INFERENCIA EN PROCESOS (J-X)

Visto bueno, por el Director

Granada, 21 de Abril de 1977

MEMORIA que, para optar al grado de Doctor, presenta la licenciado en Ciencias Matemáticas:

Dña. MARIA TERESA MIRANDA LEON

Director de TESIS: Profesor Dr. D. RAMON GUTIERREZ JAIMEZ

UNIVERSIDAD DE GRANADA. FACULTAD DE CIENCIAS. ABRIL 1977.

A G R A D E C I M I E N T O

Quiero hacer constar mi agradecimiento al profesor Dr.D. Ramón Gutierrez Jaimez por su valiosa y constante ayuda durante el tiempo de elaboración de este trabajo ,sin la cual no lo hubiera llevado a término,asi como al Director del Dpto. de Estadistica ,Pr.Dr.D. Alfonso Guiraúm ,por su magisterio durante los años de mi carrera.

A mis padres.

INDICE GENERAL

CAPITULO III . Inferencia en Procesos (J-X).Contrastes de hipótesis.

INTRODUCCION GENERAL

Esta memoria tiene por objeto el estudio de los procesos (J-X) bidimensionales, introducidos por Ronald Pyke en 1961 con el fin de tratar bajo este enfoque a los procesos semimarkovianos y de renovación markovianos y continuada después su investigación por Jacques Janssen (1969) que los aplica a la Teoria de recorridos aleatorios sobre la recta real.

Nuestro propósito es la construcción de Inferencias sobre estos procesos ,aplicando la Teoria de Anderson y Goodman , que en 1957 publicaron sus trabajos sobre Inferencia en Cadenas de Markov.

Primeramente ,se dá la definición de un proceso (J-X) como un proceso bidimensional de tripleta $(m,\bar{p},Q)$,definido sobre un espacio probabilistico (Ω ,$\mathcal{A}$,P) y con valores en IxR donde I representa el conjunto de los naturales excluido el 0. ó bien el conjunto I= (1,...,m) ,m ∞ .y R es la recta real. Se da la serie de condiciones que deben cumplir dichos procesos y que nos conduce a clasificarlos en varios tipos,siendo los procesos (J-X) positivos la base de las definiciones de los procesos markovianos de renovación y de los procesos semimarkovianos.Tiene gran importancia la definición de la variable $S_n=\sum_{k=o}^{n} X_k$, $n \in N$,llegando a la comprobación de que el proceso ; ((J_n,S_n); $n \in N$) es un proceso de Markov y (J_n; $n \in N$) una cadena de Markov ,llamada cadena inmersa en el proceso (J-X).

Pasamos despues a considerar las funciones $f: I \times I \times R \rightarrow R$,tales que $\forall\ i,j \in I$, $f(\ i,j,.)$ sea medible Lebesgue y tras establecer las medias condicionales e incondicionales ,se introducen las sumas $W_f(n)$ de tales funciones , llegando a propiedades de tales sucesiones de sumas ,establecidas en las proposiciones 2A y 2B ,llegando como consecuencia de estas a una generalización de la igualdad de Wald establecida para variables $(X_n; n \in N)$, independientes ,integrables y equidistribuidas.

Si consideramos las partes positiva y negativa del proceso $(X_n, n \in N)$,los procesos $(\ (J_n, X_n^+\ ), n \in N)$, $((J_n, X_n^-\ ), n \in N)$ asi como $((J_n, |X_n|)\ ; n \in N)$,son procesos (J-X) positivos , y se demuestra que todo proceso (J-X) se descompone en dos procesos (J-X) positivos ,donde para $\forall n \in N$: $X_n = X_n^+ - X_n^-$.

Uno de los apartados más importantes de este capitulo es aquel en el que hacemos la clasificación y estudio del proceso según su cadena inmersa $(\ J_n;\ n \in N)$,estudiando los casos siguientes:

a) Cadena irreducible recurrente positiva.

b) Cadena irreducible recurrente nula

c) Cadena aperiódica recurrente nula

Dentro del apartado a) y partiendo de que el proceso está determinado por la cuadrupleta $(\ m, \tilde{\pi}, P, F)$ donde $\tilde{\pi}$ es la distribución estacionaria única,que tomamos como distribución inicial , se introduce el proceso dual $(\ (\ \hat{J}_n, \hat{X}_n);\ 0 \leq n \leq N\ )$,determinado por la cuadrupleta $(\ m, \tilde{\pi}, \hat{P}\ , \hat{F}\ )$,donde $\hat{P}$ y $\hat{F}$ vienen dadas por la definición 4A.

En el caso b), al no existir distribución inicial apropiada , nos restringimos a clases de procesos con la misma tripleta: (m,P,F) ,llegando al establecimiento de clases duales y auto-duales.

Toda la teoria que se desarrolla a partir de aqui va encaminada al establecimiento de las Leyes de los Grandes Números y T.C.L. para estos procesos.

A partir de la cadena inmersa ,definimos para $\forall\ j \in I$,el proceso ($r_n^{(j)}$;$n \in N_o$) de los <u>indices de recurrencia</u>(Ver Chung (2)) .En el caso de cadena irreducible recurrente ,el proceso para un j fijo ($u_s^{(j)}$; $s \in N_o$) ,donde $u_s^{(j)}$ viene dado por (5.1) ,se puede considerar como un proceso de renovación en sentido amplio demostrando en la proposición 5A que $E(u_1) = \sum_{i \in I} \Pi_i \xi_i$. En el caso en que f(i,j,x)= x (recorrido aleatorio,u_1 representará la posición de la particula cuando se produce el 1$^{\underline{er}}$ retorno a j, dado que se parta del mismo j. y en este caso se verifica que $\mu_{jj} = E(u_1) =\sum_i \Pi_i \eta_i$, y que el cociente μ_{jj}/ m_{jj} es independiente de j y vale $E(X_n)$

A partir de aqui, bajo ciertas condiciones de convergencia ,se demuestra la existencia de las esperanzas $E(u_1^{(k)})$ $\forall$ k , y en particular que $E(u_1^2)$ existe y es $< \infty$,viniendo su expresión dada en funcion de los tiempos medios de primer paso por (5.7) (pág 26).

Una vez establecida la existencia y finitud de los momentos ,se deduce un caso particular de Ley Fuerte de los G.N. haciendo intervenir los momentos de orden 2 y seguidamente si se adopta el método de descomposición de Chung (ver pág 30) para cadenas tomado por Pyke(17) para los procesos (J-X) positivos se llega a una Ley Fuerte sin hipótesis sobre los momentos de orden dos, dada por la Proposición 6C.(pág 33).

Análogamente, se pueden traspasar al caso de procesos (J-X), otros Teoremas válidos para procesos (J-X) positivos.

Tras las Leyes Fuertes y basandonos nuevamente en el métodoxde descomposición de Chung, establecemos un Teorema Central del Limite para las sumas $W_f(N)$ de funciones de los procesos (J-X), cuando la cadena inmersa es irreducible recurrente positiva y var $u_1(g) < \infty$, donde $g = f - m_j / m_{jj}$, llegandose al resultado establecido en la Proposición 8A, pág 39., obteniendo el caso particular correspondiente para recorridos aleatorios. De esta proposicion 8A se deduce con facilidad un Teorema Central del Limite para el proceso bidimensional $(J_n, W_f(N))$ deducción en la que se utiliza el paso al proceso dual, ya que $W_f(N) = \hat{W}_{\hat{f}}(N)$, y cuyo resultado viene dado por la Proposicion 8B.

Como final del Capitulo 1º se realiza una aplicación de estos procesos a la práctica, en particular a la Programación Dinámica, donde X_n representa el rendimiento de la n-ésima etapa y S_n el rendimiento total en n etapas, adoptando una politica estacionaria. Howard(10) demuestra que la ganancia total esperada por etapa $E(S_n)/n$ tiende en el caso irreducible, recurrente positivo yaperiódico con $m < \infty$, hacia $g = \sum_i \tilde{\Pi}_i \eta_i$

En las aplicaciones prácticas de estos procesos,nos-tropezamos generalmente con el desconocimiento de las probabilidades de transición de la cadena de Markov inmersa y por tanto con la necesidad de obtener estimaciones de estas lo más correctas posible.

A este fin se dedican los Capitulos II y III de esta memoria basados fundamentalmente en los trabjos de Anderson y Goodman (1957) sobre Inferencia en Cadenas de Markov,dedicando el 2º capitulo a obtener estas estimaciones y el 3º a establecer contrastes de hipótesis acerca de ellas.

Supongamos que la cadena de Markov inmersa es de un nº finito de estados i : $i=1,2,..,m$.y sean los instantes de observación $t(t= 0,1,..,T)$; $p_{ij}(t)$ será la probabilidad de que la cadena esté en el estado j en el instante t ,dado que está en el i en el instante t-1, y es obvio que si las probabilidades de transición son estacionarias : $p_{ij}(t) = p_{ij}$.$t=1,...,T$ como es nuestro caso , aunque se obtendrán las estimaciones , simultáneamente para uno y otro caso .

Para llegar a estas , supongamos que existen $n_i(0)$ individuos en el estado i en el instante cero.

Una observación sobre un individuo dado consiste en la sucesión de estados que ocupa en los instantes 0,1,..,T, que llamamos $i(0),i(1),..,i(T)$,es decir,dado el estado inicial,su trayectoria tendrá m^T posibilidades.

En los estados intermedios ,llamamos $n_{ij}(t)$ al nº de individuos que pasan del estado i,al j en el instante t ,siendo evidentemente: $n_i(t-1)=\sum_{j=1}^{m} n_{ij}(t)$. De esta forma llegamos a que las variables $n_{ij}(t)$ son V.A. multinomiales con distribución(condicional en las $n_i(0)$):

$$\left(\prod_{t=1}^{T} \left(\prod_{i=1}^{m} \left(\frac{n_i(t-1)}{\prod_{j=1}^{m} n_{ij}(t)} \prod_{j=1}^{m} p_{ij}(t)^{n_{ij}(t)} \right)\right)\right)$$

Se obtiene que,para probabilidades de transición no estacionarias $p_{ij}(t)$,los $n_{ij}(t)$ forman un conjunto de estadisticos suficientes y en el caso de probabilidades estacionarias,las $n_{ij}=\sum_{t=1}^{T} n_{ij}(t)$,son las que forman un conjunto de estadisticos suficientes.

Pasamos enseguida a hallar los estimadores de máxima verosimilitud de las $p_{ij}(t)$ y p_{ij} ,en función de las $n_{ij}(t)$,siendo:

$$\hat{p}_{ij}(t)=\frac{n_{ij}(t)}{n_i(t-1)} \quad \text{y} \quad \hat{p}_{ij}=\frac{n_{ij}}{n_i^*}=\frac{\sum_{t=1}^{T} n_{ij}(t)}{\sum_{t=0}^{T-1} n_i(t)}$$

Con el fin de saber el comportamiento asintótico de las $\hat{p}_{ij}$,hay que conocer antes el de las $n_{ij}(t)$.

Consideramos que $\dfrac{n_k(0)}{\sum_j n_j(0)} \longrightarrow \eta_k \; (\eta_k > 0 \text{ y } \sum \eta_k = 1)$, cuando $\sum_j n_j(0) \longrightarrow \infty$.

Para cada estado inicial i(0) ,el conjunto de las $n_{i(0)\,i(1)..i(T)}$, son V.A. multinomiales y cuando la amplitud muestral crece se distribuyen asintóticamente segun una normal y por tanto ,las $n_{ij}(t)$,que son C.L. de estas ,están distribui-

das tambien asintóticamente según normales , calculandose sus medias, varianzas y covarianzas.
Con esta base, se comprueba, que cuando n —— ∞ ,las variables $(n)^{1/2}$ ($\hat{p}_{ij}$ - p_{ij}) siguen, en el limite ,la distribución normal, hallándose con facilidad los parámetros que la caracterizan.
La cadena inmersa de los procesos (J-X) es de probabilidades de transición estacionarias. Por tanto, se adoptarán las estimaciones propias para tal caso.

Refiriendonos al apartado 4º del Cap.I. podremos calcular la distribución estacionaria ,con la base de estas estimaciones, pudiendo pasar después ,en funcion de ellas al estudio del proceso dual.
Posteriormente pasamos a estudiar ,tomando como base la teoria desarrollada por Chia Kuei Tsao(**11** ;pág 84)(1968), funciones , probabilisticas de la cadena tratando con una sucesión especial de ellas, que son vectores aleatorios multinomiales i.i.d. bajo ciertas hipótesis acerca de sus probabilidades de transición .

En el último apartado de este capitulo ,estudiamos siguiendo los trabajos de Do Sun Bai(1975)(Ver 15,pág 85)sobre estimación de funciones de P, la eficiencia de dichas estimaciones. En el caso en que las N observaciones de las que partimos $(J_1, J_2, .., J_N)$, de la cadena inmersa del proceso (J-X), sean en nº fijo, se sabe que no existen estimadores insesgados de funciones de P. Por tanto,para que dichos estimadores insesgados existan N debe ser una variable aleatoria y todo esquema relativo a es - timaciones insesgadas será un esquema de estimación secuencial .

Como se quiere estudiar la eficiencia de estas estimaciones de funciones de P, y sabemos que un estimador es eficien te cuando su varianza alcanza la cota de FRECHET-CRAMER-RAO,el problema consiste en caracterizar tripletas eficientes (S,f,g) para la cadena de Markov de m estados inmersa en el proceso , donde S es el plan de muestreo ,$f = f(J_1, J_2, \ldots, J_N)$ el estimador y $g(P)=E(f)$ el valor esperado,estableciendo despues la desigualdad de Información para dicha cadena(Ver 3.2. Cap II)

Tratamos despues con funciones especificas de P,tales como funciones de una fila de P y funciones de dos ó más filas de P,hallando tripletas eficientes,en el primer caso y compro - bando que en el 2º caso no existen tripletas eficientes,a menos que la cadena tenga sólo dos estados.(m=2).

Como una continuación natural de la Teoria de Estima - ción,el Capitulo 3º se dedica a estudiar algunos métodos de con traste de hipótesis ,aplicandolos a los casos particulares que se nos presentan.Para ello nos basamos ,al igual que en el Capitulo anterior ,en la Teoria de Anderson y Goodman(1)(1957),so - bre Inferencia en Cadenas de Markov y en los trabajos de Chia Kuei Tsao (4) (1968) ,relativos a funciones de la Cadena. En la 1ª parte ,contrastamos primeramente la hipótesis de que las probabilidades de transición tienen valores especificos $p_{ij}^{(0)}$ es decir H_o: $p_{ij}=p_{ij}^{(o)}$; $j=1,2,..,m$. y para un i dado.

Para ello nos basamos en la distribución asintotica de los estimadores ,y en el hecho de que:

$$\sum_{j=1}^{m} n_i^{*} \frac{(\hat{p}_{ij} \quad p_{ij}^{(o)})^2}{p_{ij}^{(o)}} \longrightarrow \chi^2_{m-1}$$

grados de libertad.A partir de esto ,se hallará con facilidad la región critica, para un nivel α dado.

Aunque la cadena inmersa en el proceso (J-X) es de probabilidades de transicion estacionarias,se puede contrastar tambien la hipótesis de que dichas probabilidades son ctes. contra la alternativa H_1:$p_{ij}=p_{ij}(t)$.Para este problema se utiliza el test de razón de verosimilitudes ,hallando el cociente :

$$\lambda = \prod_{t}\prod_{i,j}\left(\frac{\hat{p}_{ij}}{\hat{p}_{ij}(t)}\right)^{n_{ij}(t)}$$

y como ya es conocido que $-2 \log \lambda \longrightarrow \chi^2_{(T-1)m(m-1)}$,cuando H_o es cierta,esto nos conduce a establecer una semejanza con los tests standard de homogeneidad utilizados en las tablas de contingencia y a establecer hipótesis equivalentes bajo este punto de vista.Usamos tambien el citado test de verosimilitudes para contrastar la hipótesis de homogeneidad para T muestras independientes de pruebas multinomiales.Para ello se calculan los correspondientes λ_i, distribuyendose $-2\log \lambda_i$ según una χ^2_i,con : $(m-1)(T-1)$ g.l. Deducimos finalmente que las χ^2_i son asintóticamente independientes y que la suma: $\chi^2 = \sum_{i=1}^{m} \chi^2_i$ es una χ^2 con $m(m-1)(T-1)$ g.l.

Una alternativa al modelo ,que se ha estudiado hasta ahora ,en que las $n_i(0)$ son fijas,se presenta considerando que sean V.A. multinomiales.A este caso dedicamos el apartado 3º de este Capitulo.(ver pág 93)

Despues de analizar los tests más importantes concernientes a las p_{ij} y como continuación de la 2ª parte del Cap. II,se estudia la <u>admisibilidad de las funciones probabilisticas de la Cadena inmersa en el proceso (J-X)</u>,como tests estadisticos acerca de la matriz de transición P.

Llamamos S_{hT} a la suma $Y_1+..+Y_T$ de funciones de la cadena,siguiendo la S_{hT} una distribución multinomial ,(binomial si h=2) según el Teorema 2(pág 63) del 2º capitulo.

Tratamos en este último apartado sólo este caso,partiendo de r_o y r_1 ,tales que $0<r_o,r_1<1$ y de 2m distribuciones , conocidas o desconocidas $\alpha^{(o)}_{ij}=(\overset{o}{\alpha}_{ij1},\overset{o}{\alpha}_{ij2},..,\overset{o}{\alpha}_{ijm_{ij}})$ para i=1,..,m y j=1,2.

Se trata entonces de contrastar la hipótesis nula H_o:$p=r_o$; $\alpha_{ij}=\overset{o}{\alpha}_{ij}$, con cada una de las alternativas:

H_1:$p=r_1$;H_2:$p>r_o$;H_3: $p<r_o$ y H_4:$p\neq r_o$,con $\alpha_{ij}=\overset{o}{\alpha}_{ij}$

Se obtienen ,con esta base 4 tests optimos para cada una de las cuatro hipótesis,siendo P_o, P_1, P_2, P_3 y P_4 las matrices de transición bajo cada una de las hipótesis expuestas.Se comprueba , posteriormente,que los test obtenidos son los U.M.P. para contrastar las hipótesis citadas(Ver Teorema 1 y Corolarios 1 y 2 págs. 97,98 y 99)

Se obtienen como consecuencias:

1ª) Si los α^{o}_{ij} se consideran como parámetros perturbadores los test φ_1, φ_2, y φ_3 son U.M.P. e insesgados.

2ª) Ya que las funciones de potencia son todas binomiales con parámetros (T,p) ,todos los tests son uniformemente consistentes.

En todos estos planteamientos se supone fija la distribución inicial P_o, y se puede contrastar H_o,contra hipótesis alternativas más generales ,suavizando las condiciones impuestas a las distribuciones α^{o}_{ij} y a P_o.

Finalmente,se estudian algunos modelos de aplicación pràctica ,considerando a S_{2T} como generalización de algunos estadisticos bien conocidos.

Los modelos estudiados son:

I) Modelo del " test de los signos "

II)Modelo del"nº total de rachas"

III)Modelo de " Recorrido aleatorio ciclico"

IV) Modelo de " Tendencia mixta " , este ultimo con aplicación a los problemas de movilidad social.

CAPITULO I

LOS PROCESOS (J - X)

CAPITULO I

ESTADO ACTUAL DE LA TEORIA

Vamos a introducir una clase de procesos estocásticos bidimensionales ,cuyo estudio fué iniciado por R.Pyke en 1961 , con motivo de sus trabajos sobre procesos semimarkovianos y de renovación markovianos, pero sin restringir las V.A. a ser c.s. positivas.

Esta generalización , que ya ha sido considerada por varios autores,como: H.D.Miller (1962) ,Keilson(1964),Whisart (1964), Hatori y Mori(1966) , es muy interesante , ya que amplia el concepto de recorrido aleatorio sobre la recta real, al no suponer que los pasos sucesivos son V.A. independientes y equidistribuidas, ya que se introduce una cierta dependencia entre los diferentes pasos.

Dada la importancia del concepto de recorrido aleatorio por sus aplicaciones en Teoria de Colas ,Teoria colectiva del Riesgo etc.. ,vemos en estos procesos un campo de investigación fructifero a la vez ,desde el punto de vista teorico,como práctico.

Además estos procesos se encuentran en la Teoria de Procesos de Decisión secuenciales y los resultados que se ob-

C A P I T U L O I

ESTADO ACTUAL DE LA TEORIA

Vamos a introducir una clase de procesos estocásticos bidimensionales ,cuyo estudio fué iniciado por R.Pyke en 1961 , con motivo de sus trabajos sobre procesos semimarkovianos y de renovación markovianos, pero sin restringir las V.A. a ser c.s. positivas.

Esta generalización , que ya ha sido considerada por varios autores,como: H.D.Miller (1962) ,Keilson(1964),Whisart (1964), Hatori y Mori(1966) , es muy interesante , ya que amplia el concepto de recorrido aleatorio sobre la recta real, al no suponer que los pasos sucesivos son V.A. independientes y equidistribuidas, ya que se introduce una cierta dependencia entre los diferentes pasos.

Dada la importancia del concepto de recorrido aleatorio por sus aplicaciones en Teoria de Colas ,Teoria colectiva del Riesgo etc.. ,vemos en estos procesos un campo de investigación fructifero a la vez ,desde el punto de vista teorico,como práctico.

Además estos procesos se encuentran en la Teoria de Procesos de Decisión secuenciales y los resultados que se obtienen son útiles a la vez en la gestión de stocks , y equipos de armamento.

