

## APLICACIÓN DE UN ANÁLISIS ESTADÍSTICO BAYESIANO OBJETIVO DE UN MODELO DE REGRESIÓN NO LINEAL A LA COMPARACIÓN DE LAS PROPORCIONES TOTALES DEGRADADAS DE VARIOS ALIMENTOS

Application of an objective statistical Bayesian analysis of a nonlinear regression model to the comparison of total degraded proportions of several foods

J. A. Cano<sup>1\*</sup>, D. Salmerón<sup>2</sup>

<sup>1</sup>Departamento de Estadística e I. O. Facultad de Veterinaria. Universidad de Murcia. Campus de Espinardo. 30100 Espinardo. Murcia y <sup>2</sup>Servicio de Epidemiología. Consejería de Sanidad de la Comunidad de Autónoma de la Región de Murcia. Ronda de Levante, 11. 30008 Murcia

\* Autor para correspondencia: Juan Antonio Cano Sánchez. Tel: +34 968 36 36 38, Fax: +34 968 36 41 82. E-mail: jacano@um.es

### RESUMEN

En este trabajo se comparan las proporciones totales degradadas en digestiones *in vitro* de varios alimentos. Para ello, la proporción de un alimento degradada hasta un tiempo  $t$  se modela como una función exponencial no linealizable de  $t$  y se realiza un análisis estadístico bayesiano objetivo sobre los parámetros que aparecen en el modelo. Estos parámetros tienen unas restricciones naturales que no se procesan adecuadamente con un análisis estadístico frecuentista mientras que un análisis estadístico bayesiano permite no sólo procesarlas adecuadamente como ya se probó en Cano y Salmerón (2007) sino que además permite comparar las curvas de degradación de varios alimentos. Se presenta una aplicación con datos reales. La metodología estadística utilizada para resolver este modelo es adaptable a la resolución de otros modelos similares en los que aparecen parámetros sometidos a ciertas restricciones.

**Palabras clave:** Digestibilidad *in vitro*, curva de degradación, análisis estadístico bayesiano objetivo, análisis estadístico frecuentista.

### ABSTRACT

In this paper the total degraded proportions of several foods subject to *in vitro* digestibility are compared. For it, the degraded proportion of a food up to time  $t$  is modelled as an exponential non linearizable function of

$t$  and for the parameters appearing in the model an objective Bayesian statistical analysis is carried out. These parameters have natural constraints which are not adequately processed with a frequentist statistical analysis while the Bayesian statistical analysis allows not only to adequately process them as was stated in Cano and Salmerón (2007) but it even allows to compare the degradation curves of several foods. An application with real data is presented. The statistical methodology we have used to solve this model can be adapted to solve similar models involving parameters subject to certain constraints.

**Key words:** *In vitro* digestibility, degradation curve, objective Bayesian statistical analysis, frequentist statistical analysis.

## INTRODUCCIÓN

En digestibilidad animal la proporción de alimento  $y(t)$  degradada hasta el instante de tiempo  $t$ , puede ser modelada con una curva no linealizable del tipo  $y(t) = a + b(1 - e^{-ct})$ , donde  $a$  representa la proporción de materia degradada casi instantáneamente y  $b$  la proporción restante a ser degradada a una velocidad controlada por  $c$ , ver Orskov y McDonald (1979) para más detalles. El interés radica en la estimación de los parámetros  $a$ ,  $b$  y  $c$ . Observamos que este modelo también se aplica a otros muchos estudios de procesos de digestión.

El problema de estimar los parámetros  $a$ ,  $b$  y  $c$  se resuelve usualmente por mínimos cuadrados. Desde un punto de vista frecuentista la distribución muestral del estimador de máxima verosimilitud, que coincide con el de mínimos cuadrados bajo la hipótesis de normalidad, es la herramienta utilizada generalmente para realizar contrastes y obtener intervalos de confianza; pero en este caso la distribución de dicho estimador no es conocida, y por lo tanto se recurre a su distribución asintótica, es decir, a la distribución normal.

