ESTADÍSTICA ESPAÑOLA
Vol. 46, Núm. 156, 2004, págs. 357 a 384

Un indicador global para la calidad del agua. Aplicación a las aguas superficiales de la Comunidad Valenciana

por
E. BEAMONTE
A.CASINO
E. VERES
Departamento de Economía Aplicada. Universitat de València

J. BERMÚDEZ
Departamento de Estadística e Investigación Operativa. Universitat de València

RESUMEN

Las Directivas de la Unión Europea (UE) definen un conjunto de características físico-químicas para determinar la calidad del agua, junto a sendas escalas que permiten categorizar en niveles dicha calidad, y que son de aplicación obligada en todos los países miembros. El presente trabajo propone un indicador global de categorización del nivel de calidad que puede ser generalizado introduciendo incertidumbre dentro de un contexto probabilístico. El trabajo finaliza aplicando ambos indicadores a la información proporcionada por unas muestras obtenidas mediante simulación.

Palabras clave: calidad del agua, indicador medioambiental, proceso de aprendizaje Dirichlet.

Clasificación AMS: 62P12, 62F15, 91B76.

1. INTRODUCCIÓN

El estudio de la calidad del agua -especialmente la superficial- es un tema de consideración muy reciente en nuestro país, no tanto en otros países de la UE que lo han abordado hace ya algunos años. Y esta circunstancia es más notoria cuando se refiere al agua prepotable, esto es, la destinada al consumo humano. En efecto, mientras que en los países del norte de la UE se han preocupado antes de las características físico-químicas del agua consumida y de sus implicaciones sobre la salud de la población, los países del sur de la UE han estado tradicionalmente más preocupados por la cantidad del agua disponible y de sus problemas derivados -captación, localización de recursos hídricos, gestión, demanda, uso, nivel de precios, etc.-, dada su habitual carestía.

La legislación básica de la UE -de obligatorio cumplimiento en todos los países miembros- enfatiza y regula más en estos momentos los aspectos relacionados con la calidad que los centrados directamente en la cantidad. También es cierto que las sociedades modernas y desarrolladas actuales -entre ellas la española- tienen una mayor sensibilidad y prestan mayor atención a la conservación medioambiental y a temas afines. Por ello la calidad del agua va a convertirse, en los próximos años, en un tema de amplia discusión y debate entre los agentes sociales, económicos y movimientos ecologistas españoles y europeos.

Sin embargo, la definición de calidad del agua es de difícil especificación por su complejidad. Son dos los principales condicionantes de la composición química y biológica de las aguas superficiales y, en particular, de las de los ríos y canales. Por una parte, la disolución y arrastre de sustancias naturales que son propias de los terrenos por los que previamente han circulado las aguas, que podríamos definir como contaminación natural; por otra, la recepción de efluentes generados por la propia actividad humana, urbana, agrícola e industrial, que constituye la contaminación artificial. Cualquier análisis químico-biológico de las aguas manifiesta en sí mismo el efecto conjunto de las dos contaminaciones anteriores, sin que resulte posible, en la mayoría de ocasiones, separarlas e identificarlas plenamente. Además, existe otro factor que influye en la valoración de las contaminaciones anteriores, el caudal circulante de los ríos, influido por los estiajes o momentos de riadas y avenidas que producen variaciones en la composición del agua en un mismo punto del cauce de los mismos. De ahí la importancia que tiene el definir un concepto de calidad del agua que permita efectuar una abstracción total, tanto del origen como de los efectos de una determinada hidraulicidad, y que resulte de su composición en sí misma.

Una calidad determinada ha de hacer referencia a un uso también preestablecido, presentando cada uno de ellos requerimientos específicos (Poch, 1999). Es por ello la consideración de distintas categorías según empleos, de los que las más usuales son la de las aguas prepotables, piscícolas y para el riego. La consideración de las aguas residuales pertenece a una categoría distinta, en cuanto que exige un previo tratamiento de reutilización para dirigirla nuevamente a un uso o final determinados. Si bien el uso del agua debe considerarse desde un punto de vista agregado, dado que lo que para una actividad puede ser una pérdida, no lo será para el sistema en su conjunto (Mateos et al., 1996), el presente trabajo se centrará en el primero de los empleos citados, el del agua prepotable.

La calidad del agua según su uso se definirá, pues, en función de un conjunto de características físico-químicas o variables, así como de sus valores de aceptación o de rechazo: son los indicadores de la calidad del agua. Aquellas aguas que cumplan con los estándares preestablecidos para el conjunto de variables o características consideradas serán aptas para la finalidad a la que se las destina. En caso contrario, deberán ser objeto de tratamiento o depuración previa. Incluso, pueden establecerse diferentes categorías de clasificación de calidad, atendiendo a la existencia de características físico-químicas con valores inadmisibles o, simplemente, mejorables.

Una clasificación de los indicadores de la calidad del agua para el consumo humano, utilizada por la Agencia de Protección de Medio Ambiente de los Estados Unidos, distingue entre indicadores primarios y secundarios. Los primarios los componen cuatro grupos:

1. Productos químicos inorgánicos (presencia de metales y compuestos).

2. Productos químicos orgánicos (por ejemplo, pesticidas).

3. Sustancias radioactivas.

4. Microorganismos.

En cuanto a los indicadores secundarios, hacen referencia a aspectos estéticos (color, turbidez, olor, sustancias en suspensión, etc.).

Una segunda clasificación de los posibles indicadores a estudiar es la empleada por la OCDE (OCDE, 1993 y 1997), que distingue entre indicadores de presión -que miden las acciones generadoras del cambio-, de estado -que definirían la situación del sistema- y de cambio o de respuesta -que evaluarían las acciones derivadas de intentar corregir los efectos de la presión generada-. Sería deseable que las características físico-químicas a utilizar fueran indicadores de estado, si bien hay que ser conscientes de los problemas que presentan para evaluar los sistemas medioambientales, y que están relacionados con la falta de linealidad, con los retardos que se producen en las respuestas y con la dificultad de establecer una causalidad clara entre el factor o factores que producen el efecto y la respuesta de los ecosistemas (Peco et al., 1998).

De ahí que la mayoría de los indicadores agroambientales propuestos en la literatura sean indicadores de presión y no, como sería deseable, de estado (Brouwer, 1995; Hammond et al., 1995).

Las Directivas 75/440/CEE y 79/869/CEE (Ministerio de Medio Ambiente, 1998) relativas a la calidad y métodos de medición, frecuencia de los muestreos y del análisis de las aguas superficiales destinadas a la producción de agua potable, especifican cuáles son aquellas características físico-químicas (parámetros, en terminología administrativa) que se consideran como indicadores para determinar la calidad del agua destinada al consumo humano, así como los criterios de clasificación -según valores admisibles- en cuatro posibles categorías: A1, A2, A3 y +A3. Además establecen las periodicidades en las tomas de muestras de agua, atendiendo a las características físico-químicas a medir, a la tipología de la estación de control y al tamaño de la población que es abastecida por esa agua. En resumen, se muestrea con más intensidad en aquellas estaciones de control cuya agua es, tradicionalmente, de peor calidad, también se intenta obtener más información de aquellas características físico-químicas consideradas de mayor relevancia por su incidencia en la calidad del agua y se muestrea con mayor intensidad en las estaciones cuya agua abastece a poblaciones de mayor tamaño.

Siguiendo con el agua prepotable, existen dos grandes tipos de características físico-químicas: las imperativas y las características guía. Los valores admisibles de aquéllas son obligatorios, mientras que los de éstas son orientativos, de ahí que el diagnóstico final de la calidad del agua deba hacerse sólo a partir de los valores imperativos, si bien las características guía establecen, en todo caso, posibles alarmas en la evolución de la calidad medida. La Tabla 1 recoge las características imperativas tal como son contempladas por la Directiva 75/440/CEE, con expresión de las unidades de medida y los valores límite para cada una de ellas.