Comenzaremos dando las definiciones y fundamentos de estos procesos , asi como las nociones de descomposición y de dualidad , para pasar después a la clasificación de estos procesos según su cadena inmersa y se establecerán resultados fundamentales sobre las Leyes Fuertes de los Grandes Números y sobre Teorema Central de Limite , para las sumas del tipo:

$$\sum_{n=1}^{N} f(J_{n-1},J_n,X_n)$$

generalizando asi los resultados de varios autores.

Se obtendrán estos Teoremas, sea a partir de razonamientos probabilisticos , como los utilizados por K.L.Chung,en cadenas de Markov o por Pyke y Schaufele(1964) en los Procesos de Renovación Markovianos , sea con la ayuda de las propiedades de sucesión centrada del proceso:

$$\left\{ f(J_{n-1},J_n,X_n) \ ; \ n \in N \right\}$$

I-1

DEFINICIONES Y PROPIEDADES FUNDAMENTALES

Designamos por I , el conjunto finito $I_m=(1,2,..,m)$ $m\in N_o$, o bién el conjunto de los naturales excluido el cero. N_o. Si $m<\infty$, significa que $I= I_m$ y $m= \infty$ que $I= N_o$.

DEFINICION 1 A

Sea un espacio probabilistico $(\Omega ,\mathcal{A}, P)$, que sin pérdida de generalidad se supone completo.Sobre este espacio llamamos proceso (J-X) a todo proceso bidimensional:

$$\left\{ (J_n, X_n) \ , n\in N \right\} \ ;(N= N_o\cup 0) \ , \text{ con valores en } I \times R$$

que satisface a :

I) $X_o= 0$ c.s.

$$P(J_o= K) = p_k \ , \quad k\in I \ , \sum_{k\in I} p_k= 1$$

II) $$P(J_n= k,X_n\leqslant x \,/\, J_o,J_1,X_1,J_2,X_2,\dots,J_{n-1} , X_{n-1}) \overset{c.s.}{=} Q_{J_{n-1},k} (x) \ ;\forall x\in R \text{ y } \forall k\in I$$

Además ,para todo $i,j\in I$, $Q_{ij}(x)$ es no decreciente, continua a la dercha , y :

$$\lim_{x\to -\infty} Q_{ij}(x) = 0 \ (1.1) \qquad \text{y} \lim_{x\to\infty} Q_{ij}(x) = p_{ij} \ (1.2)$$

donde $\sum_j p_{ij} = 1 \ ,i\in I$ (1.3)

Por tanto las funciones $H_j(x) = \sum_{k\in I} Q_{jk}(x)$, $j\in I$

son funciones de distribución definidas sobre $(-\infty, \infty)$, todo proceso (J-X), queda caracterizado por la tripleta $(m,\bar{p},Q)$ siendo:

m= un entero positivo o $+\infty$, según sea $I = I_m$ ó N_o.

$\bar{p}$= vector de probabilidades iniciales.

Q= matriz de funciones de masa, definidas sobre R y que satisfacen (1.1) y (1.3).

DEFINICION 1 B

Si para todo $i,j \in I$, $Q_{ij}(x) = Q_j(x)$, el proceso determinado por $(m,\bar{p},Q)$ se llama proceso (J-X) de "orden cero".

DEFINICION 1 C

Si para todo $i,j \in I$, $Q_{ij}(x)$ es nula para $x<0$ y es $Q_{ij}(0+)<1$, el proceso (J-X) determinado por $(m,\bar{p},Q)$ se llama proceso (J-X) positivo.

Hay que hacer constar que esta noción es, en los trabajos de Pyke, la base de la definición de los procesos semimarkovianos y de renovación markovianos.

DEFINICION 1 D

Si en la definición de proceso (J,X) positivo sustituimos la condición (1.3) por: $\sum_{j\in I} p_{ij} \leqslant 1$, $i \in I$ con al menos una desigualdad estricta, llamamos a estos procesos, positivos defectuosos.

Un proceso (J,X) positivo defectuoso es i-terminal, si y solo si, partiendo de i, se termina casi seguramente.

Un proceso (J - X) positivo defectuoso es p-terminal si se termina c.s.

Una clase de procesos (J,X) positivos defectuosos caracterizada por (m,Q) es terminal si cualquiera que sea la distribución inicial $\bar{p}$, el proceso de tripleta $(m,\bar{p},Q)$ es p-terminal.

Vamos a introducir ahora las variables aleatorias siguientes:

$$S_n = \sum_{k=0}^{n} X_k \quad , n \in N \text{ , y las funciones:}$$

$$F_{ij}(x) = \begin{cases} Q_{ij}(x)/p_{ij} & \text{si } p_{ij} > 0 \\ U_1(x) & \text{si } p_{ij} = 0 \end{cases} \qquad i,j \in I$$

donde : $$U_c(x) = \begin{cases} 0 & x < c \\ 1 & x \geq c \end{cases} \qquad c \in R$$

Bajo estas condiciones se verifican los siguiebtes lemas:

LEMA 1 A

El proceso (J_n, S_n) ,$n \in N$ es un proceso de Markov y el proceso $(J_n, n \in N)$ es una cadena de Markov.

En particular, tendremos:

$$P(J_n = k, S_n \leq y / J_0, J_1, S_1, J_2, S_2, \ldots, J_{n-1}, S_{n-1}) \overset{c.s}{=} Q_{J_{n-1}k}(y-S_{n-1})$$

$$P(J_n = k/J_0, J_1, \ldots, J_{n-1}) \overset{c.s.}{=} p_{J_{n-1}k}$$

LEMA 1 B

Todo proceso (J-X) goza de las propiedades siguientes:

a) $P(X_n \leqslant x \,/\, J_o, J_1, \ldots, J_{n-1}) \overset{c.s.}{=} H_{J_{n-1}}(x) = \sum_{j \in I} Q_{J_{n-1}j}(x)$

b) $P(X_n \leqslant x/ \, J_o, J_1, \ldots, J_n) \overset{c.s}{=} F_{J_{n-1}J_n}(x)$

c) $P(X_{n_1} \leqslant x_1, \ldots, X_{n_k} \leqslant x_k / J_n;\ n \ N) \overset{c.s.}{=} P(X_{n_1} \leqslant x_1, \ldots, X_{n_k} \leqslant$
$\leqslant x_k / J_o, J_1, \ldots, J_{n_k}) \overset{c.s.}{=} \prod_{i=1}^{k} F_{J_{n_i-1}J_{n_i}}(x_i)$, si $0 < n_1$
$< n_2 < \ldots < n_k$, $n_i \in N_o$; $i = 1, 2, \ldots, k$.

Esta última igualdad expresa la independencia condicional de $X_{n_1}, X_{n_2}, \ldots, X_{n_k}$ dados $J_{n_1-1}, J_{n_1}, \ldots, J_{n_k-1}, J_{n_k}$.

Entonces resulta que una forma equivalente de definir un proceso (J,X) consiste en dar:

P : matriz estocástica y F : matriz de funciones de distribución , donde F_{ij} representa la función de distribución condicional de X_n dado $J_{n-1} = i$, $J_n = j$. En este caso:

$$Q_{ij}(x) = p_{ij} \, F_{ij}(x)$$

Asi ,podemos definir un proceso (J,X) sea por la tripleta $(m, \bar{p}$ $Q)$, sea por la cuadrupleta$(m, \bar{p}, P, F)$

Los dos lemas anteriores han sido demostrados por Pyke (18) (1961) en su estudio de los procesos de renovación markovianos.

Posteriormente veremos la importancia que tiene la variable aleatoria S_n, en el estudio de recorridos aleatorios sobre la recta real.Dicha S_n ,representa la posición de la particula al cabo de n pasos.

II. ESTUDIO DE LAS FUNCIONES $f: IxIxR \longrightarrow R$

Vamos a considerar ahora ,las funciones reales $f(.,.,.)$ definidas sobre IxIxR tales que para todo i,j de I ,$f(i,j,.)$ es medible en sentido de Lebesgue.

Supondremos por tanto , una de las condiciones siguientes:

$$(2,1) \quad \sum_{j\in I} \int_R |f(k,j,x)| \, d\, Q_{kj}(x) < \infty \qquad \forall k \in I$$

$$(2.2) \quad \sum_{j\in I} \int_R |f(k,j,x)|^2 \, dQ_{kj}(x) < \infty \qquad \forall k \in I$$

Sea ahora :

$$\xi_{ik} = \int_R f(i,k,x) \, dQ_{ik}(x) \qquad (2.3)$$

$$\xi^{(2)}_{ik} = \int_R (f(i,k,x))^2 \, dQ_{ik}(x) \quad (2.4)$$

Asi, bajo las condiciones (2.1) y (2,2) ,las series :

$$(2.5) \qquad \xi_i = \sum_{k\in I} \xi_{ik} \quad \text{y (2.6)} \quad \xi^{(2)}_i = \sum_{k\in I} \xi^{(2)}_{ik}$$

son absolutamente convergentes para todo i I.

Si designamos ahora por $\mathcal{A}'_{n-1}$ y $\mathcal{A}''_{n-1}$ los σ-cuerpos engendrados respectivamente por (J_o ,J_1 ,X_1,..,J_{n-1},X_{n-1}) y por: (J_o,J_1,..,J_{n-1}) para $n > 1$ y ponemos $\mathcal{A}'_o = \mathcal{A}''_o = (\phi, \Omega)$, se verifica la siguiente propiedad:

PROPOSICION 2 A

Sea el proceso $\left\{ f\,(J_{n-1},J_n,X_n),\ n\in N_o \right\}$ que satisface (2,1). Entonces la sucesión :

$$\left\{ f(J_{n-1},J_n,X_n) - \xi_{J_{n-1}} \ ;\ \mathcal{A}'_n \ ;\ n\geqslant 1 \right\}$$

es una sucesión centrada.

Como la sucesión de las V.A. dadas está adaptada a la familia creciente de σ-cuerpos ($\mathcal{A}'_n$; $n\geqslant 1$) es suficiente demostrar que la esperanza matemática de dichas variables aleatorias ,adaptadas a $\mathcal{A}'_{n-1}$ es cero c.s.

La propiedad de ser centrada ,para la sucesión dada se considera en el sentido de Neveu(15)

Consideremos ahora las sumas siguientes:

$$W_f(n) = \sum_{k=1}^{n} f(\ J_{k-1},J_k,X_k) \ ;\ n\geqslant 1,\ W_f(0) = 0$$

y llamamos para simplificar la notación :

$$f\,'(\ J_{n-1},J_n\ ,X_n\ ) = f(\ J_{n-1},J_n\ ,X_n\ ) - \xi_{J_{n-1}}$$

e introducimos la V.A. $m(w)$, de valores enteros superiores o iguales a 1 ,y que verificam la propiedad siguiente:

$$\left\{ w\ :\ m(w) = n \right\} \in \mathcal{A}'_n \quad ,\text{para todo } n\in N_o .$$

De la proposición precedente y de un teorema de Neveu (15) ,extraemos la siguiente consecuencia , en forma de proposición :

PROPOSICION 2 B

Si se satisface la condición (2.1) ,entonces la sucesión : $\left\{ W_f(n) \ ; n \in N \right\}$ es una martingala con respecto a la sucesión creciente de σ-cuerpos $\left\{ \mathcal{A}'_n \ ; n \in N \right\}$.

Este resultado ,junto con un teorema de DOOB (6),nos conduce a la proposición siguiente:

PROPOSICION 2 C

Bajo la condición (2.1) y si :

I) $E\left[m(w) \right] < +\infty$

II) Existe una constante k positiva ,tal que:

$E\left[\left| f'(J_n ,J_{n+1},X_{n+1}) \right. \mathcal{A}'_n \right] \leq K$ para

$n < m(w)$ c.s.

se cumple ,entonces :

$$E\left[W_{f'}(m(w)) \right] = 0$$

Como consecuencia de la proposición anterior podemos extraer un interesante corolario , que puede interpretarse como una generalización de la igualdad de Wald , establecida para V.A. $(X_n ; n \in N_0)$,independientes ,integrables y equidistribuidas.

COROLARIO 2 A

Bajo las hipótesis de la anterior proposición ,se verifica que :

$$E\left[W_f(m(w)) \right] = E \sum_{n=1}^{m(w)} \xi_{J_{n-1}} \qquad (2.7)$$

si la primera de estas esperanzas matemáticas existe.

Es facil ver ,que la condición 2 de la anterior proposición ,se verifica cuando el supremo de la sumatoria de la condición (2.1) es finito.

Si consideramos el caso $f(i,j,x)=x$, con lo cual consideramos el caso de recorrido aleatorio sobre la recta real, resulta poniendo como expresión de las esperanzas condicional e incondicional :

$$b_{ij} = \int_R x \, d\, F_{ij}(x) \quad \text{y} \quad \eta_i = \sum_j p_{ij}\, b_{ij}$$

y suponiendo que todas las η_i son iguales a η, $i \in I$, y que $E\left[m(w)\right] < +\infty$:

$$E\left[S_{m(w)} \right] = E(X_n) \cdot E(m(w)) = \eta \, . \, E(m(w))$$

ya que $m(w)$ es una variable aleatoria.

Vemos asi que la igualdad(2.7) se puede interpretar como una generalización de la igualdad de Wald.

De forma parecida se puede generalizar el Teorema fundamental del Analisis Secuencial ,demostrado por Doob(6). Para esto consideramos la V.A. z ,compleja, y suponemos que las V.A.

$$Z_n = \frac{e^{zS_n}}{E(e^{zS_n} / J_o,J_1,..,J_n)} \quad ; \quad n \geqslant 1$$

esten definidad.Tenemos entonces la siguiente proposición:

PROPOSICION 2 D

Sea $m(w)$ de valores en N_o y tal que para todo $n \in N_o$:

$\{ w : m(w) = n \} \in \mathcal{A}_n$ siendo $\mathcal{A}_n$ el σ- cuerpo engendrado por $J_o, J_1, X_1, \ldots, J_{n-1}, X_{n-1}, J_n$. para $n > 1$.

Si :

I) $E[m(w)] < \infty$

II) Existe una constante $K > 0$, tal que :

$$E\left[|Z_{n+1} - Z_n| \,|\, \mathcal{A}_n \right] \leq K \text{ , para } n < m(w) \text{ c.s.}$$

entonces se verifica que : la esperanza matemática de la variable $Z_{m(w)}$, vale 1, es decir:

$$E\left[\frac{e^{zS_{m(w)}}}{E\left(e^{zS_{m(w)}} / J_o, \ldots, J_{m(w)} \right)} \right] = 1$$

La condición II) es equivalente a :

$$\frac{\left| e^{zS_n} \right|}{\prod_{k=1}^{n} \varphi_{J_{k-1} J_k}(z)} \cdot E\left[\left| \frac{e^{zX_{n+1}}}{\varphi_{J_n J_{n+1}}(z)} - 1 \right| \,|\, \mathcal{A}_n \right] \leq K$$

siendo $\varphi_{J_{k-1}J_k}(z)$, la función característica correspondiente a: $F_{J_{k-1}J_k}(x)$.

3.- DESCOMPOSICION

Vamos a dedicar este apartado a ibtener la descomposición de todo proceso (J-X) ,en dos procesos (J-X) positivos: $\{(J_n,X_n^+)\ ,\ n\in N\}$ y $\{(J_n,X_n^-)\ ,n\in N\}$,donde X_n^+ y X_n^- representan la parte positiva y negativa de la V.A. X_n.

Si $X^+(w)$ y $X^-(w)$, son las partes positiva y negativa de X(w),de función de distribución F(x),es fácil verificar que:

$$F^+(x) = P(X^+(w) \leq x) = U_0(x) . F(x)$$

$$F^-(x) = P(X^-(w) \leq x) = (1-F(-x-))(1-U_0(-x-))$$

$$|F|(x) = P(|X(w)| \leq x) = U(x)(F(x) - F(-x-))$$

Bajo estas condiciones se verifica la siguiente proposición:

PROPOSICION 3 A

Sea $\{(J_n,X_n)\ ,\ n\in N\}$, un proceso (J-X) de cuadrupleta $(m,\bar{p},P,F)$. Si para todo i,j de I , respectivamente ,se verifica :

$$F_{ij}(0) < 1 \quad , \quad F_{ij}(0-) > 0 \quad y \quad F_{ij}(0)-F_{ij}(0-) < 1$$

entonces los procesos :

$$\{(J_n,X_n^+),\ n\in N\},\quad \{(J_n,X_n^-),\ n\in N\}\ y\ \{(J_n,\ |X_n|),n\in N\}$$

son procesos (J,X) positivos caracterizados pobitivos caracterizados por las cuadrupletas :

$(m,\bar{p},P,F^+)$; $(m,\bar{p},P,F^-)$ y $(m,\bar{p},P,|F|)$

Se puede demostrar a titulo de ejemplo, cualquiera de las tres afirmaciones anteriores.

En lo que concierne a la relación entre las medias condicionales e incondicionales,tenemos:

$$b_{ij} = b^{+}_{ij} - b^{-}_{ij} \quad ; \quad b_{ij} = b^{+}_{ij} + b^{-}_{ij}$$

$$\eta_i = \eta^{+}_i - \eta^{-}_i \quad ; \quad {}_{\backprime}\eta'_i = \eta^{+}_i + \eta^{-}_i$$

siendo:

$$b^{+}_{ij} = E(X^{+}_n / J_{n-1}=i, J_n=j) \ ; \ b^{-}_{ij} = E(X^{-}_n/J_{n-1}=i,J_n=j)$$

$${}_{\backprime}b'_{ij} = E(X_n/ J_{n-1}=i,J_n=j)$$

$$\eta^{+}_i = E(X^{+}_n /J_{n-1}=i), \quad \eta^{-}_i = E(X^{-}_n/J_{n-1}=i)$$

$${}_{\backprime}\eta'_i = E(X_n /J_{n-1}= i)$$

las medias condicionales e incondicionales.,y siendo equivalente la convergencia absoluta de la serie $\sum_i p_{ij} b_{ij}$,a la convergencia de las dos series de términos positivos:

$$\sum_i p_{ij} b^{+}_{ij} \qquad y \qquad \sum_i p_{ij} b^{-}_{ij}$$

De todo esto podemos deducir el resultado siguiente en forma de corolario :

COROLARIO 3 A

Todo proceso (J-X) ,se descompone en dos procesos (J-X) positivos $\{(J_n, X^{+}_n) ,n\in N\}$ y $\{(J_n , X^{-}_n) ,n\in N.\}$

Estos tres procesos tienen el mismo conjunto de estados, la misma cadena de Markov inmersa y como matrices de funciones de distribución condicionales F,F^{+},y F^{-} y para todo n de N : $X_n = X^{+}_n - X^{-}_n$

4.- ESTUDIO DEL PROCESO J-X SEGUN EL TIPO DE CADENA INMERSA

En este apartado vamos a realizar un estudio de los procesos (J-X) , según sea el tipo de cadena asociada al proceso , haciendo especial hincapié en el estudio del proceso dual.

Vamos a considerar, en principio, un proceso (J-X) cuya cadena de Markov inmersa $(J_n; n \in N)$, sea irreducible, compuesta de una clase recurrente positiva de periodo d. Sean $C_1, C_2, \ldots, C_d$, las d subclases ciclicas de esta cadena. En este caso sabemos que existe distribución estacionaria única, solución del sistema :

$$\tilde{\Pi}_i = \sum_{k \in I} \tilde{\Pi}_k \; p_{ki} \quad , \qquad i \in I$$

con la condición:

$$\sum_{i \in I} \tilde{\Pi}_i = 1 \text{ ,con } \tilde{\Pi}_i \geq 0, \ \forall i \in I$$

La única solución de este sistema es: $\tilde{\Pi}_i = \dfrac{1}{m_{ii}}$, siendo m_{ii} el tiempo medio de recurrencia del estado i, y verificandose :

$$\sum_{i \in C_r} \tilde{\Pi}_i = \frac{1}{d}$$

Vamos a suponer ahora que el proceso (J-X) considerado, está determinado por la cuadrupleta $(m, \tilde{\Pi}, P, F)$, es decir tomamos como distribución de probabilidades iniciales, la distribución estacionaria, $\tilde{\Pi} = (\tilde{\Pi}_1, \tilde{\Pi}_2, \ldots\ldots, \tilde{\Pi}_k, \ldots)$, y fijamos $N \in N_0$.

Esto nos permite definir las V.A. siguientes:

$$\hat{J}_n = J_{N-n} \; , \; 0 \leq n \leq N \; ; \; \hat{X}_n = X_{N-n+1} ; 0 < n \leq N \; ; \; \hat{X}_0 = 0 \text{ c.s.} \quad (4.1)$$

Es fácil ver que, para todo n, tal que : $1 \leq n \leq N$, se verifica:

$$P(\hat{J}_n = j/\hat{J}_o, \hat{J}_1, \ldots, \hat{J}_{n-1}) = \frac{\tilde{\Pi}_j}{\tilde{\Pi}_{J_{n-1}}} p_{j\, \hat{J}_{n-1}} \quad \text{c.s.}$$

es decir, sobre $(0,N)$, el proceso se comporta como una cadena de Markov , de matriz :

$$\hat{P} = \left(\frac{\tilde{\Pi}_j}{\tilde{\Pi}_i} p_{ji}\right) \quad \text{y tal que} \quad P(\hat{J}_o = i) = \tilde{\Pi}_i, \; i \in I$$

Se verifica igualmente que la cadena de Markov de par $(\tilde{\Pi}, \hat{P})$ es irreducible recurrente positiva de periodo d, y que la única distribución estacionaria viene dada por el vector .

