UNIVERSIDAD SANTO TOMAS
PRIMER CLAUSTRO UNIVERSITARIO DE COLOMBIA

Comunicaciones en Estadística
Junio 2015, Vol. 8, No. 1, pp. 81–97

Una propuesta metodológica para elicitar el vector de parámetros π de la distribución Multinomial[1]

A methodological proposal to elicit the vector of parameters π of the multinomial distribution

Andrés Felipe Flórez Rivera[a]
afflorezr@unal.edu.co

Juan Carlos Correa Morales[b]
jccorrea@unal.edu.co

Resumen

Cuando no existe información previa o es muy costosa obtenerla, las distribuciones *a priori* informativas brindan una opción práctica y "económica" que sirven de base para el inicio de un proyecto o un estudio en particular. En este trabajo se presenta una propuesta metodológica para llevar a cabo un proceso de elicitación que permite extraer el conocimiento de un experto e incorporarlo como una distribución *a priori* del vector de parámetros π de la distribución Multinomial. La metodología propuesta es acompañada de una aplicación y se finaliza con algunas conclusiones y recomendaciones.

***Palabras clave*:** distribución *a priori*, distribución Dirichlet, estadística Bayesiana probabilidad subjetiva.

Abstract

When there is no preliminary information or is too costly to obtain it, the informative priors provide a practical option and " economic " that serving as a basis for initiating a project or a particular study. In this paper we present a methodology for carrying out elicitation process that allow derives from an expert knowledge and incorporate it as a prior distribution of the parameter vector π of the Multinomial distribution. The proposed methodology accompanied by an application and end with some conclusions and recommendations.

***Keywords*:** Bayesian statistics, Dirichlet distribution, prior distribution, subjective probability.

[1]Flórez, A. F., Correa, J. C. (2015). Una propuesta metodológica para elicitar el vector de parámetros π de la distribución Multinomial. *Comunicaciones en Estadística*, **8**(1), 81-97.

[a]Estudiante de Maestría, Escuela de Estadística, Universidad Nacional de Colombia, Sede Medellín. Colombia.

[b]Profesor Asociado, Escuela de Estadística, Universidad Nacional de Colombia, Sede Medellín. Colombia.

1. Introducción

El paradigma bayesiano es un medio natural de implementar el método científico donde se aprovecha tanto la información que nos proporcionan los datos muestrales como la información extra-muestral disponible. Una distribución de probabilidad elicitada es comúnmente usada como distribución *a priori* en el análisis bayesiano en donde esta representa las creencias iniciales que se tienen sobre los parámetros de un modelo; una vez las creencias iniciales son mezcladas con los datos muestrales a través de teorema de Bayes se obtiene la distribución posterior, la cual representa las creencias actualizadas después de ver los datos. La elicitación de expertos puede ser vista como la ingeniería del conocimiento, utilizada ampliamente en contextos donde la distribución elicitada no se combina con la evidencia de los datos, ya que la opinión de los expertos es esencialmente todo el conocimiento disponible. Lamentablemente la literatura sobre elicitación de expertos es sorprendentemente pequeña en comparación con la extensa y amplia literatura sobre estadística bayesiana en general (O'Hagan 1998, O'Hagan 2005). Algunos años atrás los desarrollos en elicitación de expertos se concentraban, unos para una clase de problemas en general y otros en problemas específicos que son útiles una sola vez (Kadane & Wolfson 1998), centrando los esfuerzos en la construcción de métodos para los modelos más populares y en la comparación de estos (Gavasakar 1988). En la actualidad los estadísticos, y en especial los estadísticos bayesianos, están construyendo modelos cada vez más complejos para aplicaciones reales (O'Hagan 2005).

La elicitación de juicios de expertos y el consenso de grupos de expertos ha sido un tema de estudio en áreas como meteorología y gestión de riesgos Budnitz et al. (1997). En contextos médicos se ha restringido en gran medida a las creencias a priori sobre los efectos de un tratamiento donde, las opiniones a priori se aplican al parámetro de interacción White et al. (2005). En ciencias políticas Gill & Walker (2005) muestran una útil aplicación de la elicitación de expertos para analizar la confianza de los ciudadanos nicaragüenses en el sistema judicial. Fox (1966), Gross (1971) y Van Noortwijk et al. (1992) ponen la elicitación de expertos en el contexto de la confiabilidad. En ecología encontramos a Gilless & Fried (2000) quienes usan la elicitación para determinar los tiempos requeridos para producir una línea de fuego cuando se producen incendios forestales, y Kynn (2006) quien desarrolla un modelo que permite elicitar distribuciones normales *a priori* para un modelo de regresión logística. En muestreo Hughes & Madden (2002) usan una técnica de elicitación para encontrar la distribución de p, la cual es usada posteriormente en la fórmula para el cálculo del tamaño muestral. En control estadístico de calidad Weiler (1965) analiza la distribución de la proporción p de unidades defectuosas que pueden haber en un lote de artículos. Adicionalmente en Jenkinson (2005) podemos encontrar un resumen de una serie de estudios en Medicina, ensayos clínicos, análisis de supervivencia, psicología, industria nuclear, veterinaria, agricultura, meteorología, economía, ecología, ingeniería, deportes, arqueología y teoría de juegos.

Cuando la opinión del experto se busca en dos o más variables desconocidas, el resultado de la elicitación debe ser la distribución de probabilidad conjunta del experto de esas variables. La tarea es ahora más compleja que cuando se elicita una distribución de una sola variable y, el facilitador debe hacer inevitablemente preguntas más complejas. En los casos en que las variables son dependientes, es difícil que el facilitador escape de la complejidad de la elicitación multivariante, donde generalmente se elicitan resúmenes de las distribuciones marginales del experto o se aplican transformaciones que den como resultado una nueva variable independiente (Garthwaite et al. 2005). Un ejemplo de una tarea de elicitación multivariada donde se correlacionan necesariamente las cantidades, es elicitar el conocimiento del experto sobre un conjunto de proporciones que deben sumar 1, como es el caso de la distribución Multinomial.

