HOMENAJE AL PROFESOR SIXTO RIOS
TRABAJOS DE ESTADISTICA Y DE INVESTIGACION OPERATIVA
Vol. 36, Núm. 3, 1985, pp. 21 a 44

ALGUNOS METODOS GENERALES PARA LA CONSTRUCCION DE TESTS PARAMETRICOS Y NO PARAMETRICOS

C.B. Bell

San Diego State University. San Diego, California 92118, USA.

Se construyen algunos tests para varias hipótesis paramétricas y no paramétricas. La metodología utiliza un nuevo sistema de coordinados, cuyos componentes son el estadístico suficiente mínimo y un estadístico independiente y complementario a tal estadístico.

Palabras Clave: Estadístico suficiente mínimo; Bondad de ajuste; Orbitas; Estadísticos de Kolmogorov; Parámetro auxiliar.

Clasificación AMS (1980): Primaria, 62F03, 62G10.

Some general methods for the construction of parametric and non-parametric tests

Tests are constructed for various parametric and non-parametric hypotheses. The methodology is based on a new coordinate system, whose components are the relevant minimal sufficient statistic, and a statistic independent of and complementary to that statistic.

Keywords: Minimal sufficient statistic; Goodness-of-fit; Orbits; Kolmogorov statistic; Nuisance parameter.

AMS Classification (1980): Primary, 62F03, 62G10.

1. INTRODUCCION

El objetivo de este trabajo es desarrollar una metodología para la construcción de tests en casos de "amplias" familias NP (no-paramétricas) y también para familias paramétricas (relativamente "pequeñas").

En este artículo se ilustra la metodología para 5 familias paramétricas y 5 familias NP . Estas familias se definen en la Sección 2.

Se dividen las hipótesis usuales en 3 conjuntos - BDA (Bondad de Ajuste); BDA con PA (parámetros auxiliares); y Muestreos Múltiples. Estos conjuntos se explican en la Sección 3.

En la Sección 4 se presenta la herramienta para la construcción y se da una regla general, que también es aplicable para familias paramétricas y familias NP . Todos los tests considerados son basados en una reordenación de datos originales por una TBD (transformación básica de datos). Las coordenadas nuevas son estadísticos suficientes y estadísticos independientes de estadísticos suficientes.

Las Secciones 5, 6 y 7 tratan los tres tipos de hipótesis en términos de la *regla general* y las reordenaciones . *Ejemplos* de tipos de tests para cada familia se encuentran en aquellas secciones.

Finalmente, en la Sección 8, se discuten las conclusiones y problemas no resueltos. La bibliografía final da solamente una parte de las referencias de interés.

2. FAMILIAS PARAMETRICAS Y NP

Una gran parte de las hipótesis estadísticas se puede dividir en tres conjuntos. Antes de discutir estos conjuntos, queremos introducir varias familias "paramétricas" y NP de funciones de distribución (FD's).

(A) *Familias Paramétricas*

(1) $\Omega(\mathrm{EXP}) = \{\mathrm{Exp}(\lambda) : \lambda > 0\}$

(2) $\Omega(\mathrm{UNI}) = \{U(0,\theta) : \theta > 0\}$

(3) $\Omega(\mathrm{GAU}) = \{N(\mu,\sigma) : \sigma > 0, |\mu| < \infty\}$

(4) $\Omega(\mathrm{PAR}) = \{P_a(A,s) : A > 0, s > 1\}$, donde

$F(z;A,s) = 1 - (Az^{-1})^s$, $x > A > 0$.

(5) $\Omega(\mathrm{GAU2}) = \{N(\mu_1, \sigma_1, \rho, \mu_2, \sigma_2) : \sigma_j > 0 , |\mu_j| < \infty\}$

Para ilustrar las ideas, vale la pena considerar distribuciones univariadas y bivariadas para los problemas NP.

(B) *Familias NP*

(1) $\Omega_2 = \{F : F(.)$ es continua $\}$

(2) $\Omega(\mathrm{SIM}) = \{F \in \Omega_2 : F(x) + F(-x) = 1$ para toda $x\}$

(3) $\Omega(\mathrm{SIM2}) = \{\mathbf{F}(.,.) : \mathbf{F}$ es continua y $\mathbf{F}(x,y) = \mathbf{F}(y,x)$ para toda x e $y\}$

(4) $\Omega(\mathrm{CSIM}) = \{\mathbf{F}(.,.)$: existe una función $h(.)$ tal que la densidad satisface $\mathbf{f}(x,y) = h(x^2 + y^2)$ para toda x e y $\}$.

(5) $\Omega(\mathrm{IND2}) = \{\mathbf{F}(.,.)$: existen $F_1, F_2 \in \Omega$ tal que $\mathbf{F}(x,y) = F_1(x)\, F_2(y)$ para toda x e y $\}$.

Como es usual, llamamos "paramétricas" las familias que tienen una parametrización de dimensión finita. Además, llamamos "NP" las familias que no poseen la propiedad anterior.

3. TIPOS DE HIPOTESIS

Como dijimos anteriormente, se pueden dividir las hipótesis usuales en tres conjuntos. El primer conjunto comprende hipótesis sobre muestreos múltiples: el problema de dos muestreos, c-muestreos, aleatoridad, etc.

Estas hipótesis están ilustradas en la Tabla 3.1. Se observa que *usualmente* estas hipótesis pertenecen a Ω(GAU) en los casos paramétricos y pertenecen a Ω_2 en los casos NP. Pero las ideas básicas de construcción de tests son casi equivalentes para las familias indicadas en Tabla 3.1. Además, valores específicos de parámetros no son de interés general para ser incluídos en las hipótesis de la Tabla 3.1.

Tabla 3.1. *Ejemplos de hipótesis de muestreos múltiples.*

Datos $\mathbf{Z} = (z_1,\ldots,z_N)$	Hipótesis NP [Familia]	Hipótesis Paramétrica [Familia]
(1) $(x_1,\ldots,x_m;\ y_1,\ldots,y_n)$	$F_1 = F_2$ [Ω(SIM)]	$\theta_1 = \theta_2$ [Ω(UNI)]
(2) $\{x_{ij}: 1 \leq j \leq n_i,\ 1 \leq i \leq c\}$	$F_1 = \ldots = F_c$ [Ω_2]	$\lambda_1 = \ldots = \lambda_c$ [Ω(EXP)]
(3) $(x_1,\ldots,x_N)$	$F_1 = \ldots = F_N$ [Ω(CSIM)]	$\beta=0$ (Regresión) [Ω(GAU)]

El segundo conjunto de hipótesis comprende casos en que todos los parámetros son específicamente determinados.

Se ve que estas hipótesis son muy diferentes; pero, las construcciones de tests son "paralelas", como se verá en la sección que sigue.

El tercer conjunto de hipótesis de interés aquí consiste en casos en que algunos parámetros son especificados y otros no son especificados. Estos parámetros no especificados se llaman "*parámetros auxiliares*".

***Tabla 3.2** Ejemplos de hipótesis con todos los parámetros especificados.*
(Datos : $\mathbf{Z} = (z_1,...,z_n)$)

Hipótesis NP BDA = Bondad de Ajuste	Hipótesis Pamétrica VAPA = Valor del Parámetro
$F = F_0\ [\Omega_\theta]$	$\theta = \theta_0\ [\Omega(UNI)]$
$F = F_0\ [\Omega(CSIM)]$	$(A,s) = (A_0,s_0)\ [\Omega(PAR)]$
$F = F_0\ [\Omega(IND2)]$	$\mu = \mu_0\ [\Omega(GAU)]$

Las obras de muchos autores, e.g. Feller (1938), Scheffé (1943), Lehmann y Stein (1949), Lehmann y Scheffé (1950,1955), Bell (1964 a,b), Lehmann (1959) indican que los tests para las hipótesis de la Tabla 3.1 son tests similares cra (con respecto a) las familias indicadas. Pero, igualmente se puede demostrar esta propiedad para los tests de las hipótesis en la Tabla 3.3. Como todos las parámetros están especificados en la Tabla 3.2, la estructura de los tests es diferente.

***Tabla 3.3** Ejemplos de hipótesis con parámetros auxiliares.*
(Datos : $\mathbf{Z} = (z_1,...,z_N)$)

Hipótesis NP [ADF = Ajuste de Familia]	Hipótesis Paramétrica [BDA con PA = Bondad de Ajuste con Parámetros Auxiliares]
$F \in \Omega(SIM)$	$F \in \Omega(GAU)$
$\mathbf{F} \in \Omega(CSIM)$	$F \in \Omega(UNI)$
$\mathbf{F} \in \Omega(SIM2)$	$F \in \Omega(EXP)$
$\mathbf{F} \in \Omega(IND2)$	$\rho=0\ [\Omega(GAU)]$

4. LA HERRAMIENTA PARA CONSTRUIR LOS TESTS

La experiencia en construir tests indica que cuando interesa el valor del parámetro se debe utilizar un estadístico suficiente. Pero, el "Teorema de Estructura de Neyman" (véase, e.g. Lehman 1959, p. 134) indica que se tiene que utilizar un estadístico independiente de un estadístico suficiente en construir tests similares. (Véase, también, Basu 1955,1958; Pathak y Kumar 1977 a,b).