En virtud de las definiciones (4.1) y (4.2) tenemos casi seguramente:

$P(\hat{X}_n \leq x, \hat{J}_n = k/ \hat{J}_o, \hat{J}_1, \hat{X}_1, \ldots, \hat{J}_{n-1}, \hat{X}_{n-1}) =$ (después de sucesivos pasos) $= P(\hat{J}_n = k / \hat{J}_o, \hat{J}_1, \ldots, \hat{J}_{n-1}) . F_{k\, J_{n-1}}(x) =$

$$= \hat{p}_{\hat{J}_{n-1}\, k} \cdot F_{k\, \hat{J}_{n-1}}(x) \qquad (4.3)$$

Además , esta claro que:

$\hat{F}_{ij}(x) = P(\hat{X}_n \leq x/ \hat{J}_{n-1} = i, \hat{J}_n = j) = P(X_{N-n+1} \leq x / J_{N-n+1} = i, J_{N-n} = j) = F_{ji}(x)$.

El resultado (4.3) , demuestra que el proceso $((\hat{J}_n, \hat{X}_n), 0 \leq n \leq N))$, obtenido invirtiendo el sentido de las N etapas, se comporta como un proceso (J,X) de cuadrupleta $(m, \tilde{\Pi}, \hat{P}, F^{t})$.

Todo esto nos lleva a la definicion siguiente:

DEFINICION 4A

Sea un proceso (J-X) ,cuya cadena de Markov inmersa sea del tipo irreducible ,recurrente ,positiva de cuadrupleta $(m, \tilde{\Pi}, P, F)$.Entonces ,el proceso (J,X) : $\{(\hat{J}_n, \hat{X}_n), n \in N\}$ caracterizado por la cuadrupleta $(m, \tilde{\Pi}, \hat{P}, \hat{F})$,donde:

$$\hat{P} = \left(\frac{\tilde{\Pi}_j}{\tilde{\Pi}_i} p_{ji} \right) \text{ ,con } \hat{F} = F^{\tau}$$

se llama proceso dual del proceso considerado.

Vamos a considerar ahora un segundo caso ,el de un proceso (J,X) ,cuya cadena inmersa $(J_n, n \in N)$ es irreducible recurrente nula(de periodo d), y no positiva como en el caso anterior.Sabemos,que en este caso,no existe distribución estacionaria de probabilidad.Sin embargo Derman ,ha demostrado que el sistema:

$$\sum_{i \in I} \Pi_i \, p_{ij} = \Pi_j \text{ , } j \in I \quad (4.4)$$

admite, en este caso ,la única solución no negativa , tal como: $\Pi_a = 1$ (donde $a \in I$ es fijo).De hecho esta solución es positiva y viene dada por:

$$(4.5) \qquad \Pi_i = \lim_{n \to \infty} \frac{\sum_{v=0}^{n} p_{ii}}{\sum_{v=0}^{n} p_{aa}}, \quad i \in I$$

En este caso, como no se puede hallar distribución inicial apropiada ,consideramos ,en vez de procesos,clases de procesos (J,X) que tengan el mismo espacio de estados ,y las mismas matrices P y F .

Esta clase vendrá caracterizada por la tripleta (m,

;P,F) , pudiendo con esto extender la definición anterior al caso recurrente nulo:

DEFINICION 4 B

Sea la clase de procesos (J,X) de tripleta (m,P,F), tal que la cadena de Markov inmersa sea del tipo recurrente nulo ; entonces la clase de tripleta $(m,\hat{P},\hat{F})$,donde :

$$\hat{P} = \frac{\Pi_j}{\Pi_i}\, p_{ji} \quad y \quad \hat{F} = F^{\tau}$$

se llama clase dual de la clase considerada.

Se demuestra, para que esta definición tenga un sentido,que la matriz $\hat{P}$, no depende de la elección del estado "a" en las relaciones (4.5).

Se nos presenta un caso particularmente interesante :el de la autodualidad,es decir aquel donde la clase dual, coincide con la clase inicial.Esto nos lleva a la definición siguiente:

DEFINICION 4 C

La clase de procesos (J,X) ,de tripleta (m,P,F),se dice que es autodual ,si y solo si:

$$P = \hat{P} \text{ y } F=\hat{F}$$

Consideremos en principio el caso particular de las matrices biestocásticas .Sabemos en este caso ,que todos los Π_j son iguales a 1/m ,si $m<\infty$ y todos los Π_j son iguales a 1,si $m=\infty$ .Con esta última hipótesis la cadena debe ser necesariamente nula recurrente , al ser $m = \infty$.

Resulta asi ,que : $\hat{p}_{ij} = p_{ji}$, $i \in I$

Obtenemos asi la siguiente proposición:

PROPOSICION 4C

Sea la clase de procesos (J,X) ,de tripleta $(m,P,\mathcal{F})$ donde P es una matriz biestocástica y determina una cadena de Markov irreducible ,recurrente (positiva ó nula ,según sea $m < \infty$ ó $m = \infty$) .Entonces, la clase $(m,P,\mathcal{F})$ es autodual si y solo si $P = P^t$ y $\mathcal{F} = \mathcal{F}^t$.

Otro caso particular se presenta ,cuando la cadena inmersa es de orden cero ,es decir:

$$p_{ij} = p_j \quad ;i,j \in I$$

$$\sum_{j \in I} p_j = 1 . \qquad p_j > 0 \quad ; j \in I$$

En estas condiciones actuales ,la cadena de Markov es aperiódica ,recurrente,positiva y:

$$\hat{p}_{ij} = p_j$$

En este caso resulta que el proceso dual está caracterizado por la cuadrupleta $(m,\bar{p},P,\mathcal{F}^t)$, y que existe autodualidad si y solo si : $\mathcal{F} = \mathcal{F}^t$.

En el caso en que además $F_{ij}(x) = F_j(x)$ y $F_j(x) = 0$,$x \leq 0$, la consideración del proceso dual ,es fructifera tanto desde el punto de vista teórico como práctico.

5.- ESTUDIO DEL PROCESO ($u_s^{(j)}$; $s \in N_o$)

Partamos ahora de la cadena de Markov inmersa : $\{ J_n, n \in N \}$, y para todo $j \in I$, definamos el proceso $\{ r_n^{(j)}, n \in N_o \}$ de los indices de recurrencia:

$$r_o^{(j)} = 0$$

$$r_n^{(j)} = \sup_k \left\{ k \in N_o \ : \ k > r_{n-1}^{(j)} , J_l \neq j, \ r_{n-1}^{(j)} < l < k \right\} \quad n \in N_o$$

siendo: $$m_{jj} = E(r_{n+1}^{(j)} - r_n^{(j)}) \qquad n \in N_o$$

Con las notaciones usuales en las cadenas de Markov(2),tenemos que:

$$P(J_n = j \ \text{para un ciertp } n) = \sum_{i \in I} p_i \ f_{ij}^{*} \quad (5.1)$$

La hipotesis de base clásica consiste en suponer que existe el estado j recurrente y que la distribución inicial $\bar{p}$ sea tal que el segundo miembro de (5.1) sea igual a la unidad.

En este caso y sin por ello perder generalidad es suficiente limitarse al caso en que la cadena de Markov $\{ J_n, n \in N \}$ es irreducible y recurrente.Estas hipótesis implican que $r_n^{(j)}$ tiende c.s. a ∞ con n , y que el proceso para j fijo:

$$\{ u_s^{(j)} , \ s \in N_o \}$$

donde

$$u_s^{(j)} = \sum_{n = r_s^{(j)} + 1}^{r_{s+1}^{(j)}} f(J_{n-1} , J_n , X_n)$$

es un proceso de renovación en sentido amplio, es decir una serie de V.A. independientes y equidistribuidas sobre $(-\infty, \infty)$. Cuando el estado j permanece fijo, no escribiremos el indice superior con el fin de aligerar las notaciones.

Según Chung, vamos a introducir unas probabilidades:

$${}_j p^{*}_{ik} = \sum_{n=1}^{\infty} {}_j p^{(n)}_{ik} \qquad \text{siendo:} \qquad {}_j p^{(n)}_{ik} = \text{Probabilidad de que}$$

partiendo de i lleguemos a k en n pasos sin pasar por j .

Evidentemente : $\quad {}_r p^{*}_{jr} = 0 \quad$ y $\quad {}_r p^{(n+1)}_{jk} = \sum_{v} {}_r p^{(n)}_{jv} \, p_{vk}$

Derman ha demostrado que : $\quad \Pi_i = {}_j p^{*}_{ji}$

siendo el vector Π la unica solución no negativa del sistema (4.4) tal que $\Pi_j=1$. Ademas la cadena es recurrente nula si y solo si $\sum_{i\in I} \Pi_i = \infty$

PROPOSICION 5 A

Si se verifica (2.1) y si $\sum_i \sum_k |\xi_{ik}| \Pi_i < \infty$, entonces:

$$E(|u_1|) < \infty$$

y

$$E(u_1) = \sum_{i\in I} \Pi_i \, \xi_i \qquad (5.2)$$

Demostración:

En el caso particular donde la cadena es recurrente positiva , la proposición resulta directamente del Corolario 2A y del Lema 4.1 de Pyke y Schaufele(20) .

En el caso recurrente nulo, la hipótesis de convergencia absoluta nos permite hacer una demostración semejante a la del Lema citado.

En el caso recurrente positivo, resulta:

$$\Pi_i =_j p^*_{ji} = \frac{m_{jj}}{m_{ii}} = \frac{\tilde{\Pi}_i}{\tilde{\Pi}_j} = m_{jj}\,\tilde{\Pi}_i \qquad (5.3)$$

donde m_{jj} es el tiempo medio de recurrencia del estado j.

De aqui resulta el:

<u>COROLARIO 5 A</u>

En el caso recurrente positivo, si se verifica (2.1) y si $\sum_i \sum_k |\xi_{ik}| \Pi_i < \infty$ entonces:

$$E(u_1) = m_{jj} \sum_i \tilde{\Pi}_i \xi_i .$$

Vamos a aplicar ahora la proposición precedente al caso particular en que : $f(i,j,x)= x$

es decir estamos en el caso de recorridos aleatorios sobre la recta real, partiendo del origen(ya que X_o= 0 c.s.), u_1 representa la posición de la particula cuando el primer retorno a j dado que se parte de $J_o = j$. pues:

$$u_1 = \sum_{n=1}^{r_1} X_n$$

Supongamos ,que para todo $k \in I$:

$$(5.4)\sum_l p_{kl}|b_{kl}| < \infty \quad \text{y que} \quad \sum_i \sum_k p_{ik}|b_{ik}|\Pi_i < \infty \quad (5.5)$$

donde: $b_{kl}= E(X_n/ J_{n-1}=k,\ J_n=l)$ y $p_{kl}=P(J_n=l/ J_{n-1}=k)$

De la proposición 5A, resulta entonces que :

$$\mu_{jj} \doteq E(u_1) = \sum_i \Pi_i \eta_i \qquad (5.6)$$

Obtenemos asi los corolarios siguientes:

COROLARIO 5B

Bajo las hipótesis (5.4) y (5.5) y las de la proposición 5A y si $\mu_{jj} \neq 0$, entonces :

$$E(u_1) = \mu_{jj} A_f$$

donde :

$$A_f = \frac{\sum_i \Pi_i \xi_i}{\sum_r \Pi_r \eta_r}$$

COROLARIO 5C

Bajo las hipótesis (5.4) y (5.5) y si la cadena incluaa es recurrente positiva ,el cociente μ_{jj}/m_{jj} es independiente de j y vale $E(X_n)$, calculado bajo la distribución estacionaria. Además :

$$\mu_{jj} \gtreqless 0 \iff \sum_i \tilde{\Pi}_i \eta_i \gtreqless 0 \iff \mu_{kk} \gtreqless 0 \quad \text{para todo } k \text{ deI}$$

es decir los μ_{jj} o son todos nulos o son del mismo signo, siendo este signo ,en el caso de que la cadena inmersa sea recurrente positiva , el de $\sum_i \tilde{\Pi}_i \eta_i$

LEMA 5A

Si $\sum_{i \in I} {}_j p^*_{ji} \xi_i$ converge absolutamente(y es diferente de 0) para un $j \in I$ entonces $\sum_{i \in I} {}_1 p^*_{1i} \xi_i$ converge absolutamente (y es diferente de 0) para todo 1 de I.Además el signo de estas series es el mismo.

Demostración :

Sabemos, según Chung, que:

$$ {}_k p^*_{ki} = {}_k p^*_{kl} \cdot {}_l p^*_{li} \quad \text{para todo } l \in I $$

Hagamos k=j y reemplazamos ${}_j p^*_{ji}$ por su valor en :

$\sum_{i \in I} {}_j p^*_{ji} \xi_i$; Obtenemos asi:

$$ \sum_{i \in I} {}_j p^*_{ji} \xi_i = {}_j p^*_{jl} \sum_{i \in I} {}_l p^*_{li} \xi_i $$

Como para todo j y todo l ,tenemos:

$$ 0 < {}_j p^*_{jl} < \infty $$

el lema queda demostrado.

De este lema y de la proposición 5A , obtenemos la

PROPOSICION 5B

Si (2,1) se satisface y si $\sum_i \sum_k |\xi_{ik}| \; {}_j p^*_{ji} < \infty$ para un j, entonces existe la $E(u_1^{(k)})$ para todo k de I.Estas esperanzas son , o bien todas nulas ,o todas diferentes de 0 y del mismo signo .

De forma más general,se puede extender una demostración ,bastante larga y complicada de Chung y demostrar asi la

PROPOSICION 5C

Si existe $E(|u_1^{(i)}|^p)$ para un j de I , entonces existe la $E(|u_1^{(k)}|^p)$ para $k \in I$ ($p \in N_0$).

Lo dicho para el momento de primer orden de u_1 , puede repetirse para el calculo de $E(u_1^2)$; obtenemos asi los resultados siguientes que generalizan los de Pyke(17) :

PROPOSICION 5D

Si (2.2) se satisface y si:

$$\sum_{i\in I} \xi_i^{(2)} \Pi_i < \infty \quad y \quad \sum_i \sum_{k\neq j} \sum_{r\neq j} |\xi_{ik}| \sum |\xi_{rl}| \Pi_i \; {}_jp^*_{kr} < \infty$$

entonces:

$$E(u_1^{(2)}) < \infty$$

y:

$$E(u_1^{(2)}) = \sum_i \xi_i^{(2)} \Pi_i + 2 \sum_i \sum_{k\neq j} \sum_{r\neq j} \xi_{ik} \xi_r \; {}_jp^*_{kr} \Pi_i \quad (5.7)$$

De esta proposición y de la relación (5.6) resulta entonces:

COROLARIO 5D

Si μ_{jj} existe y es diferente de 0 y bajo las hipótesis de la anterior proposición , tenemos :

$$E(u_1^{(2)}) = \mu_{jj} B_f$$

donde :

$$B_f = \frac{\sum_i \xi_i^{(2)} \Pi_i + 2 \sum_i \sum_{k\neq j} \sum_{r\neq j} \xi_{ik} \xi_r \Pi_i \left({}_jp^*_{kr} \right)}{\sum_i \Pi_i \eta_i}$$

Terminamos este apartado con las aclaraciones siguientes:

a) Cuando la cadena es recurrente positiva sabemos ,según Chung, que:

$$\Pi_i = {}_jp^*_{ji} = \frac{\tilde{\Pi}_i}{\tilde{\Pi}_j} = \frac{m_{jj}}{m_{ii}} \quad (5.8)$$

y

$$_j p^*_{ik} = \frac{m_{ij} + m_{jk} - m_{ik}}{m_{kk}} \quad (5.9)$$

Estas relaciones permiten determinar la $E(u_1^{(2)})$ con la ayuda de los tiempos medios de primer paso , reemplazando en (5.7) Π_i y ${}_j p^{*}_{kr}$ por las expresiones dadas antes.

b) Cuando m $< \infty$, la existencia de las cantidades ξ_{ik} y $\xi^{(2)}_{ik}$ es suficiente para poder aplicar los resultados dados.Además se pueden emplear siempre (5.8) y (5.9).

6.- LEYES FUERTES DE LOS GRANDES NUMEROS PARA PROCESOS (J-X)

Vamos a presentar primeramente en este apartado un Teorema general válido sin la hipótesis de irreducibilidad ,con la ayuda de las sucesiones centradas.Deduciremos un caso particulas de ley fuerte haciendo intrevenir los momentos de orden 2 ,y veremos despues que el metodo de descomposición de CHUNG(2), para las cadenas de Markov,utilizado por PYKE para los procesos (J-X) positivos, conduce a una ley fuerte para los procasos (J-X) sin hipotesis sobre los momentos de 2º orden , pero con una hipótesis que es en ciertos casos dificil de verificarse.

LEMA 6A

Si las series(2.5)y (2.6) convergen absolutamente y si se verifica que :

$$\sup_i \left\{ \xi_i^{(2)} - (\xi_i)^2 \right\} < \infty \qquad (6.1)$$

entonces:

$$(6.2) \quad E\left[(f(J_{n-1}, J_n, X_n) - \xi_{J_{n-1}})^2\right] = \sum_k p_k^{(n-1)} (\xi_k^{(2)} - \xi_k^2), n \geq 1$$

cualquiera que sea la naturaleza de la cadena inclusa, y siendo por definición:

$$p_k^{(n)} = P(J_n = k) = \sum_{i \in I} p_i \cdot p_{ik}^{(n)}, \quad n \in N$$

Demostración

Tenemos ;

$$E\ (f(J_{n-1}, J_n, X_n) - \xi_{J_{n-1}})^2 = E\left[E\left[(f(J_{n-1}, J_n, X_n) - \xi_{J_{n-1}})^2\right.\right. \quad (6.3)$$

$/\mathcal{A}'_{n-1}]]$ (6.3)

La esperanza condicional que figura en esta ultima igualdad se calcula aisladamente y vale:

$$\xi^{(2)}_{J_{n-1}} - \xi^{2}_{J_{n-1}}$$

El Lema resulta entonces de la igualdad (6.3).

Vamos a enunciar directamente la siguiente proposición ,resultante del Lema anterior,de la Proposición 2A y de un Teorema de Neveu(15)(pág140):

<u>PROPOSICION 6A</u>

Bajo las hipótesis del lema 6A,,para toda sucesión no decreciente ($u_n, n \in N_o$) de reales no nulos tendiendo hacia ∞, la convergencia de:

$$\sum_{n=1}^{\infty} u_n^{-2} \sum_{k \in I} p_k^{(n-1)} (\xi^{(2)}_k - \xi^{2}_k)$$

implica que la serie :

$$\frac{1}{u_N} \sum_{n=1}^{N} (f(J_{n-1}, J_n, X_n) - \xi_{J_{n-1}})$$

tiende c.s. a 0 ,cuando N tiende a ∞ .

En particular ,si:

$$\lim_N \frac{1}{u_N} \sum_{n=1}^{N} \xi_{J_{n-1}} = A \text{ c.s.}$$

tenemos:

$$\lim_N \frac{1}{u_N} \sum_{n=1}^{N} f(J_{n-1}, J_n , X_n) = A \text{ c.s.}$$

De un teorema de CHUNG(2)(t.2,pág 87) , resulta entonces el Corolario siguiente:

COROLARIO 6A

Bajo las hipótesis del lema 6A y si la cadena es recurrente positiva :

$$\lim_{N} \frac{1}{N} \sum_{n=1}^{N} f(J_{n-1},J_n,X_n) = \sum_{i\in I} \tilde{\Pi}_i \xi_i$$

siempre que esta ultima serie converja absolutamente.

Hay que tener en cuenta que cuando $m < \infty$, la existencia de los ξ_{ik} y la hipótesis de irreducibilidad son suficientes para aplicar este corolario.

Otra forma de establecer una Ley Fuerte de los grandes numeros consiste en tomar el metodo de descomposición de CHUNG ,que ha sido utilizado por PYKE (17),(20) , en el caso de Procesos (J-X) positivos,y permanece válida para los procesos (J-X) .Para recordar como es necesario proceder, introducimos las variables aleatorias siguientes, definidas unicamente sobre la cadena de Markov inmersa (J_n ,$n\in N$):

$$Z_k(v)=\begin{cases} 1, & J_v=k \\ 0, & J_v\neq k \end{cases} \qquad k\in I$$

$$N_k(n)= \sum_{v=1}^{n} Z_k(v), \quad n\in N_0 , \quad k\in I$$

Poniendo :

$$R_1(N) = \sum_{n=1}^{r_{\min}(1,N_j(N))} f_n \; ; \; f_n= f(J_{n-1},J_n,X_n)$$

$$R_2(N) = \sum_{n=r_{N_j(N)}+1}^{N} f_n$$

$$V(N)= \sum_{s=1}^{N_j(N)-1} u_s$$

con el convenio habitual de que la suma sobre un conjunto de indices de sumación vacio es nula, podemos escribir la formula de descomposición ,válida para todo $n \in N_0$:

$$W_f(N) = R_1(N) + V(N) + R_2(N)$$

Esta formula es,de hecho algo diferente a la de CHUNG ya citada .CHUNG considera los intervalos $[r_n , r_{n+1})$, en lugar de los $(r_n, r_{n+1}]$.

Se tienen entonces los resultados siguientes,análogos a los ya existentes para los Procesos(J-X) positivos((17) y(20)) y para las cadenas de Markov(2).

LEMA 6B

Si $E[|u_1|] < \infty$ y si $m_j = E(u_1)$, entonces:

(6.4) I) $$\frac{R_1(N)}{N} \to 0 \quad \text{c.s.}$$

(6.5) II) $$\frac{V(N)}{N} \to \frac{m_j}{m_{jj}} \quad \text{c.s.}$$

siendo el ultimo limite nulo si $m_{jj} = \infty$

Demostración

Tenemos: $$|R_1(N)| \leqslant \sum_{n=1}^{r_1} |f_n|$$

Ahora bién: $r_1 < \infty$ c.s. puesto que la cadena es irreducible recurrente .Por consiguiente $\sum_{n=1}^{r_1} |f_n|$ es una variable aleatoria casi seguramente finita y que además es independiente de N. Esto prueba el apartado II).