La distribución Multinomial atrae considerablemente la atención de numerosos investigadores teóricos en el área de las distribuciones multivariantes discretas. Generalmente se usa en las mismas situaciones en las que se podría utilizar una distribución Binomial, cuando hay múltiples categorías de eventos en lugar de una simple dicotomía Johnson et al. (1997). Considere una serie de n ensayos independientes, en donde tan solo uno de los k eventos mutuamente excluyentes de $E_l, E_2, \cdots, E_k$ debe ser observado, y en el cual la probabilidad de ocurrencia del evento E, en cualquier ensayo es igual a π (con $\pi_1 + \pi_2 + \cdots + \pi_k = 1$). Dejemos que $n_1, n_2, \cdots, n_k$ denoten una variable aleatoria del número de ocurrencias de los eventos $E_l, E_2, \cdots, E_k$ respectivamente en esos n ensayos, con $\sum_{i=1}^{k} n_i = n$. Entonces la función de masa de probabilidad de $n_1, n_2, \cdots, n_k$ está dada por:

$$f(n_1, n_2, \cdots, n_k; n, \pi_1, \pi_2, \cdots, \pi_k) = \begin{cases} \frac{n!}{n_1! n_2! \cdots n_k!} \pi_1^{n_1} \pi_2^{n_2} \cdots \pi_k^{n_k} & \text{cuando} \quad \sum_{i=1}^{k} n_i = n, \\ 0 & \text{en otro caso.} \end{cases} \tag{1}$$

Donde $\sum_{i=1}^{k} \pi_i = 1$ para $i = 1, 2, \ldots, k$, luego, el número esperado de veces que la categoría i es observada en n ensayos y su varianza son:

$$E(n_i) = n\pi_i; \qquad Var(n_i) = n\pi_i(1 - \pi_i); \qquad Cov(n_i, n_j) = -n\pi_i\pi_j \qquad \forall i \neq j. \tag{2}$$

En la sección 2 se da una breve introducción sobre la elicitación del vector de parámetros de la distribución Multinomial. En la sección 3 se presenta una nueva metodología de elicitación para la distribución Multinomial y finalmente en la sección 4 se desarrolla una aplicación real de la metodología propuesta.

2. Elicitación del vector de parámetros de la distribución multinomial

La distribución Multinomial es una generalización de la distribución Binomial, y por esto, gran parte de los trabajos que se han desarrollado sobre la elicitación del vector de parámetros de la distribución Multinomial han tenido su origen en metodologías originalmente construidas para su caso univariado, la distribución Binomial. Generalmente se asume que la distribución del parámetro p de la distribución Binomial sigue una distribución Beta con parámetros α y β por ser la distribución Beta su conjugada natural, de ahí que las metodologías hasta ahora propuestas se hayan centrado en elicitar los parámetros de esta distribución. Algunas de las principales propuestas para elicitar los parámetros de la distribución Beta se pueden encontrar en Weiler (1965), Fox (1966), Gross (1971), Waterman et al. (1976), Chaloner & Duncan (1983), Duran & Booker (1988), Gavasakar (1988), Joseph et al. (1995), Gilless & Fried (2000), Zapata et al. (2012), Tovar (2012) y Elfadaly & Garthwaite (2012).

Cuando se elicita la distribución del vector de parámetros $\underline{\pi}$ de la distribución Multinomial, un enfoque habitual es asumir que el conocimiento del experto puede ser representado adecuadamente por una distribución Dirichlet, ya que esta distribución es su conjugada natural. La Dirichlet es una distribución multivariable muy simple, apropiada para cuando la variable de interés es un conjunto de proporciones que suman 1. La elección de una *a priori* Dirichlet tiene algunas ventajas con respecto tanto a la tratabilidad matemática como a la representación de una rica clase de creencias. Adicionalmente permite evaluaciones analíticas de la distribución predictiva y de las distribuciones de características numéricas (Regazzini & Sazonov 1999, DeGroot 2004, Zapata et al. 2012).

Considere el caso de elicitar las creencias del experto sobre un conjunto de cantidades inciertas $\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots, \pi_k)$ las cuales están restringidas a caer en la categoría $(k-1)$, es decir, $\pi_i \geq 0$ para $i = 1, \ldots, k$ y $\sum_{i=1}^{k} \pi_i = 1$. Entonces es posible afirmar que π tiene una distribución Dirichlet con parámetros $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_k)$, denotada po r $\pi \sim Di(\alpha)$ si su función de densidad está dada por:

$$f(\pi|\alpha) = c(\alpha) \prod_{i=1}^{k} \pi_i^{\alpha_i - 1} \tag{3}$$

donde $c(\alpha) = \Gamma\left(\sum_{i=1}^{k} \alpha_i\right) / \prod_{i=1}^{k} \Gamma(\alpha_i)$, la media y la varianza de π vienen dados por:

$$E(\pi_i|\alpha_i) = \frac{\alpha_i}{n}; \;\; Var(\pi_i|\alpha_i) = \frac{\alpha_i(n-\alpha_i)}{n^2(n+1)}; \; Cov(\pi_i, \pi_j|\alpha_i) = \frac{-\alpha_i\alpha_j}{n^2(n+1)} \;\; \forall i \neq j. \tag{4}$$

donde $n = \sum_{i=1}^{k} \alpha_i$. La restricción de que la $\sum_{i=1}^{k} \pi_i = 1$, hace que los π_i's sean inevitablemente correlacionados.