Tabla 1

LÍMITES ADMISIBLES PARA LAS CARACTERÍSTICAS IMPERATIVAS

Característica	*Unidad de medida*	*A1*	*A2*	*A3*
Coloración (escala Pt)	mg/l	20(O)	100(O)	200(O)
Temperatura	ºC	25(O)	25(O)	25(O)
Nitratos (NO_3)	mg/l	50(O)	50(O)	50(O)
Fluoruros (F)	mg/l	1.5		
Hierro disuelto (Fe)	mg/l	0.3	2	
Cobre (Cu)	mg/l	0.05(O)		
Cinc (Zn)	mg/l	3	5	5
Arsénico (As)	mg/l	0.05	0.05	0.1
Cadmio (Cd)	mg/l	0.005	0.005	0.005
Cromo total (Cr)	mg/l	0.05	0.05	0.05
Plomo (Pb)	mg/l	0.05	0.05	0.05
Selenio (Se)	mg/l	0.01	0.01	0.01
Mercurio (Hg)	mg/l	0.001	0.001	0.001
Bario (Ba)	mg/l	0.1	1	1
Cianuro (CN)	mg/l	0.05	0.05	0.05
Sulfatos (SO_4)	mg/l	250	250(O)	250(O)
Fenoles (C_6H_5OH)	mg/l	0.001	0.005	0.1
Hidrocarburos disueltos	mg/l	0.05	0.2	1
Hidrocarburos aromáticos	mg/l	0.0002	0.0002	0.001
Plaguicidas totales	mg/l	0.001	0.0025	0.005
Amoníaco (NH_4)	mg/l	1.5	1.5	4(O)

(O) Valores imperativos que admiten excepciones por causas naturales

La determinación de los valores admisibles para las características físico-químicas que configuran la calidad del agua destinada al consumo público es una materia de reciente consideración en la sociedad española, a diferencia de otras asociaciones europeas que manifestaron rápidamente su preocupación por la incidencia e implicaciones de las directivas de la Unión Europea (Malcom, 1990). En cualquier caso, sigue siendo un tema abierto a continua discusión, pues la presión social determina que los criterios de clasificación sean cada vez más exigentes. Ello conduce a que cada vez sean menores los posibles recursos hídricos para el consumo de la población, lo que es especialmente delicado en los países, como el nuestro, donde dichos recursos son más bien escasos.

Una vez se dispone de información sobre ese conjunto de características o indicadores de la calidad del agua, la principal dificultad con la que nos encontramos es su integración para una valoración única. En efecto, las exigencias de control administrativo requieren definir una clasificación de la calidad del agua en la que queden integrados todos y cada uno de los posibles conjuntos de características de forma inequívoca, ante la obligatoriedad de proceder a tratamientos específicos del agua según su pertenencia a una categoría de clasificación concreta.

A pesar de la enorme influencia que en el desarrollo sostenible tiene la correcta valoración de la calidad del agua, no se conoce en la literatura la existencia de un indicador global de dicha calidad y que sea de fácil construcción y de sensible aplicación para la mayoría de situaciones y acuíferos (Hueting, 1991). Provencher y Lamontagne, del Servicio de Calidad de las Aguas del Ministerio de Riquezas Naturales del Estado de Québec (Ministerio de Obras Públicas y Urbanismo, 1983), definieron un indicador que englobaba las distintas características físico-químicas utilizando ciertas funciones de equivalencia -generalmente lineales-, definidas como el grado de concentración que los análisis efectuados especifican para cada característica. Su integración en forma de sumatorio se realiza utilizando ponderaciones de asignación espúrea. Muchos países han intentado utilizarlo -entre ellos España-, si bien no se han obtenido resultados concluyentes.

2. LOS CRITERIOS DE CLASIFICACIÓN ADMINISTRATIVOS

Las Confederaciones Hidrográficas son, en el caso español, los organismos de la Administración Española encargados de la policía y vigilancia de las aguas públicas. A fin de medir la calidad de las aguas de los ríos y embalses, han establecido una red de estaciones de control en las que, de modo sistemático y con distintas periodicidades según usos, se efectúan tomas de muestras para su posterior análisis y medición.

Por ejemplo, y en el ámbito de la Confederación Hidrográfica del Júcar, a lo largo de toda la cuenca existen 24 estaciones de control para el agua prepotable, 128 estaciones de control para el uso piscícola y una cantidad superior de estaciones de medición de la calidad del agua para el riego. Cada una de estas estaciones tiene una ubicación fija, seleccionada como un punto del cauce de relativo fácil acceso, perfectamente definido en mapa y en coordenadas geográficas, pudiendo compartir su finalidad alguna de ellas. Periódicamente se recoge información sobre las características físico-químicas que definen la calidad del agua, que tras ser analizada es remitida a los centros de cuenca para su control y seguimiento. Se dispone así de muestras de valores para cada una de las características, si bien no todas ellas tienen el mismo tamaño dados los criterios de periodicidad en la toma de datos recogidos en las Directivas de la UE.

Como se ha apuntado en la introducción, las Directivas 75/440/CEE y 79/869/CEE distinguen cuatro posibles categorías de clasificación de las aguas superficiales destinadas a la producción de agua potable según su calidad: A1, A2, A3 y +A3, y que corresponden a procesos de tratamiento tipo adecuados para la potabilización de las aguas. Así, por ejemplo, las aguas del nivel A2 exigen un tratamiento físico y químico normal de desinfección; mientras que las calificadas como A3 tienen un nivel de exigencia mayor, al ser sus tratamientos intensivos, junto al afino y desinfección. Aguas calificadas en niveles superiores al A3, la categoría +A3, sólo pueden ser de abastecimiento a la población previa elaboración de un plan de gestión de mejora en la calidad que incluya el tratamiento adecuado, incluida la mezcla.

Para formalizar el esquema que vamos a desarrollar posteriormente, consideramos que la calidad del agua se mide mediante *k* características físico-químicas (en la aplicación final de este trabajo, las 21 características imperativas) que, a su vez, se clasifican en cuatro categorías (A1, A2, A3 y +A3). Los criterios de clasificación administrativa de la calidad del agua se definen a partir de las dos etapas siguientes, en las que se tienen en cuenta los correspondientes límites que definen los intervalos de las categorías:

1. Clasificación por características

- Una característica físico-química se clasifica en la categoría +A3 cuando más del 5% de sus valores pertenece a dicha categoría.
- Una característica se clasifica en la categoría A3 cuando más del 5% de sus valores pertenece a las categorías A3 o +A3, siendo menor del 5% los valores que pertenecen a +A3.
- Una característica físico-química se clasifica en la categoría A2 cuando más del 5% de sus valores pertenece a las categorías A2, A3 o +A3, siendo menor del 5% los valores que pertenecen a las categorías A3 o +A3.
- Finalmente, una característica se clasifica en la categoría A1 cuando más del 95% de sus valores pertenece a dicha categoría.

Las Directivas admiten, para algunas características decididas por cada país miembro, ciertas excepciones en la anterior clasificación. En nuestro caso, por ejemplo, la temperatura es una característica con excepcionalidad, derivada de las altas temperaturas naturales propias de nuestro país. Otras características imperativas con excepcionalidades son la coloración, los nitratos, los sulfatos y el amoníaco. En estos casos, la clasificación directa en una categoría puede trasladarse a la inmediata anterior. De ahí que para ellos nunca exista la clasificación en +A3.

Así pues, cada característica físico-química tendrá asociado un conjunto de cuatro valores (v_1, v_2, v_3, v_4) que son todos 0 menos uno de ellos que vale 1 e indica la categoría en la que queda clasificada. Por ejemplo, la característica asociada al conjunto (0, 1, 0, 0) indica que está clasificada en la categoría A2.

2. Clasificación global o del agua en su conjunto

Utiliza el siguiente criterio general: basta que haya una característica físico-química en una categoría de nivel más alto -esto es, de peor calidad- para que todo el agua quede clasificada en ese nivel.