Para demostrar II) ,sabemos que:

$$\frac{1}{n}\sum_{s=1}^{n} u_s \longrightarrow m_j \qquad \text{c.s.} \qquad (6.6)$$

en virtud de la Ley Fuerte de los Grandes Numeros( )

Además la Teoria de la Renovación nos enseña que:

$$\frac{N_j(N)}{N} \longrightarrow \frac{1}{m_{jj}} \qquad \text{c.s.} \qquad (6.7)$$

Ahora bien ,como la cadena de Markov inmersa es irreducible recurrente ,se verifica :

$$P(\lim_{N\to\infty} N_k(N)=\infty / J_0 = i) = 1 \quad \text{para} \quad \forall\, i \in I$$

de donde :

$$P(\lim_{N\to\infty} N_k(N) = \infty) = 1$$

De (6.6) resulta entonces que:

$$\frac{1}{N_j(N)-1}\sum_{s=1}^{N_j(N)-1} u_s$$

tiende casi seguramente a m_j .

Si expresamos

$$\frac{V(N)}{N} = \frac{1}{N_j(N)-1}\sum_{s=1}^{N_j(N)-1} u_s \; \frac{N_j(N)-1}{N}$$

y sabiendo que en virtud de (6.7) el 2º miembro de esta segunda igualdad tiende c.s. a m_j/m_{jj} ,cuando $N\to\infty$, el teorema queda demostrado.

PROPOSICION 6B

Si $E(\tilde{u}_1) < \infty$,donde $u_s = \sum_{n=r_1+1}^{r_{s+1}} |f_n|$ $(s \in N_0)$,entonces :

$$\frac{W_f(N)}{N} \xrightarrow[N\to\infty]{c.s.} \frac{m_j}{m_{jj}}$$

siendo el limite nulo si m_{jj} es infinito.

Demostración

Podemos aplicar el anterior lema puesto que por hipótesis $E(\tilde{u}_1) < \infty$ y u_1 en modulo se mantiene menor que $\tilde{u}_1$:

$$|u_1| = \left| \sum_{n=r_s+1}^{r_{s+1}} f_n \right| \leqslant \hat{u}_1$$

En consecuencia es suficiente demostrar que:

$$\lim_{N \to \infty} \frac{R_2(N)}{N} = 0 \quad \text{c.s.}$$

y como : $|R_2(N)| \leqslant \tilde{u}_{N_j(N)}$

resulta:

$$\frac{R_2(N)}{N} \leqslant \frac{N_j(N)}{N} \cdot \frac{\tilde{u}_{N_j(N)}}{N_j(N)} \qquad \text{c.s.}$$

para N suficientemente grande.

La proposición resulta entonces de:

$$\frac{\tilde{u}_N}{N} \xrightarrow{\text{c.s.}} 0$$

(13. pág 173).Además : $\frac{N_j(N)}{N} \xrightarrow[\text{c.s.}]{N \to \infty} \frac{1}{m_{jj}}$

Las proposiciones 5C y 6B muestran que si para un $k \in I$, se verifica que $E(\tilde{u}_s^k) < \infty$, entonces el cociente $m_j/m_{jj} = A$,no depende del estado j considerado .

De las proposiciones 6B y 5A resulta la:

PROPOSICION 6C (LEY FUERTE DE LOS GRANDES NUMEROS PARA LOS PROCESOS (J -X))

Si un proceso (J-X) verifica las hipotesis siguientes :

i) (J_n, $n \in N$) es irreducible recurrente.

ii) Las series $\eta_k = \sum_j p_{kj} b_{kj}$, $k \in I$ y $\sum_i \pi_i \eta_i$ convergen

absolutamente ,entonces S_n/n tiende c.s. a cero en el caso recurrente nulo y hacia $\sum_i \tilde{\pi}_i \eta_i$ en el caso recurrente positivo.

Otros teoremas validos para los procesos (J-X) positivos pueden traspasarse al caso de procesos (J-X) .

7.- APLICACION A LA PROGRAMACION DINAMICA

Los procesos (J-X) se pueden utilizar en la teoria de los procesos de decisión secuenciales(10).
En efecto, si se adopta uan politica estacionaria,X_n repre - senta el rendimiento de la n-ésima etapa y S_n el rendimiento total en n etapas .Howard ha demostrado que la ganancia total esperada por etapa , es decir :

$$\frac{E(S_n)}{n}$$

tiende ,en el caso irreducible recurrente positivo y aperiódico con $m < \infty$, hacia:

$$g = \sum_i \tilde{\Pi}_i \eta_i$$

De hecho ,podemos afirmar más ,pues de la proposicion 6C,ó de la 6A si $m<\infty$,y si las series $\sum_i p_{ij} b_{ij}$, $j \in I$ y $\sum \Pi_i \eta_i$ convergen absolutamente ,podemos deducir más,puesto que:

$$\frac{S_n}{n} \xrightarrow{c.s.} \begin{cases} \sum_i \tilde{\Pi}_i \eta_i & \text{(Caso recurrente positivo)} \\ 0 & \text{(Recurrente nulo)} \end{cases}$$

La conclusión es entonces evidente: Cuando en Programación dinámica se busca una politica estacionaria que maximice g ,se maximiza no sólo $\lim_n \frac{E(S_n)}{n}$,sino también el $\lim_n \frac{S_n}{n}$ (en sentido casi seguro)

Además ,no hay ninguna razón para descartar el caso en que m = ∞ y el caso periódico.

Hay que aclarar tambien ,que en el caso recurrente nulo ,toda politica estacionaria tiene un rendimiento nulo. Esta es la razón por la cual , en este caso, es necesario buscar otros criterios de optimalidad.

8.-TEOREMA CENTRAL DEL LIMITE PARA LOS PROCESOS (J-X)

Vamos a establecer un Teorema Central limite para $W_f(N)$ para los procesos (J-X) cuya cadena inclusa es irreducible recurrente y positiva y suponiendo que la varianza de U_1, sea finita .De la formula de descomposición resulta que:

$$\frac{W_f(N)}{\sqrt{N}} = \frac{R_1(N)}{\sqrt{N}} + \frac{V(N)}{\sqrt{N}} + \frac{R_2(N)}{\sqrt{N}} \qquad (8.1)$$

Como $R_2(N)$ está mayorada por una V.A. c.s. finita e independiente de N(según 6.5′) ,tenemos:

$$\frac{R_2(N)}{\sqrt{N}} \xrightarrow[pr.]{} 0 \qquad (8.2)$$

De la misma forma se demuestra facilmente que :

$$\frac{R_1(N)}{\sqrt{N}} \xrightarrow[pr.]{} 0 \qquad (8.3)$$

Vemos asi que el estudio asintotico de $W(N)/\sqrt{N}$ se reduce al de $V(N)/\sqrt{N}$, al menos para la convergencia en probabilidad y a fortiori para la convergencia en ley.

Por definición tenemos:

$$V(N) = \sum_{s=1}^{N_j(N)-1} u_s$$

donde el proceso (u_s, $s \geqslant 1$) es un proceso de renovación de media m_j y varianza s_j^2. Aplicando el Teorema Central limite de Lindenberg-Levy ,resulta que:

$$\frac{\sum_{s=1}^{N} u_s - Nm_j}{\sqrt{N}} \xrightarrow[N\to\infty]{L} \mathcal{N}(0,s_j)$$

Puesto que $(N_j(N) - 1)/N$ tiende c.s. hacia $1/m_{jj}$, cantidad positiva y finita ,por hipótesis, podemos aplicar los resultados de Billingsley (1) sobre las sumas de numeros aleatorios de variables aleatorias y establecer asi el:

LEMA 8A

Si el proceso (J-X) es de cadena inmersa irreducible ,recurrente positiva y si:

$$\text{var } u_1 = s_j^2 < \infty$$

entonces:

$$\frac{V(N) - N_j(N)m_j}{N} \xrightarrow[N\to\infty]{L} ((0, \; m_{jj}^{-1} s_j)$$

y:

$$\frac{V(N) - N_j(N)\, m_j}{\sqrt{N_j(N)}} \xrightarrow[N\to\infty]{L} \mathcal{N}(0, \; s_j)$$

COROLARIO 8A

Bajo las hipótesis del lema 8A , tenemos:

$$\frac{W(N) - N_j(N)\, m_j}{\sqrt{N}} \xrightarrow[N\to\infty]{L} \mathcal{N}(0, \sqrt{m_{jj}^{-1}}\; s_j)$$

Este último resultado no es del todo satisfactorio ,debido a la presencia de la variable aleatoria $N_j(N)$, pero la proposición siguiente nos permitirá desembarazarnos de ella.

PROPOSICION 8A (T.C.L. para las funciones de procesos (J-X)

Si el proceso (J-X) es de cadena inmersa irreducible recurrente positiva y si var $u_1(g) < \infty$, donde:

$$g = f - \frac{m_j}{m_{jj}}$$

entonces:

$$\frac{W_f(N) - N\frac{m_j}{m_{jj}}}{\sqrt{N}} \xrightarrow[N \to \infty]{L} \mathcal{N}\left(0, D(U_1(f - \frac{m_j}{m_{jj}}))\right)$$

Si, además, μ_{jj} existe y es diferente de cero, entonces se verifica, que:

$$\frac{W_f(N) - NA_f\mu}{\sqrt{N}} \xrightarrow[N \to \infty]{} \mathcal{N}(0, \sqrt{\mu B_g})$$

con $\mu = \mu_{jj}/ m_{jj}$ y A_f y B_g definidos como en el apartado 5.

Demostración

Puesto que: $\quad g_n = f_n - \frac{m_j}{m_{jj}} \quad$ tenemos:

$$W_g(N) = W_f(N) - N\frac{m_j}{m_{jj}} \quad y \quad u_s^{(g)} = \sum_{r_s+1}^{r_{s+1}} g_n =$$

$$= u_s^{(f)} - (r_{s+1} - r_s)\frac{m_j}{m_{jj}}$$

De esta última igualdad resulta que: $\quad m_j^{(g)} = E(u_s^{(g)}) = 0$

Por aplicación del lema 8A, resulta:

$$\frac{W_g(N) - N_j(N)m_j^{(g)}}{\sqrt{N}} = \frac{W_f(N) - N\frac{m_j}{m_{jj}}}{\sqrt{N}} \xrightarrow[N \to \infty]{L} \mathcal{N}\Big(0, \sqrt{m_{jj}^{-1}} \cdot$$

$$\cdot D\left(u_1 \left(f - \frac{m_j}{m_{jj}}\right)\right)\Big) \qquad (8.4)$$

Esto demuestra la primera parte de la proposición. Si además

μ_{jj} existe y es no nulo , tenemos por una parte:

$$\frac{m_j}{m_{jj}} = \frac{m_j}{\mu_{jj}} \frac{\mu_{jj}}{m_{jj}} = A_f$$

por el corolario 5B.y 5C: $A_f = \dfrac{\sum \tilde{\pi}_i \xi_i}{\sum\limits_i \tilde{\pi}_i \eta_i}$ y $\mu = \sum\limits_i \tilde{\pi}_i \eta_i$

siendo A_f y independientes del estado j fijado.

Por otra parte el corolario 5D implica que $D^2(u_1^{(g)}) = \mu_{jj} B_g$

porque $E(U_1^{(g)}) = 0$. El resultado 8.4 se convierte entonces

en:

$$\frac{W_f(N) - NA_f\mu}{\sqrt{N}} \xrightarrow[N \to \infty]{L} (0, \sqrt{\mu B_g})$$

Como A_f y μ no dependen del estado j fijado, B_g no depende

tampoco.

Esta proposición extiende el resultado de Pyke (17) para los procesos (J-X) positivos.

En el caso particular en que: $f(i,j,x) = x$, es decir en el caso de recorrido aleatorio, si las series:

$$\sum_i p_{ij} b_{ij} \ , i \in I \quad y \quad \sum_i \tilde{\pi}_i \eta_i$$

son absolutamente convergentes y si $\mu = \sum\limits_i \tilde{\pi}_i \eta_i$ es no nulo , obtenemos que: $W_f(N) = S_N$ y $A_f = 1$

De la proposición 8A obtenemos que:

$$\frac{S_N - N\sum \tilde{\pi}_i \eta_i}{\sqrt{N}} \xrightarrow[c.s.]{} (0 , \sqrt{\mu B_g})$$

Este resultado generaliza un Teorema de Hatori y Mori(9)demostrado por un metodo algebraico pero solamente en el caso de un n° finito de estados y cuando la cadena inclusa es irreducible y aperiodica.

De la proposición 8A vamos a deducir facilmente un teorema central limite para el proceso bi-dimensional $(J_n, W_f(N), n \in N)$. Pero antes es necesario establecer los lemas siguientes:

LEMA 8B

Sea la cadena de Markov irreducible recurrente positiva, caracterizada por la pareja $(\tilde{\Pi}, P)$ y la cadena dual $(\hat{J}_n, n \in N)$ .Entonces para todo $i,j,k \in I$ se cumple:

$$ {}_k\hat{p}^{*}_{ji} = \frac{\tilde{\Pi}_i}{\tilde{\Pi}_j} \, {}_k p^{*}_{ij} \qquad (8.5) $$

Demostración

Tenemos : ${}_k p^{(n)}_{ij} = P(J_v \neq k \; ; \; v=1,..,n-1, J_n=j/J_o=i)$

$= \frac{1}{\tilde{\Pi}_i} P(J_o=i, J_v \neq k, v=1,..,n-1, J_n= j)$

Pasando al proceso dual,obtenemos:

$${}_k p^{(n)}_{ij} = \frac{1}{\tilde{\Pi}_i} P (\hat{J}_n=i, \hat{J}_v \neq k \, , \, v= 1,,. ,n-1, \hat{J}_o= j) = \frac{\tilde{\Pi}_j}{\tilde{\Pi}_i} \, {}_k \hat{p}^{(n)}_{ji}$$

Sumando sobre n de 1 a ∞ ,obtenemos(8.5).

Sean ahora el proceso (J-X) $\{(J_n, X_n), n \in N\}$ de cadena inmersa ,irreducible ,recurrente positiva de cuadrupleta $(m, \tilde{\Pi}, P, F)$ y el proceso dual asociado $\{(\hat{J}_n, \hat{X}_n), n \in N\}$ de cuadrupleta $(m, \tilde{\Pi}, \hat{P}, \hat{F})$ $(\hat{F} = F^t)$

Está claro que:

$$W_f(N)=\sum_{n=1}^{N} f(J_{n-1} ,J_n,X_n) = \sum_{n=1}^{N} f(\hat{J}_{N-(n-1)},\hat{J}_{N-n},\hat{X}_{N+1-n})$$

$$=\sum_{n=1}^{N} f(\hat{J}_n ,\hat{J}_{n-1} ,\hat{X}_n) \qquad (8.6)$$

Definiendo sobre el proceso dual la función $\hat{f}(.,.,.)$ por la relación :

$$\hat{f} (i,j,x) = f(j,i,x); \quad i,j\in I \quad ,x\in R \qquad (8.7)$$

podemos escribir: $\quad W_f(N) = \hat{W}_{\hat{f}} (N) \qquad (8.7)'$

De aqui podemos concluir con el lema siguiente:

LEMA 8C

Sean el proceso (J-X) de cadena inmersa recurrente positiva de cuadrupleta $(m, \tilde{\Pi} ,P,F)$ y su dual.Entonces:

$$\text{I) } E(\hat{u}_1) = E(u_1)$$
$$\text{II) } \hat{A}_f= A_f$$
$$\text{III) } \hat{B}_f = B_f$$

en tanto que los segundos miembros tengan un sentido.

Ahoara podemos establecer la proposición anunciada

PROPOSICION 8 B

Sea el proceso (J-X) de cadena inmersa irreducible recurrente positiva de cuadrupleta $(m,\tilde{\Pi},P,F)$; entonces ,bajo las hipotesis de la proposición 8A , resulta:

$$P(J_N=k, \frac{W_f(N) - N A_f \mu}{N} \leq x) \xrightarrow[N \to \infty]{} \tilde{\Pi}_k \rho(x)$$

donde $\rho(x)$ es la función de distribución de la ley normal $(0, \sqrt{\mu B_g})$

Demostración

Pasando al proceso dual , obtenemos en virtud de la relación (8.7)' :

$$P(J_N=k,W_f(N)\leqslant x) = P(\hat{J}_o= k, \hat{W}_{\hat{f}}(N)\leqslant x) = P (\hat{J}_o= k).$$

$$.P(\hat{W}_{\hat{f}}(N) \leqslant x/ \hat{J}_o= k) = \tilde{\Pi}_k .P(W_f(N) \leqslant x / \hat{J}_o= k)$$

La proposición resulta entonces inmediatamente de la proposición 8A y del lema 8C.

Este último resultado extiende un teorema de Keilson y Wishart (11) , que se han limitado al caso de un nº finito de estados y a una cadena irreducible recurrente aperiodica .Su método de demostración(de tipo algebraico) no puede evidentemente generalizarse al caso en que $m= \infty$.

Es fácil ver que si la cadena inclusa además de las hipotesis hechas ,es aperiodica , entonces la proposición 8B es cierta cualquiera que sea la distribución inicial $\bar{p}$. En efecto ,pasando a la cadena dual, se sabe que (12.p.163)

$$P(J_n=j / J_{n-1} =i)= \frac{p_j^{(N-m)}}{p_i^{(N-n+1)}} p_{ji}$$

Haciendo tender N hacia ∞,resulta para todo n

$$P(J_n=j / J_{n-1}= i) = \frac{\tilde{\Pi}_j}{\tilde{\Pi}_i} p_{ji}$$

en virtud de la aperiodicidad.

B I B L I O G R A F I A

(1) BILLINGSLEY.P."Limit Theorems for ramdomly selected partial Sums" Ann.Math.Stat.,33(1962), 85-92

(2) CHUNG.K.L. "Markov Chains with stationary transition probabilities." Springer.Berlin.1960.

(3) DERMAN.C."A solution to a set of fundamental equations in Markov Chains" Proc.Amer. Math. Soc.5(1954)332-334

(4) DERMAN.C."Some contributions to the theory of Denumerable Markov Chains" Trans,Amer.Math.Soc.(1955)541-545.79.

(5) DOOB.J.L.Renewal Theory from the point of view of the Theory of Probability." Trans.Amer.Math.Soc(1948)63.

(6) DOOB.J.L. Stochastic Processes.Wiley. New-York. 1953

(7) FELLER.W. " An Introducti9n to Probability Theory and its Applications "Vol I. Wiley.New York. 1957.

(8) FELLER.W. " An Introduction ..." Vol II. 1966.

(9) HATORI y MORI. " An improvement of a limit Theorem on (J-X) processes" Kodaī Math.Sem .Report.18.(1966).347-352.

(10) HOWARD.R.A.." Dynamic Programming and Markov Processes Wiley.New-York. 1960.

(11) KEILSON y WISHART.D.M.G." A Central Limit Theorem for Processes defined on a Finite Markov Chain."Proc.Camb. Phil.Soc., 60(1964) 547-567.

(12) KEMENY.J.G. SNELL J.L. y KNAPP.A.W. " Denumerable Markov

Chains,Van Nostrand . New-York. 1966.

(13) LOEVE.M. Probability Theory .Van Nostrand. N.Y. 1966

(14) MILLER .H.D. "Absorption Probabilities for Sums of Ramdom Variables defined on a finite Markov Chain. Proc.Cam.Phil.Soc. 58.(1962) 286-298.

(15) NEVEU.J. " Bases Mathematiques du calcul des Probabilités" .Masson. Paris. 1964.

(16) PARZEN .E. Stochastics Processes. Holden -Day. S.Francisco.

(17) PYKE.R."Limit Theorems for Markov Renewal Processes." Technical Report nº 24 (1961).Columbia University .

(18) PYKE.R."Markov Renewal Processes .Definitions and preliminary properties." Ann.Math.Stat. 32.(1961)1231-1242

(19) M.R.P. of Zero Order and their Application on Counter Theory" Studies in Applied Probability and Management Sciences.Edited by Arrow ,Karlin y Scarf. Stanford 1962

(20) PYKE.R. y SCHAUFELE.R. Limit Theorems for Markov Renewal Processes. Ann.Math. Stat. 35. (1964)

(21) J.JANSSEN . " Les Processus (J-X).Cahiers du C.E.R.O. 11. 1969.

(22) JANSSEN .J. "Sur une generalisaton du concept de promenade aléatoire sur la droite réelle." Ann.Inst.Henry Poincaré . Vol 6. nº 3 ,(1970) pgs. 249-269.

C A P I T U L O II

INFERENCIA SOBRE PROCESOS (J-X) . ESTIMACION

C A P I T U L O II

INFERENCIA SOBRE PROCESOS(J - X) (ESTIMACION)

II -1.- ESTIMACION

Estimación de los parámetros de la cadena de Markov de $1^{\underline{er}}$ orden inmersa en el proceso (J-X)

Introducción

Dados $I_m=(1,2,..,m)$ y $R=(-\infty,\infty)$, sabemos que se llama proceso (J,X) $:((J_n,X_n);n\in N)$, con espacio de estados $I_m xR$ a un proceso estocástico bidimensional que satisface:

I) $X_o \overset{c.s.}{=} 0$. $P(J_o=K)=p_k$,$k\in I_m$, $\sum_k p_k=1$

II) $P(J_n=k,X_n\leq x / (J_o,X_o),..,(J_{n-1},X_{n-1}))=$

$= Q_{J_{n-1}k}(x)$

para todo $(k,x)\in I_m x R$,donde $Q_{jk}(.)$,$(j,k=1,2,..,m)$,es una familia de funciones no decrecientes definidas sobre R ,tales que $Q_{jk}(-\infty)=0$ para $j,k=1,2,..,m$ y $\sum_{k=1}^{m} Q_{jk}(+\infty)=\sum_{k=1}^{m} p_{jk}=1$

con $Q_{jk}(+\infty)=p_{jk}$.