Hasta ahora el número de intentos para introducir métodos de elicitación para los parámetros de la distribución Dirichlet ha sido limitado; Dickey et al. (1983) presentan un método para elicitar los parámetros de la distribución Dirichlet usando muestras hipotéticas futuras. Elfadaly & Garthwaite (2012) presentan dos métodos, uno usando la distribución marginal y otro usando la distribución condicional, ambos métodos piden al experto estimar tres cuartiles los cuales son usados para estimar α y β por medio de una aproximación a la distribución normal. Zapata et al. (2012) presentan un método de elicitación basado en una metodología de sobre-ajuste donde se elicita la distribución Beta para cada π_i por medio del SHELF (SHELF es un paquete de documentos, plantillas y software que proporcionan protocolos elicitación estructurados). En dichas propuestas los autores hacen uso de la relación directa entre la distribución Dirichlet y su caso especial univariado, la distribución Beta. Ellos inician elicitando las distribuciones marginales y condicionales de la distribución Dirichlet, las cuales en ambos casos siguen una distribución Beta. El hecho de elicitar, ya sea la distribución marginal o condicional, pueden llevar al proceso de elicitación a obtener estimaciones incoherentes de N ($N = \sum_{i=1}^{k} \alpha_i$), lo cual implica que el proceso esté obligado a buscar opciones para reconciliar o promediar las estimaciones realizadas hasta que todos los valores de $n_i = \alpha_i^* + e_i^*$ sean iguales, adicionalmente los métodos presentados piden al experto hacer estimaciones sobre cuartiles y probabilidades, lo que implica que el facilitador deba hacer un mayor trabajo entrenando al experto en teoría de probabilidad, ya que sin un buen entendimiento de la teoría estadística el experto podría hacer estimaciones que no corresponden con sus verdaderos juicios.

3. Metodología propuesta

El método propuesto es un método indirecto, consiste de una mezcla entre las técnicas de elicitación HFS (muestras hipotéticas futuras) y método de la ruleta. Se eligieron estas dos metodologías ya que son cognitivamente más fácil de entender permitiendo al experto la posibilidad de grabar y procesar de forma intuitiva la información con mayor precisión sin importar si la experticia del experto es de tipo normativa o sustantiva, adicionalmente estas metodologías se adaptan fácilmente a la distribución Multinomial. Oakley et al. (2014) mencionan la falta de precisión en las estimaciones hechas por el experto cuando se usa el método de la ruleta, ya que si el experto distribuye un total de N fichas, entonces sus probabilidades implícitas están forzadas a ser múltiplos de $1/N$. Sin duda cuando se elicitan parámetros de una distribución continua la falta de precisión del método de la ruleta existe, pero para el caso de la distribución Multinomial el método de la ruleta se adapta fácilmente sin pérdida de precisión, gracias a que esta distribución es de tipo discreta.

El método propuesto se ejecuta a través de tres etapas. En la primera etapa se busca estimar el $n-$equivalente del experto, es decir, se busca estimar un tamaño muestral n tal que la opinión del experto sea equivalente a dicho tamaño muestral. En la segunda etapa se aplica el método de la Ruleta (propuesto porOakley et al. (2014)), esto le permite al método propuesto estimar el vector de probabilidades de ocurrencia de cada categoría. La tercera etapa hace uso de un algoritmo de simulación estadística el cual permite que los parámetros de la distribución Dirichlet sean estimados con base en el $n-$equivalente estimado en la etapa 1 y el vector de probabilidades estimado en la segunda etapa. La unión de estas tres etapas forman un método sencillo de aplicar que no requiere que el experto enfrente grandes esfuerzos para hacer estimaciones y que acompañado de retroalimentación continua para el experto podría arrojar mejores estimaciones; sin embargo, antes de llevar a cabo un proceso de elicitación es importante que el analista, facilitador o la persona quien vaya a dirigir el proceso identifique claramente cuál es su participación y cuál es la participación del experto. Para un buen funcionamiento del método propuesto se ha identificado que la participación del analista o facilitador debe incluir como mínimo:

- Preparación previa del proceso de elicitación, generalmente incluye entre otros, búsqueda del experto, diseño y evaluación de preguntas y entrenamiento en probabilidad al experto.
- Estimación del $n-$equivalente.
- Preguntar al experto sobre los valores que él cree más probables para cada categoría de la variable de interés .
- Validar que las preguntas realizadas hayan sido entendidas claramente por el experto.
- Registrar y almacenar la información entregada por el experto.
- Brindar retroalimentación al experto de manera que en lo posible los valores elicitados sean coherentes con el conocimiento del experto.

y que a su vez, la participación del experto se ve limitada a entregar su conocimiento sobre la variable de interés de forma cuantitativa.

Ahora, si X es una variable aleatoria la cual sigue una distribución Multinomial con parámetros $(n; \pi_1, \pi_2, \ldots, \pi_k)$ y considerando los resultados que se obtienen cuando se estima su distribución marginal y condicional, el método puede ser aplicado mediante los siguientes pasos:

1. **Estimación del $n-$equivalente:** la idea principal en esta etapa es que el facilitador califique *a priori* el nivel de experticia que tiene el experto sobre la variable de interés, esta calificación es dada en términos de un tamaño muestral n, es decir, el facilitador califica de acuerdo a su criterio a cuántas personas (tomadas de una muestra aleatoria de la población) está representando la opinión del experto. Por ejemplo, si el facilitador considera que la experticia del experto es baja y asigna a este un $n-$equivalente $= 5$ significa que la opinión de su experto representa una muestra de tamaño 5, contrario a esto, si el facilitador considera que la experticia del experto es alta y asigna a este un $n-$equivalente $= 500$ significa que la opinión de su experto está representando una muestra de tamaño 500. Actualmente en la literatura sobre elicitación de juicios de expertos existen pocos trabajos que orienten al facilitador en cómo calificar la experticia del experto de una forma óptima; sin embargo, a continuación se presentan dos metodologías que pueden ayudar al facilitador en la tarea de asignar una calificación objetiva al experto:

Método 1: el facilitador puede hacer uso del método de calibración propuesto por Cooke & Goossens (2008), este método se basa en realizar preguntas semilla a los expertos, las preguntas semilla son preguntas acerca del tema de interés pero para las que ya se conoce la respuesta, entonces, el facilitador podrá asignar una calificación al nivel de experticia del experto con base en el número de preguntas semilla que el experto haya contestado correctamente.