Se pueden precisar estas ideas con algunas definiciones.

Definición 4.1 Sea Ω' una familia de FD's (funciones de distribución) . Un estadístico S(Z) es un ESM (estadístico suficiente mínimo) para Ω' si para cada $S^*(Z)$ suficiente para Ω', existe ψ^* tal que $S(Z) = \psi^*(S^*(Z))$. Se observa inmediatamente que la existencia de un ESM no es necesaria para Ω' arbitrario.

Definición 4.2 Consideremos una familia Ω' para la cual (a) existe un ESM, S(Z), para Ω' ; y (b) existe un estadístico N(Z) tal que

(i) $\delta(Z) = (S(Z), N(Z))$ es una transformación 1-1 en casi todos los puntos
(ii) S(Z) y N(Z) son independientes.

Entonces,
(α) $\delta(.)$ se llama la TBD (transformación básica de datos)
(β) N(Z) se llama *una versión* del REM (ruido estadístico máximo).
(γ) cada conjunto $\{\mathbf{z} : S(\mathbf{z}) = s^*\}$ se llama una *ORBITA*.

Está claro que el ESM, S(Z) , y el REM, N(Z) , son complementarios, en el sentido de que, conociendo los dos, se pueden reconstruir los datos originales con probabilidad uno. Es decir, N(Z) es un estadístico ancilario que es también complementario de S(Z) .

Regla General
(1) Para inferencia relacionada con valores específicos del parámetro, utilizar el ESM, S(Z) . (E.g. Tests de BDA y VAPA, indicados en Tabla 3.2.)

(2) Para inferencia que no depende de valores específicos de los parámetros, utilizar una versión del REM, N(Z). (E.g. tests sobre muestreos múltiples como en la tabla 3.1; y tests de ADF y de BDA con PA indicados en Tabla 3.3.)

Esta Regla General se aplica en una forma u otra en casi todos los problemas de inferencia. (Véase Watson 1957; Pathak y Kumar, 1977 a,b.) pero, en la práctica, se necesita la herramienta para construir el ESM y el REM .

Ejemplo 4.1 Consideremos $\Omega' = \Omega(EXP)$ y datos $Z = (x_1,...,x_N)$, $V = (y_1,...,y_N)$. El cociente de FDV's (funciones de verosimilitud) L(Z) / L(V) = $\exp\{\lambda[N(m_y - m_x)]\}$ no depende de λ siss (si y solamente si) $\sum_j x_j = \sum_j y_j(Nm_w = \sum_i w_i)$. Consecuentemente, se concluye que $S(Z) = \sum_j x_j$, es el ESM para $\Omega(EXP)$.

Ejemplo 4.2 Para $\Omega' = \Omega(\mathrm{SIM})$ y datos $\mathbf{Z} = (x_1,...,x_N)$ y $\mathbf{V} = (y_1,...,y_N)$, el cociente de FDV's es, cuando existe una *densidad* f(.), $L(\mathbf{Z})/L(\mathbf{V}) = \{f(x_1),...,f(x_N) \,/\, f(y_1),...,f(y_N)\}$ y no depende del parámetro f(.) siss $[|x|_{(1)},...,|x|_{(N)}] = [|y|_{(1)},...,|y|_{(N)}]$ es decir, siss los valores absolutos ordenados son iguales para **Z** y **V**. Por eso, el ESM es $S(\mathbf{Z}) = (|X|_{(1)},...,|X|_{(N)})$.

Un principio general para la construcción de los ESM's es el siguiente.

Teorema 4.1 Sea Ω' una familia "suficientemente regular" de FD's. Cuando para el cociente de FDV's, $Q = L(\mathbf{Z})/L(\mathbf{V})$ no depende del parámetro siss $S^*(\mathbf{z}) = S^*(\mathbf{v})$, entonces $S^*(\mathbf{Z})$ es el ESM para Ω'.

Las 10 familias Ω' listadas en la Sección 2 son "suficientemente regulares". Sus ESM's se hallan en Tabla 4.1 al fin de esta sección. (No conocemos ningún método comparable al descrito en el Teorema 4.1 para la construcción de REM .)

Ejemplo 4.3 para $\Omega(\mathrm{EXP})$ y datos $\mathbf{Z} = (X_1,...,X_n)$. Consideremos $N_1(\mathbf{Z}) = [V_1,...,V_{N-1}]$, donde $V_r = [\Sigma_i X_i]\,[\Sigma_j X_j]^{-1}$; i= 1,...,r, j = 1,...,N y $N_2(\mathbf{Z}) = [U_1,...U_{N-1}]$, donde $U_j = X_j\,[\Sigma_i X_i]^{-1}$. $N_1(\mathbf{Z})$ y $N_2(\mathbf{Z})$ son ambos versiones del REM , porque satifacen las condiciones de la Def. 4.2. Es decir, los dos son independientes de $S(\mathbf{Z})$, y complementarios de $S(\mathbf{Z})$.

Ejemplo 4.4 Para $\Omega(\mathrm{SIM})$ datos $\mathbf{Z} = (x_1,...,x_N)$, consideremos $N_1(\mathbf{Z}) = (\mathbf{R}^*, \boldsymbol{\varepsilon})$, donde $\mathbf{R}^* = (R(|x_1|),...,R(|x_N|))$, y $\boldsymbol{\varepsilon} = (\varepsilon(x_1),...,\varepsilon(x_N))$ con $\varepsilon(u) = 1$ si $u \geq 0$; y $\varepsilon(u) = 0$ si $u < 0$. $N_1(\mathbf{Z})$ satisface las condiciones de la Definición 4.1 y, por lo tanto, es una versión del REM.

Se define $S' = \{\gamma\}$, el conjunto de $k^* = (N!)\,(2^N)$ permutaciones de signos y posiciones de$(x_1,...,x_N)$; (b) $h(\mathbf{Z}) = \Sigma_j\, j\, x_j$ (j = 1,...,n) ; y $R(h(\mathbf{Z})) = \Sigma\, \boldsymbol{\varepsilon}\{h(\mathbf{Z}) - h(\gamma(\mathbf{Z}))\}$, donde la suma es sobre las permutaciones de S'. $N_2(\mathbf{Z}) = R(h(\mathbf{Z}))$ es también una versión del REM para $\Omega(\mathrm{SIM})$

De estos ejemplos uno puede comenzar a inferir lo siguiente.

Teorema 4.2 Sea Ω' una familia con ESM, $S_1(\mathbf{Z})$, y REM, $N_1(\mathbf{Z})$.

(i) Para cualquier otro ESM, $S_2(\mathbf{Z})$, es válido que $S_1(\mathbf{Z}) \equiv S_2(\mathbf{Z})$.

(ii) Es posible que exista otra versión del REM, $N_2(\mathbf{Z})$, tal que $N_1(\mathbf{Z}) \neq N_2(\mathbf{Z})$.

(iii) Para cada función $\psi(.)$, el estadístico $T = \psi(N(Z))$ tiene una FD $Q(.)$, tal que $P\{\psi(N(Z)) \leq w \mid J\} = Q(w)$ para toda w y toda J en la familia Ω'.

Es importante ahora introducir los tipos de estadísticos que se utilizan en los tests deseados. (ver las definiciones correspondientes, e.g., Bell, 1964a. Bell y Haller, 1969; Bell y Kurotschka, 1971; Bell y Mason 1985).

Definición 4.3

(i) Un estadístico T(z) es NPDL (no paramétrico y de distribución libre) con respecto a Ω', si existe una FD Q, tal que, para cada w y cada J en Ω', $P\{T(Z) \leq w \mid J\} = Q(w)$.

(ii) Un conjunto $\{T^*(J,Z) : J \in \Omega'\}$ es PDL con respecto es Ω', si existe una FD,Q^*, tal que, para cada w y cada $J \in \Omega'$, $P\{T^*(J;Z) \leq w \mid J\} = Q^*(w)$.

Ejemplo 4.5 En el Ejemplo 4.1, definir $T^*(F_\lambda, Z) = 2\lambda\, S(Z)$. Entonces $\{T^*(J,Z) : J \in \Omega(EXP)\}$ es PDL cra $\Omega(EXP)$ con $Q^* = \chi^2(2N)$. Si se define en el Ejemplo 4.3, $T(Z) = \sup |(N-1)^{-1} \sum_j \boldsymbol{\varepsilon}(w-V_j) - w|; j=1,\ldots,N-1$ sobre $0<w<1$, entonces T(Z) es NPDL cra $\Omega(EXP)$, con $Q = K - S(N-1)$, la FD del estadístico de Kolmogorov, con parámetro N-1.

