta}) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \cdot e^{-\frac{1}{2} \sum_{r=1}^n (a_0 x_r + b_0)^2 |a_0|^{-n}}$$

$$= \left(\frac{|a_0|}{\sqrt{2\pi}}\right)^n e^{-\frac{a_0^2}{2} \sum_{r=1}^n (x_r + i)^2}$$

Ahora bien, si \bar{x} es la media $\sum x_r/n$, tendremos:

$$\begin{aligned} \sum_{r=1}^n (x_r - \bar{x} + \bar{x} + i)^2 &= \sum_{r=1}^n (x_r - \bar{x})^2 + n(\bar{x} + i)^2 + 2(\bar{x} + i) \sum (x_i - \bar{x}) = \\ &= \sum (x_r - \bar{x})^2 + n(\bar{x} + i)^2 \end{aligned}$$

ya que: $\sum (x_i - \bar{x}) = 0$

En definitiva, tendremos

$$L(x_1, \dots, x_n, \bar{\theta}) = k e^{-a_0^2/2 n (\bar{x} + i)^2}$$

Se trata de una distribución normal de i con \bar{x} como valor medio.

De aquí se deduce que la estimación del modelo estructural viene dada por: $a = a_0$
 $b = -a_0 \bar{x}$.

3.- Estimación del modelo multiecuacional.

Consideremos el modelo multiecuacional

$$BY_t + \Gamma Z_t = \epsilon_t$$

en el que $Y_t = \begin{pmatrix} Y_{1t} \\ Y_{2t} \\ \vdots \\ Y_{mt} \end{pmatrix}$, $t = 1, 2, \dots, T$

es el vector de variables endógenas;

$$Z_t = \begin{pmatrix} Z_{1t} \\ Z_{2t} \\ \vdots \\ Z_{nt} \end{pmatrix}, \quad t = 1, 2, \dots, T$$

es el vector cuyas componentes son variables exógenas,

$$\epsilon_t = \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \\ \vdots \\ \epsilon_{mt} \end{pmatrix}, \quad t = 1, 2, \dots, T$$

es el vector cuyas componentes son variables aleatorias. Finalmente B y Γ son matrices (m.m) y Γ (m.n), respectivamente, cuyos elementos son parámetros (B no singular).

Efectuando la transformación

$$B^{-1} \cdot \varepsilon_t = \eta_t$$

el modelo queda expresado en su forma reducida, es decir,

$$Y_t = A \cdot Z_t + \eta_t ; \quad A = -B^{-1} \cdot \Gamma$$

Vamos a hacer la hipótesis de que las variables aleatorias para cada valor de t tienen una distribución conjunta normal multivariante con media 0 y matriz de covarianzas

$$\Sigma = E[\varepsilon_t \cdot \varepsilon_t'] = E \left[\begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \vdots \\ \varepsilon_{mt} \end{pmatrix} \cdot (\varepsilon_{1t} \ \varepsilon_{2t} \ \dots \ \varepsilon_{mt}) \right] = \begin{pmatrix} \lambda_{11} & \lambda_{12} & \dots & \lambda_{1m} \\ \lambda_{21} & \lambda_{22} & \dots & \lambda_{2m} \\ \dots & \dots & \dots & \dots \\ \lambda_{m1} & \lambda_{m2} & \dots & \lambda_{mm} \end{pmatrix}$$

En consecuencia, la función de densidad conjunta para cada valor de t será:

$$\phi(\varepsilon_t) = \left(\frac{1}{\sqrt{2\pi}} \right)^m |\Sigma|^{-\frac{1}{2}} \cdot e^{-\varepsilon_t' \cdot \Sigma^{-1} \cdot \varepsilon_t}$$

Además, supondremos que la distribución conjunta de $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_T$ es también normal multivariante con valor medio igual a 0, la matriz de covarianzas (Σ) es independiente de t , y no existe autocorrelación ni correlación serial, por lo que se verifica

$$E[\varepsilon_1 \varepsilon_1'] = \dots = E[\varepsilon_T \varepsilon_T'] = \Sigma, \quad \text{y} \quad E[\varepsilon_{it} \varepsilon_{j+h}] = 0 ; \quad \forall i \neq j, \quad \forall h \neq 0$$