Método 2: otra forma práctica de estimar el valor del $n-$equivalente es haciendo uso del método propuesto por Bromaghin (1993) para calcular un tamaño muestral que permita estimar un parámetro de una proporción que sigue una distribución Multinomial, este método está basado en el enfoque dado por Tortora (1978) y propone que un tamaño de muestra puede ser estimado mediante la siguiente ecuación:

$$n = \left(\frac{z^2_{(1-(\alpha_i/2))}}{2d_i^2} \right) \left[\pi_i(1-\pi_i) - 2d_i^2 + \sqrt{\pi_i^2(1-\pi_i)^2 - d_i^2\left[4\pi_i(1-\pi_i)-1\right]} \right] \tag{5}$$

donde α_i corresponde al nivel de significancia, d_i al porcentaje de error y π_i a la probabilidad de pertenecer a la categoría i, con $i = 1, 2, \ldots, k$. Ahora, si no se cuenta con información *a priori* sobre la probabilidad de cada categoría, el tamaño de muestra puede ser estimado como:

$$n = 1 + int \left(\underbrace{max}_{i \in \{1,2,\cdots,k\}} \left[\frac{0.25 z^2_{(1-(\alpha_i/2))}}{2d_i^2} - z^2_{(1-(\alpha_i/2))} \right] \right) \tag{6}$$

entonces el facilitador puede estimar el valor del $n-$equivalente fijando un valor para d_i y dejando e l conocimiento del experto sobre la variable de interés en términos del nivel de significancia α. Por ejemplo, el facilitador puede asignar un valor de $\alpha = 0.20$ si considera que el conocimiento del experto es alto, asignar un valor de $\alpha = 0.40$ si considera que el nivel de conocimiento del experto es medio o asignar un valor de $\alpha = 0.60$ si considera que el nivel de conocimiento del experto bajo. En la tabla 1 se muestran los valores estimados del $n-$equivalente para diferentes valores de α, d y k, donde, k corresponde al número de categorías de la variable de interés.

2. **Estimación de la frecuencia relativa:** el facilitador da al experto una muestra hipotética y le pide distribuir dicha muestra en las diferentes categorías de la variable a elicitar, es decir, el facilitador da una muestra hipotética N y pide al experto estimar según su conocimiento el número de ocurrencias n_i para cada evento $E_1, E_2, \cdots, E_k$ donde $N = \sum_{i=1}^{k} n_i$, de manera que la probabilidad típica de ocurrencia de cada categoría pueda ser estimada como $n_1/N, n_2/N, \cdots, n_k/N$. Como se hizo mención anteriormente, antes de pedir al experto estimar el número de ocurrencias n_i es ideal que él haya sido contextualizado sobre el tema de interés, el objetivo de la investigación y haya tenido algún tipo de entrenamiento sobre la teoría de

Tabla 1: *Tamaño muestral equivalente con un nivel de significancia α y un porcentaje de error d para probabilidades Multinomiales descrito por Bromaghin (1993). Funte: elaboración propia.*

	α								
	0.20			0.40			0.60		
	d			d			d		
k	0.05	0.075	0.10	0.05	0.075	0.10	0.05	0.075	0.10
2	268	118	65	163	72	40	107	47	26
3	333	147	81	224	98	55	163	72	40
4	381	167	93	268	118	65	206	91	50
5	418	184	102	304	134	74	240	106	59
6	449	197	109	333	147	81	268	118	65
7	475	209	116	359	158	87	293	129	71
8	498	219	121	381	167	93	314	138	77
9	518	228	126	400	176	97	333	147	81
10	536	236	130	418	184	102	351	154	85
11	553	243	134	434	191	106	366	161	89
12	568	249	138	449	197	109	381	167	93
13	582	256	141	462	203	112	394	173	96
14	595	261	145	475	209	116	407	179	99
15	607	267	147	487	214	118	418	184	102

probabilidad. Después de que el experto ha sido entrenado, el facilitador puede proceder a elicitar los conocimientos del experto, una forma sencilla de abordarlo con preguntas como "Experto, sé que usted es una persona conocedora del tema y po r eso su opinión es de gran utilidad para la investigación, usted pudiera decirme según su conocimiento para una muestra hipotética de $N = 100$ (de la variable de interés) cuantos podrían pertenecer a la categoría 1, cuantos a la categoría 2, ... cuantos a la categoría k". Sin duda el interés es realizar preguntas en las cuales las respuestas del experto puedan proporcionar la mayor cantidad de información posible. El facilitador puede encontrar mayor información sobre elaboración de preguntas en Fowler (1995), donde se tratan factores influyentes y tips para la elaboración de una "buena" pregunta.