Esta etapa es claramente conservadora. En efecto, es suficiente que un agua tenga una característica con el 6% de sus muestras pertenecientes a la categoría +A3 para que, aún siendo la clasificación del resto de características la A1, el agua correspondiente se clasifique en la peor categoría, con la consiguiente exigencia de tratamiento para su correcta potabilización.

Además, el criterio administrativo expuesto es muy sensible a la existencia de valores atípicos dentro de la muestra de valores de cada característica. Y este hecho se agrava considerablemente cuando el número de datos no es suficientemente elevado. Por ejemplo, en una muestra de 20 valores, basta que uno de ellos pertenezca a una categoría de clasificación superior para que toda el agua en su conjunto quede clasificada en ella.

La postura claramente prudente del criterio de clasificación administrativo es coherente con la exigencia de salvaguardar al máximo la salud pública de la población. Sin embargo, hace perder totalmente la continuidad que debiera tener cualquier indicador de calidad, al discretizar doblemente -por las dos etapas que la definen- y radicalmente, la calidad de un agua concreta. Además, desde un punto de vista estrictamente económico, su eficiencia es claramente cuestionable. De ahí que sea necesario introducir un índice que, globalizando todas las características físico-químicas, respete esa necesaria continuidad que permita clasificar y ordenar las calidades de aguas distintas.

3. UN ÍNDICE GLOBAL PARA LA CALIDAD DEL AGUA

Como hemos comentado en la sección anterior, la calidad del agua se mide mediante *k* características físico-químicas que, a su vez, se clasifican en A1, A2, A3 y +A3. Una vez medida una muestra de agua M, vamos a definir el $\text{valor}(M)$ como el vector (a, b, c, d), siendo *a* el número de características de calidad igual a A1, *b* el número de características físico-químicas de calidad igual a A2, y de manera similar *c* y *d*.

Definición 1 *Dadas dos muestras de agua,* M_1 *y* M_2*, cuyas calidades respectivas son* $\text{valor}(M_1) = (a_1, b_1, c_1, d_1)$ *y* $\text{valor}(M_2) = (a_2, b_2, c_2, d_2)$*, decimos que* M_1 *es de mejor calidad que* M_2*,* $M_1 > M_2$*, si y sólo si se da una de las tres condiciones siguientes*:

1. $d_1 < d_2$
2. $d_1 = d_2$ y $c_1 < c_2$
3. $d_1 = d_2$, $c_1 < c_2$ y $b_1 < b_2$

Serán de igual calidad, $M_1 = M_2$*, si y sólo si* $\text{valor}(M_1) = \text{valor}(M_2)$*.*

La definición anterior permite ordenar, de peor a mejor, todas las posibles calidades del agua. Esto permite crear un índice de calidad que coincida con el rango que ocupa el valor de la muestra estudiada, al ordenar de peor a mejor todas las posibles calidades del agua.

Esa idea es muy intuitiva, pero no es operativa: no permite calcular rápidamente cuál es el índice de calidad de una muestra dada. Sin embargo, en algunos casos especiales sí puede hacerse. Por ejemplo, la peor calidad posible está dada por el valor (0, 0, 0, k), todas las características físico-químicas son de calidad +A3, que ocupará la primera posición en la ordenación de las calidades, luego el índice de calidad de (0, 0, 0, k) debe ser 1.

Los resultados algebraicos desarrollados en el Anexo y, en concreto, el teorema 1, permiten enunciar esa idea intuitiva mediante la siguiente definición.

Definición 2. *Sea una muestra de agua* M *con valor de calidad v = (a, b, c, d), se define su índice de calidad como*:

$$I(a,b,c,d) = \frac{1}{6}(s_1^3 + 3s_1^2 + 2s_1) + \frac{1}{2}(s_2^2 + s_2) + a + 1$$

siendo $s_1 = a + b + c$ *y* $s_2 = a + b$*.*

Con esa definición, el cálculo de los índices de calidad es inmediato, por ejemplo:

$$I(8,3,5,1)=\frac{1}{6}(16^2+3\times 16^2+2\times 16)+\frac{1}{2}(11^2+11)+9=816+66+9=891\,.$$

Sin embargo, no parece nada fácil deshacer el índice. Esto es, obtener el número de características físico-químicas que hay en cada uno de los niveles de calidad a partir del valor del índice $I(v)$. El teorema 3, incluido en el Anexo, justifica el siguiente algoritmo para el cálculo inverso del índice.

Algoritmo inverso. *Dada una muestra de agua de la que se desconoce su valor* v *pero se sabe que su índice de calidad es* i, $I(v)=i$, *se pueden calcular las componentes del vector* v= (a, b, c, d) *a partir del valor i de la siguiente forma*:

Paso 1 *Calcular* x, *la parte entera por exceso de* $\sqrt[3]{6i}$, $x=\lceil\sqrt[3]{6i}\rceil$, *y calcular* $f=x^3-x$. *Si* $f<6i$ *hacer* $x=x+1$. *Se obtiene* $d=k-x+2$, *siendo* $k=a+b+c+d$.

Paso 2 *Calcular* $j=i-\frac{1}{6}(x^3-3x^2+2x)$. *Calcular y, la parte entera por exceso de* $\frac{1}{2}(\sqrt{1+8j}-1)$, $y=\lceil\frac{1}{2}(\sqrt{1+8j}-1)\rceil$. *Se obtiene* $c=x-y-1$.

Paso 3 Se obtiene $a=j-\frac{y^2-y}{2}-1$ y $b=k-a-c-d$. *Con lo que ya se completa el valor de la calidad de la muestra (a, b, c, d).*

Así por ejemplo, si hay k = 17 características físico-químicas y el índice de calidad $I(v)=553$ corresponde al vector v = (a, b, c, d), entonces éste puede calcularse de la siguiente forma:

En el primer paso se calcula $x=\lceil\sqrt[3]{6\times 553}\rceil=\lceil 14.9\rceil=15$, se calcula $f=15^3-15=3360$ que es mayor que $6\times 553=3318$, por lo que el valor de x se deja inalterado. Se obtiene $d=17-15+2=4$.

En el segundo paso se calcula, $j=553-\frac{15^3-3\times 15^2+2\times 15}{6}=553-455=98$,

y se calcula $y=\lceil\frac{\sqrt{1+8\times 98}-1}{2}\rceil=\lceil 13.5\rceil=14$. Se obtiene $c=15-14-1=0$.

Por último, $a=98-\frac{14^2-14}{2}-1=6$ y $b=17-6-0-4=7$. Con lo que ya se completa el valor de la calidad de la muestra, v = (6, 7, 0, 4).

La clasificación global del agua, introducida en la sección anterior, se puede obtener de manera sencilla a partir del índice de calidad dado en la definición 2, como se muestra a continuación.

Cálculo de la clasificación global. *Una muestra de agua cuyo índice de calidad es* i, $I(v) = i$, *corresponde a una clasificación global obtenida mediante el siguiente algoritmo:*

Calidad A1 *Sólo si* $i = \frac{1}{6}(k+1)(k+2)(k+3)$, *siendo k = a + b + c + d.*

Calidad A2 *Si* $\frac{1}{6}k(k+1)(k+5) < i < \frac{1}{6}(k+1)(k+2)(k+3)$.

Calidad A3 *Si* $\frac{1}{6}k(k+1)(k+2) < i \leq \frac{1}{6}k(k+1)(k+5)$.

Calidad +A3 *Si* $i \leq \frac{1}{6}k(k+1)(k+2)$.

4. LA DIMENSIÓN ESTOCÁSTICA DEL ÍNDICE DE CALIDAD DEL AGUA

Sea Y la variable aleatoria que representa el valor de una determinada característica físico-química en una estación de medida concreta. Para clasificar esa característica en uno de los cuatro niveles de calidad habrá que comparar Y con los límites de calidad establecidos por la administración.