Sin embargo, desde un punto de vista estadístico, es interesante realizar contrastes y obtener intervalos de confianza (intervalos creíbles en un tratamiento bayesiano), en relación con las estimaciones obtenidas aun cuando el tamaño muestral sea pequeño o moderado. Proponemos entonces llevar a cabo un análisis bayesiano bajo el supuesto de que se han realizado  $n$  experimentos independientes en las mismas condiciones y que disponemos de la observación de

las proporciones de alimento degradado  $y_1, \dots, y_n$ , en los instantes de tiempo  $0 < t_1 < \dots < t_n$ ; este supuesto se adopta en vez de suponer que estas observaciones son las componentes de una serie temporal debido a las limitaciones de las técnicas actuales para medir la proporción de comida degradada hasta un tiempo  $t$  que no permiten continuar con un experimento una vez que se ha tomado una medida. Suponemos también que cada variable  $Y_i$  sigue una distribución normal con media  $a + b(1 - e^{-ct_i})$  y varianza desconocida  $\sigma^2$ . Una dificultad añadida cuando se lleva a cabo un análisis estadístico frecuentista consiste en que los parámetros tienen las siguientes restricciones naturales:  $0 < a$ ,  $b < I$ ,  $c > 0$  y  $0 < a + b < I$ , que no se procesan adecuadamente ya que algunos de los paquetes estadísticos comerciales que llevan a cabo estos análisis a veces no tienen en cuenta estas restricciones y producen estimaciones fuera de rango mientras que los que si las consideran buscan el óptimo en la frontera del conjunto de soluciones posibles dando estimaciones del tipo  $a = 0$ ,  $a + b = I$  u otras de este estilo que son claramente inadmisibles. Esta dificultad se supera automáticamente con un análisis estadístico bayesiano; no obstante, para llevar a cabo un análisis bayesiano es necesario no sólo la función de verosimilitud sino también una distribución a priori para los parámetros del modelo. En Cano y Salmerón (2007) se proponen distribuciones a priori objetivas para obtener la distribución a posteriori de  $a$ ,  $b$  y  $c$ , y, en consecuencia, la estimación de los parámetros y los intervalos creíbles. Ahora, además comparamos la proporción total degradada,  $a + b$ , de varios alimentos distintos.

Por otra parte, se han usado modelos similares desde un punto de vista frecuentista en Beuvink y Kogut (1993) y Schofield *et al.* (1994) sobre la cinética de la digestión de fibra presentando el mismo tipo de deficiencias provenientes de no tener en cuenta la presencia de un espacio paramétrico restringido; sin embargo, nuestra metodología puede adaptarse para resolver estos modelos similares desde la perspectiva bayesiana.

## MATERIAL Y MÉTODOS

Con las distribuciones a priori objetivas propuestas en Cano y Salmerón (2007) ni la media de la distribución a posteriori ni los conjuntos creíbles se obtienen de manera exacta sino que se adapta un método de Monte Carlo, el denominado SIR (Sampling Importance Resampling), que permite muestrear de forma aproximada de la distribución a posteriori para obtenerlos de forma aproximada. En este trabajo damos un paso más y utilizamos el SIR para comparar la proporción total degradada,  $a + b$ , de varios alimentos distintos. La comparación de las proporciones totales degradadas se puede llevar a cabo entre dos o más alimentos, obteniéndose en todo caso como resultado para cada alimento la probabilidad de que su proporción total degradada sea mayor que la de los demás con los que es comparado.

Las simulaciones aproximadas se han obtenido utilizando el programa estadístico Mathematica, versión 4.0 para Windows (2000) copyright© Wolfram Research, Inc. En Cano y Salmerón (2007) dichas simulaciones se utilizaron para obtener estimaciones de los parámetros y conjuntos creíbles mientras que ahora, utilizando el hecho de que las simulaciones llevadas a cabo para cada experimento son independientes, se han utilizado para comparar las proporciones totales degradadas en varios alimentos.