3. **Simulación:** una vez el facilitador ha estimado el valor del $n-$equivalente en el paso 1 y el vector de probabilidad típica en el paso 2 se procede con la simulación de una distribución Multinomial por medio de la función rmultinom del software R. La función rmultinom recibe tres parámetros. El primero corresponde al número de simulaciones que deseamos realizar. El segundo es el tamaño de muestra equivalente (estimado en el paso 1). El tercero es el vector de probabilidades de cada categoría (estimado en el paso 2) (R Core Team (2014)). Por ejemplo, si en el paso 1 se estimó para el experto un valor muestral equivalente de $n = 500$, el vector de probabilidades hallado en el paso 2 es $Prob = (0.1, 0.2, 0.3, 0.2, 0.2)$ y se desea realizar 1000 simulaciones,

el comando en R debería lucir como:

```
>rmultinom(1000, 500, prob=c(0.1, 0.2, 0.3, 0.2, 0.2))
```

El comando retorna una matriz de 1000 columnas y 5 filas, donde cada columna representa una simulación.

4. **Estimación de α_i:** en esta etapa finalmente se hace la estimación del vector de parámetros $\underline{\boldsymbol{\alpha}} = (\alpha_1, \alpha_2, \ldots, \alpha_k)$ de la distribución Dirichlet. Dado que cada X_i simulado en el paso 3 sigue una distribución Multinomial, entonces $Y_i = X_i/n$ tiene una distribución Dirichlet con un vector de parámetros $\underline{\boldsymbol{\alpha}} = (\alpha_1, \alpha_2, \cdots, \alpha_k)$, donde n es el tamaño muestral equivalente estimado en el paso 1, $\alpha_0 = \sum_{i=1}^{k} \alpha_i$, su primer y segundo momento vienen dados por:

$$E[y_i] = \frac{\alpha_i}{\alpha_0}. \tag{7}$$

$$Var[Y_i] = \frac{\alpha_i(\alpha_0 - \alpha_i)}{\alpha_0^2(\alpha_0 + 1)}. \tag{8}$$

y la moda de Y_i viene dada por:

$$Moda = \frac{\alpha_i - 1}{\alpha_0 - k}; \qquad \alpha_i > 1. \tag{9}$$

Note que los valores de (7), (8) y (9) pueden ser fácilmente estimados a partir de los valores simulados en el paso anterior, entonces el problema se reduce a resolver las ecuaciones (7) y (8) en términos de α_0 y α_i. De (7) tenemos que:

$$\alpha_i = \alpha_0 E[Y_i] \tag{10}$$

Ahora usamos (8) para dejar a α_0 en términos de $E[Y_i]$ y $Var[Y_i]$:

$$\alpha_0 = \frac{(E[Y_i] - E[Y_i]^2)}{Var[Y_i]} - 1 \tag{11}$$

Para encontrar los valores de cada α_i reemplazamos (11) en (10) y obtenemos que:

$$\alpha_i = \left(\frac{(E[Y_i] - E[Y_i]^2)}{Var[Y_i]} - 1 \right) E[Y_i] \tag{12}$$

Ahora, antes de reemplazar en (12) los valores de $E[Y_i]$ y $Var[Y_i]$ por sus respectivas estimaciones $\bar{y}_i$ y s_i^2 se debe normalizar Y_i de manera que se cumpla la restricción $\sum_{i=1}^{k} \bar{y}_i = 1$:

$$\bar{\bar{y}}_i = \frac{\bar{y}_i}{\sum_{i=1}^{k} \bar{y}_i} \tag{13}$$

Finalmente cada α_i puede ser estimado reemplazando los valores de $\bar{\bar{y}}_i$ y S_i^2 en (12):

$$\alpha_i = \left(\frac{(\bar{\bar{y}}_i - \bar{\bar{y}}_i^2)}{S_i^2} - 1 \right) \bar{\bar{y}}_i \tag{14}$$

5. **Estimación de α_0 y α_i usando su valor modal:** un enfoque que puede llevarse en paralelo es despejar α_i en términos de su valor modal. De (9) encontramos que:

$$\alpha_i = Moda_i \left[\frac{(\bar{\bar{y}}_i - \bar{\bar{y}}_i^2)}{S_i^2} - 1 \right] - Moda_i * k + 1. \tag{15}$$

   Luego el facilitador podrá escoger el valor de α_i que más le convenga, es decir, usar la estimación de α_0 y α_i proveniente de su valor modal, usar la estimación de α_0 y α_i proveniente de su valor esperado o simplemente promediar los dos valores de α_i.

Note que los pasos 4 y 5 dependen mucho del paso 1, ya que los valores de α_i se ven afectados directamente por cualquier modificación de n.

4. Aplicación: estimación del Score de Gleason en pacientes diagnosticados con cáncer de próstata en Colombia

El Score de Gleason es un sistema que se emplea para medir el grado de agresividad de un cáncer, basándose en la observación al microscopio de las características que presentan las células de la muestra obtenida en una biopsia del órgano. El procedimiento consiste en seleccionar dos zonas de la muestra y asignar a cada una de ellas un número del 1 al 5. El 1 corresponde a un tumor bien diferenciado y por lo tanto poco agresivo y el 5 a un tumor escasamente diferenciado. Los valores comprendidos entre el 2 y el 4 se asignan a grados de diferenciación intermedia. Posteriormente se suman las cifras obtenidas en las dos zonas y se obtiene un número comprendido entre el 2 y el 10. Este valor es el Score Gleason (Zollo 2005, Gleason et al. 2002). Los posibles resultados son:

- Escala de Gleason entre 2 y 4: cáncer con escasa agresividad, crecimiento lento y por lo tanto de mejor pronóstico.
- Escala de Gleason de 5 y 7: cáncer con agresividad intermedia.
- Escala de Gleason entre 8 y 10: cáncer de alta agresividad, y peor pronóstico.