Ejemplo 4.6 Véanse los Ejemplos 4.2 y 4.4. El estadístico de Wilcoxon, $T(Z) = \sum_j \varepsilon(x_j)\, R(|x_j|)$ $(j=1,\ldots,n)$ es NPDL cra $\Omega(SIM)$. También $N^{-1/2} \sum_j \Phi^{-1}(G_F(|x_j|))$; $j = 1,\ldots N$, es PDL cra $\Omega(SIM)$ con $Q^* = \Phi$, cuando $G_F(w) = P\{|x_1| \leq w \mid F\}$ y Φ es la FD de una distribución N(0,1).

En muchos casos NP se pueden construir estadísticos NPDL utilizando permutaciones como se ilustra en el Ejemplo 4.4 (Véase e.g., Bell y Sen 1984.).

Definición 4.4 Sea Ω' una familia de FD's teniendo un ESM, $S(\mathbf{Z})$, tal que cada órbita $\{S(\mathbf{z}) = s^*\}$ es finita con k^* puntos. Sea S' un conjunto de k^* permutaciones $\{\gamma\}$ tal que, para casi cada $\mathbf{z}$, $S'(\mathbf{z}) = \{\gamma(\mathbf{z}) : \gamma \in S'\}$ es una órbita. Sea h(.) una función, tal que, $P\{h(\mathbf{z}) = h(\gamma(\mathbf{z}))\} = 0$ si $\gamma(\mathbf{z}) \neq \mathbf{z}$. Se define, $R(h(\mathbf{Z})) = \sum \boldsymbol{\varepsilon}\{h(\mathbf{Z}) - h(\gamma(\mathbf{Z}))\}$ donde la suma es sobre todas las permutaciones de S'.

(i) h(.) se llama una *función de Pitman* (FP), cra S' y Ω', y
(ii) $R(h(\mathbf{Z}))$ se llama un *estadístico de Pitman* (EP) cra S' y Ω'. (Véase Bell y Bellot (1965).)

Teorema 4.3

(i) Sea h(.) una función de Pitman cra S' y Ω'(como en Def. 4.4). Entonces, R(h(Z)) es NPDL cra Ω', con FD, Q, tal que $Q(r) = r[k^*]^{-1}$, para $r = 1, 2,...,k^*$, es decir que R(h)Z)) tiene una FD discreta uniforme.

(ii) R(h(Z)) es una versión del REM cra Ω'.

(iii) Además, si el ESM S(Z) es completo cra funciones acotadas, un estadístico T(Z) es NPDL cra Ω' siss existen una función Pitman h(.) y una función W(.), tal que $T(Z) \equiv W[R(h(Z))]$.

Se observa que las propiedades NPDL y PDL son igualmente aplicables a familias "paramétricas" como Ω(GAU) y a familias "no-paramétricas" como Ω_2. "DL"= "distribución-libre" indica que existe una FD, Q(.), que no depende en la FD, J(.), de Ω'. Análogamente, "NP"= " no paramétrico" indica que el parámetro no aparece en la definición del estadístico. Esta última propiedad es muy importante en los casos en que el parámetro es desconocido.

Otro caso importante a considerar es que, para Ω_2, Ω(SIM), Ω(SIM2) y Ω(IND2), las órbitas son finitas con $k^* = N!$, $[N!]\ [2^N]$, $[N!]\ [2^N]$ y $[N!]^2$, respectivamente, para muestreos aleatorios de tamaño N . Por eso, el Teorema 4.2 es aplicable a esas familias, que también son completas.

Para Ω(CSIM) y las familias paramétricas Ω(EXP), Ω(UNI), Ω(GAU), Ω(PAR) y Ω(GAU2) , las órbitas son infinitas y el Teorema 4.2 no es aplicable, pero para esas familias se puede demostrar el siguiente:

Teorema 4.4 Si Ω' es una de las 6 familias indicadas en el párrafo anterior. Entonces

(1) El ESM, S(Z), es completo para funciones acotadas cra Ω'.

(2) Un estadístico T(Z) es NPDL cra Ω' siss existe una versión, N(Z), del REM y una función ψ(.) tal que $T(Z) \equiv \psi[N(Z)]$.

(3) T(Z) es NPDL cra Ω' siss para cada w, y para casi cada s* $P\{T(Z) \leq w\} = P\{T(Z) \leq w \mid S(Z) = s^*\}$.

En general, el autor no ha podido hallar un conjunto "manejable" S' de permutaciones para las órbitas en los casos de familias paramétricas. Sin embargo, para datos $\mathbf{Z} = (x_1,y_1,...,x_N,y_N) = (z_1,z_2,...z_N)$ y Ω(CSIM) se tiene (Bell y Smith, 1972; Bell, 1975).

Teorema 4.5 Sea $\Omega' = \Omega(\text{CSIM})$.

(1) El ESM, $S(\mathbf{Z}) = [R_{(1)},...,R_{(n)}]$, los radios ordenados, donde $R_j^2 = x_j^2 + y_j^2$.

(2) Una versión del REM es $N(\mathbf{Z}) = [\mathbf{R}^{**}, \boldsymbol{\theta}^{**}]$ donde $\mathbf{R}^{**} = [R(R_1),...,R(R_n)]$ y $\boldsymbol{\theta}^{**} = [\theta_1,...\theta_n]$ con $(\tan\theta_j) = y_j[x_j]^{-1}$.

(3) $S(\mathbf{Z})$ es completo para funciones acotadas cra Ω'.

Para "permutaciones" se puede considerar $S^* = \{$todas las permutaciones de subíndices $(1,...,n)$ y todas las rotaciones $(\gamma_1,...,\gamma_n)$ donde $0 \le \gamma_j < 2\pi$ y γ_j es una transformación de $\theta_j\} = \{\gamma^*\}$

Definición 4.5

(1) $h^*(.)$ es una *función generalizada de Pitman* (FGP) si $P\{h^*(\mathbf{Z}) = h^*(\gamma^*(\mathbf{Z}))\} = 0$ cuando $\mathbf{Z} \neq \gamma^*(\mathbf{Z})$.

(2) $R^*(h^*(\mathbf{Z}))$ es un *estadístico generalizado de Pitman* (EGP) cra S^* y $\Omega(\text{CSIM})$, si $R^*(h^*(\mathbf{Z})) = P\{h^*(\gamma^*(\mathbf{Z})) \le h^*(\mathbf{Z}) \mid S^*(\mathbf{Z})\}$ para casi cada $\mathbf{Z}$.

Ejemplo 4.7 Sea $h^*(z) = \sum_j j^5(x_j^3 + 2y_j^3)$; $j=1,...,n$

(1) h^* es una FGP cra S^* y $\Omega(\text{CSIM})$; y

(2) $R^*(h^*(\mathbf{Z}))$ es un EGP cra S^* y $\Omega(\text{CSIM})$.

(3) $R^*(h^*(\mathbf{Z}))$ es una versión del REM para $\Omega(\text{CSIM})$.

(4) $R^*(h^*(\mathbf{Z})) \sim Q = U(0,1)$.

Teorema 4.6 Sea h^* una FGP cra S^* y $\Omega(\text{CSIM})$.

(1) $R^*(h^*(\mathbf{Z}))$ es una versión del REM para $\Omega(\text{CSIM})$.

(2) $R^*(h^*(\mathbf{Z})) \sim Q = U(0,1)$.

Además, $T(\mathbf{Z})$ es NPDL cra $\Omega(\text{CSIM})$ siss existen $\psi^*(.)$ y FGP, $g^*(.)$, tal que $T(\mathbf{Z}) \equiv \psi^*(R^*(g^*(\mathbf{Z})))$.

Utilizando las herramientas desarrolladas, se puede comenzar a dar tests para varias hipótesis de interés. Un sumario de los ESM's y REM's, se puede encontrar en la Tabla 4.1.