Por todo lo anterior, la Matriz de Covarianzas de $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_T$ será de la forma:

$$\Sigma^0 = E(\varepsilon \varepsilon') = E \begin{pmatrix} \varepsilon_1 \varepsilon_1' & \varepsilon_1 \varepsilon_2' & \dots & \varepsilon_1 \varepsilon_T' \\ \dots & \dots & \dots & \dots \\ \varepsilon_T \varepsilon_1' & \varepsilon_T \varepsilon_2' & \dots & \varepsilon_T \varepsilon_T' \end{pmatrix} = \begin{pmatrix} \Sigma & 0 & \dots & 0 \\ 0 & \Sigma & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \Sigma \end{pmatrix}$$

La distribución conjunta de los vectores aleatorios $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_T$ será pues:

$$N[\varepsilon/0; \Sigma^0]$$

y su función de densidad:

$$f_{\varepsilon}(\varepsilon) = \left(\frac{1}{\sqrt{2\pi}} \right)^T (|\dot{\Sigma}|)^{-\frac{1}{2}} e^{-\frac{1}{2} \varepsilon' \dot{\Sigma}^{-1} \varepsilon} = \left(\frac{1}{\sqrt{2\pi}} \right)^T |\Sigma|^{-\frac{T}{2}} e^{-\frac{1}{2} \varepsilon' \Sigma^{-1} \varepsilon}$$

Para calcular la función de verosimilitud correspondiente al modelo reducido

$$Y_t = A Z_t + \eta_t, \quad \text{por ser} \quad E(\varepsilon_t) = 0 \quad \text{y} \quad \eta_t = B^{-1} \cdot \varepsilon_t,$$

resulta $E(\eta_t) = 0$, $\forall t = 1, 2, \dots, T$.

Análogamente, la Matriz de covarianzas de η_t será:

$$\Omega = E(\eta_t, \eta_t') = B^{-1} E(\varepsilon_t, \varepsilon_t') B^{-1} = B^{-1} \cdot \Sigma \cdot B^{-1}$$

y la de $\eta_1, \eta_2, \dots, \eta_T$

$$\dot{\Omega} = E(\eta, \eta') = E \left\{ \begin{pmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_T \end{pmatrix} (\eta_1' \eta_2' \dots \eta_T') \right\} = \begin{pmatrix} \Omega & 0 & 0 & \dots & 0 \\ 0 & \Omega & \dots & & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & & \Omega \end{pmatrix}$$

Lo que supone afirmar que los vectores aleatorios $\eta_1, \eta_2, \dots, \eta_T$ tienen una distribución conjunta normal multivariante con media 0 y matriz de covarianzas $\dot{\Omega}$.

Si ahora efectuamos la sustitución $\eta_t = Y_t - A Z_t$, llegamos a la conclusión que la distribución condicionada $P(Y/A, Z, \Omega) \equiv N[A \cdot Z; \dot{\Omega}]$, es decir, normal multivariante con media $A \cdot Z$ y matriz de covarianzas $\dot{\Omega}$.

Ahora bien, considerando a Ω como elemento estructural, cada estructura vendrá dada, en general, por los parámetros estructurales, es decir, los elementos de las matrices B y Γ y por la matriz de dispersión Ω . Es decir, $\theta = (B, \Gamma, \Omega)$.

Por tanto, $P(Y/A, Z, \Omega)$ es la función de verosimilitud de la forma reducida del modelo.

Entonces, dadas dos estructuras θ' y θ'' tales que

$$\left. \begin{aligned} \theta' &= (B, \Gamma, \Omega) \\ \theta'' &= (\bar{B}, \bar{\Gamma}, \bar{\Omega}) \end{aligned} \right\} \theta' \neq \theta''$$

se verificará para las funciones de verosimilitud correspondientes:

$$l(Y/\theta') \neq l(Y/\theta''),$$

pues al ser $\bar{A} \neq A$ y $\bar{\Omega} \neq \Omega$ se obtendrán dos funciones de densidad normales con distinto valor medio y distintas matrices de dispersión.

De aquí obtenemos la conclusión de que "si dos estructuras θ' y θ'' son observacionalmente equivalentes, han de ser iguales", ya que, si no lo fueran, tendrían distinta función de verosimilitud.

Por otra parte, como ya sabemos, si las clases de equivalencia están formadas por un solo elemento, la estructura θ' es identificable y ha de ser función del identificante, por lo que los elementos que integran la estructura podrán ser determinados una vez conocido el identificante.