La aplicación busca estimar la prevalencia de cada nivel del Score de Gleason en pacientes que han sido diagnosticados con cáncer de próstata. Conocer la distribución de los diferentes niveles de Gleason en los pacientes ya diagnosticados con cáncer de próstata en Colombia es de gran ayuda en temas de salud pública, dado que la trascendencia y significado del Score de Gleason en esta patología es invaluable, ya que no solo permite diagramar mejor una estrategia diagnóstica, clínica y quirúrgica, sino también permitirá, sin duda alguna, pronosticar la evolución de los casos estudiados en el tiempo y las posibilidades de recaída bioquímica que se traducen en estrategias de tratamientos asociados y mayor alerta para los médicos a cargo de dichos casos Potenziani (2014).

4.1. Metodología

Con el fin de que el proceso de elicitación satisfaga un escrutinio profesional, además de contextualizar al experto sobre el tema, es necesario incluir los siguientes pasos durante el proceso:

- **Diseño y validación de preguntas**: dado que el método de elicitación propuesto usa una técnica de elicitación indirecta, se preguntó al experto directamente por el número de pacientes (ya diagnosticados con cáncer de próstata) que él considera que están en cada uno de los niveles (2-4, 5-7 y 8-10) del Score de Gleason dada una muestra hipotética de pacientes. Antes de preguntar por el número de pacientes en cada nivel del Score de Gleason, se dio al experto una introducción del problema, donde se validó que la pregunta fuera entendida correctamente y que el experto en efecto tuviera un respuesta para dicha pregunta.
- **Identificación y contratación del experto**: como experto se cuenta con la participación del profesor de la Universidad Sur Colombiana, Dr. Manuel García, es considerado como experto por sus diferentes investigaciones sobre la próstata y cáncer de próstata dentro de las cuales se destaca un Posdoctorado el cual titula "Analisis de microRNA que regulan el Receptor de Andrógeno en el cáncer de próstata" realizado en la Universidad de Sao Paulo.
- **Estructuración, descomposición y entrenamiento en probabilidad**: se da al experto una introducción, inicialmente sobre la distribución de probabilidad Multinomial y su relación con el Score de Gleason. Posteriormente sobre elicitación de juicios de expertos. Finalmente se especifica al experto con exactitud que la cantidad incierta que se desea estimar es la prevalencia de cada uno de los niveles del Score de Gleason en la población de pacientes diagnosticados con cáncer de próstata y, se determina que la escala de medición es el número de individuos en cada nivel del Score. Con el fin de que el experto en sus estimaciones tenga presente la causalidad y factores de riesgo que pueden influir en la adquisición de la enfermedad y en la distribución del Score, se muestra al experto durante todo el proceso de elicitación una representación gráfica (mapa mental) que refleja los principales factores de riesgo y la causistica del problema.

- **Aplicación del método**:

  1. Estimación del $n-$equivalente: con base en en los diferentes estudios sobre el cáncer de próstata que ha realizado y su cercanía con el área de urología, se considera al Dr. García como un experto con alto nivel de experticia por lo que se le asigna un $n-$equivalente de 333, el cual según el método 2 (presentado en sección anterior) corresponde a un $\alpha = 0.20$ y un error de estimación igual al 5 % (ver tabla 1; note que aquí α tiene que ver con el nivel de significancia en la estimación del tamaño muestral y no con el vector de parámetros de la distribución Multinomial).

  2. Estimación de la frecuencia relativa: por medio de preguntas como " Dr García, si se seleccionan aleatoriamente 100 pacientes diagnosticados con cáncer de próstata en Colombia, cuántos de ellos según su conocimiento, se encuentran en el nivel 1 del Score de Gleason (Gleason entre 2 y 4), cuántos en el nivel 2 (Gleason entre 5 y 7) y finalmente cuántos en el nivel 3 (Gleason entre 8 y 10)" este proceso se repite seis veces, donde se le pide al experto distribuir seis muestras hipotéticas diferentes en los tres niveles del Score de Gleason. Una vez el experto ha realizado la distribución de las seis muestras, se calcula la frecuencia relativa para cada una de ellas. Si las frecuencias son cercanas se procede con la etapa de simulación tomando como frecuencia relativa el promedio de las seis frecuencias estimadas. Si las frecuencias estimadas divergen una de otra se procede a validar con el experto (imitando el método Delphi) la coherencia de dichas estimaciones, se muestran las frecuencias estimadas para las seis muestras hipotéticas y se pide al experto que redistribuya nuevamente las muestras hipotéticas de manera que las frecuencias relativas estimadas estén lo más cerca posible.

  3. Simulación y estimación del vector de parámetros: una vez se valida la coherencia de las estimaciones realizadas por el Dr. García, se lleva a cabo el proceso de simulación y estimación del vector de parámetros $\underline{\alpha} = (\alpha_1, \alpha_2, \alpha_3)$. Para dicho proceso solo es necesario introducir los valores estimados dentro del software R, el cual se encarga de realizar la simulación y entregar los valores estimados de α_1, α_2 y α_3.

- **Retroalimentación**: luego de ser elicitadas las muestras hipotéticas estipuladas, se mostró al experto una gráfica con la distribución de cada uno de los niveles del Score de Gleason de manera que él validara que dichas distribuciones si corresponden con sus creencias. Si el experto estaba de acuerdo con los resultados se daba por finalizado el proceso de elicitación, de lo contrario, se procedía a actualizar las proporciones estimadas para cada muestra hipotética. Paralelo a esto se usó durante todo el proceso de elicitación un mapa mental que ilustrara los factores más influyentes en los pacientes con cáncer de próstata, de manera que cada estimación fuera analizada bajo dichos factores.