Nuestro análisis estadístico bayesiano se ha aplicado para comparar las proporciones totales degradadas correspondientes a subproductos de

limón y distintos tipos de pajas tratadas y no tratadas (cinco subproductos en total) por incubación *in vitro* con líquido ruminal extraído de dos rebaños de cabras murciano-granadina alimentadas con alfalfa y paja, respectivamente. Los alimentos fueron incubados durante 12, 24, 36, 48 y 72 horas con cada tipo de líquido ruminal. Una descripción completa del material y métodos utilizados desde un punto de vista biológico se puede encontrar en Martínez-Teruel *et al.* (2007), donde para cada una de las diez combinaciones posibles tipo de subproducto por tipo de rebaño se estimaron además las correspondientes curvas de degradación. Aquí nos centramos en la comparación de las proporciones totales degradadas, llevándolas a cabo entre los cinco subproductos para cada tipo de líquido ruminal fijo y para cada uno de los subproductos variando el tipo de líquido ruminal, y en todo caso describimos el material y métodos utilizados desde un punto de vista estadístico. En consecuencia, en lo que sigue y en aras de la simplicidad etiquetamos los distintos subproductos como subproducto 1, subproducto 2, ... y subproducto 5 y los distintos tipos de líquido ruminal como líquido ruminal 1 y líquido ruminal 2.

Para empezar llevamos a cabo un análisis estadístico bayesiano objetivo usando como distribución a priori la de referencia sugerida en Berger y Bernardo (1992), tomando como parámetros de mayor importancia el grupo formando por  $a$ ,  $b$  y  $c$ , y dejando a  $\sigma^2$  como parámetro de ruido. Esta distribución a priori se obtiene en Cano y Salmerón (2007), donde además se calcula la correspondiente distribución a posteriori y se adapta el método SIR para simular aproximadamente de ella. Todo esto nos permite estimar cada una de las curvas de degradación y construir intervalos creíbles para sus parámetros, además ahora aprovechamos las simulaciones de las distintas curvas para compararlas entre ellas y dar como resultado de la comparación entre varias curvas la probabilidad para cada una de ellas de ser la que presente una proporción total degradada mayor, es decir, la probabilidad

de que su correspondiente  $a + b$  sea mayor que el de las demás, probabilidad que designamos por  $P(a + b)$ . Estas probabilidades se estiman a partir de las simulaciones de las distribuciones a posteriori, igualando el número de simulaciones de los parámetros de cada una de las curvas a comparar, y calculando las proporciones de estimaciones que dan un mayor  $a + b$  para cada una de las curvas, las cuales sirven para estimar las probabilidades arriba mencionadas.

## RESULTADOS Y DISCUSIÓN

En el cuadro 1 se presentan los resultados relativos a la comparación de las proporciones totales degradadas entre los cinco subproductos para cada tipo de líquido ruminal fijo, así como los parámetros correspondientes a las curvas de degradación estimadas en cada caso. Por otra parte, en el cuadro 2 se presentan los resultados relativos a la comparación de las proporciones totales degradadas para cada subproducto cuando varía el tipo de líquido ruminal, indicando en este caso los valores de  $P(a + b)$  la probabilidad de que con el líquido ruminal 2 haya una mayor degradación.