Sea θ_i, i = 1, 2, 3, 4, la probabilidad de que el valor de Y se encuentre entre los límites del nivel de calidad i, considerando el nivel +A3 como nivel 4. Los cuatro valores θ_i deben sumar 1, pero alguno de ellos puede valer 0 pues, como se observa en la Tabla 1, para algunas características no todos los niveles de calidad son posibles.

El criterio utilizado por la administración consiste en estimar los valores desconocidos de θ_i mediante las correspondientes frecuencias relativas de los datos observados en los últimos tres años: $(\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3, \hat{\theta}_4)$. Si el estimador de θ_4, $\hat{\theta}_4$, es superior o igual a 0.05, esa característica se cataloga como +A3. En caso contrario, pero si $\hat{\theta}_3 + \hat{\theta}_4 \geq 0.05$, se clasifica como A3. Si ninguna de esas condiciones se ha dado pero $\hat{\theta}_2 + \hat{\theta}_3 + \hat{\theta}_4 \geq 0.05$, la clasificación es en A2. En otro caso, la clasificación es A1.

Este criterio administrativo admite que existe incertidumbre en el problema, al estimar el vector desconocido è $= (\theta_1, \theta_2, \theta_3, \theta_4)'$, pero se olvida de la incertidumbre inherente en los estimadores puntuales dando una clasificación rotunda, que parece no admitir la más mínima duda.

Alternativamente, nuestra propuesta consiste en proponer una clasificación dada por un vector de probabilidades, **p** = $(p_1, p_2, p_3, p_4)'$, admitiendo la incertidumbre presente en el problema. Así, la componente i-ésima del vector de clasificación, p_i, será la probabilidad de que el parámetro estudiado deba ser clasificado en el nivel de calidad i, i = 1, 2, 3, 4. Esto es fácil de abordar desde una perspectiva bayesiana, al considerar al vector desconocido $\boldsymbol{\theta}$ como un vector aleatorio, por lo que:

$$p_4 = \Pr(\theta_4 > 0.05)$$

$$p_3 + p_4 = \Pr(\theta_3 + \theta_4 > 0.05)$$

$$p_2 + p_3 + p_4 = \Pr(\theta_2 + \theta_3 + \theta_4 > 0.05)$$

$$p_1 + p_2 + p_3 + p_4 = 1$$

El problema se reduce, por tanto, a construir la distribución de probabilidades sobre el vector θ.

4.1 Una propuesta no paramétrica

Una primera solución no paramétrica se puede obtener discretizando la variable aleatoria original Y en cuatro categorías, según los límites de calidad. La variable así transformada, X, debe seguir una distribución Multinomial cuyos parámetros son precisamente las componentes del vector θ. Utilizando la distribución previa de Jeffreys (ver, por ejemplo, Box y Tiao, 1973, página 55), la distribución a posteriori sobre θ será Dirichlet con vector de parámetros $(\alpha_1 + r_1,\ \alpha_2 + r_2,\ \alpha_3 + r_3, \alpha_4 + r_4)'$, donde r_i, i = 1, 2, 3, 4, es el número de datos observados en el nivel de calidad i, y donde α_i vale 0.5 si el nivel i es posible para el parámetro estudiado pero vale 0 en otro caso. Utilizando las propiedades de la distribución Dirichlet (ver, por ejemplo, DeGroot, 1970, página 50), se obtienen los siguientes resultados.

La distribución marginal a posteriori sobre θ_4 es Beta con parámetros $a = \alpha_4 + r_4$ y $b = \alpha_1 + r_1 + \alpha_2 + r_2 + \alpha_3 + r_3$. Por tanto:

$$p_4 = \Pr(\theta_4 > 0.05) = \int_{0.05}^{1} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}dx .$$

La distribución marginal a posteriori sobre $\theta_3 + \theta_4$ también es Beta, pero ahora con parámetros $a = \alpha_3 + r_3 + \alpha_4 + r_4$ y $b = \alpha_1 + r_1 + \alpha_2 + r_2$. Por lo que $p_3 + p_4$ puede calcularse de manera similar a p_4, obteniendo luego p_3 como la diferencia. De forma similar se calcula p_2 y luego p_1.

De este modo, una vez observado un conjunto de datos, a cada característica físico-química se le asocia un vector de probabilidades de clasificación **p** = $(p_1, p_2, p_3, p_4)'$. Todos esos vectores, colocados como filas, forman una matriz estocástica, M, que permite valorar la calidad global del agua a través del vector suma de sus columnas, **v = 1'M** Nuestra definición del índice estocástico de calidad del agua consiste en aplicar la definición 2 al vector suma **v**.

4.2 Implicaciones de esta aproximación

Supongamos que se han observado n = 36 datos de la variable aleatoria Y que, tras discretizarla, se resume en el vector de estadísticos suficientes **r** = (30, 3, 2, 1)'. Esto es, 30 de los datos observados de la variable *Y* pertenecen al nivel de calidad A1, 3 al nivel A2, 2 al nivel A3 y tan sólo 1 al nivel +A3. Con esos datos, el criterio de la administración sería A3. Sin embargo, la distribución a posteriori sobre $\boldsymbol{\theta}$ será Dirichlet con parámetros (30.5*,* 3.5*,* 2.5*,* 1.5), por tanto:

$$p_4 = \Pr(\theta_4 > 0.05) = \int_{0.05}^{1} Be(x\,|\,1.5\,,36.5)dx = 0.2874\,,$$

$$p_3 + p_4 = \int_{0.05}^{1} Be(x\,|\,4.34)dx = 0.8881\,,$$

$$p_2 + p_3 + p_4 = \int_{0.05}^{1} Be(x\,|\,7.5\,,30.5)dx = 0.9990\,,$$

por lo que, en este caso, **p** = (0.0010, 0.1109, 0.6007, 0.2874)'. Es decir, la clasificación más probable es A3, pero no es descartable +A3, pues esta última categoría también tiene asociada una probabilidad muy elevada.

Esta situación puede ser incluso contradictoria si el número de datos es más pequeño, que es lo que ocurre en la realidad. Así por ejemplo, si se han observado tan sólo n = 4 datos de un parámetro para el que las categorías A2 y A3 no son posibles, obteniéndose como vector de estadísticos suficientes **r** = (4, 0, 0, 0)', la distribución previa debe ser Dirichlet de parámetros (0.5, 0, 0, 0.5)' (de hecho es una Be(½,½), pues sólo hay dos categorías posibles), con lo que la distribución a posteriori sobre θ_4 es Beta de parámetros $a = 0.5$ y $b = 4.5$, y el valor de p_4 es 0.5086. Como las categorías A2 y A3 son imposibles, $p_2 = p_3 = 0$, y el vector completo de probabilidades es (0.4914, 0, 0, 0.5086)'. En este caso la categoría más probable es +A3, aunque los cuatro datos observados estén en A1.

Esa aparente paradoja es debida a que la clasificación en +A3 se produce en cuanto la probabilidad de que una observación pertenezca a esa categoría supere el valor 0.05. Con tan sólo 4 datos no es posible asegurar que la probabilidad de que un nuevo dato esté en esa última categoría sea inferior a 0.05. De hecho, el intervalo de confianza frecuencialista exacto al 95% sobre la proporción de éxitos en pruebas Bernoulli, si se han observado 4 fracasos en las 4 pruebas realizadas, es (0.0000, 0.6024) y el intervalo de Wald modificado, menos conservador pero más aconsejable según Agresti y Coull (1998), es (0.0000, 0.5460), lo que en ningún caso permite descartar el valor 0.05.

5. APLICACIÓN AL AGUA PREPOTABLE EN LA CONFEDERACIÓN HIDROGRÁFICA DEL JÚCAR

Los datos utilizados en esta sección han sido proporcionados por la Confederación Hidrográfica del Júcar y han sido obtenidos en sus 24 estaciones de agua prepotable desde marzo de 1994 hasta junio de 2002. Antes hay que hacer alguna salvedad: no vamos a utilizar la característica imperativa cobre, pues casi no existen datos sobre ella; tampoco vamos a tener en cuenta los criterios de excepcionalidad permitidos por las Directivas, y utilizamos todos los datos disponibles, a pesar de que según las Directivas sólo habría que utilizar los últimos tres años, aún así el número de datos disponible es muy pequeño en la mayoría de las estaciones. De hecho, es raro que se hagan mediciones mensuales, lo habitual son trimestrales e incluso anuales. La aplicación de ambos índices a ese banco de datos proporciona la Tabla 2 en la que la primera columna muestra el código de la estación, la segunda es el índice administrativo, I^A, y la tercera el índice estocástico, I^E.

