AJUSTE DE DISTRIBUCIONES DIAMÉTRICAS POR LOS MÉTODOS EMPÍRICOS DE JOHNSON

Alvaro Lema T.[1]

RESUMEN

Se revisa la posibilidad de uso de los métodos empíricos de Johnson para el ajuste de las distribuciones diamétricas en silvicultura, ante la falta de indicios acerca del tipo de modelo a utilizar. Se presentan las cartas gráficas necesarias para la toma de decisiones acerca del modelo y se concluye acerca de la bondad del proceso. Con ello se pretende ganar objetividad frente al uso de la presentación de histogramas únicamente ante la falta de modelos.

Palabras claves: *Distribuciones empíricas, distribuciones diamétricas, ajuste, funciones de distribución diamétrica.*

ABSTRACT

A review on the posibility of the use of Johnson empirical methods for the fitting of the diametrical distributions in silviculture was made, in lack of clues about usable models, with desire of most objectivity in that circusnstances.

Key words: *Empirical distributions, diametrical distributions, fit of diametrical distributions models.*

En silvicultura es muy importante el ajuste de distribuciones diamétricas

[1] Profesor Titular. Universidad Nacional de Colombia. Facultad de Ciencias Agropecuarias. Departamento de Ciencias Forestales. e-mail: adjlema@perseus.unalmed.edu.co

(Lema, 1995; Uribe, 1985) aunque a veces es notoria la dificultad para lograrlo, debido a diversos factores como la falta de indicios teóricos para la selección de algún modelo o la baja información de campo, acudiendo a veces sólo a los respectivos histogramas, para lograr una visualización del concepto. Pero por objetividad y necesidad de parámetros (generalmente fuente importante de conclusiones acerca de la historia del bosque o rodal) que posibiliten alguna manipulación sistemática de los datos se considera deseable una representación formal, aunque sea empírica, de ellos.

Al intentar modelar frecuencias diamétricas existen algunos indicios como la coetaneidad y disetaneidad que hacen oscilar las curvas entre distribuciones unimodales y jotas invertidas. Al estudiar los comportamientos entre estas distribuciones; Pearson y Hartley (1954) proponen un análisis del sesgo evaluado

como, de la ecuación:

(1)

y la curtosis evaluados de acuerdo con:

(2)

para identificar, en un gráfico como el de la Figura 1, el tipo posible de modelo por ajustar.

Figura 1. Regiones para varias distribuciones, con base en

Por ejemplo la distribución normal con queda representada por ese punto. La distribución exponencial queda representada por y entre líneas las regiones de las diferentes distribuciones identificadas en él, para ayudar en la selección adecuada. No se debe ignorar, sin embargo, que los valores anteriores son usualmente desconocidos por lo cual se acude a la estimación de los momentos como:

con lo cual

y que la forma de una distribución no siempre queda bien definida por estos dos estadísticos. Por lo anterior se acudirá a las distribuciones empíricas (DJ) de (Johnson, 1949), documentadas para otras ciencias por Hahn y Shapiro (1968), las cuales permiten la estimación de percentiles con una tabla de distribución de probabilidades de la normal estándar.

Sea X una variable susceptible de ser ajustada por una DJ empírica. La forma general de la transformación será:

en que θ es una función arbitraria; α y β, parámetros de forma, ε un parámetro de localización, γ un parámetro de escala, y z una variable aleatoria normal estándar. Tres alternativas constituyen las familias para θ:

Para 1). se encuentra de (5) y (6) que

con la cual se obtiene la función de

que no es más que la lognormal triparamétrica.

en que:

densidad de probabilidades:

Haciendo:

se llega a:

conocida como la distribución familia S_L de Johnson. Para 2). se encuetra en

forma parecida:

conocida como la distribución tetraparamétrica, o familia de Johnson con todos los parámetros

Por último para la función 3). se encuentra:

conocida como la segunda distribución tetraparamétrica, o familia de Johnson, con iguales definiciones paramétricas que la anterior.

Para estas familias Johnson propone un plano como la Figura 2 para determinar la distribución aproximada

con base en .

Figura 2. Carta de Johnson para aproximación a varias distribuciones.