4.2. Resultados

Las muestras hipotéticas elicitadas fueron:

Tabla 2: *Muestras hipotéticas elicitadas y estimaciones para cada nivel del Score de Gleason. Fuente: elaboración propia.*

Muestra	Nivel 1	Nivel 2	Nivel 3
50	10	25	15
150	35	80	35
250	80	150	20
340	80	180	80
510	90	300	120
713	150	400	163

Las estimaciones obtenidas para la distribución Dirichlet son:

Tabla 3: *Vector de parámetros $\underline{\alpha}$ con su respectiva media y varianza para los valores elicitados. Fuente: elaboración propia.*

	Alfa	Media	Varianza
α_1	141.64	0.221	0.0002
α_2	417.74	0.652	0.0003
α_3	81.186	0.126	0.0001

donde los intervalos de probabilidad con una región de credibilidad de $100(1-\alpha)\,\%$ y $\alpha = 0.05$ para cada nivel del Score están dados por:

Tabla 4: *Intervalos de probabilidad con credibilidad de* $100(1-\alpha)\,\%$*. Fuente: elaboración propia.*

Nivel	LI	LS
1	0.073	0.137
2	0.577	0.678
3	0.223	0.316

El hecho de que los pacientes con menor proporción sean los del nivel 1 nos da indicio que en Colombia la detección temprana del cáncer es un problema latente, por otro lado, el resultado de que alrededor del 65 % de los pacientes diagnosticados se encuentren en el nivel 2 es un insumo para validar si las políticas de salud pública se orientan a atender pacientes con este nivel de agresividad y si las clínicas y hospitales en Colombia están preparadas para atenderlos, vale la pena resaltar que el tratamiento del cáncer de próstata, incluso en la enfermedad clínicamente localizada, es cada vez más complejo debido a las diversas opciones terapéuticas disponibles que presentan una eficacia oncológica equivalente, pero efectos secundarios relacionados con el tratamiento significativamente diferentes Heidenreich et al. (2010), por esto, conocer si las clínicas y hospitales están preparadas para brindar un tratamiento adecuado a pacientes con un Score de Gleason en el nivel 2 puede generar un impacto significativo en la población diagnosticada con la enfermedad. Finalmente encontramos que una proporción no despreciable se encuentra en el nivel 3, esta proporción requiere de tratamientos más agresivos por lo que al igual que con los pacientes que se encuentran en el nivel 2 vale la pena validar si la orientación de las políticas de salud pública, la clínicas y hospitales están preparados para brindar atención oportuna y tratamientos adecuados a éstos pacientes, ya que, las decisiones relativas al tratamiento en cada nivel del Score de Gleason deben basarse en guías clínicas, indicando claramente la que se utilice en el proceso de toma de decisiones, donde adicionalmente es aconsejable una estrategia multidisciplinar Heidenreich et al. (2010).

5. Conclusiones

El método propuesto es una forma sencilla de elicitar la distribución del vector de parámetros π de la distribución Multinomial, note que con tan solo preguntar una vez al experto por la distribución de una muestra hipotética en las diferentes categorías de la variable de interés es posible hacer una estimación de la distribución del vector de parámetros π; sin embargo, resaltamos que el facilitador debe aumentar su atención en la validación de la concordancia entre la distribución obtenida y los juicios del experto, ya que por la naturaleza del método es fácil que el experto caiga en la heurística de anclaje y sobre ajuste Tversky & Kahneman (1974). Por otro lado, es importante que el facilitador realice un trabajo cuidadoso en la selección del experto, en lo posible siguiendo los criterios dados en Hora & Von Winterfeldt (1997), ya que esta es fundamental para la estimación del $n-$equivalente, el cual juega un papel vital en el rendimiento del método propuesto. Para dar una "mejor" calificación *a priori* sobre la experticia del experto el facilitador puede iniciar diagnosticando si el conocimiento es de tipo nominal, sustantivo o una mezcla de los dos, posteriormente hacer uso de alguno de los dos métodos propuestos anteriormente o de cualquier otro método que le permita cuantificar dicha experticia de manera que pueda ser puesta en términos de una muestra aleatoria. Actualmente en la literatura sobre elicitación de juicios de expertos existen pocos trabajos que orienten al facilitador en como calificar la experticia del experto de una forma óptima, entonces, queda el camino abierto para que otros investigadores interesados en el método hagan sus propuestas sobre la forma "más óptima" de estimar el $n-$equivalente.

Finalmente vale la pena resaltar que si comparamos el procedimiento de elicitación aplicado con seleccionar una muestra aleatoria representativa de la población de pacientes diagnosticados con cáncer de próstata y realizar en dicha muestra una biopsia que determine el nivel del Score de Gleason, podemos ver que el proceso de elcitación es más económico y relativamente más rápido. Ahora si ya se tienen datos muestrales sobre la distribución del Score de Gleason, la distribución estimada por el método de elicitación también puede ser actualizada por estos datos, de manera que el resultado será una distribución más cercana o la real.

Recibido: 27 de noviembre de 2014
Aceptado: 17 de marzo de 2015

Referencias

Bromaghin, J. F. (1993), 'Sample size determination for interval estimation of multinomial probabilities', *The American Statistician* **47**(3), 203–206.

Budnitz, J., Boore, M., Apostolakis, G., Cluff, S., Coppersmith, J., Cornell, A. & Morris, A. (1997), 'Recommendations for probabilistic seismic hazard analysis: Guidance on uncertainty and use of experts', **1**, 1–278. *http://www.osti.gov/scitech/servlets/purl/479072

Chaloner, K. M. & Duncan, G. T. (1983), 'Assessment of a beta prior distribution: PM elicitation', *The Statistician* pp. 174–180.

Cooke, R. M. & Goossens, L. L. (2008), 'TU Delft expert judgment data base', *Reliability Engineering & System Safety* **93**(5), 657–674.

DeGroot, M. H. (2004), *Optimal Statistical Decisions*, John Wiley & Sons.