Tabla 2

ÍNDICES DE CALIDAD PARA LAS ESTACIONES DE CONTROL DE AGUA PREPOTABLE DE LA CONFEDERACIÓN HIDROGRÁFICA DEL JÚCAR.

	I^A	I^E		I^A	I^E
N603	1291	580.08	M607	1517	823.95
L703	1308	677.23	K520	1540	223.80
K507	1310	725.70	K503	1539	633.66
K202	1538	969.30	I603	1327	794.46
I301	1538	840.70	H603	1330	291.78
H602	1310	701.90	H601	1539	1102.05
H402	1540	817.53	H102	1540	613.09
H101	1538	925.75	G501	1520	762.61
G402	1540	621.30	G007	1540	419.47
F702	1539	873.99	E516	1771	272.86
C403	1520	648.19	C402	1520	466.95
C301	1138	843.45	A801	1540	433.10

En algunas estaciones los resultados de los dos índices son bastante parecidos, siempre inferior el índice estocástico por incorporar más incertidumbre, pero en alguna estación, como la E516, los resultados son muy distintos. Esto es debido a la existencia de muy pocos datos, todos ellos de calidad buena. En concreto, para la estación E516 se dispone tan sólo de 3 datos en total; el índice administrativo le asigna una clasificación muy buena cuando la incertidumbre presente en el problema no permite ese excesivo optimismo.

Por ello, hemos investigado lo que podría ocurrir si dispusiéramos de más datos. Utilizando las distribuciones finales de cada característica y cada estación, hemos simulado muestras de tamaño *n* con las que construir una tabla similar a la Tabla 2 y, a partir de ella, calcular la correlación entre los dos índices. Repitiendo este proceso 1000 veces, primero con $n = 36$ y posteriormente con $n = 100$, pretendemos estudiar, por simulación, la distribución de las correlaciones entre los dos índices bajo las condiciones actuales en las cuencas de la Confederación Hidrográfica del Júcar, así como la posible influencia del tamaño muestral en dichas distribuciones. En la Tabla 3 se recogen algunas características de las distribuciones empíricas obtenidas en la simulación. La Figura 1 muestra el resultado de la estimación no paramétrica de las densidades de dichas distribuciones.

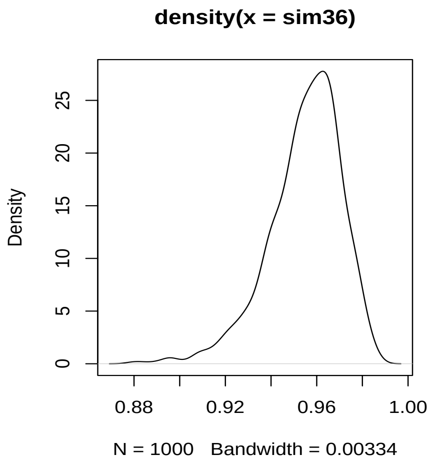
Tabla 3

ESTIMACIONES POR MONTE CARLO DE ALGUNAS CARACTERÍSTICAS DE LA DISTRIBUCIÓN DE LA CORRELACIÓN ENTRE LOS ÍNDICES ADMINISTRATIVO Y ESTOCÁSTICO

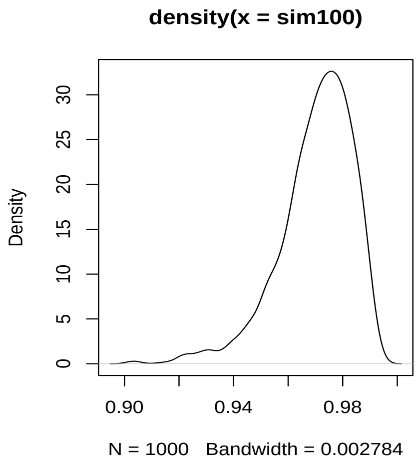
	n = 36	*n = 100*
Media	0.9548	0.9703
Desv. típica	0.0160	0.0136
Intervalo al 95%	(0.9175, 0.9792)	(0.9371, 0.9896)

FIGURA 1

ESTIMACION DE LAS DENSIDADES DE LA CORRELACIÓN ENTRE LOS ÍNDICES ADMINISTRATIVO Y ESTOCÁSTICO



n = 36



n = 100

En esos resultados se aprecia cómo la correlación crece, aunque sólo ligeramente, al aumentar el número de datos (información muestral), mientras que su dispersión disminuye.

6. CONCLUSIONES

Primeramente, la aplicación anterior confirma la utilidad del indicador estocástico propuesto en el trabajo. En efecto, independientemente de la satisfacción del conjunto de propiedades teóricas recogidas en los distintos teoremas, la alta correlación existente entre el indicador administrativo y el indicador estocástico expresada en el correspondiente intervalo de confianza, cuando existe suficiente información muestral, confirma que ambos están realizando de idéntico modo la medición de la calidad de un agua concreta. Deducimos, por tanto, su intercambiabilidad, presentando el indicador estocástico la ventaja adicional de estar concebido introduciendo en el análisis elementos de incertidumbre.

No obstante, también los comentarios realizados en el trabajo han puesto de manifiesto la sensibilidad del indicador estocástico respecto al número de datos disponibles, pues al tener en cuenta la incertidumbre presente en las estimaciones, le cuesta descartar niveles de baja calidad comportándose de manera más conservadora que el indicador administrativo. Por eso, en la exposición de la sección 4.2 se ha considerado como ejemplo, para una característica concreta, la observación de 36 datos, número que obedece al hecho administrativo de que los informes evacuados por la Administración española a la UE, obligatorios por la aplicación de las Directivas 75/440/CEE y 79/869/CEE, corresponden a períodos trianuales. La periodicidad máxima en la toma de información para ciertas características físico-químicas es mensual, por lo que dicho número respondería al máximo de datos disponibles por característica. Número -como ha quedado ya indicado- a todas luces insuficiente en el análisis estadístico para el objetivo de clasificación administrativa. De ahí que, como consecuencia, manifestemos el necesario incremento en la periodicidad de las tomas de información, y para todas las características imperativas.

Al respecto, hay que señalar que es a partir de 72 datos por característica -esto es, tomas quincenales para todas las características imperativas en el trienio de referencia de los informes citados- cuando la inexistencia de observaciones clasificadas en una categoría concreta, con una confianza del 95 %, da lugar a un vector **p** con componente nula en esa categoría, respetando por tanto el vector probabilístico de clasificación la estructura de las frecuencias relativas observadas. De ahí que en la simulación utilizada en el trabajo las correlaciones se hayan obtenido de bases de datos en las que las características disponen de 100 observaciones cada una.

La aproximación no paramétrica aquí presentada es tan sólo una primera aproximación al problema. La distribución Multinomial ni tan siquiera tiene en consideración el orden existente entre las categorías, por ello necesita mucha información muestral y sólo sería útil si se incrementa considerablemente el número de muestras realizadas. Al ser conscientes de la dificultad administrativa de incrementar el número de mediciones de las características físico-químicas del agua, una alternativa manifestada en este trabajo -conveniente en cualquier caso- es la de introducir la información inicial de la que disponen los técnicos, en forma de una distribución previa informativa y estudiar, también, cómo la adición de nuevos datos van influyendo en el resultado final del indicador.