Selección de una familia empírica de Johnson. Se deben estimar con los estadísticos definidos por (4), como si fueran (1) y (2). Sí por ejemplo (b_1, b_2) cae razonablemente cercanos a S_L esta será la familia escogida, si cae en la región por encima de la línea, entonces S_β y sí muy por debajo entonces S_U. Las distribuciones anotadas posibilitan prácticamente todas las formas asumibles por una curva diamétrica, pues cubren variables ilimitadas o acotadas por uno o ambos extremos.

Los rangos de las líneas en la Figura 2 pueden extenderse, en caso de necesidad, escogiendo valores de v y encontrado los respectivos β_1 y β_2, de acuerdo con las ecuaciones paramétricas:

A manera de ilustración para el procedimiento se acude como ejemplo a los valores observados agrupados por hectárea en un rodal de *P. patula*. (Tabla 1), de los cuales se obtuvieron al ser agrupados: y b_2=3.3475, similares a los de los datos sin agrupar: , con lo cual se deberá llegar a la familia S_L .

Estimación de los parámetros de una

Johnson S_L. Se deben considerar dos casos de acuerdo con el conocimiento o no del parámetro de localización ε. cuando es posible conocer el diámetro mínimo, o cuando este apenas se puede estimar de los datos, cuando se hacen agrupaciones desde el campo por ejemplo.

Valor de ε conocido. De acuerdo con una lognormal con parámetros μ y θ, los siguientes son los estimadores para tales parámetros:

en que x_i, variable observada, para i=1, 2,.. n. Los estimados de α* y β, para (12) surgen de:

Para los datos de la Tabla 1 se obtuvieron los valores:

;y c

on sus respectivos m_3=0.00045018 y m_4=0.0002285. Al intentar ajustar la distribución diamétrica recomponiendo clases se obtuvo un buen ajuste a la lognormal por medio de la prueba X^2, como muestran la Tabla 2 y la Figura 3 que respaldan la decisión de evaluarla como una lognormal. Al estimar la distribución diamétrica de acuerdo con las ecuaciones (12) y (13), los ajustes no fueron satisfactorios pues se supuso un ε=0.012, por ser el valor inferior de la primera clase, cuyos estadísticos fueron

y su _*=4.288217.

Tabla 1. Datos de DAP y frecuencias observadas agrupadas para *P. Patula*

DAP	DAPc	FREC	DAP	DAPc	FREC
0.05 ≤ DAP < 0.075	0.625	1	0.350 ≤ DAP<0.375	0.3625	22
0.075 ≤ DAP < 0.100	0.875	1	0.375 ≤ DAP < 0.400	0.3875	27
0.100 ≤ DAP <0.125	0.1125	2	0.400 ≤ DAP < 0.425	0.4125	18
0.125 ≤ DAP < 0.150	0.1375	10	0.425 ≤ DAP <0.450	0.4375	11
0.150 ≤ DAP <0.175	0.1625	23	0.450 ≤ DAP < 0.475	0.4625	3
0.175 ≤ DAP <0.200	0.1875	36	0.475 ≤ DAP <0.500	0.4875	2
0.200 ≤ DAP <0.225	0.2125	42	0.500 ≤ DAP < 0.525	0.5125	4
0.225 ≤ DAP < 0.250	0.2375	43	0.525 ≤ DAP < 0.550	0.5375	3
0.250 ≤ DAP < 0.275	0.2625	55	0.550 ≤ DAP < 0.575	0.5625	0
0.275 ≤ DAP < 0.300	0.2875	40	0.575 ≤ DAP < 0.600	0.5875	2
0.300 ≤ DAP < 0.325	0.3125	43	0.600 ≤ DAP < 0.625	0.6125	1
0.325 ≤ DAP < 0.350	0.3375	31	0.625 ≤ DAP < 0.650	0.6375	0

Tabla 2. Prueba de Bondad de ajuste para el DAP y prueba de X^2.

LIMITE INFERIOR	LIMITE SUPERIOR	FRECUENCIA OBSERVADA	FRECUENCIA ESPERADA	X^2
	< 0.200	72	69.35	0.10
0.200	< 0.275	140	146.36	0.28
0.275	< 0.350	115	114.97	0.00
0.350	< 0.425	67	56.13	2.11
0.425	< 0.500	16	21.83	1.56
> 0.500		10	11.37	0.16

X^2=4.2059,con 3 gl. Valor de p=0.240068.

