Pub. Mat. UAB
Nº 22 Nov. 1980
Actes VII JMHL

NOTA SOBRE LA OBTENCION DEL TEST DE KUIPER USANDO UN RECORRIDO ALEATORIO CIRCULAR

Concepción Fernández Vivas

Dpto. de Estadística Matemática
Universidad de Granada

The suggestion pointed by Gnedenko-Koroliuk to interpret the distance tests from the view-point of Random Walks has been the basic idea to develop in this note a parallel derivation for the Kuiper test. The difference hier rests on the consideration of a random walk over the circle,reaching a proof more simplified than the clasic one.

1.- En algunos problemas biológicos y geofísicos se presentan situaciones en las que los datos básicos aparecen en forma de direcciones,deseandose contrastar la hipótesis de si éstas están o no aleatoriamente orientadas. Imaginando un círculo de circunferencia unidad,se pueden representar los datos como un conjunto de n puntos sobre el círculo.El test para la aleatoriedad de direcciones equivale entonces a contratar la hipótesis de que los n puntos estén distribuidos al azar sobre la circunferencia.

2.- Sean $0\leq t_1\leq t_2 \ldots \leq t_n \leq 1$ las distancias de los n puntos desde un punto P(arbitrario sobre la circunferencia) que se toma como referencia. Se prefija un sentido,por ejemplo,el contrario al de las agujas del reloj. Fijado P , $(t_1,\ldots,t_n)$ determina la muestra completamente,pero como P es arbitrario resulta deseable que el estadístico-test prrmanezca invariante frente a elecciones de P.

De los varios estadísticos que se podrían sugerir,nos vamos a fijar en el de Kuiper $V_n = D_n^+ + D_n^-$, donde $D_n^+ = \max_{1\leq j\leq n}(j/n - t_j)$ y $D_n^- = \max_{1\leq j\leq n}(t_j - (j-1)/n)$ son los estadísticos unilaterales de Kolmogorov-Smirnov.(En términos de la función de distribución empírica: $D_n = \sup_{t\in[0,1]}\{F_n(t)-t\}$ y $D_n = \sup_{t\in[0,1]}\{t-F_n(t)\}$).

<u>LEMA</u>.-"V_n es independiente de la elección de P."

Dem. : Sea $\tilde{P}$ un nuevo origen, y $\tilde{D}_n^+$ y $\tilde{D}_n^-$ los nuevos valores de D_n^+ y D_n^-. Llamando $\tilde{t}$ a la longitud del arco $P\tilde{P}$,consideremos los cambios en $\tilde{D}_n^+$ y $\tilde{D}_n^-$ cuando $\tilde{t}$ se mueve de 0 a 1, manteniendo P fijo.Obsérvese que para $t_i \leq \tilde{t} \leq t_{i+1}$ se tiene

$$\tilde{D}_n^+ = D_n^+ - i/n + \tilde{t}$$
$$\tilde{D}_n^- = D_n^- + i/n - \tilde{t}$$

para $i=0,\ldots,n$; $t_0=0$, $t_{n+1}=1$.

Sumando miembro a miembro: $\tilde{V}_n = V_n$,c.q.d.

3.- Queremos ahora abordar el problema del cálculo de la distribución exacta de V_n de un modo que permita aprovechar los resultados de la teoría de Recorridos Aleatorios circulares en forma análoga a como se aplican los recorridos sobre la línea en el test de Kolmogorov-Smirnov.

Sean $A_1,\ldots,A_n$ n puntos sobre el círculo,determinados por una muestra particular.Sea A^* el puntoúnico/(con prob.1) en donde D_n^- es alcanzable.Consideremos un mecanismo aleatorio para la elección de P,mediante el cual,P se situe en A_1 $\ldots,A_n$ con igual probabilidad. Teniendo en cuenta que la elección de P no afecta a V_n,se puede escribir:

$$
\begin{aligned}
P(V_n \leqslant v) &= P(V_n \leqslant v/P=A^*) \\
&= P(V_n \leqslant v, P=A^*/P=A_1 o \ldots o A_n)/P(\ P=A^*/P=A_1 o \ldots A_n) \\
&= nP(V_n \leqslant v, P=A^*/P=A_1 o \ldots o\ A_n).
\end{aligned}
$$

Consideremos ahora la gráfica de $F_n(t)$. Si $P=A_1 o \ldots o\ A_n$ hay un salto en el origen de amplitud 1/n. Fijado P, los n-1 puntos restantes forman una muestra de tamaño n-1 sobre el intervalo (0,1). Estos dos hechos combinados implican que la gráfica desde el punto(0,1/n) al punto (1,1) corresponde a la de $F_{n-1}(t)$ con la excepción de que la amplitud de los saltos es ahora 1/n-1 .

Nótese que cuando $P=A^*$, la gráfica de $F_n(t)$ no debe estar bajo la línea y=t; además,si $V_n \leqslant v$,la gráfica de $F_n(t)$ no debe cruzar la línea y=v+t . Por tanto $P(V_n \leqslant v, P=A^*/P=A\ o \ldots A\ )$ es la probabilidad de que la gráfica de $F_n(t)$ no toque ni cruce las barreras y=t e y=t+v,dado que hay un salto de amplitud 1/n en t=0.

Considerando la muestra de tamaño n-1 y re-escalando los saltos de 1/n a 1/n-1 , la probabilidad anterior es la misma que la de que la gráfica de $F_{n-1}(t)$ fluctue entre las barreras $y+(1/n-1)=(n/n-1)t$ e $y+(1/n-1)=(n/n-1)(v+t)$. Pero és to es sencillamente $p_{n-1}(nv-1,1,1)$, donde $p_n(a,b,c)$ representa la prob. de que la gráfica de $F_n(t)$, basada en una muestra de n puntos en $(0,1)$, quede entre las líneas $ny=a+(n+c)t$ y $ny = -b + (n+c)t$ con $a,b,c+a,b-c > 0$.

NOTA.- Brunk sugirió el estadístico $B_n= C_n^+ + C_n^-$, donde C_n^+ y C_n^- son los estadísticos de Pyke, como una alternativa a V_n. Stephens ha probado que $P(B_n \leq b) = P(V_n \leq b+1/n+1)$. Por tanto el método anterior se podría usar para calcular la distribución de B_n.

Referencias

BRUNK: On the range of the difference between hypothetical distribution function and Pyke s modified empirical distrib. function. A.M.S.33,1962.525-532

STEPHANS: Results from the relation between two statistics of the Kolmogorov-Smirnov type. A.M.S.(1969),1833-1837.

WATSON: Goodness of fit tests on a circle.Biometrica 48(1961)

------: Some problems in the statistics of directions.Bull.In. Internat.Stat. 42(1969),374-375