Con una estimación paramétrica se puede disminuir considerablemente el tamaño muestral requerido para extraer consecuencias asociadas a probabilidades de error pequeñas. Posiblemente, en estos problemas podría funcionar adecuadamente el siguiente modelo: suponer que la característica físico-química en estudio puede estar presente con probabilidad desconocida y, condicionado a su presencia, el logaritmo de la cantidad de esa característica presente en el agua sea Normal, con parámetros desconocidos. Aunque no disponemos de suficientes datos para realizar un test de bondad de ajuste fiable, las pruebas que hemos realizado hasta ahora han dado resultados compatibles con este modelo.

Finalmente, la consideración de criterios de eficiencia económica -con indicadores definidos como valores esperados en una estructura de costes- podría dar lugar a una crítica sobre el conservadurismo del criterio de clasificación administrativo. La existencia de una sola característica físico-química clasificada en un mal nivel arrastra consigo la calidad del agua en su conjunto, hecho que es respetado por los indicadores administrativo y estocástico propuestos. Determinar las probabilidades de clasificación de cada uno de los niveles A1, A2, A. y +A3 es relevante en un análisis de eficiencia. Pero además, si el criterio predominante es de eficiencia social -en línea a la salvaguardia de la salud pública de la población-, la comparación de resultados permitiría cuantificar el coste adicional exigido por los criterios de salud, frente a los estrictamente económicos.

AGRADECIMIENTOS

Agradecemos a la Confederación Hidrográfica del Júcar la disponibilidad de los datos empleados en este trabajo, así como las valiosas indicaciones y sugerencias de su personal técnico.

ANEXO

Sea C_k^p el subconjunto de N^p cuyas componentes suman k. Esto es, $C_k^p = \left\{(a_1, a_2,\ldots, a_p) : \sum_{i=1}^{p} a_i = k,\ a_i \in N\right\}$. Nuestro objetivo es establecer una relación de orden total en ese conjunto, y construir la biyección que asigne a cada elemento de C_k^p su número de orden. Estamos interesados particularmente en C_k^4.

La relación de orden se puede establecer mediante la siguiente definición.

Definición 3 *Dados dos elementos cualesquiera del conjunto* C_k^p*, a = (a₁,…,aₚ) y b = (b₁,…,bₚ), diremos que a es menor que b, a<b, si y sólo si:*

$$(a_p > b_p) \vee (a_p = b_p \wedge a_{p-1} > b_{p-1}) \vee \ldots \vee (a_p = b_p \wedge \ldots \wedge a_3 = b_3 \wedge a_2 > b_2).$$

La propia definición anterior muestra el algoritmo para realizar las comparaciones, y es evidente que cualquier par de elementos de C_k^p es comparable. Por tanto es una relación de orden total y definida sobre un conjunto finito, por lo que existe máximo y mínimo en el conjunto C_k^p. El mínimo es el elemento $e_{(1)} \in C_k^p$ tal que $e_{(1)} < e \quad \forall e \in C_k^p$, y es inmediato comprobar que $e_{(1)} = (0,\ldots,0,k)$. De manera similar se puede definir el máximo, que resulta ser $(k,\ldots,0,0)$.

La Definición 3 permite construir una aplicación biyectiva entre el conjunto C_k^p y el subconjunto de números naturales {$i \in N$, tal que $1 \le i \le \# C_k^p$}, siendo $\# C_k^p$ el cardinal del conjunto C_k^p.

Definición 4 *Se define el índice del elemento* $e \in C_k^p$*,* $I_{p,k}(e)$*, como el rango que ocupa al ordenar, de menor a mayor, todos los elementos del conjunto* C_k^p *siguiendo la Definición 3.*

Esta última definición no es operativa, no incorpora un algoritmo que permita calcular el valor de $I_{p,k}(e)$ para un vector dado $e \in C_k^p$. Mucho menos la aplicación inversa para un entero *i* dado, $I_{p,k}^{-1}(i)$, con $0 \le i \le \# C_k^p$, que coincide con el rango del mayor elemento de dicho conjunto.

Para construir esos algoritmos necesitamos algunos resultados sencillos previos. Como nos interesa especialmente el estudio del conjunto C_k^4, a partir de ahora vamos a interesarnos tan sólo en dimensiones menores o iguales a 4. Vamos a empezar obteniendo el cardinal de esos conjuntos.

Obviamente $\#C_k^1$ = 1 pues el conjunto C_k^1 está constituido por un único elemento, C_k^1={(k)}. También de manera inmediata se obtiene $\#C_k^2$ = k + 1, dado que C_k^2 = {(0,k), (1,k-1),…, (k,0)}, donde ya se presentan ordenados de menor a mayor todos los elementos de C_k^2. El cálculo de los cardinales de los conjuntos C_k^3 y C_k^4 se muestra en las dos siguientes proposiciones.

Proposición 1 *El cardinal del conjunto* C_k^3 *es* $\#C_k^3 = \frac{(k+1)(k+2)}{2}$.

Demostración. Hemos de contar el número posible de vectores (a_1, a_2, a_3), donde las tres componentes son números naturales que suman k.

El valor de a_1 puede variar en el conjunto {0,1,…,k}, pero si a_1 = n, a_2 sólo puede tomar uno de los valores {0,1,…,k-n}, y a_3 sólo puede tomar un valor: k - a_1 - a_2. Así pues, con a_1 = n sólo hay k - n + 1 elementos en el conjunto C_k^3, por lo que su cardinal es:

$$\sum_{n=0}^{k}(k-n+1) = \frac{1}{2}(k+2)(k+1)$$

lo que demuestra la proposición (c.q.d.).

Proposición 2 *El cardinal del conjunto* C_k^4 *es* $\#C_k^4 = \frac{(k+1)(k+2)(k+3)}{6}$.

Demostración. El número posible de vectores (a_1, a_2, a_3, a_4), donde todas las componentes son números naturales y suman k, y que empiezan con a_1 = n (n =0,…,k) coincide con el cardinal de C_{k-n}^3. En efecto, pues el subvector (a_2, a_3, a_4) habrá que completarlo con números naturales que sumen k - n, lo que implica que (a_2, a_3, a_4) debe pertenecer a C_{k-n}^3. Por tanto, el cardinal de C_k^4 es:

$$\#C_k^4 = \sum_{n=0}^{k} \#C_{k-n}^3 = \sum_{n=0}^{k}\frac{1}{2}(n+2)(n+1) = \frac{1}{2}\sum_{n=0}^{k}(n^2+3n+2)$$

y, teniendo en cuenta que $\sum_{n=0}^{k} n^2 = k(k+1)(2k+1)/6$:

$$\#C_k^4 = \frac{1}{6}(k+3)(k+2)(k+1)$$

lo que demuestra la proposición (c.q.d.).

El máximo del conjunto C_k^4 es el vector (*k*,0,0,0), cuyo rango debe coincidir con el cardinal de C_k^4. Por tanto, la proposición 2 nos permite calcular el índice de ese elemento, que es el mayor índice posible:

$$I_{4,k}(k,0,0,0) = \frac{1}{6}(k+3)(k+2)(k+1)$$

A continuación vamos a calcular rangos de algunos elementos intermedios de esos dos conjuntos.

Proposición 3. *Sea* $C_{n,k}^3$ *el subconjunto de* C_k^3 *formado por todos aquellos elementos* e = (a_1, a_2, a_3) *para los que* $a_3 \geq n$. *El cardinal de* $C_{n,k}^3$ *es:*

$$\# C_{n,k}^3 = \frac{1}{2}(k-n+1)(k-n+2).$$

Demostración. Con $a_3 = r$ existen k – r + 1 elementos en C_k^3, como ya vimos en la demostración de la proposición 1. Por tanto, el *cardinal* de $C_{n,k}^3$ se puede calcular como:

$$\sum_{r=n}^{k}(k-r+1) = (k-n+1)(k+1) - \frac{1}{2}(k-n+1)(k+n) = \frac{1}{2}(k-n+1)(k-n+2)$$

lo que demuestra la proposición (c.q.d.).

