

# Problema n-alcenos para estimar la dieta de los herbívoros

A continuación vamos a resumir el problema planteado en el artículo [1] escrito por P.Barcia, M. N. Bugalho, M. L. Campagnolo y J. O. Cerdeira.

Los n-alcenos son hidrocarburos que se encuentran en las cutículas de las plantas, y que pueden ser usados para estimar la composición de la dieta de herbívoros a partir de la composición de sus heces. Una limitación del uso de n-alcenos es que para realizar esta estimación por medio de procedimientos matemáticos, el número de marcadores (i.e. diferentes n-alcenos que se pueden detectar en las heces de un animal) debe ser mayor que el número de plantas diferentes incluidas en la dieta. En este caso tratamos con herbívoros en libertad que se alimentan de plantas complejas. Para ello usaremos un enfoque para estimar la dieta con n-alcenos que se aplica en todos los casos.

El modelo usa programación lineal para estimar el mínimo y el máximo de las proporciones de cada planta de la dieta. Se ilustra el modelo con dos conjuntos de datos de plantas y heces, obtenidos en experimentos de pastoreos de ovejas en Australia y el estudio de ciervos rojos en Portugal. En estos experimentos podemos ver qué proporciones de sustancias forman parte de la dieta de estos dos tipos de animales. (no hablaremos en este artículo de los beneficios que trae este estudio, permitiéndonos añadir sustancias beneficiosas para los animales en plantas que comen o eliminar otras que sean perjudiciales para su salud.)

## 0.1 Introducción

Conocer la composición de la dieta de los herbívoros es importante para comprender su ecología y alimentación, lo cual es útil para medir sus efectos sobre la vegetación y los ecosistemas.

Las especies de plantas se caracterizan por diferentes concentraciones de n-alcenos. Los marcadores químicos recuperados en las heces pueden ser utilizados para cuantificar las plantas ingeridas por un animal.

A partir del estudio de Dove y Mayes (1991), se planteó resolver el problema con un sistema restringido de ecuaciones lineales.

El requisito de que el número máximo de componentes que se pueden detectar en la dieta está limitado por el número de marcadores disponibles es común en todos los casos. En el caso donde haya

más plantas que n-alcanos. Se trata de agrupar las plantas en grupos bien definidos funcionalmente (por ejemplo, pastor, leguminosas, ramoneo) o estadísticamente.

Al ser nuestro objetivo estimar la composición de la dieta de animales herbívoros en libertad a partir de n-alcanos. Donde el número de las especies de plantas va a ser siempre superior a los n-alcanos. Es importante realizar grupos coherentes de especies.

Describimos un modelo en el que la composición de la dieta se identifica con (posibles) infinitas combinaciones de especies. Los límites superior e inferior de las concentraciones de cada especie en la dieta pueden determinarse mediante programación lineal. El enfoque se puede aplicar sin importar si el número de marcadores es menor, igual o mayor que el número de especies, evitando así la necesidad de agrupar especies

Probamos el modelo en dos conjuntos de datos que constan de registros de los perfiles de n-alcanos de diferentes especies de plantas y heces de ovejas y ciervos recolectadas en experimentos llevados a cabo en Australia y Portugal.

## 0.2 Material y métodos

Supongamos que  $d$  es el número de n-alcanos en planta y hez. Cada hez de la muestra puede ser considerada un **punto** en el espacio d-Euclideo, que describe la concentración de cada  $d$  n-alcanos. Cada especie de planta puede ser interpretada como un vector en este espacio. Definiendo el valor de cada  $d$  n-alcano en la planta (Figura 1).

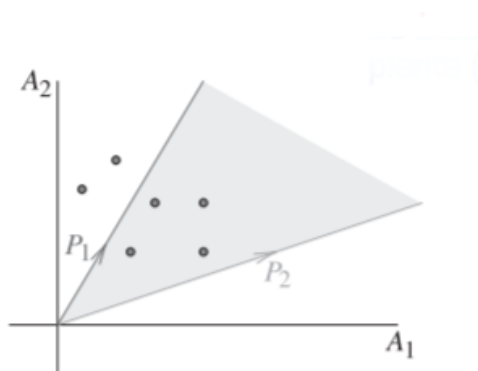


Figure 1: Ejemplo representa dos alcanos ( $A_1, A_2$ ), dos especies de plantas (vectores  $P_1$  y  $P_2$ ), y seis muestras de heces (puntos de la figura). El área sombreada es el cono bidimensional generado por los vectores  $P_1$  y  $P_2$ .