Tabla 4.1 *ESM's y REM's para la construcción de tests*

Familias	Datos	ESM , S(Z)	Versiones Invariantes	FP	Permutaciones	Notación
$\Omega(EXP)$	$X_1,\ldots,X_N$	$N m_X$	$(V_1^*,\ldots,V_N^*)$	N-C	N-C	$V_r^*=(\Sigma_j X_j)(Nm_X)^{-1}$
$\Omega(UNI)$	$X_1,\ldots,X_N$	$X_{(N)}$	$(U_1^*,\ldots,U_{N-1}^*,\mathbf{R(X)})$	N-C	N-C	$\mathbf{R(X)})=((R(X_1),\ldots R(X_N))$ $U_r^*= X_{(r)}(X_{(N)})^{-1}$
$\Omega(GAU)$	$X_1,\ldots,X_N$	(m_X, s^2)	$(W_1,\ldots,W_N)$	N-C	N-C	$W_r = (X_r - m_X)s^{-1}$
$\Omega(PAR)$	$X_1,\ldots X_N$	$(X_{(1)}, \hat{s})$	$(D_{(1)}^*,\ldots,D_{(N-2)}^*)$	N-C	N-C	$D_{(r)}^*= D_r D_{N-1}^{-1}$; $\hat{s}= ND_N^{-1}$ $D_r= \Sigma_j \ln X_{(j)} +(N-r)\ln X_{(r+1)} - N\ln X_{(1)}$
$\Omega(GAU2)$	$\mathbf{U}_1,\ldots,\mathbf{U}_n$ $(\mathbf{U}_r=(X_r,Y_r)^T)$	$(\mathbf{m}_U, \hat{s})$	$(\mathbf{V}_1,\ldots,\mathbf{V}_n)$	N-C	N-C	$\mathbf{V}_r= (\mathbf{U}_r- \mathbf{m}_U)^t(\hat{s})^{-1}(\mathbf{U}- \mathbf{m}_U)$ $\hat{s}= n^{-1}\Sigma_r(\mathbf{U}_r - \mathbf{m}_U) (\mathbf{U}_r - \mathbf{m}_U)^t$
Ω_2	$X_1,\ldots,X_N$	$(X_{(1)},\ldots,X_{(N)})$	$\mathbf{R(X)}$	$\Sigma_r rX_r$	N! permutaciones de índices.	
$\Omega(SIM)$	$X_1,\ldots,X_N$	$(\lvert X\rvert_{(1)},\ldots,\lvert X\rvert_{(N)})$	$(\mathbf{R^*(X)},\boldsymbol{\varepsilon}\mathbf{(X)})$	$\Sigma_r X_r$	2^N permutaciones de signos.	$\mathbf{R^*(X)} = (R(\lvert X_1\rvert),\ldots,R(\lvert X_N\rvert))$ $\boldsymbol{\varepsilon}\mathbf{(X)} = (\boldsymbol{\varepsilon}(X_1),\ldots,\boldsymbol{\varepsilon}(X_N))$
$\Omega(SIM2)$	$U_1,V_1,\ldots,U_n,V_n$	$(U_{(1)},\lvert V^{(1)}\rvert,\ldots,U_{(n)},\lvert V^{(n)}\rvert)$	$(\mathbf{R(U)},\boldsymbol{\varepsilon}\mathbf{(V)})$	$\Sigma_r V_r$	2^n permutaciones de signos.	$\mathbf{R(U)} = (R(U_1),\ldots,R(U_n))$
	$U_r= X_r + Y_r$ $V_r= Y_r - X_r)$			$\Sigma_r U_r V_r$	n! permutaciones de v's	$\boldsymbol{\varepsilon}\mathbf{(V)} = (\varepsilon(V_1),\ldots\varepsilon(V_n))$; $V^{(r)}$ la V asociada con $U_{(r)}$.
$\Omega(CSIM)$	$R_1,\ldots R_n,\theta_1,\ldots,\theta_n$	$(R_{(1)},\ldots R_{(n)})$	$(R(R_1),\ldots,R(R_n))$, $(\theta_0,\ldots.\theta_n)$	$\Sigma_r r^5(X_r^3+2Y_r^3)$	rotaciones y perm. de índices.	$R_j^2= X_j^2+Y_j^2$; $\tan\theta_j= Y_j(X_j)^{-1}$
$\Omega(IND2)$	$X_1,\ldots X_n,Y_1,\ldots,Y_n$	$(\mathbf{X^*}, \mathbf{Y^*})$	$(\mathbf{R(X)},\mathbf{R(Y)})$	$\Sigma_j X_j Y_j$	n! perm. de y's	$\mathbf{X^*}=(X_{(1)},\ldots,X_{(n)})$;$\mathbf{Y^*}= (Y_{(1)},\ldots,Y_{(n)})$ $\mathbf{R(X)}=(R_X(X_1),\ldots,R_X(X_n))$

N-C = No Conocido.

5. TESTS BDA (BONDAD DE AJUSTE) $H_0 : F = F_0$

5.1 VAPA $H_0 : \gamma = \gamma_0$ PARA FAMILIAS PARAMETRICAS .

Aquí, los estadísticos deben contener el parámetro y el ESM, S(**Z**) .

[Ω(EXP)] $H_0 : \lambda = \lambda_0$.

El estadístico es $2\lambda_0 \sum_j x_j$; j=1,...,N, ~ (H_0) $\chi^2(2N)$

[Ω(UNI)] $H_0 : \theta = \theta_0$.

El estadístico es $X(N)\ \theta_0^{-1}$, ~ (H_0) $P_d(N)$, es decir $P\{Z(N)\theta_0^{-1} \leq v\} = v^N$, para $0 \leq v \leq 1$.

[Ω(GAU)] $H_0 : (\mu,\sigma) = (\mu_0,\sigma_0)$.

El ESM es (m_x, s) donde m_x , s son independientes. Aquí, se utilizan tests independientes basados en $(N-1)\ s^2\sigma_0^{-2}$ y en $[m_x - \mu_0]\ s^{-1}\ N^{-1/2}$.

[Ω(PAR)] $H_0 : (A,s) = (A_0,s_0)$, donde la FD es $F(z) = 1 - (Az^{-1})$, $z > A > 0$.

El ESM aquí (Véase la Tabla 4.1 y Bell, Ahmad, Park, y Lui, 1982) es $(\hat{A}, \hat{s}) = (X(1), D_N^{-1})$. Puesto que $\hat{A}$ y $\hat{s}$ son independientes, se recomiendan tests independientes basados en $2\,N\hat{s}\,s^{-1}$ ~(H_0) $\chi^2(2N-2)$ y $\hat{A}\,A_0^{-1}$ ~ (H_0) $P_a(1,1)$.

[Ω(GAU2)]H_0: $(\mu,\Sigma) = (\mu_0,\Sigma_0)$.

El ESM es $(\mathbf{m}_U,\mathbf{S})$ (Véase Tabla 4.1). $n[\ \mathbf{m}_U - \mu_0]^T\Sigma_0^{-1}[\ \mathbf{m}_U - \mu_0]$ ~ (H_0) $\chi^2(2)$ y S ~ (H_0) $W(n^{-1}\Sigma_0,\ n)$ la FD de Wishart.

Cada uno de estos estadísticos anterior puede ser obtenido utilizando el cociente de FDV's (funciones de verosimilitud).

Para las familias NP , algunos estadísticos son "reflexiones" de los casos de Ω(EXP) y Ω(GAU), en el sentido de que un conjunto de estadísticos importantes son de la forma $\sum H^{-1}(F_0(Z_{(j)})) = \psi(H,F_0,\mathbf{Z})$. Algunas veces tales estadísticos tienen FD's "manejables".

5.2 BDA H_0: $F = F_0$

Teorema 5.1

(1) Si F_0 y $H \in \Omega_2$, entonces $\psi \equiv \psi(H,F_0,Z) = (d)\ \Sigma_j W_j$, donde $W_1,\ldots,W_r$ son i.i.d. H. En particular ,

(2) Si $H = Exp\ (0.5)$, entonces $\psi \sim \chi^2(2N)$; y

(3) Si $H = \Phi$, entonces $N^{-1/2}\psi \sim \Phi$.

El otro conjunto de estadísticos que se consideran aquí, no tiene relación con los estadísticos paramétricos; estan basados en una de las ideas de Kolmogorov (1933).

Definición 5.1

(1) $(V_1,\ldots,V_k) \sim$ E - 0 - H(k) , la FD de estadísticos de orden si $(V_1,\ldots,V_k) = (d)\ (Y_{(1)},\ldots,Y_{(k)})$, donde $Y_1,\ldots,Y_k$ son i.i.d. H(.).

(2) Si $(V_1,\ldots,V_k) \sim$ E - 0 - H(k) , donde $H(.) \in \Omega_2$, entonces $D(H,V) = \sup |k^{-1} \Sigma_j \varepsilon(z - V_j) - H(z)| \sim$ K - S(k) , la FD del estadístico de Kolmogorov con parámetro k.

Teorema 5.2 (Kolmogorov) La FD de D(H,V) no depende de $H \in \Omega_2$, y por tanto es PDL cra Ω_2.

Ahora, se pueden dar varios tests para BDA.

$[\Omega_2]$ $H_0 : F= F_0 \in \Omega_2$.

Aquí, $D(F_0,Z) \sim (H_0)$ K - S(N); y $[- 2\ \Sigma_j \ln F_0(x_{(j)})] \sim (H_0)\ \chi^2(2N)$.

$[\Omega(SIM)]$ H_0: $F= F_0 \in \Omega(SIM)$.

El ESM es $S(\mathbf{Z}) = \mathbf{Z}^* = (|x|_{(1)},\ldots,|x|_{(N)})$; y $G(F_0,w) = P\{|x_1| \leq |w|\ F_0\}$. En este caso, se puede considerar, p.e.:

$D((G(F_0,.); \mathbf{Z}^*) \sim (H_0)$ K - S(N) y $N^{-1/2}\Sigma_j \Phi^{-1}(G(F_0,|x|_{(j)})) \sim (H_0)\ \Phi$.