El subconjunto $C_{n,k}^3$ considerado en la última proposición incluye a todos los elementos con los menores índices del conjunto C_k^3. Por tanto nos permite calcular los índices de algunos elementos concretos de C_k^3. Así, de la definición 3 se deduce que, si *n* es tal que $0 \leq n \leq k$, (k - n, 0, n) es el máximo del conjunto $C_{n,k}^3$, por tanto su índice debe coincidir con el cardinal de dicho conjunto:

$$I_{3,k}(k-n,0,n) = \frac{1}{2}(k-n+1)(k-n+2), \qquad \text{si} \quad 0 \leq n \leq k.$$

Podemos construir un resultado similar en dimensión 4, como muestra la siguiente proposición.

Proposición 4. *Sea* $C^4_{n,k}$ *el subconjunto de* C^4_k *formado por todos aquellos elementos* e = (a_1, a_2, a_3, a_4) *para los que* $a_4 \geq n$. *El cardinal del subconjunto* $C^4_{n,k}$ *es:*

$$\# C^4_{n,k} = \frac{1}{6}(k-n+1)(k-n+2)(k-n+3).$$

Demostración. Con $a_4 = r$ existen $\# C^3_{k-r}$ elementos en C^4_k, como ya vimos en la demostración de la proposición 2. Por tanto, aplicando la proposición 1, el cardinal de $C^4_{n,k}$ se puede calcular como:

$$\sum_{r=n}^{k} \frac{1}{2}(k-r+1)(k-r+2) = \frac{1}{2}\sum_{s=1}^{k-n+1} s(s+1)\ ,$$

donde se ha hecho la transformación s = k – r + 1 para obtener la última igualdad. Ahora, utilizando las fórmulas para calcular la suma de los primeros enteros y de sus cuadrados, la proposición se demuestra inmediatamente (c.q.d.).

De la proposición anterior, y por el mismo razonamiento empleado tras la proposición 3, se deduce inmediatamente que

$$I_{4,k}(k-n,0,0,n) = \frac{1}{6}(k-n+1)(k-n+2)(k-n+3)\,, \quad \text{si } 0 \leq n \leq k$$

y también que

$$I_{4,k}(0,0,k-n+1,n-1) = \frac{1}{6}(k-n+1)(k-n+2)(k-n+3)+1\,, \quad \text{si } 1 \leq n \leq k+1.$$

Estos resultados ya nos permiten obtener un método sencillo para el cálculo del índice de C^4_k.

Teorema 1. *Sea* e = (e_1, e_2, e_3, e_4) *un elemento cualquiera de* C^4_k. *Si definimos* $k_1 = k - e_4$ *y* $k_2 = k_1 - e_3$ *entonces*:

$$I_{4,k}(e) = \frac{1}{6}(k_1^3 + 3k_1^2 + 2k_1) + \frac{1}{2}(k_2^2 + k_2) + e_1 + 1\,.$$

Demostración. $I_{4,k}(e)$ puede calcularse contando todos los elementos menores a *e*, que son: los que tienen cuarta componente mayor que e_4, que son todos los del conjunto $C^4_{e_4+1,k}$; los que tienen en la cuarta componente e_4 pero la tercera es mayor que e_3, que forman un conjunto equivalente a $C^3_{e_3+1,k-e_4}$, y los que, teniendo las dos últimas componentes iguales a las de *e*, tienen segunda componente mayor que e_2, que forman un conjunto equivalente a $C^2_{e_2+1,k-e_4-e_3}$. Por tanto:

$$I_{4,k}(e) = \# C^4_{e_4+1,k} + \# C^3_{e_3+1,k-e_4} + \# C^2_{e_2+1,k-e_4-e_3} + 1 =$$

$$= \frac{k_1(k_1+1)(k_1+2)}{6} + \frac{k_2(k_2+1)}{2} + e_1 + 1$$

lo que demuestra el teorema (c.q.d.).

Los resultados obtenidos hasta ahora también nos dan las claves para construir la aplicación inversa del índice. Esto es, la aplicación que relacione a cada número natural i, $1 \leq i \leq \# C^p_k$, con el elemento de C^p_k cuyo rango sea *i*: a esta aplicación la denotaremos por $I^{-1}_{p,k}(i)$. Algunos valores son obvios, así por ejemplo $I^{-1}_{p,k}(1) = (0,...,0,k)$, pero su fórmula general es bastante compleja. Otro resultado también inmediato, a partir de las proposiciones anteriores, es el siguiente.

Teorema 2. *Sea i un número natural entre 1 y* $\# C^4_k$*, su imagen inversa por la aplicación índice,* $I^{-1}_{4,k}(i) = (a(i),\ b(i),\ c(i),\ d(i))$*, es tal que*:

(i) Si $i \leq \frac{1}{6}k(k+1)(k+2)$ entonces $d(i) > 0$.

(ii) Si $\frac{1}{6}k(k+1)(k+2) < i \leq \frac{1}{6}k(k+1)(k+5)$ entonces $d(i) = 0$ y $c(i) > 0$.

(iii) Si $\frac{1}{6}k(k+1)(k+5) < i \leq \frac{1}{6}(k+1)(k+2)(k+3) - 1$ entonces $c(i) = d(i) = 0$ y $b(i) > 0$.

(iv) Si $i = \frac{1}{6}(k+1)(k+2)(k+3)$ entonces $b(i) = c(i) = d(i) = 0$ y $a(i) = k$.

Demostración. La proposición 4, aplicada para n = 1, implica que $d(i) \geq 1$ si y sólo si $i \leq \frac{1}{6}k(k+1)(k+2)$. Esto demuestra (i) y la cota inferior de (ii). Si a esa cantidad, $\frac{1}{6}k(k+1)(k+2)$, se le añade el valor de la proposición 3, aplicada para n = 1, obtenemos la cota superior de (ii). El apartado (iv) se deduce inmediatamente de la proposición 2 y, por exclusión, queda demostrado también *(iii)* (c.q.d.).

Para C_k^1 la aplicación inversa del índice es trivial, pues C_k^1 *sólo* contiene un elemento, y para C_k^2 también es sencilla: si $1 \leq i \leq \#C_k^2 = k+1$, $I_{2,k}^{-1}(i) = (i-1, k-i+1)$. A continuación se estudia la extensión de esos resultados a C_k^3 y a C_k^4.

Proposición 5. *Sea i un número natural tal que* $1 \leq i \leq \#C_k^3$, *entonces la tercera componente de* $e = I_{3,k}^{-1}(i)$, e_3, *resulta ser*:

$$e_3 = k - \left\lceil \frac{\sqrt{1+8i}-1}{2} \right\rceil + 1,$$

siendo $\lceil x \rceil$ la parte entera por exceso de x.

Demostración. Sea *f(x)= x(x* + 1)/2. Esa es la función de la proposición 3, con x = k – n + 1, que proporciona el rango del elemento máximo de $C_{n,k}^3$. Por tanto, todo elemento de C_k^3 con tercera componente igual a n tendrá un rango i tal que f(k - n) < i ≤ f(k – n + 1).

Como f(x) es una función monótona creciente sobre la semirrecta real positiva, las desigualdades anteriores se traducen a:

$$k-n < f^{-1}(i) \leq k-n+1 \Rightarrow \lceil f^{-1}(i) \rceil = k-n+1,$$

donde $\lceil \cdot \rceil$ representa la función parte entera por exceso. Es necesario considerar la parte entera por exceso, para que el resultado siga siendo válido aunque $f^{-1}(i)$ sea ya un número entero. La proposición queda demostrada al notar que $f^{-1}(i) = \dfrac{\sqrt{1+8i}-1}{2}$ (c.q.d.).