En efecto, en primer lugar el identificante será la siguiente función de θ' .

$$l(\theta') = l(B, \Gamma, \Omega) = (-B^{-1} \cdot \Gamma, B^{-1} \cdot \Sigma \cdot B^{-1})$$

El problema consiste en determinar la estructura $\theta' = (B, \Gamma, \Sigma)$ conocidos A y Ω . Sin embargo, el sistema $A = -B^{-1} \cdot \Gamma$ tiene mayor número de incógnitas ($m \cdot m + m \cdot n$) que de ecuaciones ($m \cdot n$) por lo que será necesario introducir restricciones a priori, por ejemplo que $m \cdot m$ parámetros estructurales sean nulos, a fin de evitar que el sistema sea indeterminado.

Una vez obtenido B , la determinación de Σ es inmediata a partir de la relación $B \cdot \Omega \cdot B' = \Sigma$.

En cuanto a la estimación del modelo multiecuacional, como sabemos, lo fundamental es el conocimiento del identificante ya que de él se desprende la determinación de la estructura. Así pues, dado el modelo

$$B Y_t + \Gamma Z_t = \varepsilon_t, \quad t = 1, 2, \dots, T$$

el problema de su estimación se apoyará en la estimación del identificante, es decir, del modelo reducido.

La forma reducida es:

$$\begin{aligned} Y &= A \cdot Z + \eta \\ Y' &= Z' \cdot A' + \eta' \\ \eta &= B^{-1} \cdot \varepsilon \end{aligned}$$

en la que:

$$Y = (Y_1, Y_2, \dots, Y_T) \quad \text{e} \quad Y_t = \begin{pmatrix} Y_{1t} \\ Y_{2t} \\ \vdots \\ Y_{mt} \end{pmatrix}, \quad t = 1, 2, \dots, T$$

$$Z = (Z_1, Z_2, \dots, Z_T) \quad \text{y} \quad Z_t = \begin{pmatrix} z_{1t} \\ \vdots \\ z_{nt} \end{pmatrix}, \quad t = 1, 2, \dots, T$$

$$\varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_T) \quad \text{y} \quad \varepsilon_t = \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \vdots \\ \varepsilon_{nt} \end{pmatrix}, \quad t = 1, 2, \dots, T$$

Si expresamos las filas de Y , A y η por $Y_{(r)}$, $A_{(r)}$, $\eta_{(r)}$; ($r = 1, 2, \dots, m$)

tendremos: $Y'_{(r)} = Z' A'_{(r)} + \eta'_{(r)}$.

Si ahora definimos los vectores:

$$Y^{\circ} = \begin{pmatrix} Y'_{(1)} \\ Y'_{(2)} \\ \vdots \\ Y'_{(m)} \end{pmatrix}, \quad A^{\circ} = \begin{pmatrix} A'_{(1)} \\ A'_{(2)} \\ \vdots \\ A'_{(m)} \end{pmatrix}, \quad \eta^{\circ} = \begin{pmatrix} \eta'_{(1)} \\ \eta'_{(2)} \\ \vdots \\ \eta'_{(m)} \end{pmatrix}$$



la expresión $Y' = Z' \cdot A' + \eta'$ queda

$$Y^{\circ} = \begin{pmatrix} Y'(1) \\ Y'(2) \\ \vdots \\ Y'(m) \end{pmatrix} = \begin{pmatrix} Z' A'(1) + \eta'(1) \\ Z' A'(2) + \eta'(2) \\ \dots \\ Z' A'(m) + \eta'(m) \end{pmatrix}$$

Expresión equivalente a

$$Y^{\circ} = \begin{pmatrix} z' & \phi & \phi & \dots & \phi \\ \phi & z' & \phi & \dots & \phi \\ \dots & \dots & \dots & \dots & \dots \\ \phi & \phi & \phi & \dots & z' \end{pmatrix} \cdot A^{\circ} + \eta^{\circ}$$

que puede ponerse de la forma $Y^{\circ} = M \cdot A^{\circ} + \eta^{\circ}$, siendo

$$M = \begin{pmatrix} z' & \phi & \phi & \dots & \phi \\ \phi & z' & \phi & \dots & \phi \\ \dots & \dots & \dots & \dots & \dots \\ \phi & \phi & \phi & \dots & z' \end{pmatrix}$$

Se trata, por tanto, de estimar los parámetros

$$A^{\circ} = \begin{pmatrix} a_{11} \\ a_{12} \\ \vdots \\ a_{1n} \\ a_{21} \\ \vdots \\ a_{2n} \\ \vdots \\ a_{m1} \\ \vdots \\ a_{mn} \end{pmatrix}$$

a partir del conocimiento de Y° y Z .