Las familias bidimensionales presentan situaciones algo diferentes.

$[\Omega(CSIM)]$ $H_0 : F = F_0 \in \Omega(CSIM)$.

Sea $R_j^2 = X_j^2 + Y_j^2$; $\tan \theta_j = Y_j / X_j$ para los datos $\mathbf{Z} = (X_1,Y_1,X_2,Y_2,\ldots X_n,Y_n)$.

El ESM es $\mathbf{R}^{**} = [R_{(1)},...,R_{(n)}]$, y dos estadísticos de interés son los siguientes. $D(J(\mathbf{F}_0,\mathbf{R}^{**}) \sim (H_0)$ K - S(n), y $-2\sum_j \ln J(\mathbf{F}_0, R_{(j)}) \sim (H_0)\ \chi^2(2N)$, donde $J(\mathbf{F}_0 ; w) = P\{R_1 \leq w|\mathbf{F}_0\}$.

$[\Omega(\text{IND2})]\ H_0 : \mathbf{F} = \mathbf{F}_0 \in \Omega(\text{IND2})$.

$\mathbf{F}_0 \in \Omega(\text{IND2})$ implica que existen $F_1(.)$ y $F_2(.)$ tales que $\mathbf{F}_0(x,y) = F_1(x)F_2(y)$ para toda (x,y). Pueden también considerarse los tests basados en $D(F_1,\mathbf{X})$ y $D(F_2,\mathbf{Y})$; ambos tienen la FD K - S(n). También, se pueden combinar los dos estadísticos de Fisher-Pearson para formar $T^* = -2\,[\sum_j \ln F_1(x_{(j)}) + \sum_j \ln F_2(y_{(j)})] \sim (H_0)\ \chi^2(4N)$.

$[\Omega(\text{SIM2})]\ H_0 : \mathbf{F} = \mathbf{F} \in \Omega(\text{SIM2})$.

Como se ha indicado anteriormente, vale la pena transformar los datos. $\tau(x_j,y_j) = (U_j, V_j)$ donde $U_j = x_j + y_j$ y $V_j = y_j - x_j$.

Entonces se tiene el siguiente

Teorema 5.3 Sea $\mathbf{Z} = [x_1,y_1,...x_n,y_n]$ i.i.d. $\mathbf{F} \in \Omega(\text{SIM2})$ con función de densidad $\mathbf{f}(.,.)$; y $\tau(x_j,y_j) = (U_j,V_j)$.
Entonces,

(1) $\mathbf{W} = [U_1,V_1,...,U_n,V_n]$ son i.i.d. $\mathbf{G}$ tal que la función de densidad $\mathbf{g}(.,.)$ satisface $\mathbf{g}(u,v) = \mathbf{g}(u,-v)$ para toda (u,v).
(2) El ESM, $S(\mathbf{W}) = [U_{(1)}, |V|^{(1)}, U_{(2)}, |V|^{(2)}, ..., U_{(n)}, |V|^{(n)}]$ donde $V^{(r)}$ es el valor de V asociado con $U_{(r)}$ Es decir, e.g., si $U_{(r)} = U_2$ entonces $V^{(r)} = V_2$.
(3) Las FD's condicionales satifacen: $G(.|u) \in \Omega(\text{SIM})$ para cada u. Por lo tanto la FD marginal de V satisface $G^2(.) \in \Omega(\text{SIM})$.

Consecuentemente, se pueden introducir los siguientes estadísticos de contraste.

(a) $D(G_{10},U) \sim (H_0)$ K - S(n), donde $U_1,...,U_n$ son i.i.d. G_{10}.
(b) $D(G(G_{20}),V^*) \sim (H_0)$ K - S(n), donde $|V_1|,...,|V_n|$ son i.i.d. $G(G_{20})$.
(c) $-2\sum_j \ln G_{10}(U_{(j)}) \sim (H_0)\ \chi^2(2n)$.
(d) $-2\sum_j \ln G_{20}(|V|_{(j)}) \sim (H_0)\ \chi^2(2n)$.

6. BDA CON PA; AJUSTE DE FAMILIAS. H_0: $F \in \Omega'$

6.1 FAMILIAS PARAMETRICAS.

Aquí los estadísticos presentados se dividen en dos conjuntos: (a) estadísticos para BDA para el REM, y (b) modificaciones del estadístico original de Kolmogorov (1933).

(a) *Estadísticos de BDA para el REM*

Teorema 6.1 Sean $X_1,...,X_N$ i.i.d. F.

(1) Si $F \in \Omega(EXP)$, entonces una versión del REM es $N_1(\mathbf{Z}) = (V_1^*,...,V_{N-1}^*) \sim$ E - 0 - U(N-1), donde $V_j^* = (\Sigma_r X_r)(Nm_X)^{-1}$ (r= 1,...,j).

(2) Si $F \in \Omega(UNI)$, entonces una versión del REM es $N_2(\mathbf{Z}) = (\mathbf{U}^*,\mathbf{R})$ donde $\mathbf{U}^* = (U_1^*,...,U_{N-1}^*) \sim (H_0)$ E - 0 - U(N-1); $U_j^* = X_{(j)}[X_{(N)}]^{-1}$, $1 \le j \le N-1$; y $\mathbf{R} = [R(X_1),..., R(X_N)]$.

(3) Sea $F \in \Omega(GAU)$, y $\xi_1,...,\xi_N$ son i.i.d. Φ e independiente de $\mathbf{Z}$. Entonces, $N(\mathbf{Z}) = \mathbf{W} = (W_1,...,W_N)$ es una versión del REM, donde $W_r = (X_r - m_X)S^{-1}$. Además, $\xi'_1,...,\xi'_N$ son i.i.d. Φ, si $\xi'_r = m_\xi + S_\xi W_r$, donde S_ξ es la varianza de $(\xi_1,...,\xi_N)$.

(4) Sea $F \in \Omega(PAR)$ y $N_4(\mathbf{Z}) = \mathbf{D}^* = (D_1^*,..., D_{N-2}^*)$, donde $D_r^* = D_r(D_N)^{-1}$ y $D_r = \Sigma_j \ln X_{(j)} + (N - r) \ln X_{(r+1)} - n \ln X_{(1)}$ (j=1,...,r). Entonces, una versión del REM es $\mathbf{D}^* \sim$ E - 0 - U(N-2).

(5) Sea $\mathbf{F} \in \Omega(GAU2)$ y $N_5(\mathbf{Z}) = (V_1, V_2,...,V_n)$ de la Tabla 4.1. Entonces, una versión del REM es $N_5(\mathbf{Z})$.

Para $H_0^{(1)}$: $F \in \Omega(EXP)$, se recomienda $\Pi(1) = -2\, \Sigma_j \ln V_j^* \sim (H_0)\ \chi^2(2(N-1))$ y $D(U_0,\mathbf{V}^*) \sim (H_0)$ K - S(N-1), donde $U_0 = U(0,1)$ y $\mathbf{V}^* = (V_1^*,...,V_{N-1}^*)$.

Para $H_0^{(2)}$: $F \in \Omega(UNI)$, se recomienda $\Pi(2) = -2\, \Sigma_j \ln U_j^* \sim (H_0)\ \chi^2(2(N-1))$ y $D(U_0,\mathbf{U}^*) \sim (H_0)$ K - S(N-1).

Para $H_0^{(3)}$: $F \in \Omega(GAU)$, se puede utilizar $\Pi(3) = -2\, \Sigma_r \ln \Phi(\xi'_r) \sim (H_0)\ \chi^2(2N)$ y $D(\Phi,\mathbf{W}) \sim (H_0)$ K - S(N).

Para $H_0^{(4)}$: $F \in \Omega(PAR)$, dos estadísticos útiles son $\Pi(4) = -2\, \Sigma_j \ln D_j^* \sim (H_0)$ $\chi^2(2(N-2))$, y $D(U_0,\mathbf{D}^*) \sim (H_0)$ K - S(N-2).

Cuando $F \in \Omega(GAU2)$ y se tiene interés en $H_0^{(5)}: \rho = 0$, se usa el estadístico $T^{**} = r(n-2)^{1/2}(1-r^2)^{-1/2} \sim (H_0)\ t'(n-2)$, la FD de Student.

(b) *Estadísticos Modificados de Kolmogorov*

Recuérdese que, para $H_0: F = F_0 \in \Omega_2$, Kolmogorov consideró el estadístico $D(F_0,\mathbf{Z})$, en la Def. 5.1, donde $\mathbf{Z} = (X_1,\ldots,X_N)$. Lilliefors (1967, 69), Srinivasan (1970), Choi (1980), y Bell y Mason (1985) han remplazado F_0 por un estimador, cuando F_0 es desconocida, como en el caso de BDA con PA.

Se presentan aquí solamente los estimadores de tipo Lilliefors.