Las dietas son combinaciones lineales, es decir, las sumas ponderadas de los vectores representan las especies vegetales. El peso o coeficiente de cada vector define la cantidad correspondiente de la especie en la dieta. Cada coeficiente dividido por el sumatorio de los coeficientes, da la proporción

de cada planta en la dieta. Si  $P_1, P_2, \dots, P_p$  son los d-vectores que representan p plantas, cada vector indica la concentración de n-alcanos en esa especie, la combinación lineal con coeficientes  $c_1, c_2, \dots, c_p$  es  $c_1P_1 + c_2P_2 + \dots + c_pP_p$ .

Las combinaciones lineales con coeficientes negativos son "*dietas sin sentido*". Por lo tanto, debemos centrar nuestra atención en las combinaciones lineales que tienen coeficientes no negativos. Este conjunto de combinaciones lineales no negativas forma el cono generado por esos dos vectores. En la Figura 1 el cono generado por los vectores  $P_1$  y  $P_2$  es la región sombreada. Usamos  $C$  para denotar el cono generado, es decir:

$$C = c_1P_1 + c_2P_2 + \dots + c_pP_p, \text{ con } c_1, c_2, \dots, c_p \geq 0$$

La estimación de dietas a partir de pruebas fecales  $F_1, F_2, \dots, F_q$  toma una muestra  $F$  y buscar sí:

- (i)  $F$  pertenece al cono  $C$  o
- (ii)  $F$  es un punto fuera de  $C$ .

Esto es determinado resolviendo el sistema de ecuaciones lineal

$$Ax = F.$$

Donde  $A$  es  $d \times p$  cuyas columnas son los vectores  $P_1, P_2, \dots, P_p$  que representan las plantas. Si todos los componentes de la solución  $x$  son no-negativos, estaremos en el caso (i), y  $x$  especifica la composición correspondiente de especies en las heces. Si hay un componente negativo, estamos en la situación (ii), indicando que son "*dietas sin sentido*". Esto puede resultar de errores en la medición de las concentraciones de n-alcanos, o la existencia de plantas no incluidas en la muestra. Una posible solución es encontrar un vector no negativo de coeficientes que se aproximen a la  $x$ . Un posible enfoque consiste en:

- (iii) identificar un punto  $F_0$  en el cono  $C$  considerado "*similar a*"  $F$ , y
- (iv) usando  $F_0$  en lugar de  $F$ , proceder resolviendo como en (i) el sistema de ecuaciones (S).

Normalmente la proyección de  $F$  en  $C$ , es decir el punto en  $C$  más cercano a  $F$ , es seleccionado para ser  $F_0$ . En la figura 1 hay dos puntos en el caso (ii). Cada uno será remplazado por su proyección

en  $C$ , por lo que se encontrará definido por el vector  $P_1$ . La dieta estimada del animal se calcula finalmente a partir de las soluciones obtenidas de las muestras fecales  $F_1, F_2, \dots, F_q$ . A menudo, la dieta es definida por el promedio de las  $q$  soluciones, es decir, la dieta son las soluciones del sistema de ecuaciones lineal  $(S)$ .

En este ejemplo estamos asumiendo que el número de n-alcanos es igual al número de especies de plantas, es decir, que  $A$  es una matriz cuadrada (no-singular). Si hubiera más n-alcanos que especies la situación no cambiaría (Figura 2). Pero si hubiese más plantas que n-alcanos (Figura 3) la situación cambia considerablemente. Ahora cualquier combinación lineal no-negativa de vectores  $P_1$  y  $P_3$  define un nuevo vector  $P$  contenido en el cono generado por estos vectores. Cualquier punto  $F$  encontrado en el cono formado por  $P$  y  $P_2$  es una combinación lineal no negativa de  $P$  (en consecuencia de  $P$  y  $P_3$ ) y  $P_2$ . Al ser infinitos los vectores  $P$  definidos a partir  $P$  y  $P_3$  tales que  $F$  esta dentro del cono generado en  $P$  y  $P_2$ , el sistema  $(S)$  puede tener un número infinito de soluciones no negativas. O lo que es lo mismo, la concentración de n-alcanos en las muestras fecales puede resultar de infinidad de mezclas de plantas diferentes.