De aqui resulta que las funciones $H_j(x)=\sum_{k\in I_m} Q_{jk}(x)$, $j\in I_m$ son funciones de distribución definidas sobre$(-\infty,\infty)$ y que todo proceso (J-X) está definido por la tripleta$(m,\bar{p},Q)$, siendo:m ,un entero positivo ó $+\infty$,según sea $I=I_m$ ó N_o , $\bar{p}$,el vector de probabilidades iniciales y Q,la matriz de funciones de masa definidas sobre R y que satisfacen las condiciones dichas anteriormente.

Construyendo las variables $S_n = \sum_{k=1}^{n} X_k$, se demuestra que el proceso ((J_n, S_n); $n \in N$) es un proceso de Markov y que el proceso (J_n , $n \in N$) es una cadena de Markov de $1^{\underline{er}}$ orden , llamada cadena de Markov inmersa en el proceso, y que con probabilidades de transición estacionarias.

Vamos a aplicar a esta cadena los resultados obtenidos por Anderson y Goodman ,sobre inferencia en cadenas de Markov.

MODELO

Sean los estados $i = 1,2,..,m$ .El estado i se entiende usualmente como un entero que va de 1 a m ,aunque i puede ser tanto un lugar geográfico, un partido politico etc...

Sean los instantes de observación $t=0,1,..,T$.

Sean $p_{ij}(t)$ = las probabilidades de transición de dicha cadena ,para $i,j= 1,2,..,m$ y $t= 0,1,..,T$, es decir :

$$p_{ij}(t) = P(J_t= j /J_{t-1}= i) =p_{ij}$$

ya que tratamos con probabilidades de transición estacionarias

Suponemos que :Hay $n_i(0)$ individuos en el estado i en t=0, y que los $n_i(0)$,en principio ,no son aleatorios aunque se discutirá posteriormente el caso en que sean V.A.

Una observación sobre un individuo dado,consiste en la sucesión de estados en que el individuo está en $t= 0,1,..,T$,o sea $i(0),i(1),..,i(T)$.

Dado el estado inicial $i(0)$ existen m^T posibles sucesiones .Estas representan sucesos mutuamente exclusivos,con probabilidades:

$$(I.1) \qquad p_{i(0)i(1)}p_{i(1)i(2)}\cdots p_{i(T-1)i(T)}$$

Sean $n_{ij}(t)$,el nº de individuos que están en el estado i en t-1 y j en t.Podemos ver que el conjunto de los $n_{ij}($ (t) $(i,j=1,2,..,m)(t=1,2,..,T)$, un conjunto de m^2T números , forman un conjunto de estadisticos suficientes para las sucesiones observadas .

Sea $n_{i(0)i(1)..i(T)}$ el nº de individuos cuya sucesión de estados es $i(0)i(1)...i(T)$ .Entonces :

$$n_{gj}(t) = \sum n_{i(0)i(1)..i(T)} \tag{I.2}$$

donde la suma es sobre todos los valores de los i ,con : $i(t-1)=g$ e $i(t)=j$.

En el espacio nmT dimensional que describen todas las sucesiones para todos los n individuos(para cada estado inicial existen nT dimensiones), la probabilidad de un conjunto ordenado dado de sucesiones para los n individuos es:

$$\prod \left(p_{i(0)i(1)}(1)\, p_{i(1)i(2)}(2) \,..\, p_{i(T-1)i(T)}(T)\right)^{n_{i(0)i(1)..i(T)}} = \left(\prod_{i(0)i(1)} p_{i(0)i(1)}(1)^{n_{i(0)i(1)}(1)}\right)...\left(\prod_{i(T-1)i(T)} p_{i(T-1)i(T)}(T)^{n_{i(T-1)i(T)}(T)}\right) = \prod_{t=1}^{T} \prod_{g,j} p_{gj}(t)^{n_{gj}(t)} \tag{I.3}$$

en el caso general de probabilidades de transición no estacionarias .

En nuestro caso ,tenemos: que la expresión anterior se transforma en :

$$\prod_{t=1}^{T} \prod_{g,j} p_{gj}^{n_{gj}(t)} = \prod_{i,j} p_{ij}^{n_{ij}} \tag{I.3'}$$

haciendose la sumatoria en (I.3)sobre todos los valores delos T+ 1 indices.

Entonces el conjunto de los $n_{ij}(t)$ forman un conjunto de estadisticos suficientes,como se habia anunciado.

La distribución de los $n_{ij}(t)$ es , en nuestro caso, (I.3)' multiplicada por una cierta función de factoriales.Dado que $n_i(t-1) = \sum_{j=1}^{m} n_{ij}(t)$, la distribución condicional de los $n_{ij}(t)$,j=1,...,m, dados los $n_i(t-1)$,es pues:

$$\text{(I.4)} \qquad \frac{n_i(t-1)!}{\prod_{j=1}^{m} n_{ij}(t)!} \prod_{j=1}^{m} p_{ij}^{n_{ij}(t)}$$

Ahora bien,esta es la misma distribución que se obtiene si se tienen $n_i(t-1)$ observaciones de una distribución multinomial con probabilidades p_{ij} y con números resultantes : $n_{ij}(t)$. La distribución de los n_{ij} , condicional de los $n_i(0)$ es:

$$\text{(I.5)} \qquad \prod_{t=1}^{T} \left(\prod_{i=1}^{m} \left(\frac{n_i(t-1)!}{\prod_{j=1}^{m} n_{ij}(t)!} \prod_{j=1}^{m} p_{ij}^{n_{ij}(t)} \right) \right)$$

En nuestro caso de probabilidades de transición estacionarias el conjunto $n_{ij} = \sum_{t=1}^{T} n_{ij}(t)$ forma un conjunto de estadisticos suficientes.

Para probabilidades de transición no necesariamente estacionarias $p_{ij}(t)$, los $n_{ij}(t)$ forman un conjunto minimal de estadisticos suficientes.

Vamos a construir ahora estimadores de máxima verosimilitud para las probabilidades de transición p_{ij}.

ESTIMACIONES DE MAXIMA VEROSIMILITUD

Las probabilidades de transición estacionarias de la cadena de Markov inmersa en el proceso (J-X) , $(J_n ; n \in N)$,se pueden estimar ,maximizando la probabilidad (I.3)′ ,con respecto a las p_{ij} , sujetas desde luego a las restricciones :

$$(I.6) \qquad p_{ij} \geqslant 0 \quad y \quad \sum_{j=1}^{m} p_{ij} = 1 \ , \ i=1,2,..,m$$

cuando las n_{ij} son las observaciones actuales. Esta probabilidad es precisamente de la misma forma ,excepto para un factor no dependiente de las p_{ij} ,tal como el que se obtiene para m muestras independientes, donde la i-ésima muestra (i=1, ..,m) , consiste en n_i^* pruebas multinomiales ($n_i^* = \sum_j n_{ij}$) con probabilidades p_{ij} (i,j=1,2,..,m) .Para tales muestras es bien conocido y facilmente verificable,que las estimaciones de máxima verosimilitud para las p_{ij} son :

$$\hat{p}_{ij} = n_{ij} / n_i^* = \sum_{t=1}^{T} n_{ij}(t) / \sum_{k=1}^{m} \sum_{t=1}^{T} n_{ik}(t) =$$

$$(I.7) \qquad = \sum_{t=1}^{T} n_{ij}(t) / \sum_{t=0}^{T-1} n_{i}(t)$$

y por tanto esto es tambien cierto para cualquier otra distribución , en la cual la probabilidad elemental sea de la misma forma y cuyas restricciones sobre las p_{ij} sean las mismas.

Las estimaciones se pueden describir gráficamente en la forma siguiente: Las $n_{ij}(t)$ se pueden disponer en una doble tabla $m \times m$. En general, el estimador de $p_{ij}(t)$ es la i,j-ésima componente de la tabla ,dividida por la suma de las componentes de la i-ésima fila.En orden a estimar las p_{ij} para nuestra cadena estacionaria , se suman las correspondientes componentes en las tablas de doble entrada para $t=1,..,T$,obteniendo una tabla doble,con entradas $n_{ij}= \sum_t n_{ij}(t)$. El estimador de p_{ij} es la i,j-ésima componente de la tabla dividida por la suma de las componentes de la i-ésima fila.

Vamos a estudiar ahora el comportamiento **asintótico** de los $n_{ij}(t)$, para poder estudiar después el de las $\hat{p}_{ij}$.

COMPORTAMIENTO ASINTOTICO DE LOS $m_{ij}(t)$

Vamos a considerar que el cociente $n_k(0)/\sum n_j(0) \longrightarrow \eta_k$ ($\eta_k > 0$, $\sum \eta_k = 1$) , cuando $\sum n_j(0) \longrightarrow \infty$.

Para cada $i(0)$, el conjunto $n_{i(0)i(1)..i(T)}$ son variables multinomiales simples ,con amplitud muestral $n_{i(0)}(0)$ y parámetros $p_{i(0)i(1)}, p_{i(1)i(2)}, \dots, p_{i(T-1)i(T)}$, y por tanto están asintoticamente normalmente distribuidas ,cuando la amplitud muestral crece . Al ser las $n_{ij}(t)$ combinaciones lineales de estas variables multinomiales , están, por tanto, también normalmente distribuidas asintoticamente.

Sea $P=(p_{ij})$ y sean $p_{ij}^{(t)}$ los elementos de la matriz P^t. Entonces $p_{ij}^{(t)}$ es la probabilidad del estado j en el instante t dado el estado i en el instante 0.Si designamos por : $n_{k;ij}(t)$ el nº de sucesiones que incluyen el estado k en el instante cero, i en el instante t-1 y j en el instante t, vamos a calcular los primeros momentos de :

$$n_{ij}(t) = \sum_{k=1}^{m} n_{k,ij}(t) \tag{I.8}$$

La probabilidad asociada con las $n_{k;ij}(t)$ es $p_{ki}^{(t-1)} p_{ij}$ con una amplitud muestral de $n_k(0)$.

Entonces :

$$E(n_{k;ij}(t)) = n_k(0)\, p_{ki}^{(t-1)} p_{ij} \tag{I.9}$$

$$\mathrm{Var}(n_{k;ij}(t)) = n_k(0)\, p_{ki}^{(t-1)} p_{ij}(1 - p_{ki}^{(t-1)} p_{ij}) \tag{I.10}$$

$$\mathrm{Cov}(n_{k;ij}(t), n_{k;gh}(t)) = -\, n_k(0)\, p_{ki}^{(t-1)} p_{ij}\, p_{kg}^{(t-1)} p_{gh} \tag{I.11}$$

para $(i,j) \neq (g,h)$

ya que el conjunto de los $n_{k;ij}(t)$ siguen una distribución multinomial.

Vamos a estudiar ahora los momentos de $n_{k;ij}(t) - n_{k;i}(t-1)\, p_{ij}$, donde los $n_{k;i}(t-1) = \sum_j n_{k;ij}(t)$; nos van a ser necesarios para obtener una teoria asintotica para procedimientos de test.La distribución condicional de los $n_{k;ij}(t)$ dados los $n_{k;i}(t-1)$ se vé facilmente que es multinomial ,con probabilidades p_{ij}. Entonces se verifica:

(I.12) $E(n_{k;ij}(t)/ n_{k;i}(t-1))= p_{ij}\, n_{k;i}(t-1)$

(I.13) $E(n_{k;ij}(t) - n_{k;i}(t-1)p_{ij}) = EE((n_{k;ij}(t)- n_{k;i}(t-1)$

$p_{ij})/ n_{k;i}(t-1)) = 0$

(I.14) La varianza de esta cantidad será:

$$E(n_{k;ij}(t) - n_{k;i}(t-1)\, p_{ij})^2 = EE((n_{k;ij}(t)-n_{k;i}(t-1)$$

$$\cdot p_{ij})^2/ n_{k;i}(t-1))= E (n_{k;i}(t-1)p_{ij}(1-p_{ij}) =$$

$$= n_k(0)\, p_{ki}^{(t-1)}\, p_{ij}(1-p_{ij})$$

Las covarianzas de los pares de **tales** cantidades son:

(I.15) $E(n_{k;ij}(t) - n_{k;i}(t-1)\, p_{ij})(n_{k;ih}(t) - n_{k;i}(t-1)p_{ih})$

$=E(- n_{k;i}(t-1)\, p_{ij}p_{ih}) = - n_k(0)\, p_{ki}^{(t-1)}\, p_{ij}p_{ih} \qquad j\neq h$

(I.16) $E(n_{k;ij}(t) - n_{k;i}(t-1)p_{ij})(n_{k;gh}(t) - n_{k;g}(t-1)p_{gh})=0$

$i \neq g$

(I.17)

$$E(n_{k;ij}(t)- n_{k;i}(t-1)p_{ij})(n_{k;gh}(t+r)- n_{k;g}(t+r-1)p_{gh})$$

$$= 0 \quad , \quad r>0$$

Para resumir ,las variables aleatorias $n_{k;ij}(t)- n_{k;i}(t-1)p_{ij}$ para $j=1,2,..,m$, tienen medias 0 y varianzas y covarianzas de variables multinomiales con probabilidades p_{ij} y amplitud muestral $n_k(0)\, p_{ki}^{(t-1)}$. Las variables $n_{k;ij}(t) - n_{k;i}(t-1)\, p_{ij}$ y $n_{k;gh}(s) - n_{k;g}(s-1)p_{gh}$ son incorreladas si $t\neq s$ ó $i \neq g$.

Ya que suponemos que los $n_k(0)$ son fijos, $n_{k;ij}(t)$ y $n_{l;gh}(t)$ son independientes si $k \neq l$.

Por consiguiente:

$$(I.18) \qquad E(n_{ij}(t) - n_i(t-1)p_{ij}) = 0$$

$$(I.19) \qquad E(n_{ij}(t) - n_i(t-1)p_{ij})^2 = \sum_{k=1}^{m} n_k(0)p_{ki}^{(t-1)} p_{ij}(1-p_{ij})$$

$$(I.20) \qquad E(n_{ij}(t)-n_i(t-1)p_{ij})(n_{ih}(t)-n_i(t-1)p_{ih}) =$$

$$= - \sum_{k=1}^{m} n_k(0)\, p_{ki}^{(t-1)} p_{ij}\, p_{ih}\,, \quad j \neq h$$

y finalmente:

$$(I.21) \qquad E(n_{ij}(t)-n_i(t-1)p_{ij})(n_{gh}(s)- n_g(s-1)p_{gh}) = 0$$

para: $\quad t \neq s \;$ ó $\; i \neq g$

Vamos a ver ahora como se distribuyen asintóticamente, los estimadores de las p_{ij}.

DISTRIBUCION ASINTOTICA DE LOS ESTIMADORES

Vamos a ver ahora que cuando $n \longrightarrow \infty$,

$$(I.22) \qquad \sqrt{n}\,(\hat{p}_{ij} - p_{ij}) = \sqrt{n}\,\left(\frac{\sum_{t=1}^{T} n_{ij}(t)}{\sum_{t=1}^{T} n_i(t-1)} - p_{ij}\right) =$$

$$= \sqrt{n}\,\left(\frac{\sum_{t=1}^{T} (n_{ij}(t) - p_{ij}n_i(t-1))}{\sum_{t=1}^{T} n_i(t-1)} \right) =$$

$$= n \left(\frac{\sum_{k=1}^{m} \sum_{t=1}^{T} (n_{k;ij}(t) - p_{ij} n_{k;i}(t-1))}{\sum_{t=1}^{T} n_i(t-1)} \right)$$

tiene como limite la distribución normal , y vamos a calcular las medias ,varianzas y covarianzas . Ya conocemos que ,puesto que $n_{k;ij}(t)$ es una variable multinomial :

$$n_{k;ij}(t)/ n \approx (n_{k;ij}(t)/n_k(0))\eta_k \tag{I.23}$$

converge en probabilidad a su esperanza,cuando $n_k(0)/n \longrightarrow \eta_k$

Entonces :

$$\underset{n \to \infty}{\text{p lim}} \frac{1}{n} \sum_{t=1}^{T} n_i(t-1) = \lim_{n \to \infty} \frac{1}{n} E \sum_{t=1}^{T} n_i(t-1) = \sum_{k=1}^{m} \eta_k \sum_{t=1}^{T} p_{ki}^{(t-1)} \tag{I.24}$$

Por lo tanto $n^{1/2} (\hat{p}_{ij} - p_{ij})$ tiene la misma distribución limite que:

$$\frac{\sum_{t=1}^{T} (n_{ij}(t) - p_{ij} n_i(t-1))/n^{1/2}}{\sum_{k=1}^{m} \sum_{t=1}^{T} \eta_k \; p_{kj}^{(t-1)}} \tag{I.25}$$

Da las conclusiones de la sección anterior , el numerador de (I.25) tiene media cero y varianza :

$$\sum_{k=1}^{m} \sum_{t=1}^{T} n_k(0) \; p_{ki}^{(t-1)} p_{ij} (1 - p_{ij})/ n \tag{I.26}$$

y las covarianzas entre dos diferentes numeradores ,será:

$$(I.27) \quad - \delta_{ig} \sum_{k=1}^{m} \sum_{t=1}^{T} n_k(0)\, p_{ki}^{(t-1)} p_{ij}\, p_{gh}/n$$

donde $\delta_{ig} = 0$ si $i \neq g$ y $\delta_{ii} = 1$

Si llamamos ϕ_i a $\sum_{k=1}^{m} \sum_{t=1}^{T} {}_k\, p_{ki}^{(t-1)}$

entonces la varianza limite del numerador de (I.25) es:

$\phi_i\, p_{ij}(1-p_{ij})$, y la covarianza limite entre dos diferentes numeradores es $- \delta_{ig}\, \phi_i\, p_{ij}\, p_{gh}$.

Ya que los numeradores de (I.25) son combinaciones lineales de variables multinomiales normalizadas, con probabilidades fijas y amplitud muestral creciente , tienen como limite la distribución normal y las varianzas y covarianzas de esta distribución limite son los limites de las respectivas varianzas y covarianzas.

Ya que $n^{1/2}(\hat{p}_{ij}-p_{ij})$, tiene la misma distribución limite que (I.25) , dichas variables tiene la distribución limite normal conjunta con medias 0,varianzas $p_{ij}(1-p_{ij})/\phi_i$ y las covarianzas $-\delta_{ig} p_{ij} p_{gh}$ .Tambien,el conjunto : $(n_i^*)^{1/2}(\hat{p}_{ij} - p_{ij})$ tiene distribución limite normal conjunta con medias cero, varianzas $p_{ij}(1-p_{ij})$ y covarianzas $-\delta_{ig}\, p_{ij}\, p_{gh}$,donde $n_i^* = \sum_{t=0}^{T-1} n_i(t)$.

En otras palabras,el conjunto $(n\,\phi_i)^{1/2}(\hat{p}_{ij} - p_{ij})$ para un i dado ,tiene la misma distribución limite que los estimadores de las probabilidades multinomiales p_{ij} ,con amplitud muestral $n\phi_i$, que es el nº total esperado de observaciones n_i^* en el i-ésimo estado para t=0,1,..,T-1.

Las variables $(n\,\phi_i)^{1/2}\ (\hat{p}_{ij} - p_{ij})$ para m valores diferentes de i.(i=1,2,..,m) ,son asintoticamente independientes , y por tanto tienen tambien la misma distribución limite conjunta que la obtenida a partir de funciones similares de los estimadores de las probabilidades multinomiales p_{ij} de m muestras independientes con amplitudes muestrales $n\,\phi_i$.(i=1,2,...,m).

Podemos hacer tambien uso del hecho de que las variables $\hat{p}_{ij}(t)= n_{ij}(t)/n_i(t-1)$ para un dado i y t tienen la misma distribución asintótica que los estimadores de las probabilidades multinomiales con amplitudes muestrales $E\,n_i\,(t-1)$ y las variables $\hat{p}_{ij}(t)$ para dos valores diferentes de i ,ò dos valores diferentes de t son asintoticamente independientes .Por tanto, cuando se contrasten hipótesis concernientes a los $p_{ij}(t)$,será posible reformular estas hipótesis en terminos de m x T muestras independientes,consistentes en pruebas multinomiales, y los procedimientes normales de contrastes se podrán aplicar.

II. 2

ESTUDIO DE FUNCIONES PROBABILISTICAS DE LA CADENA DE MARKOV INMERSA EN EL PROCESO (J-X)

Despues de estudiar la Inferencia Estadistica, en particular la estimación de las probabilidades de transición y la distribución asintotica de los estimadores de máxima verosimilitud para la cadena ($J_n; n \in N$) ,basandonos en los resultados de Anderson y Goodman(2) , vamos a tratar con las distribuciones de ciertas funciones probabilisticas de la cadena (considerada finita) y la admisibilidad de estas funciones como test estadisticos.

Consideraremos que el espacio de estados de la cadena es finito y que el parametro tiempo ,como hasta ahora ,es discreto .

Para nuestro estudio haremos uso de la representación vectorial de la cadena de Markov y trataremos después de una sucesión especial de funciones de la cadena , comprobando que son vectores aleatorios multinomiales independientes e identicamente distribuidos, bajo ciertas hipótesis acerca de sus probabilidades de transición .