$[\Omega(EXP)]$: $\hat{F}(z) = 1 - \exp[-\hat{\lambda} z]$, $z > 0$, donde $\hat{\lambda} = (m_X)^{-1}$;

$[\Omega(UNI)]$: $\hat{F}(z) = z(X(N))^{-1}$, $0 < z < X(N)$;

$[\Omega(GAU)]$: $\hat{F}(z) = \Phi((z - m_X)S^{-1})$; y

$[\Omega(PAR)]$: $\hat{F}(z) = 1 - (\hat{A} z^{-1})^{\hat{S}}$, $z > \hat{A} > 0$, donde $\hat{A} = X(1)$, $\hat{S} = ND_N^{-1}$ que provienen de la Tabla 4.1.

En cada caso, se remplazan los parámetros por sus EVM's (estimadores de máxima verosimilitud).

Los estadísticos son $D(\hat{F}, \mathbf{Z})$ (Véase la Def. 5.1) y sus distribuciones nulas se hallan en Bell y Mason (1985).

Ejemplo 6.1 Para H_0: $F \in \Omega(EXP)$, el estadístico correspondiente es $D(\hat{F}, \mathbf{Z}) = \sup |F_n(z) - (1 - \exp[-\hat{\lambda} z])|$; $(z > 0)$. Las tablas de este estadístico se puede hallar en Lilliefors (1969) y Bell y Mason (1985). En los casos no-paramétricos, las formas de los estadísticos son, en general, diferentes.

6.2. FAMILIAS NP

Los estadísticos que se presentan aquí son principalmente relacionados a tests de independencia o cero-correlación. El estadístico, correlación de Rangos de Spearman, para datos $(X_1,Y_1),\ldots,(X_N,Y_N)$ es equivalente al estadístico $T''' = \sum_j R_x(X_j)R_Y(X_j)$. T''' se modifica de varias maneras como se indica abajo.

(a) *Modificaciones de* T'''_0 *Correlación de Rangos.*

[Ω(SIM)] $\mathbf{Z} = (X_1,...,X_N)$.

Se prueba la independencia de signo de $\varepsilon(X_j)$, y valor absoluto. El estadístico es $T(\mathbf{Z}) = \sum_j \varepsilon(X_j)R(|X_j|)$ debido a Wilcoxon (1945). Aquí también se utiliza el estadístico de signos, $S(N) = \sum_j \varepsilon(X_j)$.

[Ω(CSIM)] Los datos $\mathbf{Z} = (X_1,Y_1,...,X_n,Y_n)$ son transformados a $W = (R_1,\theta_1,...,R_n,\theta_n)$, donde $R_j^2 = X_j^2 + Y_j^2$ y $\tan \theta_j = Y_j / X_j$.

Los cuestiones aquí son la independencia de R y θ , y la uniformidad de θ sobre $(0, 2\pi)$. Para la independencia, se puede utilizar $T(S) = \sum_j R_R(R_j)R(\theta_j)$.

Para la uniformidad, un estadístico $h'' = \sum_j(M_j - nk^{-1})^2kn^{-1}$ tiene como distribución nula, aproximadamente la distribución $\chi^2(k\text{-}1)$, donde M_j es el número de θ's en el intervalo $((j\text{-}1)/2\pi, j/2\pi]$.

[Ω(IND2)] los datos son $\mathbf{Z} = (X_1, Y_1,..., X_n,Y_n)$ y T''' se usa directamente.

[Ω(SIM2)] Se transforma $(X_1,Y_1,...,X_n,Y_n)$ a $(U_1,V_1,...,U_n,V_n)$ donde $U_j = X_j+Y_j$ y $V_j = Y_j - X_j$.

Si $F \in \Omega$(SIM2) y momentos del segundo orden existen, entonces la correlación de X y Y es 0 . También, la FD, $G_2(.)$ de V , es un elemento de Ω(SIM). Por eso, se consideran los estadísticos siguientes: $\sum_j \varepsilon(V_j)$, $\sum_j \varepsilon(V_j)R(|V_j|)$, y $\sum_j R_V (V_j)R_U (U_j)$.

Para Ω(IND2), Pitman (1938) introdujo un test de permutaciones.(Véase Def. 4.4) Se puede extender su idea a otros casos.

(b) *Estadísticos de Permutaciones.*

Nos referimos aquí a los estadísticos de Pitman $R(h(\mathbf{Z}))$. (Véase Bell y Sen, 1984).

[Ω(SIM)] : $h = \sum_j z_j$, y permutaciones de signos.

[Ω(CSIM)] : $h = \sum_j r_j \theta_j$, y permutaciones de las r's.

$[\Omega(\mathrm{IND2})] : h = \Sigma_j x_j y_j$, y permutaciones de las y's.

$[\Omega(\mathrm{SIM2})] : h_1 = \Sigma_j v_j$, y permutaciones de los signos; y

$[\Omega(\mathrm{SIM2})] : h_2 = \Sigma_j u_j v_j$, y permutaciones de las v's.

Se observa que no se consideró "H_0: $F \in \Omega_2$ " porque no tiene sentido en este contexto. También se tiene el siguiente resultado.

Teorema 6.2 Cada estadístico de esta sección es NPDL cra la propia familia, y puede ser representado como $T(\mathbf{Z}) = \psi(N(\mathbf{Z}))$, donde $N(\mathbf{Z})$ es el propio REM. En la próxima sección, Ω_2 es la familia mas importante.

7. TESTS PARA MUESTREOS MULTIPLES

Las hipótesis que se consideran aquí son las siguientes

$H_0^{(1)} : F_1 = F_2$, donde $\mathbf{Z} = (X_1,...,X_m; Y_1,...,Y_n) = (\mathbf{X},\mathbf{Y})$ con $N = m + n$ (2-muestreos);
$H_0^{(2)} : F_1 = F_2 = .. = F_c$, donde $\mathbf{Z} = (X_{11},...,X_{1n_1},...,X_{c1},..., X_{cn_c})$ con $N = n_1+...+ n_c$ (c-muestreos); y
$H_0^{(3)} : F_1 = F_2 = .. = F_N$, donde $\mathbf{Z} = (X_1,...,X_N)$ (aleatoridad); y se fijan principalmente en $H_0^{(1)}$ y $H_0^{(3)}$.

Los estadísticos que se presentan aquí no varían mucho entre las familias $\Omega_2,\Omega(\mathrm{SIM})$, etc. Esos estadísticos fueron desarrollados principalmente para Ω_2 y luego adaptados para otras familias, utilizando una versión del REM, $N(\mathbf{Z})$, en cada caso.
Son de importancia aquí:

(a) El estadístico de Kolmogorov para dos muestras $D(\mathbf{V},\mathbf{W}) = \sup |m^{-1} \Sigma_j \varepsilon(z - V_j) - n^{-1} \Sigma_i \varepsilon(z - W_i)|$ $(j=1,...,m;\ i=1,...,n)$ $(-\infty < z < \infty)$ cuando $\mathbf{V} = (V_1,...,V_m)$ y $\mathbf{W} = (W_1,...,W_n)$;
(b) $S^*(\mathbf{U},\mathbf{W})$ = el conjunto de permutaciones que intercambian a lo menos una V_i y una W_j ; y
(c) $S'(N)$, el conjunto de $N!$ permutaciones de $(Z_1,...,Z_N)$.

[Ω_2] Para $H_0^{(1)}$, se pueden utilizar

(i) $\Sigma_j R(X_j)$, el estadístico de Mann-Whitney-Wilcoxon;
(ii) $R(h(\mathbf{Z}))$, el EP cra $S^*(\mathbf{X},\mathbf{Y})$ y Ω_2, donde $h(\mathbf{Z}) = m_X - m_Y$ o equivalentemente $h_1(\mathbf{Z}) = \Sigma_j X_j$ (Véase Def 4.4);
(iii) $D(\mathbf{X},\mathbf{Y})$.

Se nota aquí que los primeros dos estadísticos para $H_0^{(1)}$ fueron construidos para alternativas de diferencias en "localización." Para diferencias en "dispersión", Siegel y Tukey (1960) han hecho una modificación del test de Mann-Whitney-Wilcoxon. Bell y Chicarro (1966) investigaron generalizaciones del estadístico de Siegel-Tukey.

Para $H_0^{(3)}$, se puede utilizar $\Sigma_j\, j\, R(X_j)$, que está relacionada con T''' que parece en la Sección 6, y $R(h'(\mathbf{Z}))$ y EP cra $S'(N)$ y Ω_2, donde $h'(\mathbf{Z}) = \Sigma_j\, j\, X_j$.

[Ω(SIM)]
(i) $\Sigma_j R(|X_j|)$,
(ii) $D(\mathbf{X}^*,\mathbf{Y}^*)$, donde $\mathbf{X}^* = (X_{(1)},\ldots,X_{(m)})$, y $\mathbf{Y}^* = (Y_{(1)},\ldots,Y_{(n)})$; y
(iii) $R(h(\mathbf{Z}))$, donde $h(\mathbf{Z}) = \Sigma_j |X_j|$, son tres estadísticos para $H_0^{(1)}$.