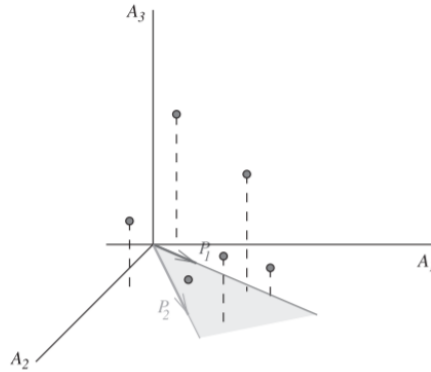


Figure 2: Ejemplo representa tres n-alcanos ( $A_1, A_2, A_3$ ), dos especies de plantas (vectores  $P_1$  y  $P_2$ ), y seis muestras de heces (puntos de la figura), dando lugar a un cono bidimensional generado por los dos vectores de plantas.

Por tanto, siempre que el número de especies supere al número de n-alcanos se suele enfocar el problema agrupando las especies en categorías. Sin embargo, no existe un claro criterio satisfactorio para decidir como agrupar las especies (cómo definir una partición  $d$  al conjunto de vectores  $P_1, P_2, \dots, P_p$ ) ni cómo ponderarlas dentro de cada grupo.

Para resolver este problema se sugiere un enfoque alternativo que usa números arbitrarios de especies de plantas y n-alcanos.

Partiendo de  $P_1, P_2, \dots, P_p$  vectores de  $d$  representan  $p$  potenciales plantas, donde  $d$  es el número de

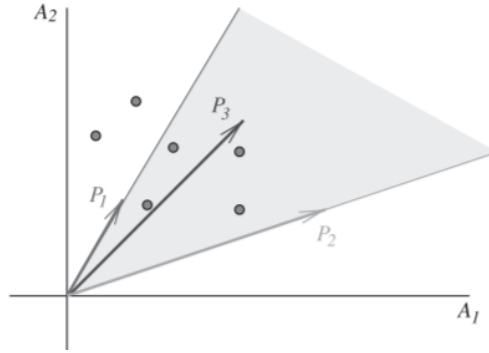


Figure 3: Ejemplo representa dos n-alcanos ( $A_1, A_2$ ), tres especies de plantas (vectores  $P_1$ ,  $P_2$  y  $P_3$ ), y seis muestras de heces (puntos de la figura). El área sombreada es el cono bidimensional generado por los vectores  $P_1$ ,  $P_2$  y  $P_2$ ..

n-alcanos usados, y  $C$  el cono generado por los vectores, Denotamos  $F_i = 1, 2, \dots, q$  la proyección en  $C$  del punto en  $d$ , y definimos  $\phi$  como el conjunto de muestras fecales "correctas".

Definimos una dieta factible aquella solución no negativa del sistema lineal  $Ax = F$ , para cualquier punto de  $F$  en  $\phi$ , donde  $A$  es  $d \times p$  matriz con columnas  $P_1, P_2, \dots, P_p$ . Hay muchas dietas factibles. No obstante, el conjunto de dietas está acotado por **límites superiores e inferiores**, la programación lineal es una herramienta adecuada para abordar algunas de las siguientes cuestiones sobre la dieta:

- ¿Cuáles son las proporciones máximas y mínimas de cada especie en la dieta?
- ¿Cuáles son las proporciones máximas y mínimas de cada grupo en la dieta?
- ¿Cuál es la proporción mínima de una planta, en concreto en aquellas dietas que cumplen un requisito ( por ejemplo, los niveles de ingesta de nitrógeno o tanino contenido)

En particular, Si  $x_p$  denota el componente de  $x$  correspondiente a la planta  $P$ , entonces  $M(i, P) = \max\{x_p, \text{ sujeto a } A_x = F_i \text{ y } x \geq 0\}$  y  $m(i, P) = \min\{x_p, \text{ sujeto a } A_x = F_i \text{ y } x \geq 0\}$ , for  $i = 1, 2, \dots, q$ , nos permite responder a las preguntas sobre la presencia o ausencia de las especies vegetales en la dieta.

- Especies nunca se comen - especies ausentes (absent species)
- Especies se comen siempre - especies obligatorias (mandatory species)

- Especies identificables en algunas muestras fecales - especies obligatorias condicionales (conditional mandatory species)

En caso de que la especie vegetal no coincida con ninguna de las categorías anteriores (especies opcionales) no podemos asegurar si se comen o no, sería necesario realizar más recopilación de datos en busca de esa especie no encontrada.

**Especies ausentes** (Absent species): son las plantas no incluidas en ninguna dieta. La especie  $P$  está ausente si el correspondiente coeficiente en cada solución no negativa de la ecuación  $A_x = F$ , para cada  $F$  en  $\phi$ , es igual a 0. Esto puede determinarse comprobando  $M(i, P) = 0$ , for  $i = 1, 2, \dots, q$ . Este caso coincidirá con los vectores  $P$  que no permiten llegar a una muestra.