EXPRESION VECTORIAL DE LA FUNCION DE DISTRIBUCION DE PROBABILIDAD DE UNA CADENA DE MARKOV DISCRETA DE NUMERO FINITO DE ESTADOS.

Supongamos $m \geqslant 2$ y llamemos $I_m = (z_1, z_2, \ldots, z_m)$, al espacio de estados finito de la cadena.

Cada estado se puede expresar vectorialmente como :

$$z_i = (\delta_{i1}, \delta_{i2}, \ldots, \delta_{im}) \quad ; \; i = 1,2,\ldots,m \;.\; y \quad \delta_{ij} = \begin{matrix} 1 & i=j \\ 0 & i \neq j \end{matrix}$$

es decir: $z_1 = (1,o,o,\ldots,o); z_2 = (o,1,o,\ldots,o); \ldots; z_m = (o,o,\ldots,1)$.

Sea $J_o, J_1, \ldots, J_T$ la cadena definida sobre I_m, esto es: Cada J_t expresada vectorialmente será de la forma:

$$J_t = (J_{t1}, J_{t2}, \ldots, J_{tm}); \quad t = o,1,\ldots,T$$

donde para cada t ,una y sólo una de las componentes toma el valor uno y las restantes toman el valor cero.

Sean $P_o = (p_1, p_2, \ldots, p_m)$ y $P = (p_{ij})$ la distribución de probabilidades iniciales y la matriz de probabilidades de transición estacionarias de la cadena de Markov $J_o, J_1, J_2, \ldots, J_T$

Podemos expresar los elementos de P_o y P como funciones exponenciales de $z_1, z_2, \ldots, z_m$.

Esto es, tenemos:

$$(2.1) \qquad p_i = Pr(J_o = z_i) = \exp(L(P_o)\, z_i'). = e^{\ln p_i} = p_i$$

para $i=1,2,..,m$ donde $L(P_o)=(\ln p_1, \ln p_2, .., \ln p_m)$ y z_i' es z_i escrita como vector columna(traspuesta de z_i).

Igualmente tenemos para las p_{ij}:

$$p_{ij}= Pr(J_t= z_j/J_{t-1}= z_i) = \exp(z_i \; L(P) \; z_j'); \; i,j=1,2,..,m$$

donde $L(P) = (\ln p_{ij})$ (2.2)

Por conveniencias de la notación ponemos aqui:

$(\ln 0)0= 0$ y $x(\ln 0)y \neq 0$ bién sea $x = 0$ ó $y=0$

Con esta representación vectorial para la cadena de Markov podemos establecer el siguiente :

TEOREMA 1

Sean $P_o=(p_1, p_2, ..., p_m)$ y $P=(p_{ij})$ la distribucion inicial y la matriz de transición de la cadena de Markov finita $J_o, J_1, .., J_T$. Entonces la densidad de probabilidad de la cadena se pueden escribir como:

$$f(z_{io}, z_{i1}, .., z_{iT})= \exp(L(P_o) \; z_{io}' + \sum_{t=1}^{T} z_{i(t-1)} \; L(P) \; z_{it}')$$

con $z_{it} \in I_m$ y $t=0,1,..,T$. (2.3)

Vista esta representación vectorial de las probabilidades iniciales ,de transición y de la función de densidad de la cadena de Markov ,vamos a estudiar cierto tipo de funciones probabilisticas de dicha cadena y su distribución.

DISTRIBUCION DE FUNCIONES PROBABILISTICAS DE LA CADENA DE MARKOV (J_n; $n \in N$)

Es ya conocido,que si las filas de la matriz de probabilidades de transicion $P=(p_{ij})$ son todas iguales ,entonces J_1, $J_2,..,J_T$ son vectores aleatorios multinomiales independientes e identicamente distribuidos .Esta propiedad se comprueba también para otras funciones de algunas cadenas de Markov .

En este apartado se definen una clase de funciones más generales de $J_1,J_2,..,J_T$ y damos una condición suficiente para que estas funciones probabilisticas sean estocásticamente independientes.

Para llegar a estas funciones ,vemos antes unas matrices que juegan un papel fundamental en nuestro estudio.

Sea $A = (a(u,v))$ una matriz mxm de enteros positivos ,tal que las filas de A son permutaciones m-arias de los enteros positivos (1,2,,.,m) .

Sea h un entero positivo ,tal que : $2 < h \leq m$ y la matriz $M= (m_{ij})$,m x h ,con las propiedades:

(2.4)
a) Cada m_{ij} es un entero positivo.
b) $m_{i1}+....+ m_{ih}= m$; $i = 1,2,..,m$

Sean $Y_1,Y_2,..,Y_T$, T funciones probabilisticas vector-valuadas de $J_o,J_1;..,J_T$, definidas de la siguiente forma:

(2.5) $$Y_t=(Y_{t1}, Y_{t2},....,Y_{th}) \quad ; t= 1,2,..,T$$

donde (2.6) $Y_{tj} = \sum_{u=1}^{m} \sum_{v=M_{u(j-1)}+1}^{M_{uj}} J_{(t-1)u} \, J_{ta(u,v)}$; $j = 1,..,h$

y donde :

$$M_{uo} = 0 \quad y \quad M_{uj} = m_{u1} + m_{u2} + \ldots + m_{uj}$$

$$j = 1,..,h$$

Se vé facilmente que $Y_1,..,Y_T$ es una sucesion de vectores aleatorios,cada uno definido en el espacio de estados I_h.

Estos vectores aleatorios no son en general estocásticamente independientes ,pero ,si las probabilidades de transición satisfacen ciertas condiciones, no muy fuertes,pueden llegar a ser vectores aleatorios independientes.

TEOREMA 2

Supongamos que la matriz de probabilidades de transición $P = (p_{ij})$ de la cadena $J_o,J_1,..,J_T$ satisface la propiedad de que para todo u (u=1,,..,m) y cada j (j=1,..,h) ,el conjunto de probabilidades de transición $p_{ua(u,M_{u(j-1)}+1)},\ldots,$ $p_{ua(u,M_{uj})}$,es un escalar , múltiplo de una distribución multinomial m_{uj}- variante $\alpha_{uj} = (\alpha_{uj1},\ldots, \alpha_{ujm_{uj}})$.

En otras palabras,tenemos:

$$(p_{ua(u,M_{u(j-1)}+1)},\ldots,p_{ua(u,M_{uj})}) = q_j(\alpha_{uj1},..,\alpha_{ujm_{uj}}) =$$

$$= q_j \, \alpha_{uj} \quad ; \; j=1,\ldots,h \quad y \quad u=1,\ldots,m \qquad (2.7)$$

donde $(q_1,\ldots,q_h)$, es una distribución multinomial h-variante con miembros no nulos .

Entonces $Y_1,Y_2,..,Y_T$ son vectores aleatorios independientes e identicamente distribuidos según una distribución multinomial común $(q_1,\ldots,q_h)$

DEMOSTRACION

Es suficiente probar que,para todo entero positivo $\alpha \geqslant 2$ y todo subconjunto $(t_1,t_2,..,t_\alpha)$ de $(1,..,T)$,se satisface la siguiente identidad:

$$P(Y_{t_1} = z_{i_1},\ldots,Y_{t_\alpha} = z_{i_\alpha}) = \prod_{j=1}^{\alpha} P(Y_{t_j} = z_{i_j}) \qquad (2.8)$$

donde cada i_k es uno de los números $1,\ldots,h$.

Podemos probar esta identidad por inducción matemática.
En primer lugar ,por la definición de Y_t, tenemos para todo t $(t=1,..,T)$ y todo i $(i=1,..,h)$ que:

$$(2.9)\ P(Y_t = z_i) = P(Y_{t_i} = 1) = \sum_{u=1}^{m} \sum_{v=M_{u(i-1)}+1}^{M_{ui}} P(J_{t-1u} J_{ta(u,v)}=1) = \sum_u Pr(J_{t-1u} = 1)\cdot \sum_v p_{ua(u,v)} = p_i$$

Vamos a ver ahora que ,para todo $1 \leqslant t_1 < t_2 \leqslant T$ y todo par $(i,k);i,k= 1,..,h$, se verifica :

$$(2.10)\ Pr(Y_{t_1}=z_i,Y_{t_2}=z_k) = Pr(Y_{t_1}=z_i)\cdot Pr(Y_{t_2}=z_k).$$

Ahora bién ,si: $p_{ij}^{(o)} = \delta_{ij}$ y $p_{ij}^{(n)}$ = probabilidad de transición de n pasos del estado z_i al z_j. para i,j= 1, 2,..,m; n= 1,2,..

Entonces tenemos :

$$\Pr(Y_{t_1}=z_i, Y_{t_2} = z_k) = \Pr(Y_{t_1 i}\, Y_{t_2 k} = 1) = \sum_{u=1}^{m} \sum_{v=M_{u(i-1)}+1}^{M_{ui}} \sum_{c=1}^{m} \sum_{d=M_{c(k-1)}+1}^{M_{ck}} \Pr (J_{t_1-1u} J_{t_1 a(u,v)} \; J_{t_2-1,c} \; J_{t_2 a(c,d)} = 1)$$

$$= \sum_{u} \Pr(J_{t_1-1,u} = 1) \cdot \sum_{v} p_{ua(u,v)} \cdot \sum_{c} p_{a(u,v)c}^{(t_2-t_1-1)} \sum_{d} p_{ca(c,d)} = q_i \cdot q_k = \Pr(Y_{t_1} = z_i) \Pr(Y_{t_2} = z_k) \qquad (2.11)$$

con lo cual se completa la demostración.

Finalmente, y suponiendo que (2.8) es cierta para = 2,..,n , necesitamos probar que es tambien cierto para = = n+1 .Sin embargo puede probarse facilmente ,usando el mismo procedimiento que en (2.1o).

COROLARIO 1

Bajo las suposiciones del Teorema anterior ,la distribución de probabilidad de la suma :

$$S_{hT} = Y_1 + Y_2 + \ldots + Y_T$$

viene dada por la siguiente densidad multinomial:

$$Pr\left(S_{hT}= \left(T_1,\ T_2 ,\ldots,T_{h-1},T-T_1-\ldots-T_{h-1}\right)\right) = T!\cdot \left(T_1!T_2!\ldots\right.$$
$$\left.\cdot(T-T_1-\ldots-T_{h-1})!\ \right)^{-1}\cdot\ q_1^{T_1}\ q_2^{T_2}\ldots q_h^{T-T_1-\ldots-T_{h-1}} \qquad (2.12)$$

En el caso en que h = 2 ,tenemos la distribución binomial :

$$Pr\left(S_{2T} =\left(x,T-x\right)\right) = \binom{T}{x}\ p^x\ q^{T-x}, \qquad x= 0,1,2,\ldots,T$$

donde $p=q_1$ y $q=q_2$.

II.3.ESTIMACION EFICIENTE DE FUNCIONES DE P

Supongamos que tenemos N observaciones $(J_1, J_2, .., J_n)$ de una cadena de Markov de m estados,con matriz de probabilidades de transición estacionaria $P = (p_{ij})$ con las $p_{ij} > 0$ $i, j=1,2,..,m$.

N debe ser una variable aleatoria , ya que si permanece fijo ,se sabe que no existen estimadores insesgados de funciones de P. Entonces, todo esquema que trate con estimadores insesgados de funciones de P , será un esquema de estimación secuencial.

El criterio más común de optimalidad ,cuando se trabaja con estimación insesgada , viene dado en función de las varianzas de los estimadores y sabemos que un extremo inferior de la varianza de un estimador insesgado viene dado por la desigualdad de información de Cramer-Rao (ya sea con muestra fija ó secuencial).

La igualdad y con esto la eficiencia del estimador , se alcanza ,si y sólo si, el plan de muestreo,S, el estimador $f=f(J_1, J_2, ..., J_N)$ y el valor esperado $E(f) = g(P)$ son eficientes , en el sentido que indicaremos.

Entonces el problema está en caracterizar las tripletas eficientes (S,f,g) , para la cadena de Markov de m estados.

DESIGUALDAD DE INFORMACION PARA LA CADENA DE MARKOV $(J_n; n \in N)$ Y C.N. y S. PARA LA EFICIENCIA DE UN ESTIMADOR

Partimos de nuestra cadena de Markov $(J_n; n\in N)$, de m estados con matriz de probabilidades de transición estacionarias $P = (p_{ij})$, $0 < p_{ij} < 1$; $\sum_{j=1}^{m} p_{ij} = 1$ y con distribución inicial $p_k = Pr (J_1 = k)$, $0<p_k<1$; $k=1,2,..,m$ y $\sum_{k=1}^{m} p_k = 1$.

Sea $P_i = (p_{i1}, p_{i2}, \ldots, p_{im})$, $i= 1,..,m$, el vector fila que indica la fila i-ésima de P y $P_o = (p_1, p_2, \ldots, p_m)$ el vector de probabilidades iniciales.

Supongamos que se observa $(J_1, J_2, .., J_N)$, donde N es una variable aleatoria cuya distribución está completamente especificada por el plan de muestreo en consideración. Definimos la V.A. de recuento de transiciones por:

$$N_{ij} = n_{ij}(J_1, .., J_N) = \sum_{t=2}^{N} N_{ij}(t) \quad , \quad \text{donde:}$$

$$N_{ij}(t) = 1 \quad \text{si} \quad J_{t-1} = i \text{ y } J_t = j$$
$$= 0 \quad \text{en los demás casos}$$

Expresamos por $N_{i.} = \sum_{j=1}^{m} N_{ij}$; $N_i = (N_{i1}, \ldots, N_{im})$ y $N = (N_{ij})$. Además definimos la V.A. V_k por:

$$V_k = 1 \quad \text{si} \quad J_1 = k$$
$$= 0 \quad \text{en los demás casos}$$

ya que : $Pr(V = e_k) = Pr(J_1 = k) = p_k$, donde $V=(V_1,..,V_m)$ y e_k es el vector fila de m componentes ,con un uno en la k-ésima componente y ceros en las demás.

Para llegar a establecer la desigualdad de información , vamos a ver primeramente en que consiste un plan de muestreo cerrado y no trivial. Definimos:

a) $A(S) = (n : Pr(N=n/S) > 0)$

b) $$B(S) = \bigcup_{n \in A(S)} (\mathbf{n} = (n_{ij}) : n_{ij}=n_{ij}(j_1,..,j_n)) \quad i,j=1,..,m$$

c) $B^*(S) = ((\mathbf{v},\mathbf{n}) : \mathbf{v} \in (e_1 ,..,e_m), \mathbf{n} \in B(S))$

La distribución conjunta de (V,N) bajo S es la dada por:

$$p(v,n/S) = Pr(V=\mathbf{v} ,N=n/S)= k(n/\mathbf{v};S).(\prod_{k=1}^{m} p_k^{v_k}).(\prod_{i,j=1}^{m} p_{ij}^{n_{ij}}) \quad (\mathbf{v},\mathbf{n}) \in B^*(S) \tag{3.1}$$

$$= 0 \quad \text{en los demás casos}$$

donde $k(n/\mathbf{v};S)$ es el nº de todas las sucesiones posibles $(j_1, ...,j_n)$; $n \in A(S)$ que dan los mismos recuentos de transición N=n dado V=v.

A partir de estas definiciones se define un plan de muestreo cerrado y no trivial ,como :

I) CERRADO si $\sum_{B^*(S)} p(\mathbf{v},\mathbf{n}/S) = 1$

II) NO TRIVIAL si $Pr(N \leq 1 / S) = 0$

(Se adoptan las mismas definiciones de Girshick y De Groot) (Ver (14))

Bajo el plan de muestreo S ,un estimador f($\mathbf{v},\mathbf{n}$) es una función real- valuada definida sobre B^* (S) .Es insesgado ya que su valor esperado es: $g(P_o,P) = Ef(V,N)= \sum_{B^*(S)} f(v,n)p(v,n/S)$. Esta serie es absolutamente convergente y diferenciable termino a termino con una serie derivada absolutamente convergente para todos los valores de (P_o,P) .

El objetivo es caracterizar todos los planes de muestreo ,que admiten estimadores insesgados de algunas funciones paramétricas cuyas varianzas alcanzan los extremos inferiores de la desigualdad de información.

Para la cadena de Markov ($J_n;n\in N$) ,la desigualdad de información para un estimador $f=f(V,N)$ de $g(P_o,P)=E(f)$ se demuestra ,que toma la forma :

$$\operatorname{Var}(f) \geqslant \left(\sum_{k=1}^{m-1} p_k g_k'^2 - \left(\sum_{k=1}^{m-1} p_k g_k'\right)^2 \right) + \tag{3.2}$$
$$+ \sum_{i=1}^{m} (EN_{i.})^{-1} . \left(\sum_{j=1}^{m-1} p_{ij} g_{ij}'^2 - \left(\sum_{j=1}^{m-1} p_{ij} g_{ij}' \right)^2 \right)$$

donde $g_k'= \partial g/ \partial p_k$ y $g_{ij}'= \partial g/ \partial p_{ij}$ son las derivadas parciales de $g(P_o,P)$ respecto a p_k y p_{ij} respectivamente,para $k,j=1,..,m-1;i =1,...,m$, siendo $p_m= 1 - \sum_{k=1}^{m-1} p_k$ y $p_{im}=1- \sum_{j=1}^{m-1} p_{ij}$.

Por ejemplo , para todo estimador insesgado de una probabilidad de transición p_{ij}, tendremos:

$\mathrm{Var}(f) \geq p_{ij}(1-p_{ij}) / (E(N_{i.}))$

Para la demostración ,de esta desigualdad nos remitimos a : (Do Sun Bai)(15) , resultando entonces el siguiente Teorema sobre la desigualdad de información de una cadena de Markov:

TEOREMA 1

Bajo todo plan de muestreo cerrado y no trivial S, la varianza de todo estimador insesgado $f=f(V,N)$ de $E(f)=g(P_o,P)$ está acotada inferiormente por (3.2)

La igualdad se verifica si y solo si $f(v,n)$ es una función lineal de las u_k y las w_{ij} ,para todo $(v,n)\in B^*(S)$, donde u_k y w_{ij} son las derivadas parciales del logaritmo de la función de verosimilitud; es decir:

$$(3.3)\quad U_k = \partial \log p(V,N/S) / \partial p_k = (p_m V_k - p_k V_m)/p_k\, p_m$$

$$(3.4)\quad W_{ij} = \partial \log p(V,N/S) / \partial p_{ij} = (p_{im} N_{ij} - p_{ij} N_{im}) / p_{ij}\, p_{im} .$$

A partir de esta introducción,vamos a caracterizar los estimadores eficientes . La noción de estimador eficiente que se utiliza aqui es la misma utilizada por Fisher ,en relación con la estimación insesgada.

ESTIMADORES EFICIENTES

DEFINICION 1

I) Para un plan de muestreo dado S, un estimador no constante $f = f(V,N)$ se dice que es eficiente para $E(f)=g(P_o,P)$ en (P_o^*,P^*) si la igualdad se verifica en (3.2) ,cuando $P_o=P_o^*$ y $P=P^*$. g se dice entonces eficientemente estimable en(P_o^*,P^*)

II) Si f es eficiente en todos los valores de (P_o,P) , entonces se dice que es eficiente para $g(P_o,P)$, y g se dice , que es eficientemente estimable.

III) Un plan de muestreo S ,se dice eficiente si admite al menos, un estimador eficiente no constante.

El siguiente corolario, que es una consecuencia inme - diata del Teorema (1) caracteriza a los estimadores efici- entes.

COROLARIO 1

" Bajo un plan de muestreo S ,cerrado y no trivial , un estimador no constante $f=f(V,N)$ es eficiente para $E(f)$ si y sólo si ,existen constantes a_k, b_{ij} ,$k,j=1,\dots,m-1$;$i=1,\dots,m$,no todas cero , y d , tal que :

$$f(v,n)=\sum_{k=1}^{m-1} a_k(p_m v_k - p_k v_m) + \sum_{i=1}^{m}\sum_{j=1}^{m-1} b_{ij}(p_{im}n_{ij} - p_{ij}n_{im})$$

$+ d$, para todo (v,n) perteneciente a $B^*(S)$. (3.5)

ESTIMACION EFICIENTE DE FUNCIONES DE P

Vamos ahora a considerar funciones especificas de P y le aplicaremos lo visto hasta ahora. En primer lugar vamos a considerar funciones de una fila de P, hallando tripletas eficientes para este caso particular.

Posteriormente, se tratarán funciones de dos ó más filas de P, y se comprobará que, a menos que la cadena tenga solo dos estados($m = 2$) no existen tripletas eficientes, tratando después este caso particular.

ESTIMACION EFICIENTE DE FUNCIONES DE UNA FILA DE P

Vamos a caracterizar todas las tripletas eficientes (S,f,g), donde las g son funciones de una fila de P.

Tomamos, sin por ello perder generalidad, la primera fila de P ; sea $P_1 = (p_{11}, p_{12}, \ldots, p_{1m})$.
Dos valores distintos de P_1, $P_1^{(o)}$ y $P_1^{(1)}$ se dice que son equivalentes con respecto a $g(P_1)$ si $g(P_1^{(o)}) = g(P_1^{(1)})$, y las probabilidades iniciales $P_o = (p_1, \ldots, p_m)$, se consideran parámetros perturbadores.

Bajo estas suposiones vamos a establecer el siguiente :

LEMA 1

Sea S, un plan de muestreo cerrado y no trivial para el cual existe un estimador no constante f, que es eficiente para alguna función $g(\ P_1\ )$,en dos valores de P_1 que no sean equivalentes con respecto a $g(P_1)$.