Al remplazar las X's por |X|'s en las expresiones para Ω_2, se obtiene $\Sigma_j R(|X_j|)$ y $R(h''(\mathbf{Z}))$, donde $h''(\mathbf{Z}) = \Sigma j|X_j|$ para $H_0^{(3)}$.

[Ω(CSIM)]
Aquí, se remplaza X_j, en las expresiones para Ω_2, por R_j, el radio. Los estadísticos se convierten en

$\Sigma_j R(R_j)$, $D(\mathbf{R}_1,\mathbf{R}_2)$, $R(h_3(\mathbf{Z}))$, $\Sigma_j R(R_j)$ y $R(h_4(\mathbf{Z}))$ donde

$\mathbf{Z} = (X_1,Y_1,\ldots,X_n,Y_n)$, $\mathbf{R}_1 = (R_1,\ldots,R_m)$, $\mathbf{R}_2 = (R_{m+1},\ldots,R_N)$,

$h_3(\mathbf{Z}) = \Sigma_j R_j$, y $h_4(\mathbf{Z}) = \Sigma_j\, j\, R_j$.

[Ω(IND2)]
Aquí, $H_0^{(1)}$ equivale a $H_0 : \mathbf{F} = \mathbf{G}$, donde $\mathbf{F}(x,y) = F_1(x).\,F_2(y)$ y $\mathbf{G}(x,y) = G_1(x).\,G_2(y)$ para toda (x,y). Equivalentemente, se tienen dos hipótesis $H_0^* : F_1 = G_1$ y H_0^{**}: $F_2 = G_2$. Para H_0^* y también para H_0^{**} se utilizan los estadísticos de Ω_2 indicados arriba. El mismo tipo de solución vale para $H_0^{(3)}$.

[Ω(SIM2)]

Cuando F_1 y $F_2 \in \Omega$(SIM2), la hipótesis $H_0^{(1)}: F_1 = F_2$, resulta $H_0': F_V = G_V$ y $H_0'': F_U = G_U$ donde $U = X + Y$, y $V = Y - X$. Esta situación se trata como en el caso de Ω_2. Por otra parte, la situación es "paralela" para $H_0^{(3)}$.

La situación con $H_0^{(2)}$: $F_1 = F_2 = ... = F_c$ no se ha discutido hasta ahora. En cierto sentido, las soluciones relacionadas con $H_0^{(2)}$ son "intermedias entre" soluciones para $H_0^{(1)}$ y $H_0^{(3)}$. Por ejemplo, para $H_0^{(2)}$ y Ω_2, un estadístico usual es el de Kruskal - Wallis, a $\Sigma_j R_j^2 (n_j)^{-1} + b$ donde $R_j = n_i^{-1} \Sigma_j R(X_{ij})$; $(j=1,...n_i)$ y un EP es R(h(Z)), donde $h(Z) = \Sigma_i(m_{i.} - m_{..})^2$; $m_{i.} = n_i^{-1} \Sigma_j x_{ij}$, $m_{..} = \Sigma_{ij} x_{ij} / \Sigma_i n_i$.

Se pueden considerar a $H_0^{(1)}$, $H_0^{(2)}$, $H_0^{(3)}$ como hipótesis de independencia del número del muestreo y valores. Algunos de los estadísticos mencionados arriba son "reflejos" de aquella independencia.

Para las familias NP arriba indicadas, cada versión del REM está expresada en términos de rangos (o invariantes máximos) o permutaciones. La situación es diferente en las familias paramétricas. Los estadísticos son listados a continuación.

[Ω(EXP)]

(1) Para $H_0^{(1)}$, se puede utilizar $m_X(m_Y)^{-1} \sim (H_0)\ F(2m,2n)$, la distribución de Fisher.

(2) Para $H_0^{(3)}$, si $n_1 = ... = n_c = n = Nc^{-1}$, un estadístico importante es $\max(m_{1\cdot},..., m_{c\cdot})\ [\min(m_{1\cdot},...,m_{c\cdot})]^{-1}$; (donde $m_t = (\Sigma_j X_{tj})/n_i$; con $j = 1,...n_i$, $t = 1,...c$) que tiene la FD de Hartley (1950) como FD nula. (Nota: Hartley desarolló esa tabla para la igualdad de Varianzas en Ω(GAU).

[Ω(UNI)]

(1) Para $H_0^{(1)}$, se utiliza $X_{(m)}[Y_{(n)}]^{-1}$. (Véase : Choi, Bell, Ahmad, Park (1982)).

(2) Para $H_0^{(3)}$, si $n_1 = ... = n_c = n = Nc^{-1}$, se puede utilizar $\max[X_{1(n)},...,X_{c(n)}]\ (\min[X_{1(n)},...,X_{c(n)}])^{-1}$.

[Ω(GAU)]

$H_0^{(1)}$ adopta la forma $H_0': \mu_1 = \mu_2$; $H_0^{(2)}$ se transforma en $H_0'': \mu_1 = .. = \mu_c$ y $H_0^{(3)}$ equivale a $H_0''': \beta = 0$. Algunos estadísticos "optimales" son, respectivamente, $[m_X - m_Y]\ S_p^{-1}\ [m^{-1} + n^{-1}]^{-1/2}$, $C^*\ [\Sigma_i(m_i - m_{..})^2][\Sigma_j(X_{ij} - m_{i\cdot})^2]^{-1}$, el EVM de la inclinación de la línea recta de regresión.

[Ω(PAR)]

Para esta familia la teoría de las distribuciones no está tan desarrollada como para Ω(GAU) . Pero , Bell, Ahmad, Park y Lui (1982) han tratado varios casos al respecto. Para $H_0^{(1)}$, sea $X_1,...,X_m$ i.i.d. $P_a(A_1,s_1)$ y el muestreo independente $Y_1,...,Y_m$ sea i.i.d. $P_a(A_2,s_2)$. Se definen

(a) $\eta_1 = \Sigma_j \ln X_j - m \ln X_{(1)}$, (j=1,...,m) ;
(b) $\eta_2 = \Sigma_j \ln Y_j = n \ln Y_{(1)}$, (j=1,...,n) ; y
(c) $\eta = \Sigma_j \ln X_j + \Sigma_i \ln Y_i - N \min\{\ln X_{(1)}, \ln Y_{(1)}\}$; (j=1,...,m ; i=1,...,n)

Algunos estadísticos de interés para $H_0^{(1)}$ (dependiendo de cuales son los parámetros conocidos) son

$V_1^* = \ln Y_{(1)} - \ln X_{(1)}$;
$V_2^* = [n\Sigma_j \ln X_j - m \ln A_1]\,[m\, \Sigma_i \ln Y_i - n \ln A_2]^{-1}$; (i=1,...n ; j=1,...,m)
$V_3^* = [(n-2)\eta_1]\,[(m-2)\eta_2]^{-1}$.

Todos los casos tratados en esta sección tienen algunos factores en común .

Teorema 7.1 Cada estadístico de esta sección es
(1) NPDL cra la propia familia; y
(2) de la forma $\psi[N(\mathbf{Z})]$, donde $N(\mathbf{Z})$ es una versión del REM para dicha familia.

8. CONCLUSIONES Y PROBLEMAS NO RESUELTOS

La *regla general* de la Sección 4 es válida para todos los contrastes de hipótesis. Se puede utilizar esta regla para inferencia sobre procesos estocásticos, diseño de experimientos, etc. (Véase Bell et.al. (1970) y Bell (1975b). Sin embargo, hay algunos problemas prácticos que necesitan investigación:

(A) *¿Cómo se construyen las versiones del REM en general?* Se ha visto que en ciertos casos NP especiales se pueden construir los EP's (Estadísticos de Pitman). (Véanse los Teoremas 4.3 y 4.6.). También en los casos paramétricos se ve que todos los REM's son invariantes máximos cra un grupo "natural" de transformaciones. Bell y Kurotschka (1971); Junge (1976) y Bell y Sen (1984) han indicado que los EP's también son invariantes máximos cra grupos "naturales" de transformaciones en los casos considerados. ¿Existe un método general para la construcción de REM's ?

(B) *¿Cuál es la situación sobre optimalidad?.* En los casos en que se puede hallar la familia de todos los tests PDL y todos los tests NPDL, (Véase la Def. 4.3), se necesitan reglas prácticas para la construcción de contrastes optimales. Bell y Donoghue (1969) discuten tests NP que son más potentes versus algunas alternativas paramétricas, pero las alternativas "naturales" en casos NP son familias NP.

En general se necesitan cálculos y teoremas sobre las potencias de los tests de interés. El autor conoce muy pocos resultados en esas direcciones. (Véase Chapman, 1958).

(C) *Tablas de estadísticos.* Es obvio que pueden construirse millares de estadísticos usando la *regla general* cuando se conoce el ESM y el REM . Existen tablas para muchos estadísticos comunes, y existen mas teoremas de límites. (Véase, por ejemplo, Kolmogorov, 1933; Chernoff y Savage, 1958; Puri y Sen ,1970; y Bell y Mason, 1985).