En el cuadro 1 se observa como todas las estimaciones están dentro de su rango, es decir, no se observan valores ni de  $a$  ni de  $b$  negativos, ni de  $a$ ,  $b$  o  $a + b$  mayores que uno, ni tan siquiera valores de esas cantidades iguales a 0 o a 1, situaciones a las que —siendo claramente inadmisibles e inaceptables— se llega muchas veces en la práctica cuando se lleva a cabo un análisis estadístico frecuentista. Esto se pone de manifiesto en el ejemplo 5.1 en Cano y Salmerón (2007), donde además se observa como ni tan siquiera los conjuntos creíbles se salen del rango establecido para los parámetros correspondientes cosa que no ocurre con los intervalos de confianza, que son su contrapartida frecuentista; siendo los conjuntos creíbles de menor longitud que los intervalos de confianza, sin que ello suponga una reducción de sus probabilidades de cubrimiento cuando se someten a una prueba de validación frecuentista. Además, la estadística bayesiana proporciona como ya hemos visto un marco para la comparación de dos curvas de degradabilidad, posibilidad que no ofrece la estadística frecuentista desde un punto de vista que se pueda cuantificar.

**Cuadro 1. Valores ajustados de  $a$ ,  $b$  y  $c$  y de  $P(a + b)$  en las comparaciones con tipo de líquido ruminal fijo**

Líquido ruminal 1	$a$	$b$	$c$	$P(a + b)$
Subproducto 1	0.11	0.70	0.07	0.706
Subproducto 2	0.03	0.64	0.02	0.094
Subproducto 3	0.06	0.49	0.03	0.024
Subproducto 4	0.04	0.65	0.02	0.133
Subproducto 5	0.05	0.62	0.03	0.044
Líquido ruminal 2	$a$	$b$	$c$	$P(a + b)$
Subproducto 1	0.07	0.80	0.02	0.603
Subproducto 2	0.08	0.70	0.01	0.229
Subproducto 3	0.08	0.38	0.03	0.004
Subproducto 4	0.09	0.48	0.02	0.029
Subproducto 5	0.03	0.74	0.02	0.134

**Cuadro 2. Valores de  $P(a + b)$  para cada subproducto variando el tipo de líquido ruminal**

Subproducto 1	Subproducto 2	Subproducto 3	Subproducto 4	Subproducto 5
0.604	0.229	0.004	0.029	0.134

Finalmente, queremos poner el énfasis en que lo que hemos hecho, que es lo que la estadística bayesiana permite hacer, es llevar a cabo un análisis estadístico a medida para una situación concreta, para lo cual nos ha hecho falta la teoría para este modelo desarrollada en Cano y Salmerón (2007) y el uso del programa Mathematica para llevar a cabo los cálculos que allí se indican. En consecuencia, en la actualidad disponemos de los recursos necesarios para aplicar un análisis bayesiano objetivo del modelo  $y(t) = a + b(I - e^{-ct})$  a otras situaciones con las ventajas indicadas que esto supone; mientras que para modelos similares, es decir, con parámetros sometidos a restricciones, de forma paralela se podría desarrollar la teoría correspondiente para llevar a la práctica su análisis bayesiano.

## BIBLIOGRAFÍA

- Berger J., Bernardo J. M. 1992. Ordered group reference priors with applications to the multinomial problem. *Biometrika* 79: 25-37.
- Beuvinck J. M. W., Kogut J. 1993. Modelling gas production kinetics of grass silages incubated with buffered ruminal fluid. *J. Anim. Sci.* 71: 1041-1046.
- Cano J. A., Salmerón, D. 2007. Objective Bayesian analysis of an exponential regression model with constrained parameters applied to animal digestibility. *Comm. Statist. Theory and Methods* 36(13): 2463-2473.
- Martínez-Teruel A., Hernández F., Megías M. D., Madrid M. J., Cano J. A., Salmerón D. 2007. Bayesian estimation for by-product degradability in the rumen. To be submitted to *J. Anim. Biosci.*
- Orskov E. R., McDonald I. 1979. The estimation of protein degradability in rumen from incubation measurements weighted according to rate of passage. *J. Agric. Sci. Camb.* 92: 499-503.
- Schofield P., Pitt R. E., Pell A. N. 1994. Kinetics of fiber digestion from in vitro gas production. *J. Anim. Sci.* 72: 2980-2991.