Entonces existen constantes $\mu_1, \mu_2, \ldots, \mu_m$,no todas cero , y $\xi \neq 0$ tal que :

$$(3.6) \quad \mu_1\, n_{11} + \ldots\ldots + \mu_m\, n_{1m} = \xi \text{ para } \forall\ \mathbf{n}_1 = (n_{11}, \ldots, n_{1m})$$

perteneciente a $B_1(\ S\ )$, donde :

$$B_i(S) = (\ \mathbf{n}_i = (\ n_{i1}, \ldots, n_{im})\ : \mathbf{n} \in B(S)); \ i = 1, \ldots, m)$$

Demostración

Como g es una función de P_1 solamente,el Teorema( ) demuestra que para todo estimador insesgado f de g:

$$(3.7) \mathrm{Var}(f) \geqslant (EN_{i.})^{-1} \cdot \left(\sum_{j=1}^{m-1} p_{ij}\, g_{ij}'^{2} - \left(\sum_{j=1}^{m-1} p_{ij}\, g_{ij}'\right)^2 \right)$$

y por el corolario I f es eficiente en P_1,si y sólo si , existen constantes $a_1, \ldots, a_{m-1}$ no todas cero y b tal que :

$$(3.8) \quad f = \sum_{j=1}^{m-1} a_j (p_{1m}\ n_{1j} - p_{1j} n_{1m}) + b \text{ ,para todo } \mathbf{n}_1 \in B_1(S)$$

Supongamos que f es eficiente en $P_1^{(o)}$ y $P_1^{(1)}$ .Entonces existen constantes $a_j^{(o)}$,$a_j^{(1)}$,$j=1,\ldots,m-1$,$b^{(o)}$y $b^{(1)}$,tales que

$$f = \sum_{j=1}^{m-1} a_j^{(o)} \cdot (\ p_{1m}^{(o)} n_{1j} - p_{1j}^{(o)}\ n_{1m}) + b^{(o)}$$

$$= \sum_{j=1}^{m-1} a_j^{(1)} \cdot (\ p_{1m}^{(1)}\ n_{1j} - p_{1j}^{(1)}\ n_{1m}) + b^{(1)}$$

De esto resulta directamente (3.6) .Ya que $b^{(i)}=g(P_1^{(i)})=E_i(f)$;i=0,1. donde $E_i(f)$ es la esperanza de f ,cuando $P_1=P_1^{(i)}$, y $P_1^{(o)}$ y $P_1^{(1)}$ no son equivalentes respecto a $g(P_1)$, tenemos $\xi \neq 0$, y que no todos los μ_j son cero.

El lema anterior demuestra que si la condición ( ) no se satisface ,entonces f no puede ser eficiente en dos ó mas valores no equivalentes de P_1.

<u>LEMA 2</u>

Sea S un plan de muestreo cerrado y no trivial para el que se verifica que: $Pr(N_{1.} = 0 /S) = 0$.Entonces $N_1=(N_{11}, \ldots,N_{1m})$ se puede escribir como:

(3.9) $N_1 =\sum_{k=1}^{N_1} z_k$ donde los z_k son vectores aleatorios independientes e identicamente distribuidos con:

(3.1o) $Pr(z_1= e_j) = p_{1j} \qquad j=1,..,m$

<u>Demostración</u>

Definimos las V.A. (M_k) por:

$$M_1 = \min (t: J_t = 1)$$

(3.11) $$M_k = \min (t: J_t = 1 , t > M_{k-1}) \qquad k=2,3,..$$

Entonces,ya que $Pr(N_{1.} = 0/S)=0$,$p_k > 0$ y $p_{1j} > 0$,j,k=1,.. ..,m tenemos que $Pr(M_k < \infty /S)= 1$,k=1,2,..

Sea ahora $z_k=e_j$ si $J_{M_k+ 1} = j$, j=1,..,m y k= 1,2,..

Es decir $z_k=e_j$ si la k-ésima visita al estado 1 por la cadena es seguida por una visita al estado j.Entonces obtenemos:

$$(3.12)\ Pr(z_k = e_j) = Pr(J_{M_k+1} = j) = \sum_{m_k} Pr(J_{M_k+1}=j/M_k=m_k).Pr(M_k = m_k) = p_{1j}.\sum_{m_k} Pr(M_k = m_k) = p_{1j};\ j=1,..,m$$

y se puede ver facilmente que las (z_k) son independientes.

Además se puede ver que N_1, puede representarse como la suma de N_1. observaciones independientes de un vector aleatorio z ,con la misma distribución que la dada para las (z_k).

Vamos ahora a introducir algunas notaciones y definiciones que se utilizarán para caracterizar planes de muestreo eficientes :

Sean c y c^* dos enteros positivos prefijados y $W = w(J_1,\dots,J_n)$ una variable aleatoria, tal que $w= w(j_1,\dots,j_n); n\in A(S)$ es una función positiva entero -valuada de $(j_1,\dots,j_n)$ no decreciente en n. Sea W^* otra variable aleatoria tal que: $W^* = w^*(J_1,\dots,J_n)$. $S(W,c)$ nos indica el plan de muestreo en el cual las observaciones se han realizado hasta que: $W=c$ se alcanza .

DEFINICION 2

Un plan de muestreo $S(c^* ; W^*)$ se dirá que es similar a $S(c;W)$ si $w = w(j_1,\dots,j_n) = c$ pra todo $(j_1,\dots,j_n)$, siempre que $w^* = w^*(j_1,\dots,j_n)= c^*$ y $n\in A(S(c^*;W^*))$;esto es, si $W = c$ a la terminación del muestreo bajo $(S(c^*,W^*))$.

EJEMPLO I

Sea $J^{(i)} = j^{(i)}(J_1, J_2, \ldots, J_N) =$ nº de veces en que

$J_t = i$, $t = 1, \ldots, N$

Entonces tenemos:

$$J^{(i)} = \sum_{t=1}^{N} I_{(i)}(J_t) = N_{1.} + I_{(i)}(J_N)$$

donde I. es una función indicador.

Ahora se puede ver que el plan de muestreo $S(c+1, J^{(i)})$ tiene la propiedad de :

1) $j_n = i$ para todo $n \in A(S(c+1, J^{(i)}))$

2) $n_{i.} = c$ para todo $n_i \in B_i(S(c+1; J^{(i)}))$

Entonces $S(c+1; J^{(i)})$ es similar a $S(c; N_{i.})$

EJEMPLO II

Sea $R^{(i)} = r^{(i)}(J_1, \ldots, J_N) =$ nº de rachas de i en N observaciones.

Entonces tenemos:

$$R^{(i)} = \sum_{j=1, j \neq i}^{m} N_{ij} + I_{(i)}(J_N)$$

El plan $S(c+1; R^{(i)})$ se puede ver que tiene la propiedad de que :

I) $j_n = i$ para todo $n \in A(S(c+1; R^{(i)}))$

II) $\sum_{j=1, j \neq i}^{m} n_{ij} = c$ para todo $n_i \in B_i(S(c+1; R^{(i)}))$

Entonces $S(c+1; R^{(i)})$ es similar a $S(c; \sum_{j=1, j\neq i}^{m} N_{ij})$

Bhat y Kulkarni (1966) extendieron los resultados de de Groot (1959) para poblaciones binimiales al caso multinomial, y han caracterizado planes de muestreo eficientes ,para tales poblaciones.El siguiente Teorema es una consecuencia de los lemas 2 y 3 y de los resultados de Bhat y Kulkarni.

TEOREMA 2

" Sea S un plan de muestreo cerrado y no trivial ,tal que:

I) $Pr(N_{1.} = 0 / S) = 0$

II) Existen constantes $\mu_1, \ldots\ldots, \mu_m$ no todas cero y $\xi \neq 0$ tal que :

$$\mu_1 \; n_{11} + \; \ldots . + \mu_m \; m_{1m} = \xi \quad \text{para todo } n_1 \in B_1(S)$$

Entonces S es ,bién $S(c; \sum_{k=1}^{k_o} N_{1,\sigma(k)})$ para algún entero positivo c ,ó un plan de muestreo similar a este ,donde ($\sigma(1)$, ,..., $\sigma(m)$) es una permutación de $(1,\ldots,m)$ y k_o es un entero tal que $1 \leq k_o \leq m$."

De los lemas 1 y2 y del Teorema anterior ,se deduce que solamente los planes de muestreo que pueden ser eficientes para alguna función $g(P_1)$ son $S(c; \sum_{k=1}^{k_o} N_{1,\sigma(k)})$,ó planes similares a ellos.

Vamos a ver en el siguiente Teorema que estos planes son verdaderamente eficientes .Todo esto ,unido al lema implica que un estimador no constante f es eficiente para $g(P_1)$, para todos los valores de P_1 si y sólo si es eficiente en dos valores distintos,no equivalentes de P_1.

TEOREMA 3

El plan de muestreo $S(c;\sum_{k=1}^{k_o} N_{1,\sigma(k)}), 1 \leq k_o \leq m$ ó planes similares a él son eficientes si toda función no constante f de la forma:

$$(3.13) \quad f=\mu_1 N_{11}+ \dots +\mu_m N_{1m}$$

es eficiente para :

$$(3.14) \quad c(\mu_1 p_{11}+ \dots +\mu_m p_{1m}) / (p_{1,\sigma(1)}+\dots+ p_{1,\sigma(k_o)})$$

Demostración

Habrá que probar que $S(c;\sum_{j=1}^{k_o} N_{1j})$, $1\leq k_o \leq m$, es eficiente y (3,13) es el estimador eficiente para:

$$g(P_1) =c(\mu_1 p_{11}+\dots+\mu_m p_{1m})/ (p_{11}+\dots+ p_{1k_o})$$

demostrando que existen constantes a_j ,j=1,..,m-1,y b ,tal que la ecuación (3.2) se verifica para todo $m_1 \in B_1(S(c;\sum_{j=1}^{k_o} N_{1j}))$.Escribimos :

$$(3.15) \quad p_{ij}n_{1m}=-(p_{1m}n_{1j}-p_{1j}n_{1m})+p_{1m}n_{1j}\ ;\ j=1,\dots,k_o$$

Ya que: $\sum_{j=1}^{k_o} n_{1j} = c$ para todo $m_1 \in B_1(S(c;\sum_{j=1}^{k_o} N_{1j}))$,sumando ambos miembros de (3.15),obtenemos:

$$(p_{11}+ \dots + p_{1k_o}) \cdot n_{1m} = - \sum_{j=1}^{k_o} (p_{1m} n_{1j} - p_{1j} n_{1m}) + cp_{1m}$$

Entonces:

$$(3.16) \qquad n_{1m} = - \sum_{j=1}^{k_o} (p_{11}+ \dots +p_{1k_o})^{-1} \cdot (p_{1m} n_{1j} - p_{1j} n_{1m})$$

$$+ c\, p_{1m}(p_{11}+ \dots + p_{1k_o})^{-1} .$$

Si escribimos:

$$n_{1j} = p_{1m}^{-1} \cdot (p_{1m} n_{1j} - p_{1j} n_{1m}) + p_{1j} n_{1m} p_{1m}^{-1}$$

De donde resulta finalmente , la expresión de f ,después de algunas simplificaciones:

$$f = \sum_{j=1}^{k_o} p_{1m}^{-1} (\mu_j -(p_{11}+ \dots + p_{1k_o})^{-1} \cdot \sum_{k=1}^{m} \mu_k p_{1k})(p_{1m} n_{1j} -$$

$$-p_{1j} n_{1m}) + \sum_{j=k_o+1}^{m-1} \mu_j p_{1m}^{-1} (p_{1m} n_{1j} - p_{1j} n_{1m}) + c(p_{11}+ \dots +p_{1k_o})^{-1}$$

$\cdot \sum_{j=1}^{m} \mu_j p_{1j}$ que es de la forma (3.17).Por tanto f es eficiente para :

$$E(f) = c(p_{11}+ \dots + p_{1k_o})^{-1} \cdot \sum_{j=1}^{m} \mu_j p_{1j} = g(P_1)$$

Aclaremos ,que en particular, $S(c;N_1.)$ es eficiente y (   ) es un estimador eficiente para $c \cdot \sum_{j=1}^{m} \mu_j p_{1j}$.

Vamos ahora a considerar el problema de caracterizar tripletas eficientes (S, f,g) ,cuando las g son funciones de dos filas de P ,sean P_1 y P_2.

Cuando las p_{ij} y $p_{i'j'}$ ($i \neq i'$) no estan funcionalmente relacionadas:

a) Los terminos que contienen a las p_{ij} en la probabilidad conjunta desarrollada en (3.1)se pueden factorizar de forma que los terminos que contengan a las $P_i = (p_{i1},\ldots,p_{im})$,contengan sólo a las $n_i = (n_{i1},\ldots,n_{im})$ y:

b) El extremo inferior en la desigualdad de información y la forma de los estimadores eficientes (3.5) tienen una propiedad aditiva ,en el sentido de que son ,en parte,sumas de términos que contengan solamente a las P_i y n_i.

Estos hechos, junto con el Teorema 2 , indican que , en orden a que el plan de muestreo S sea eficiente para alguna $g(P_1,P_2)$, S debe tener la propiedad de que ,para algunos enteros positivos fijos c_1 y c_2 :

$$(3.18) \qquad \sum_{k=1}^{k_o} n_{1,\sigma(k)} = c_1 \quad y \quad \sum_{\delta=1}^{\delta_o} n_{2,\sigma(\delta)} = c_2$$

para todo (n_1,n_2) para los cuales $n \in B(S)$,donde k_o y δ_o son enteros positivos ,$1 \leq k_o$, $\delta_o \leq m$.

Sin embargo ,es fácil ver que la condición (3.18) no se puede satisfacer si $m > 2$.

Cuando m=2,estamos en el caso de pruebas de Bernouilli dependientes, ($J_1, \ldots$),tales que:

$$\Pr(J_1=1)=1-\Pr(J_1=0)= p \qquad 0<p<1$$

$$\Pr (J_t=1/J_{t-1}=1) = \alpha \qquad 0<\alpha<1$$

$$\Pr (J_t=1/J_{t-1}=0) = \beta \qquad 0<\beta<1$$

con $1-\alpha=\bar\alpha$; $1-\beta = \bar\beta$; $1-p= q$

En este caso la desigualdad de información se reduce a:

$$(3.19) \quad Var(f) \geq pq g_p'^{2} (p,\alpha,\beta) + \alpha\bar\alpha\, g_\alpha'^{2} (p,\alpha,\beta)/E(N_{1.}) + \beta\bar\beta \quad g_\beta'^{2} (p,\alpha,\beta) /E(N_{o.})$$

y los estimadores eficientes tienen la forma:

$$(3,20) \quad f(j_1,n) = a(j_1-p) + b_1(\bar\alpha n_{11} -\alpha n_{1o}) + b_2(\bar\beta n_{o1}-\beta n_{oo}) + d$$

Entonces se demuestra que: (Ver :Do SUN BAI ,(15))

TEOREMA 3

"El único plan de muestreo eficiente , para funciones de ambos parámetros α y β ,es S(2c; $N_{1o}+ N_{o1}$) o planes similares a este, y toda función no constante f de la forma :

$f= aN_{1.} + b\, N_{o.} + d$ es un estimador eficiente para :

$$g(\alpha, \beta) = c(a/\bar\alpha + b/\beta) + d"$$

En particular, se verifica que f es eficiente para:

$f=N$ es eficiente para : $g(\alpha ,\beta) =c(1/\bar\alpha + 1/\beta)+1$

BIBLIOGRAFIA

(1) T.W. ANDERSON . " Probability models for analyzing time changes in attitudes" .RAND.Research Memorandum No. 455. 1951.

(2) T.W.ANDERSON. y LEO.A. GOODMAN." Statistical Inference about Markov Chains." Anm Math.Stat.28.(89 - 110) .1957.

(3) M.S.BARTLETT . " The frequency goodness of fit test for probability chains." Proc. Cambridge Philos.Soc. Vol. 47 (1951) ,pp. 86-95.

(4) H.CHERNOFF. "Large sample theory: parametric case" Ann.Math.Stat. Vol 27 (1956) pp 1- 22.

(5) W.G.COCHRAN." The χ^2 test of goodness of fit".Ann . Math. Stat. Vol 23. pp 315-345. (1952).

(6) H.CRAMER . Mathematical Methods of Statistics.(1946)

(7) W.FELLER. An Introduction to probability Theory and its Applications.Vol I . J.Wiley.New York (1950).

(8) L.A.GOODMAN. "On the Statistical Analysis of Markov Chains". Ann.Math.Stat. Vol 26.(1955) p 771.

(9) P.G.HOEL. " A test for Markov Chains"Biometrica.V.41

(10) J.NEYMAN."Contribution to the Theory of the χ^2 test". Symp.on Math.Stat.and Pr. Univ. de California Press.

(11) CHIA KUEI TSAO. "Admissibility and distribution of some probabilistics functions of discrete finite sta-

te Markov Chains" Ann.Math.Stat. Vol 39. No 5 .(1968)

(12) BARTON.D.E.DAVID,F.N. y FIX.E. (1962)." Persistence in a chain of multiple events when there is simple dependence." Biometrica 49. 351-357.

(13) BHAT y KULKARNI . (1966) ."On efficient multinomial estimation." J.Roy.Stat. Soc.Serv. B 28 .45-52.

(14) DE GROOT.M.H. Unbiased sequential estimation for binomial population".

(15) DO SUN BAI." Efficient estimation of transition probabilities in a Markov Chain".Ann.Math.Stat.(1975) Vol 3.nº 6. p 13o5.

C A P I T U L O III

INFERENCIA EN PROCESOS (J - X). CONTRASTES DE HIPOTESIS

3.1 . TEST DE HIPOTESIS ACERCA DE PROBABILIDADES ESPECIFICAS Y REGIONES DE CONFIANZA.

Sobre la base del capitulo anterior,referente a distribuciones asintoticas de los estimadores de las probabilidades de transición ,podemos estudiar algunos métodos de inferencia estadistica y aplicarlos en cualquier caso particular que se presente.

Para esto,vamos a suponer que todo p_{ij} es mayor que cero,y en primer lugar vamos a contrastar la hipótesis de que ciertas probabilidades de transición p_{ij} tienen valores especificos p^{o}_{ij} .

Con este fin,hacemos uso de que, bajo la hipótesis nula ,las variables: $(n^{*}_{i})^{1/2} \cdot (\hat{p}_{ij} - p^{o}_{ij})$, siguen una distribución limite normal ,con medias cero y varianzas y covarianzas dependientes de los p^{o}_{ij}, de la misma forma que se obtenia para los estimadores multinomiales.

Entonces,podemos usar la Teoria asintótica standard,referente a distribuciones normales o multinomiales,para contrastar una hipótesis acerca de una ó mas de las p_{ij},o determinar regiones de confianza para estas.

Como un ejemplo especifico ,vamos a contrastar la hipótesis :

$p_{ij} = p^{o}_{ij}$, $j = 1,2,..,m$,para un i dado.

Bajo la hipótesis nula tenemos que:

$$(1.1) \qquad \sum_{j=1}^{m} n_i^{*} \frac{(\hat{p}_{ij} - p_{ij}^{o})^2}{p_{ij}^{o}}$$

tiene asintoticamente una distribución χ^2 con m-1 grados de libertad .Entonces,la región critica de un contraste de esta hipótesis al nivel de significación α ,consta del conjunto de los $\hat{p}_{ij}$, para los cuales la suma anterior es mayor que el punto de α-significación de la distribución χ^2,con m-1 grados de libertad.

Una región de confianza ,de coeficiente de confianza α ,consta del conjunto de los p_{ij}^{o} , para los cuales la suma es menor que el punto de α-significancia.

Ya que las variables $n_i^{*} (\hat{p}_{ij} - p_{ij})^2$,para i diferentes,son asintóticamente independientes,las formas (1.1) para diferentes i ,son asintoticamente independientes ,y por tanto se pueden sumar para obtener otras variables χ^2 .

Un test para todas las $p_{ij}(i,j = 1,2,..,m)$,se puede obtener sumando (1.1) para todo i ,resultando una variable χ^2 con m(m-1) grados de libertad.

3.2 . CONTRASTE DE LA HIPOTESIS DE QUE LAS PROBABILIDADES DE TRANSICION SON CONSTANTES.

En nuestra cadena de Markov ($J_n ; n \in N$) suponemos que las probabilidades de transición p_{ij}, son estacionarias, es decir no dependen del tiempo .

Una alternativa general a esta suposición ,es que las probabilidades de transición ,dependan de t, es decir son de la forma $p_{ij}(t)$.

Contrastamos la hipótesis nula : $H_0 : p_{ij}(t)=p_{ij}$

$t=1,..,T$

Bajo la hipótesis alternativa, los estimadores de las probabilidades de transición para el instante t son:

$$(2.1) \qquad \hat{p}_{ij}(t)= \frac{n_{ij}(t)}{n_i(t-1)}$$

La funcion de verosimilitud maximizada ,bajo la hipótesis nula ,es:

$$(2.2) \quad \prod_{t=1}^{T} \prod_{i,j} p_{ij}^{n_{ij}(t)}$$

La función de verosimilitud maximizada bajo la alternativa , será :

$$(2.3) \quad \prod_{t} \prod_{i,j} \hat{p}_{ij}(t)^{n_{ij}(t)}$$

El cociente nos dá el criterio de razón de verosimilitudes:

$$(2.4) \qquad \lambda = \prod_{t} \prod_{i,j} \left(\frac{\hat{p}_{ij}}{\hat{p}_{ij}(t)} \right)^{n_{ij}(t)}$$

Ahora bién ,ya sabemos que $-2 \log \lambda$ se distribuye según una χ^2 con $(T-1)(m(m-1))$ grados de libertad cuando la hipótesis nula es cierta .