Las distribuciones son sencillas también para los EP's (estadísticos de Pitman), pero el estadístico es difícil de calcular para muestras no muy pequeñas. Aquí se necesitan aproximaciones (Véase Bell y Sen, 1984) .

En general, se necesitan algoritmos para el cálculo de tablas exactas, y tablas aproximadas para los estadísticos de interés.

(D) *Tests Aleatorizados* Para evitar los problemas discutidos en (C) , Durbin (1961) esencialmente introdujo el estadístico tratado en el Teorema 6.1 (iii) para familias NP. Bell y Doksum (1965) emplearon esa idea para familias NP. Bell (1984) puso este resultado en términos de ESM y REM en su "Randomized Noise Theorem".

De las referencias citadas se puede ver que los tests aleatorizados resuelven algunos problemas de distribuciones y tablas. Pero muy poco se conoce sobre potencia y optimalidad. (Véase Jogdeo, 1966).

REFERENCIAS

BASU, D. (1955). On statistics independent of a complete sufficient statistic. *Sankhya* ***15***, 377.

BASU, D.(1958) Correction to "On statistics independent of a complete sufficient statistic", *Sankhya*. **20**, 233.

BELL, C. B., (1964a). Some basic theorems of distribution-free statistics. *Ann. Math. Statist.* **35**, 150-156

BELL, C. B., (1964b). A characterization of multisample distribution-free statistics. *Ann. Math. Statist.* **35**, 735-738.

BELL, C.B., AHMAD, R., PARK, C. J., LUI, R.(1982). Signal detection for Pareto renewal processes. *Series in Statistics and Biostatistics*, **8-82** Dept. of Math. Sciences, San Diego State Univ.

BELL, C. B.y BELLOT, F.,(1965). Nota sobre los estadísticos no paramétricos de Pitman, *Trab. Estadist.* **16**, 25-39.

BELL, C. B. y CHICARRO, M. F. (1966). Algunas generalizaciones del test no paramétrico de Siegel-Tukey.*Trab. Estadist.* **17**, 33-43.

BELL, C. B. y DOKSUM, K. A. (1965). Some new distribution-free statistics. *Ann. Math. Statist.* **36**, 203-214.

BELL, C. B. y DOKSUM, K. A.(1967). Distribution-free tests of independence. *Ann. Math. Statist.* **38**, 429-446.

BELL, C. B. y DONOGHUE, J. F. (1969). Distribution-free tests of randomness, *Sankhya*, **A 31**, 157-176.

BELL, C. B., y HALLER, H. SMITH, (1969). Bivariate symmetry tests: parametric and nonparametric. *Ann. Math. Statist.* **40**, 259-269.

BELL, C. B. y KUROTSCHKA, V. (1971). Einige prinzipien zur behandlung nicht parameterischer hypothesen. *Studi di Probabilita, Statistica, e Ricera Operative in onore di Giuseppe Pompilj*. Oderrisi-Buggio, 164-186

BELL, C.B. y SEN, P. K., (1984). Randomization procedures. *Handbook of Statistics*. **4**, (P.R. Kushaiah and P.K. Sen. eds.) Amsterdam: North-Holland.

BELL, C.B. y MASON, A. L., (1985). Nuisance Parameters, Goodness-of-fit parameters and Kolmogorov-type statistics. *Proceedings of 1984 Conference on Goodness-of-fit*. Hungarian Academy of Sciences.

BELL, C.B., y SMITH, P. J. (1969). Some nonparametric tests for the multivariate goodness-of-fit, multisample, independence, and symmetry problems. *Multivariate Analysis II*, (P.R. Krishnaiah, ed.) New York: Academic Press, 3-23.

BELL, C. B., M. F. WOODROFFE A y T.V. AVADHANI (1970). Nonparametric tests for stochastic processes. *Nonparametric Techniques in Statistical Inference*, (M. L. Puri. ed.) Cambridge: University Press.

BELL, C.B., y SMITH, P. J., (1972). Some aspects of the concept of symmetry in nonparametric statistics. *Symmetric Functions in Statistics*, (D.S. Tracy, ed.) Univ. of Windsor. Press. Ontario, Canada.

BELL, C.B. (1975a). Circularidad en Estadística. *Trab. Estadist.*, **26**, 61-81.

BELL, C.B.(1975b). Statistical inference of some special stochastic processes. *Statisitcal Inference and Related Topics* ,(M.L.Puri. ed.) New York: Academic Press, 273-290.

BELL, C.B., (1984). Inference for goodness-of-fit problems with nuisance parameters. *J. Stat. Plan. Inf.* **9**, 273-284.

CHAPMAN, D. G., (1958). A comparative study of several one sided goodness-of-fit tests. *Ann. Math. Statist.* **29**, 655-674.

CHERNOFF, H., y SAVAGE, I. R. (1958), Asymptotic normality and efficiency of certain nonparametric tests. *Ann. Math. Statist.* **29**, 972-994.

CHOI, Y. J. (1980). Kolmogorov-Smirnov test with nuisance parameters in the uniform case. *University of Washington.* M. Sc. Thesis.

CHOI, Y., BELL, C.B., AHMAD, R., y PARK, C. J. (1982). Signal detection for uniform renewal processes. *Series in Statistics and Biostatistics.* **1-82**, Dept. of Math. Sciences, San Diego State University.

DURBIN, J. (1961). Some methods of constructing exact tests. *Biometrika.* **48**, 41-55.

FELLER,W. (1938). Note on regions similar to the sample space. *Statist. Res. Men* . **2**, 117.

HARTLEY, H. O. (1950).The maximum F-ratio as a short-cut test for heterogeneity of variance, *Biometrika.* **37**, 308-312.

JOGDEO, K. (1966). On the randomized rank score procedure of Bell and Doksum, *Ann. Math. Statist.* **37**, 1697-1703.

JUNGE, K. (1976). Charakterisierungssatze verteilungs freier statistiken und ahnlicher megen. *Univ. Gottingen.* Diplomarbeit.

KOLMOGOROV, A. (1933). Sulla determinazione empirica di una legge di distribuzione. *Giorn. Ist. ital. Atturi.* **4**, 83-91.

KUMAR, A. y PATHAK, P. K. (1977a). Two applications of Basu's Lemma. *Scand. J. Statist.* **4**, 37-38.

KUMAR, A. y PATHAK, P.K. (1977b). Sufficiency and tests of goodness-of-fit, *Scand. J. Statist.* **4**, 39-43.

LEHMAN, E. L. y STEIN, C. (1949). On the theory of some nonparametric hypotheses. *Ann. Math. Statist.* **20**, 28-45.

LEHMAN, E. L. y SCHEFFE, H. (1950). Completeness, similar regions and unbiased estimation, I, *Sankhya.***10**, 305.

LEHMAN, E. L. y SCHEFFE, H. (1955). Completeness, similar regions and unbiased estimation, II, *Sankhya.* **15**, 219.

LEHMAN, E. L. (1959). *Testing Statistical Hypotheses.* New York: Wiley.

LILLIEFORS, H.W. (1967). On the Kolmogorov-Smirnov test for normality with mean and variace unknown. *J. Amer. Statist. Assoc.* **62**, 399-402.

LILLIEFORS, H.W. (1969).On the Kolmogorov-Smirnov test for exponential distribution with mean unknown. *J. Amer. Statist. Assoc.* **64**, 387-389.

MASON, A. L. (1984). Kolmogorov-Smirnov type statistics for goodness-of-fit problems with nuisance parameters. *San Diego State Univ.* M.Sc. Thesis.

PITMAN, E. J. G. (1937a). Significance tests which may be applied to samples from any population I. *Suppl. JRSS* .**4**, 119-130.

PITMAN, E. J. G. (1937b). Significance tests which may be applied to samples from any population II. *Suppl. JRSS* .**4**, 225-232.

PITMAN, E. J. G. (1938), Significance tests which may be applied to samples from any population III. The analysis of variance tests. *Biometrika.* **29**, 322.

PURI, M. L. SEN, P. K. (1971). *Nonparametric Methods in Multivariate Analysis.* New York: Wiley.

ROSENBLATT, M. (1952). Remarks on a multivariate transformation. *Ann. Math Statist.* **23**, 470-472.

SCHEFFE, H. (1943), On a measure problem arising in the theory of nonparametric tests. *Ann. Math. Statist.* **14**, 227-233

SIEGEL, S. y TUKEY, J. W., (1960), A nonparametric sum of ranks procedure for relative spread in unpaired samples. *J. Amer. Statist. Assoc.* **55**, 429.

SRINIVASAN, R. (1970). An approach to testing the goodness-of-fit of incompletely specified distributions, *Biometrika.* **57**(3), 605-661.

WILCOXON, F. WILCOX, R. A.(1945/1964). Some rapid approximate statistical procedures. *Lederle Laboratories.* Pearl River, New York.