**Especies obligatorias** (Mandatory species): son las presentes en todas la dietas. La especie  $P$  es obligatoria en el coeficiente de cada solución no negativa  $A_x = F$ , por cada  $F$  en  $\phi$ , es mayor que 0. Desde un punto de vista geométrico,  $P$  es obligatorio si no hay  $F_i$  en el cono  $C_p$ . Para decidir si una especie es obligatoria tenemos que comprobar si  $m(iP) > 0$ , for  $i = 1, 2, \dots, q$ . Son obligatorias las muestras que se encuentren justo en el vector de la especie.

**Especies obligatorias condicionales** (Conditional mandatory species): son las plantas obligatorias para algunas dietas pero no para todas. La especie  $P$  es condicional obligatoria si para algunos es obligatoria y para otros no.  $F$  en  $\phi$  no hay soluciones no negativas en  $A_x = F$  con coeficiente correspondiente a  $P$  igual a 0. Equivalentemente, especies  $P$  son condicional obligatorias si hay puntos en  $\phi$  que están en  $C_p$ , y otros que no. Para comprobar si las especies son condicionales obligatorias tenemos que pasar en alguna  $i = 1, 2, \dots, q$ ,  $m(i, P) > 0$ .

**Especies opcionales** (Optional species): son las plantas que no coinciden con ninguno de los casos anteriores. Especies  $P$  son opcionales si para cada  $F$  en  $\phi$  hay soluciones no-negativas de  $A_x = F$  para que el coeficiente  $P$  sea igual a 0, al menos para el lado derecho de  $F$  existe una solución no negativa, para cuyo coeficiente de  $P$  es mayor que 0. La especie  $P$  es opcional si ( $P$  no está ausente y)  $F$  está contenida en  $C_p$ . Si especie  $P$  es opcional entonces  $m(i, P) = 0$ , para cada  $i = 1, 2, \dots, q$  y  $M(i, P) > 0$ , para al menos una  $i$ . Las especies opcionales son las únicas de las que no podemos asegurar si comieron o no determinadas plantas. Por lo tanto, puede ser interesante identificar un subconjunto de plantas de tamaño mínimo sin especies opcionales, Generando un cono que contiene  $F$  (problema entero mixto - Programación lineal).

En la Figura 5 se muestra un ejemplo aplicando el modelo explicado anteriormente sobre un conjunto de datos sobre un rebaño de ovejas en Australia.

		F <sub>1</sub>		F <sub>2</sub>		F <sub>3</sub>		F <sub>4</sub>	
Status <sup>†</sup>		Min.	Max.	Min.	Max.	Min.	Max.	Min.	Max.
<i>L. corniculatus</i> Prostrate	absent	0	0	0	0	0	0	0	0
<i>L. corniculatus</i> Goldie	absent	0	0	0	0	0	0	0	0
<i>L. pedunculatus</i> Sharnae	cond. mand.	0.04 518	0.04 518	0	0	0.01 374	0.01 374	0.01 812	0.01 812
<i>L. pedunculatus</i> Maku	cond. mand.	0	0	0	0	0.01 972	0.01 972	0	0
<i>A. richardsonii</i>	absent	0	0	0	0	0	0	0	0
<i>P. aquatica</i>	cond. mand.	0.04 810	0.04 810	0	0	0	0	0	0
<i>V. myuros</i>	mand.	0.22 702	0.22 702	0.09 375	0.09 375	0.12 632	0.12 632	0.12 293	0.12 293
<i>Bromus catharticus</i>	mand.	0.25 366	0.25 366	0.26 611	0.26 611	0.13 777	0.13 777	0.21 798	0.21 798
<i>F. arundinacea</i>	mand.	0.11 879	0.11 879	0.18 392	0.18 392	0.23 013	0.23 013	0.23 105	0.23 105
<i>T. striatus</i>	mand.	0.05 146	0.05 146	0.08 413	0.08 413	0.09 908	0.09 908	0.04 962	0.04 962
<i>T. glomeratum</i>	cond. mand.	0	0	0	0	0	0	0.03 343	0.03 343
<i>T. repens</i>	mand.	0.24 747	0.24 747	0.36 785	0.36 785	0.37 324	0.37 324	0.32 686	0.32 686

<sup>†</sup> Cond. = conditional; mand = mandatory.