Dickey, J., Jiang, J., Kadane, J., at Albany, S. U. o. N. Y. & of America, U. S. (1983), 'Bayesian Methods for Multinomial Sampling With Missing Data Using Multiple Hypergeometric Functions'.

Duran, B. S. & Booker, J. M. (1988), 'A Bayes sensitivity analysis when using the beta distribution as a prior', *Reliability, IEEE Transactions on* **37**(2), 239–247.

Elfadaly, F. & Garthwaite, P. (2012), *On Eliciting Some Prior Distributions for Multinomial Models*, Department of Mathematics and Statistics.

Fowler, F. J. (1995), *Improving survey questions: Design and evaluation*, Vol. 38, Sage.

Fox, B. L. (1966), 'A Bayesian approach to reliability assessment'.

Garthwaite, P. H., Kadane, J. B. & O'Hagan, A. (2005), 'Statistical methods for eliciting probability distributions', *Journal of the American Statistical Association* **100**(470), 680–701.

Gavasakar, U. (1988), 'A comparison of two elicitation methods for a prior distribution for a binomial parameter', *Management Science* **34**(6), 784–790.

Gill, J. & Walker, L. D. (2005), 'Elicited priors for Bayesian model specifications in political science research', *Journal of Politics* **67**(3), 841–872.

Gilless, J. K. & Fried, J. S. (2000), 'Generating beta random rate variables from probabilistic estimates of fireline production times', *Annals of Operations Research* **95**(1-4), 205–215.

Gleason, D. F., Mellinger, G. T., Group, V. A. C. U. R. et al. (2002), 'Prediction of prognosis for prostatic adenocarcinoma by combined histological grading and clinical staging', *The Journal of urology* **167**(2), 953–958.

Gross, A. J. (1971), 'The application of exponential smoothing to reliability assessment', *Technometrics* **13**(4), 877–883.

Heidenreich, A., Bolla, M., Joniau, S., Mason, M., Matveev, V., Mottet, N., Schmid, H. et al. (2010), 'Guía clínica sobre el cáncer de próstata', *Asociación Europea de Urología* .

Hora, S. C. & Von Winterfeldt, D. (1997), 'Nuclear waste and future societies: A look into the deep future', *Technological Forecasting and Social Change* **56**(2), 155–170.

Hughes, G. & Madden, L. (2002), 'Some methods for eliciting expert knowledge of plant disease epidemics and their application in cluster sampling for disease incidence', *Crop Protection* **21**(3), 203–215.

Jenkinson, D. (2005), The elicitation of probabilities: A review of the statistical literature, Technical report, Citeseer.

Johnson, N. L., Kotz, S. & Balakrishnan, N. (1997), *Discrete multivariate distributions*, Vol. 165, Wiley New York.

Joseph, L., Gyorkos, T. W. & Coupal, L. (1995), 'Bayesian estimation of disease prevalence and the parameters of diagnostic tests in the absence of a gold standard', *American Journal of Epidemiology* **141**(3), 263–272.

Kadane, J. B. & Wolfson, L. J. (1998), 'Experiences in elicitation', *The Statistician* pp. 3–19.

Kynn, M. (2006), 'Designing elicitor: Software to graphically elicit expert priors for logistic regression models in ecology', *Available from www. winbugs-development. org. uk* .

Oakley, J. E., Daneshkhah, A. & O'Hagan, A. (2014), Nonparametric Prior Elicitation using the Roulette Method, Technical report, Technical report, available at http://www. tonyohagan. co. uk/academic/pdf/elic-roulette. pdf. Accessed July 1.

O'Hagan, A. (1998), 'Eliciting expert beliefs in substantial practical applications', *The Statistician* pp. 21–35.

O'Hagan, A. (2005), *Research in elicitation*, University of Sheffield, U.K.

Potenziani, J. (2014), Significado del Grado de Gleason y del Score de Gleason en pacientes con Cáncer Prostático.
*http://www.sexarchive.info/BIB/Potenziani/gleason.htm

R Core Team (2014), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
*http://www.R-project.org/

Regazzini, E. & Sazonov, V. V. (1999), 'Approximation of laws of multinomial parameters by mixtures of Dirichlet distributions with applications to Bayesian inference', *Acta Applicandae Mathematica* **58**(1-3), 247–264.

Tortora, R. D. (1978), 'A note on sample size estimation for multinomial populations', *The American Statistician* **32**(3), 100–102.

Tovar, J. R. (2012), 'Eliciting beta prior distributions for binomial sampling', *Rev. Bras. Biom* **30**(1), 159–172.

Tversky, A. & Kahneman, D. (1974), 'Judgment under uncertainty: Heuristics and biases', *science* **185**(4157), 1124–1131.

Van Noortwijk, J. M., Dekker, R., Cooke, R. M., Mazzuchi, T. et al. (1992), 'Expert judgment in maintenance optimization', *Reliability, IEEE Transactions on* **41**(3), 427–432.

Waterman, M., Martz, H. & Waller, R. (1976), Fitting beta prior distributions in Bayesian reliability analysis, Technical report, Los Alamos Scientific Laboratory Report.

Weiler, H. (1965), 'The use of incomplete beta functions for prior distributions in binomial sampling', *Technometrics* **7**(3), 335–347.

White, I. R., Pocock, S. J. & Wang, D. (2005), 'Eliciting and using expert opinions about influence of patient characteristics on treatment effects: a Bayesian analysis of the CHARM trials', *Statistics in medicine* **24**(24), 3805–3821.

Zapata, R. E., O'Hagan, A. & Soares Bastos, L. (2012), 'Eliciting expert judgements about a set of proportions', *Journal of Applied Statistics* **41**(9), 1919–1933.

Zollo, A. J. (2005), *Medicina interna*, Elsevier España.