La proposición anterior permite, de manera sencilla, calcular la aplicación inversa del índice en el conjunto C_k^3. En efecto, puesto que la proposición 5 ya proporciona directamente la componente e_3 de $e = I_{3,k}^{-1}(i)$ y sus otras dos componentes, (e_1, e_2), se pueden calcular de la siguiente manera:

Sea $i^* = i - \#C_{e_3+1,k}^3$ si $e_3 < k$ o $i^* = i$ si $e_3 = k$, siendo e_3 el valor obtenido por la proposición 5. El valor de i* es el rango que ocupa el vector formado por las dos primeras componentes de $e = I^{-1}_{3,k}(i)$, (e_1,e_2), entre todos los elementos de C^3_k cuya tercera componente es igual a e_3. Por tanto,

$$(e_1, e_2) = I^{-1}_{2,k-e_3}(i^*) = (i^* - 1, k - e_3 - i^* + 1)\,.$$

La siguiente proposición extiende el resultado anterior a 4 dimensiones.

Proposición 6. *Sea i un número natural tal que 1 ≤ i ≤ #*C^4_k*, entonces la última componente de* $e = I^{-1}_{4,k}(i), e_4$*, es tal que*:

$$e_4 = \begin{cases} k - \left\lceil \sqrt[3]{6i} \right\rceil + 2 & \text{si } \left\lceil \sqrt[3]{6i} \right\rceil^3 - \left\lceil \sqrt[3]{6i} \right\rceil \geq 6i \\ k - \left\lceil \sqrt[3]{6i} \right\rceil + 3 & \text{en otro caso} \end{cases},$$

siendo $\lceil x \rceil$ la parte entera por exceso de x.

Demostración. Sea $f(x) = \dfrac{(x-1)x(x+1)}{6} = \dfrac{x^3 - x}{6}$. Esa es la función de la proposición 4, con $x = k - n + 2$, que proporciona el rango del elemento máximo de $C^4_{n,k}$. Como f(x) es una función monótona creciente sobre la semirrecta $\left(\dfrac{1}{\sqrt{3}}, +\infty\right)$, podemos utilizar el mismo razonamiento que en la demostración de la proposición 5, con lo que:

$$\lceil f^{-1}(i) \rceil = k - n + 2 \;\Rightarrow\; n = k - \lceil f^{-1}(i) \rceil + 2\,.$$

Pero ahora el cálculo de $f^{-1}(i)$ no es inmediato, por lo que vamos a utilizar la función auxiliar $g(x) = \dfrac{x^3}{6}$. Como $f(x) < g(x) < f(x+1)$ para todo $x > \dfrac{1}{\sqrt{3}}$, considerando esas desigualdades en $x_0 = g^{-1}(i)$:

$$f(g^{-1}(i)) < i < f(g^{-1}(i) + 1) \;\Rightarrow\; g^{-1}(i) < f^{-1}(i) < g^{-1}(i) + 1 \leq \lceil g^{-1}(i) \rceil + 1 \;\Rightarrow$$

$$\Rightarrow \lceil \sqrt[3]{6i} \rceil \leq \lceil f^{-1}(i) \rceil \leq \lceil \sqrt[3]{6i} \rceil + 1$$

lo que demuestra la proposición (c.q.d.).

Empleando la proposición 6 y un razonamiento análogo al utilizado al comentar los resultados de la proposición 5, no resulta difícil demostrar el siguiente teorema.

Teorema 3. Sea i un número natural entre 1 y $\# C_k^4$, su imagen inversa por la aplicación índice es:

$$I_{4,k}^{-1}(i)=\left(j-\frac{y^2-y}{2}+1,\quad \frac{y^2+y}{2}-j,\quad x-y-1,\quad k-x+2\right),$$

siendo

$$x=\begin{cases}\left\lceil\sqrt[3]{6i}\right\rceil & \text{si } \left\lceil\sqrt[3]{6i}\right\rceil^3-\left\lceil\sqrt[3]{6i}\right\rceil\geq 6i\\ \left\lceil\sqrt[3]{6i}\right\rceil+1 & \text{en otro caso}\end{cases}$$

donde $\lceil\cdot\rceil$ representa la parte entera por exceso, $j=i-\dfrac{x^3-3x^2+2x}{6}$ e $y=\left\lceil\dfrac{\sqrt{1+8j}-1}{2}\right\rceil$.

Demostración. De la proposición 6 se deduce inmediatamente el valor de la cuarta componente, $k - x + 2$. Ahora vamos a contar cuantos elementos tienen una cuarta componente estrictamente mayor que la del elemento con índice i. Por la proposición 4, el número de elementos con última componente mayor o igual a $k - x + 3$ es $\dfrac{x^3-3x^2+2x}{6}$.

Por tanto, haciendo $j=i-\dfrac{x^3-3x^2+2x}{6}$, la solución buscada coincide con la solución de un problema en tres dimensiones: encontrar $I_{3,x-2}^{-1}(j)$. De la proposición 5 se deduce inmediatamente el valor de la tercera componente, $x - y - 1$. Las dos primeras componentes se obtienen siguiendo el procedimiento explicado tras la proposición 5, y el teorema queda demostrado (c.q.d.).

REFERENCIAS

AGRESTI, A. Y COULL, B.A. (1998). «Approximate is better than *exact* for interval estimation of Binomial proportions». The American Statistician 52, 119–126.

BOX, G.E.P Y TIAO, G.C. (1973). «Bayesian *Inference* in Statistical Analysis». Reading: Addison-Wesley.

BROUWER, F. (1995). «Indicators to *monitor* agri-environmental policy in the Netherlands». The Hague: Agriculture Economics Research Institute.

DEGROOT, M.H. (1970). «Optimal Statistical Decisions». New York: McGraw-Hill.

HAMMOND, A., ADRIAANSE, A., RODENBURG, E., BRYANT, D. Y WOODWARD, R. (1995). «Environmental indicators: a systematic approach to measuring and reporting on environmental policy performance in the context Sustainable Development». Washington: World Resources Institute.

HUETING, R. (1991). «Correcting national income for environmental losses: a practical solution for a theoretical dilemma». En Constanza, R. (ed): Ecological Economics. The Science and Management of Sustainability, 194–213. New York: Columbia University Press.

MALCOM, J. (1990). «Nitrates in water. The UK farmers'view». En Calvet, R. (ed.): Nitrates, agriculture, eau. Paris.

MATEOS, J., FERERES, E. Y LOSADA, A. (1996). «Eficiencia del riego y modernización de regadíos». En Actas del XIV Congreso Nacional de Riegos. AERYD, 481–488. Almería.

MINISTERIO DE MEDIO AMBIENTE (1998). «Manual de interpretación y elaboración de informes». Directivas 75/440/CEE y 79/869/CEE relativas a la calidad y métodos de medición, frecuencia de los muestreos y del análisis de las aguas superficiales destinadas a la producción de agua potable. Madrid.

MINISTERIO DE OBRAS PÚBLICAS Y URBANISMO (1983). «La vigilancia de la contaminación fluvial». Dirección General de Obras Hidráulicas. Madrid.

OCDE (1993). «OECD Core Set of Indicators for Environmental Performance Reviews». Environmental Monograph 83. Paris.

OCDE (1997). «Environmental indicators for agriculture». OECD. Paris.

PECO, B., SUÁREZ, F., OÑATE, J.J., MALO, J.E., AGUIRRE, J. Y CUMMINGS, C. (1998). «Definición y utilización de indicadores agroambientales: la experiencia de un proyecto FAIR». Agricultura y Sociedad 86, 207–220.

POCH, M. (1999). «Las calidades del agua». Barcelona: Rubes Editorial S.L.

A GLOBAL INDICATOR FOR QUALITY WATER. APPLICATION TO SUPERFICIAL WATER OF VALENCIA COMMUNITY

ABSTRACT

The environmental laws of the European Union (EU) define an ensemble of characteristics physical-chemistry to determine the quality of the water, together with scales that permit to make levels of categories for it, and that are of obliged application for all the member countries. This work propose a global indicator of categorization of the quality level that can be generalized introducing uncertainty in a probabilistic context. The work ends applying both indicators to the information provided by some simulated samples.

Key words: water quality, environmental indicator, Dirichlet learning process.

AMS Classification: 62P12, 62F15, 91B76.