Figure 4: Resultado de aplicar el modelo anterior sobre un conjunto de datos de ovejas en Australia, clasificando las especies de plantas y dando limites máximos y mínimos para las muestras fecales.

### 0.3 Ejercicio a mano:

A continuación vamos a plantear la ecuación que nos permite calcular los n-alcanos que consumieron un determinado conjunto de animales. Para este ejemplo vamos a trabajar con dos especies de plantas ( $p$ ), y dos tipos de n-alcanos ( $d$ ).

$P_1 = \{0.3, 0.4\}$  tiene 0.3 del primer n-alcano y 0.4 del segundo n-alcano.

$d = 2$  es el número de n-alcanos que definen las plantas.

$$P_1 = \{0.3, 0.4\} = \begin{bmatrix} 0.3 \\ 0.4 \end{bmatrix}$$

$$P_2 = \{0.1, 0\} = \begin{bmatrix} 0.1 \\ 0 \end{bmatrix}$$

Y los gramos de cada planta:

$$x_1 \text{ gramos de } P_1$$

$$x_2 \text{ gramos de } P_2$$

Partiendo de la ecuación definida anteriormente  $Ax = F$  sabemos que  $A$  es  $\begin{bmatrix} P_1 & P_2 \end{bmatrix} x = F$ . E introduciendo los valores de este ejemplo obtenemos:

$$\begin{bmatrix} 0.3 & 0.1 \\ 0.4 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} f_1 \\ f_2 \end{bmatrix}$$

$$\begin{bmatrix} 0.3 \\ 0.4 \end{bmatrix} x_1 + \begin{bmatrix} 0.1 \\ 0 \end{bmatrix} x_2 = \begin{bmatrix} f_1 \\ f_2 \end{bmatrix}$$

$$F_1 = \begin{bmatrix} 0.3 \\ 0.2 \end{bmatrix}, F_2 = \begin{bmatrix} 0.1 \\ 0 \end{bmatrix}, F_3 = \begin{bmatrix} 0 \\ 0.2 \end{bmatrix}$$

Las  $F$  son las muestras fecales, partiendo de  $F_2$  que no tiene el segundo n-alcano, se puede deducir que  $F_2$  **no** puede haber ingerido de  $P_1$  (al tener  $P_1$  gramos del primer n-alcano). En cuanto a  $F_3$  no tiene al primer n-alcano. Por tanto, no puede haber ingerido ninguna de las especies vegetales, al tener todas valores del primer n-alcano  $> 0$ .

En la ecuación anterior para calcular  $F_2$   $x_1$  tendría que valer  $= 0$  y  $x_2$  exactamente 1, ya que  $F_2$  coincide con la cantidad de n-alcanos de  $P_2$ .

$$\begin{bmatrix} 0.3 \\ 0.4 \end{bmatrix} 0 + \begin{bmatrix} 0.1 \\ 0 \end{bmatrix} 1 = F_2$$

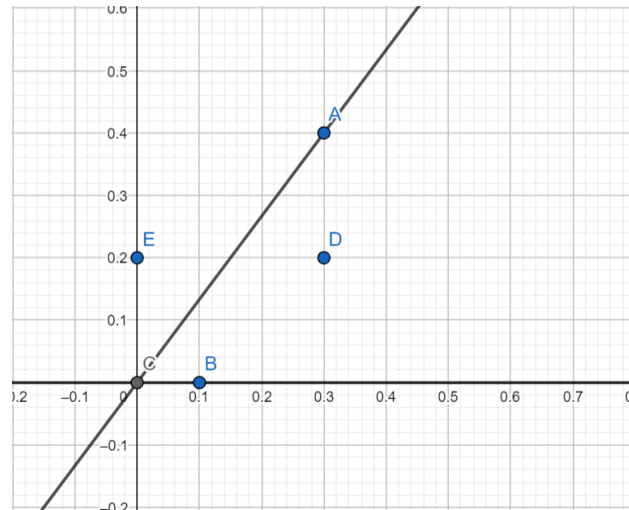


Figure 5: En esta imagen podemos apreciar los vectores que pasan por A y B (que son las especies vegetales) y los puntos B y D (son muestras fecales que se pueden formar a partir de combinaciones entre estas dos plantas), en cuanto al punto E se desconoce de que especies puede provenir el número de n-alcanos encontrado.



## References

- [1] P. Barcia, M. N. Bugalho, M. L. Campagnolo, and J. O. Cerdeira. Using n-alkanes to estimate diet composition of herbivores: A novel mathematical approach. volume 1, 2007.