La razón de verosimilitud anterior, es semejante a las razones de verosimilitud obtenidas para tests standard de homogeneidad en las tablas de contingencia.Considerando esta semejanza,sea para un i dado ,el conjunto $\hat{p}_{ij}(t)$ que tiene la misma distribución asintotica que los estimadores de las probabilidades multinomiales $p_{ij}(t)$,para T muestras independientes .

Se puede usar una tabla m x T ,que tiene la misma apariencia formal que una tabla de contingencia, para representar los estimadores conjuntos $\hat{p}_{ij}(t)$ para un i dado y para $j=1,2,..,m$ y $t=1,2,..,T$

t \ j	1	2..........		m
1	$\hat{p}_{i1}(1)$	$\hat{p}_{i2}(1)$.....		$\hat{p}_{im}(1)$
2	$\hat{p}_{i1}(2)$	$\hat{p}_{i2}(2)$		$\hat{p}_{im}(2)$
:	:	:	:::::	:
T	$\hat{p}_{i1}(T)$	$\hat{p}_{i2}(T)$.....		$\hat{p}_{im}(T)$

La hipotesis de interés es,que las variables aleatorias representadas por las T filas,tienen la misma distribución, ya que los datos son homogéneos a este respecto.

Esto es equivalente a la hipótesis de que existen m constantes $p_{i1},p_{i2},\ldots,p_{im}$, con $\sum_j p_{ij}=1$,tal que la probabilidad asociada con la columna j-ésima es igual a p_{ij} ,en todas las filas(T);esto es ,$p_{ij}(t) = p_{ij}$,para $t=1,2,\ldots,T$.

El test χ^2 de homogeneidad ,nos parece el más apropiado aqui.Por tanto ,en orden a contrastar esta hipótesis,calculamos :

$$(2.5)\quad \chi^2_i = \prod_{i,j} n_i(t-1)(\hat{p}_{ij}(t) - \hat{p}_{ij})^2 / \hat{p}_{ij}$$

si la hipótesis nula es cierta ,χ^2_i tiene la distribución límite usual con (m-1)(T-1) grados de libertad.

Otro test de la hipótesis de homogeneidad para T muestras independientes de pruebas multinomiales se puede obtener usando el criterio de razón de verosimilitudes; esto es ,en orden a contrastsr esta hipótesis para los datos dados en la tabla m x T ,calculamos :

$$(2.6)\qquad \lambda_i = \prod_{t,j} (\hat{p}_{ij}/ \hat{p}_{ij}(t))^{n_{ij}(t)}$$

que es formalmente similar al criterio de razón de verosimilitudes .La distribución asintótica de $-2 \log \lambda_i$ es la χ^2_i con (m-1)(T-1) grados de libertad .

La aclaración precedente relativa a la semejanza con las tablas de contingencia ,trata con un valor dado de i. Por tanto,la hipótesis se puede contrastar separadamente para cada valor de i.

Vamos a considerar ahora la hipótesis conjunta de que $p_{ij}(t) = p_{ij}$ para todo i,j= 1,2,..,m y t = 1,2,..,T.

Un test para esta hipótesis nula conjunta, resulta directamente del hecho de que las variables aleatorias $\hat{p}_{ij}(t)$ y $\hat{p}_{ij}$,para dos valores diferentes de i,son asintoticamente independientes.Por tanto,bajo la hipótesis nula ,el conjunto de los χ^2_i calculados para cada i=1,2,..,m , son asintoticamente independientes, y la suma:

$$(2.7)\quad \chi^2 = \sum_{i=1}^{m} \chi^2_i = \sum_{i} \sum_{t,j} n_i(t-1)\,(\hat{p}_{ij}(t) - \hat{p}_{ij})^2/\,\hat{p}_{ij}$$

tiene la distribución limite usual con m(m-1)(T-1) grados de libertad .

De igual,forma,el test del cociente de verosimilitudes como en (2.4))se puede expresar de la forma:

$$(2.8)\qquad \sum_{i=1}^{m} -2\log\lambda_i = -2\log\lambda$$

3.3 UN MODELO MODIFICADO

En los apartados anteriores hemos supuesto que los $n_i(0)$,(nº de individuos en el estado i en el instante cero) eran no aleatorios .Una alternativa a esta suposición es que los $n_i(0)$,estén distribuidos multinomialmente,con probabilidades η_i y amplitud muestral n. Entonces ,la distribución del conjunto de las $n_{ij}(t)$ es (1.5) multiplicada por la distribución marginal del conjunto de los $n_i(0)$ que es:

$$\frac{n\,!}{\prod_{i=1}^{m} n_i(0)\,!} \prod_{i=1}^{m} \eta_i^{\,n_i(0)} \tag{3.1}$$

En este modelo ,el estimador de máxima verosimilitud de p_{ij} es otra vez (1.8),y el estimador de máxima verosimilitud de η_i es:

$$\hat{\eta}_i = n_i(0) / n \tag{3.2}$$

Las medias ,varianzas y covarianzas de $n_{ij}(t) - n_i(t-1)p_{ij}$,se calculan ,tomando los valores esperados de(1.20) a (1.23) . Las mismas formulas se aplican con $n_k(0)$ reemplazado por $n\eta_k$= Tambien $n_{ij}(t) - n_i(t-1)p_{ij}$ son incorreladas con $n_i(0)$.

Ya que $n_k(0)/n$ estima a η_k,consistentemente,las varianzas y covarianzas asintoticas de $n^{1/2}(\hat{p}_{ij} - p_{ij})$ son iguales a los del apartado .

Resulta de estos hechos que la teoria asintótica de los tests dada en el apartado anterior ,continua siendo util para este modelo modificado.

Se puede simplificar algo la expresión de las varianzas y covarianzas asintóticas , si la cadena parte de un estado estacionario;esto es ,si:

$$\sum_{k=1}^{m} \eta_k \; p_{ki} = \eta_i \qquad (3.3)$$

Entonces $\sum \eta_k \, p_{ki}^{(t-1)} = \eta_i$ y $\phi_i = T\,\eta_i$.

Se sabe,que si la cadena parte de un estado estacionario,las ecuaciones (4.3) pueden ser de algún uso adicional para la estimación de las p_{ki} ,cuando el conocimiento de las η_i,ó bien los estimadores de las η_i ,son calculables.

Tratamos aqui ,con el caso más general ,donde no se conoce cuando (3.3) se verifica,y se usan en este caso los estimadores de máxima verosimilitud .

En el caso más general ,los estimadores obtenidos no son eficientes ,en el caso especial de que la cadena esté en un estado estacionario, ya que no se conoce información de relieve.En el caso especial ,las estimaciones de M.V. para las $_i$ y las p_{ij} se obtienen maximizando:

$$\log L = \sum n_{ij} \log p_{ij} + n_i(0) \log \eta_i$$ sujeta a las restricciones:

$$\sum_j p_{ij} = 1 \ , \ \sum_i \eta_i \, p_{ij} = \eta_j \ , \ \sum_j \eta_j = 1 \quad \text{con}$$

$$p_{ij} \geqslant 0 \qquad y \qquad \eta_i \geqslant 0.$$

En el caso de que la cadena parta de un estado estacionario con los η_i conocidos ,las estimaciones de máxima verosimilitud de los p_{ij} se obtienen maximizando $\sum n_{ij} \log. p_{ij}$ sujeta a las restricciones : $\sum_j p_{ij} = 1$, $\sum_i \eta_i p_{ij} = \eta_j$ con $p_{ij} \geqslant 0$.

Para obtener las ecuaciones para los estimadores de maxima vero similitud se pueden usar los multiplicadores de Lagrange.

3.4

TESTS ADMISIBLES BASADOS EN S_{2T}

Deciamos en el capitulo anterior que las funciones probabilisticas de la cadena de Markov $(J_n ;\ n \in N)$,considerada finita,se podian admitir como test estadisticos.

Basándonos en esto ,vamos a considerar nuevamente, el problema de contrastar hipótesis acerca de la matriz de probabilidades de transición de la cadena.

Utilizando la misma notación ,llamamos S_{hT} a la suma $Y_1 + Y_2 + .. + Y_T$, y sabemos que la ley de probabilidad de dicha suma viene dada por la densidad multinomial,siguiendo la ley binomial ,cuando $h = 2$.

Vamos a tratar sólo este ultimo caso ,cumpliendose las hipotesis del Teorema .

Sean r_o y r_1 dos números dados ,tales que $0 < r_o , r_1 < 1$ y sean :

$$\alpha^{o}_{ij} = (\alpha^{o}_{ij1}, \alpha^{o}_{ij2}, \ldots\ldots, \alpha^{o}_{ijm_{ij}}) \ , \quad i=1,..,m \ , \quad j=1,2 \tag{4.1}$$

2m distribuciones(conocidas o desconocidas)

Llamemos H_k; k=0,1,2,3,4. a las siguientes hipótesis:

$$(4.2)\quad \begin{array}{llll} H_o : p=r_o \ , & \alpha_{ij} = \alpha^{o}_{ij} & & i=1,2,..,m, j=1,2 \\ H_1 : p=r_1 \ , & \alpha_{ij} = \alpha^{o}_{ij} & & i=1,2,..,m, \ j=1,2 \end{array}$$

$$(4.2)\quad \begin{array}{lll} H_2: p > r_o & \alpha_{ij} = \alpha^o_{ij} , & i=1,2,..,m , \; j= 1,2. \\ H_3: p < r_o & \alpha_{ij} = \alpha^o_{ij} , & i=1,2,..,m , \; j= 1,2. \\ H_4: p \neq r_o & \alpha_{ij} = \alpha^o_{ij} , & i=1,2,..,m , \; j= 1,2. \end{array}$$

A partir de estas hipótesis podemos extraer cuatro tests optimos para contrastar H_o, contra las cuatro hipótesis alternativas H_1, H_2, H_3 y H_4, respectivamente. En lo que sigue, las matrices de probabilidades de transición , bajo las cinco hipótesis expuestas se notarán por P_o, P_1, P_2, P_3 y P_4.

TEOREMA 1

"Supongamos que la distribución inicial P_o, permanece fija(sea conocida o desconocida).Entonces el test más potente para contrastar H_o contra H_1 ,viene dado por:

$$(4.3)\quad \begin{array}{rll} \varphi_1(x) = 1 & \text{para} & x > c_1 \\ = d_1 & \text{para} & x = c_1 \\ = 0 & \text{para} & x < c_1 \end{array}$$

si $r_o < r_1$, ó:

$$(4.4)\quad \begin{array}{rll} \varphi_2(x) = 1 & \text{para} & x < c_2 \\ = d_2 & \text{para} & x = c_2 \\ = 0 & \text{para} & x > c_2 \end{array}$$

si $r_o > r_1$, donde (c_1,d_1) y (c_2,d_2) estan determinados por un nivel de significación preasignado , y por el valor r_o y donde x viene dado por: $x = x_{11} + x_{21} + .. + x_{T1}$

con :

$$x_{t1} = \sum_{u=1}^{m} \sum_{v=1}^{m_{u1}} \delta_{i_{t-1}u} \; \delta_{i_t a(u,v)} \; ; t=1,..,T$$

DEMOSTRACION

Vamos a probarlo ,solo para el caso r_o r_1.La demostración de la segunda parte es,en todo,similar.

Por el Teorema 1 y el lema fundamental de Neyman y Pearson, se sabe que para cualquier α dado, el test más potente viene dado por:

$$\begin{aligned} \varphi(z_{i_o},..,z_{i_T}) &= 1 \quad \text{para} \quad R(z_{i_o},..,z_{i_T}) > c \\ &= d \quad \text{"} \quad R(z_{i_o},..,z_{i_T}) = c \qquad (4.5) \\ &= 0 \quad \text{"} \quad R(z_{i_o},..,z_{i_T}) < c \end{aligned}$$

donde (c,d) esta determinado por (α ,r_o) y donde:

$$R(z_{i_o},..,z_{i_T}) = \sum_{t=1}^{T} z_{i_{t-1}} (L(P_1) - L(P_o)) z'_{i_t} = (\ln r_1 - \ln r_o)x + (\ln (1-r_1) - \ln (1-r_o))(T - x) = (\ln r_1(1-r_o)/ /r_o(1-r_1))x + T(\ln (1-r_1)/(1-r_o)).$$

Entonces el test más potente esta dado por (4.5)y es equivalente al test dado por (4.3).

Con esto se completa la prueba.

COROLARIO 1

Si la distribución inicial P_o permanece fija, entonces :

I) El test φ_1 es uniformemente más potente para contrastar H_o contra H_2 y :

II) El test φ_2 es uniformemente más potente para contrastar H_o contra H_3.

COROLARIO 2

Si P_o permanece fijo, entonces, el test :

$$\begin{aligned} \varphi_3(x) &= 1 \text{ para } x < c_1' \text{ ó } x > c_2' \\ &= d_3 \text{ para } x = c_1' \text{ ó } c_2' \\ &= 0 \text{ para } c_1' < x < c_2' \end{aligned} \tag{4.6}$$

donde (c_1', c_2', d_3) está determinada por (α, r_o), es uniformemente más potente e insesgado para contrastar H_o contra la hipótesis alternativa H_4.

Notemos que si miramos (α_{ij}^{o}) como parámetros adicionales entonces los tres tests φ_1, φ_2, φ_3, son tests similares UMP e insesgados. En segundo lugar, ya que las funciones de potencia son todas binomiales con parámetros (T,p), todos los tests son uniformemente consistentes.

En tercer lugar, podemos usar los mismos tests para contrastar la hipótesis nula H_o, contra hipótesis alternativas más generales, suavizando las condiciones sobre las distribuciones (α_{ij}) y/o sobre la distribución inicial $P_o=(p_1,p_2, ,\dots,p_m)$,como veremos ahora.

COROLARIO 3

Sea λ un nº positivo ,tal que $0< \lambda \leq 1/s$ .Sea P_o la distribución inicial de la cadena de Markov inmersa $(J_n; n\in N)$, tal que : $P_o \geqslant \lambda = (\lambda, \lambda,\dots,\lambda)$, esto es p_i , $i=1,2,..,m$.

Entonces φ_1, φ_2, φ_3, son tests uniformemente más potentes e insesgados para contrastar la hipótesis nula H_o,contra las hipótesis H_2,H_3 y H_4,respectivamente.

Vamos ahora a considerar el problema de contrastar la hipótesis H_o contra una de las siguientes hipótesis alternativas extendidas :

$$H_1' : p = r_1 \quad ; \quad H_2' : p > r_o$$

$$(4.7) \qquad H_3' : p < r_o \quad ; \quad H_4' : p \neq r_o$$

Obviamente,no existen tests insesgados uniformemente más potentes para contrastar H_o contra estas hipótesis alternativas .Sin embargo, ya que tenemos:

$$H_i \subset H_i' \qquad i= 1,2,3,4.$$

podemos extraer el siguiente resultado:

TEOREMA 2

Bajo las suposiciones del teorema y suponiendo que P_o permanece fijo ,los tests φ_1, φ_2 y φ_3 son admisibles para contrastar H_o contra la hipótesis alternativas : H_1', H_2', H_3' y H_4' respectivamente ,y tienen las mismas funciones de potencia ,que las que sirven para contrastar H_o contra H_1 H_2, H_3 y H_4 respectivamente.

ALGUNOS MODELOS ESTADISTICOS

Ya que la distribución de S_{2T} es la binomial ,bien bajo la hipótesis nula ó cualquiera de las alternativas,nos es de gran conveniencia usar este hecho para las aplicaciones prácticas.

Es más ,el estadistico S_{2T}, se puede considerar como una generalización de algunos estadisticos bien conocidos, especialmente en el caso,en que:

$$m_{11} = m_{21} = \dots\dots = m_{m1} = 1.$$

Vamos a exponer ,algunos ejemplos :

I) MODELO " TEST DE LOS SIGNOS "

Si se verifica el caso anterior ,es decir todas las $m_{i1} = 1$, y $a(u,1)= 1$, $u = 1,..,m$, entonces S_{2T} se reduce a :

$$S' = \left(\sum_{t=1}^{T} J_{t1} \, , \, T - \sum_{t=1}^{T} J_{t1} \right)$$

y el test correspondiente se puede considerar como una generalización del"test de los signos", ya que,para el caso m=2 ,es el ya conocido test de los signos.

II) MODELO DEL"NUMERO TOTAL DE RACHAS"

Si se verifica y $a(u,1) = u$, $u=1,2,..,m$

Entonces S_{2T} , se reduce a :

$$S'' = \left(\sum_{t=1}^{T} \sum_{u=1}^{m} J_{t-1u} J_{tu} , T - \sum_{t=1}^{T} \sum_{u=1}^{m} J_{t-1u} J_{tu} \right)$$

y este es justamente el estadistico " nº total de rachas"de los m estados para nuestra cadena de Markov, ya que :

$$T + 1 - \sum_{t=1}^{T} \sum_{u=1}^{m} J_{t-1\,u} J_{tu}$$

es exactamente el nº total de rachas de los m estados.

Este estadistico ,fué primero introducido por Goodman ( ) y más tarde usado por Barton, David y Fix.

Relativo a este 2º apartado se pueden encontrar fácilmente ejemplos de persistencia en una cadena de sucesos múltiples.

III) <u>MODELO DE " RECORRIDO ALEATORIO CICLICO "</u>

Si se verifica ( ) y $a(u,1) = u + 1$, $u=1,2,..,$ $,m-1$ y $a(m,1) = 1$, entonces S_{2T} se reduce a :

$$S''' \equiv \left(\sum_{t=1}^{T} \sum_{u=1}^{m} J_{t-1u} J_{tu+1} ; T - \sum_{t=1}^{T} \sum_{u=1}^{m} J_{t-1u} J_{tu+1} \right)$$

donde $J_{tm+1} = J_{t1}$ .Este es un test UMP e insesgado para bién sea una hipótesis unilateral ó bilateral , acerca de p, en una cadena de Markov , con,por ejemplo,la siguiente matriz 4 x 4 de transición :

$$P = \begin{bmatrix} 0 & p & 0 & q \\ q & 0 & p & 0 \\ 0 & q & 0 & p \\ p & 0 & q & 0 \end{bmatrix}$$

que es justamente la matriz estocástica, para un recorrido aleatorio ciclico:

IV) MODELO DE "TENDENCIA MIXTA "

Para explicar este modelo,nos valemos de un simple ejemplo.Sea P la siguiente matriz de probabilidades de transición de una cadena de Markov con tres estados :

$$P = \begin{bmatrix} \alpha_1 p & (1-\alpha_1)p & q \\ q & \alpha_2 p & (1-\alpha_2)p \\ q & \alpha_3 p & (1-\alpha_3)p \end{bmatrix}$$

Si se quiere contrastar la hipótesis de aleatoriedad:

$$H_0: \alpha_1 = \alpha_2 = \alpha_3 = 1/2 \quad y \quad p= 2/3$$

contra la hipotesis alternativa: de tendencia mixta:

$$H_1 : p > 2/3$$

Entonces se puede usar el test estadistico:

$$S^{iv} = (Y, T-Y)$$

donde:

$$Y = \sum_{t=1}^{T} (J_{t-1\,1} J_{t1} + J_{t-1\,1} J_{t2} + J_{t-12} J_{t2} + J_{t-12} J_{t3}$$

$$+ J_{t-13} J_{t2} + J_{t-1\,3} J_{t3})$$

Como ejemplo de aplicación de este modelo , tenemos un problema de interés en Sociologia, contrastando el efecto de la clase social del padre en la clase social del hijo.

Si consideramos las clases sociales alta ,media y baja como los tres estados de una cadena de Markov,se ve en la matriz de transición ,por su estructura, que dicho modelo es el más conveniente para contrastar dicha hipótesis.

BIBLIOGRAFIA

(1) ANDERSON y GOODMAN. Statistical Inference about Markov Chains. (1957) Ann.Math.Stat. 28. 89-11o

(2) BARTON.D.E.,DAVID y FIX. Persistence in a chain of multiple events when there is simple dependence" Biometrica 49. 351-357 .(1962)

(3) BAUM y PETRIE. (1966) Statistical Inference for probabilistics functions of finite states Markov chains. Ann.Math,Stat. 37. 1554-1563

(4) CHIA KUEI TSAO. Admissibility and distribution of some probabilistic functions of discrete finite state Markov Chains .Ann.Math.Stat.(1968) Vol 39. pp.1646

(5) FISZ.M. Probability Theory and Mathematical Statistics

(6) GOODMAN.LEO.A. Simplified run tests and likelihood ratio tests for Markov Chains. Biometrica 45.181-197

(7) GOODMAN .LEO.A.(1964)The analysis of persistence in a chain of multiple events.Biometrica 51.4o5-411.

(8) MOOD.A.M. The distribution Theory of runs. Ann.Math. Stat. 11. 367-392:.(1940)

(9) PARZEN.E. Stochastic processes.Holden Day.S.Francisco.

(10) PRAIS.S.J. Measuring Social mobility. (1955)J.Roy.Sta. Soc.Ser. A.118. 56-66.

DILIGENCIA:

Reunido el Trib[illegible] en el día de la fecha, constituido por:

D. Alfonso Guiraum Martín

D. Mariano Gasca González

D. Rafael Infante Macías

D. Ramón Gutiérrez Jáimez

D. Takayuki Kawada

para juzgar la Tesis Doctoral [illegible] Licenciado Don(s)

María Teresa Miranda León

acordó por unanimidad otorg[illegible] califica-

n de Sobresaliente Cum laude

para que conste, [illegible] por los

componentes del [illegible] diligen-

[illegible].

[illegible] 18 Mayo de 1977.

El Presidente,

El Vocal, El Vocal, El Vocal,