Valor de desconocido. De acuerdo con (6) y (11) es posible reescribir (5) como

Al utilizar tres percentiles cualesquiera de los datos reales y los correspondientes de una normal estándar, por (21), es posible estimar los parámetros desconocidos α,β y ε de tres ecuaciones simultáneas de:

en que, z_p el p-ésimo percentil de la normal y x_p el correspondiente a los datos α (n+1)-ésimo valor de los datos agrupados; para los 420 datos el percentil 5% sería 0,05*421, entre los valores 21 y 22 y haciendo las interpolaciones que sean necesarias y las siguientes ecuaciones:

observados.

Por ejemplo, si se acude a los percentiles observados: para el 5%,

0,1625, para el 50%, 0,2625 y para el 95%, 0,4375, obtenidos como los

Rev.Fac.Nal.Agr.Medellín.Vol.55,No.1.p.1415-1423.2002.

Figura 3. Distribución e histograma de frecuencias para los datos de la Tabla 2.

se obtienen:

Para calcular los valores estimados se acude a la ecuación (22), recomponiendo las clases si es necesario y se evalúan por cualquier prueba conocida. Por ejemplo con los datos de la Tabla 1 y con recomposición de clases se obtuvieron los resultados de la Tabla 3, con los cuales se obtuvo la Figura 4, que permite visualizar

con lo cual

el ajuste logrado.

Anotación final: Es posible intentar varias estimaciones con diversos juegos de percentiles simétricos y escoger el mejor ajuste, dependiendo de X^2.

Para las familias S_U se sigue un procedimiento similar en cualquiera de las tres situaciones en que los dos extremos son conocidos, solo uno de ellos lo es o ninguno lo es, pero se pueden prácticamente estimar como se anotó antes.

Para las familias S_U no acotadas en general sus cuatro parámetros son desconocidos, pero existen algunas tablas para facilitar la labor (Johnson, 1965), aunque los procedimientos no se alejan mucho de los descritos. Por último el autor es consciente de la necesidad de simulaciones para producir datos que garanticen la proveniencia del tipo de distribuciones a estudiar, pero el carácter de este artículo por ahora es sólo de espíritu divulgativo.

Tabla 3. Valores estimados de las frecuencias diamétricas de la Tabla 1 con parámetro de localización desconocido

Limite Superior De clase	Z_α	Probabilidad acumulada	Frecuencia estimada acumulada	Frecuencia estimada por clase	Frecuencia observada	Categoría diamétrica
0.10	-3.5044	0.000229	0.096134	1	1	1
0.15	-1.9344	0.026530	11.14277	11.0466	12	2
0.20	-0.9165	0.179691	75.47022	64.3274	59	3
0.25	-0.1619	0.435687	182.9885	107.5183	85	4
0.30	0.4380	0.669321	281.1148	98.12630	95	5
0.35	0.9360	0.825375	346.6575	65.54268	74	6
0.40	1.3618	0.913360	383.6112	36.953770	49	7
0.45	1.7336	0.958504	402.5717	18.96048	29	8
0.50	2.0663	0.980472	411.7982	9.226560	5	9
0.55	2.3603	0.990869	416.1650	4.36674	7	10
0.60	2.6297	0.995729	418.2062	2.04120	2	11
0.65	2.8765	0.997990	419.1560	0.94962	1	12

Figura 4. Comparación de frecuencias observadas y estimadas con recomposición de clases y valores de

BIBLIOGRAFíA

HAHN, G. J. and SHAPIRO, S.S. Statistical models in engineering. London: John Willey, 1968. 355p.

JOHNSON, N.L. Systems of frecuency curves generated by methods of translation. *En*: Biometrika. Vol. 36 (1949); p.149.

________. Tables to facilitate fitting frecuency curves. *En*: Biometrika, Vol. 52 (1965); p.547-588.

Aprobado para su publicación
Marzo 20 de 2002.

LEMA,A. Dasometría: algunas aproximaciones estadística a la medición forestal. Medellín: Centro de Publicaciones Universidad Nacional de Colombia, 1995. 401p.

PEARSON, E.S. and HARTLEY, H.O. Tables for staticians. *En:* Biometrika. Vol. 1 (1954); 967-873..

URIBE, A. . Comportamiento de las distribuciones diamétricas de frecuencias de bosques disetáneos. Medellín: Universidad Nacional de Colombia. Departamento de Ciencias Forestales, 1985. 90p.