PROPUESTA PARA OBTENER DISTRIBUCIONES PREVIAS PARA LOS PARÁMETROS DE LA DISTRIBUCIÓN BETA

Luis Gabriel Arroyo Bravo[1], Fabian Alejandro Lasso Balanta, José Rafael Tovar Cuevas
Estadísticos. Escuela de Estadística, Universidad del Valle, Santiago de Cali, Colombia

ABSTRACT
The Beta distribution has gained importance in the field of applied statistics, because it is a probability function that allows us to model the natural behavior of random variables that can only take values within a delimited range of real numbers with finite length. A special case quite common in practice is that of the variables whose domain is within the interval (0.1), the case of proportions or percentages. In this article, a methodology is proposed to obtain a priori distributions for the shape parameters of the Beta distribution, which entails a new approach for these distribution from the Bayesian paradigm, since, most of the time, it is used as a prior distribution when the sample data can be adjusted to a binomial distribution, but in the case of interest for this work, the sample data can be adjusted with a Beta distribution which would imply having a likelihood function that is the product of n densities Beta. A joint distribution for the distribution parameters was initially obtained from the information for the mean and the variance of the prior distribution. Finally, the a priori distributions for the shape parameters were obtained marginalizing the joint distribution.

KEYWORDS: Beta distribution, a priori distribution, 60E05, marginal distribution, Bayesian inference.

MSC: 62F15, 60E05

RESUMEN
La distribución Beta ha ganado importancia en el campo de la estadística aplicada, debido a que es una función de probabilidad que permite modelar el comportamiento natural de variables aleatorias que solo pueden tomar valores dentro de un intervalo delimitado de números reales con longitud finita. Un caso especial bastante común en la práctica, es el de las variables cuyo dominio se encuentra dentro del intervalo (0,1), caso de las proporciones o porcentajes. En este artículo, se propone una metodología para obtener distribuciones a priori para los parámetros de forma de la distribución Beta, lo cual conlleva un nuevo enfoque para dicha distribución desde el paradigma Bayesiano, pues, la mayoría de las veces es utilizada como distribución previa cuando los datos de la muestra pueden ser ajustados a una distribución binomial, pero en el caso de interés para este trabajo, los datos de la muestra pueden ser ajustados con una distribución Beta lo que implicaría tener una función de verosimilitud que es el producto de n densidades Beta. Se obtuvo inicialmente una distribución conjunta para los parámetros de la distribución a partir de la información externa sobre la media y la varianza de la distribución a priori. Finalmente, las distribuciones a priori para los parámetros de forma, fueron obtenidas a partir de las marginales de la distribución conjunta.
PALABRAS CLAVE: Distribución Beta, distribución a priori, distribución conjunta, distribución marginal, Inferencia Bayesiana.

[1] luis.arroyo@correounivalle.edu.co

1. **INTRODUCCIÓN**

En la práctica, muchas de las variables aleatorias asociadas a los problemas de investigación, toman valores en espacios acotados, es decir intervalos cerrados, semiabiertos o semicerrados con límites inferior y superior en los números reales. Para modelar probabilísticamente el comportamiento de este tipo de variables, en la literatura se pueden encontrar funciones de probabilidad como la triangular, la Kumaraswamy y la distribución Beta, además del truncamiento que se puede realizar sobre modelos diseñados para variables aleatorias que tienen un recorrido en la recta real o en el subconjunto de los números reales positivos. La distribución Beta con parámetros de forma α y β es quizá, la alternativa más comúnmente usada para representar el comportamiento natural de variables que toman valores dentro de intervalos (c, d) siendo c y d números reales positivos. En términos más específicos, cuando se trata de modelar proporciones, indicadores construidos a partir de la teoría de las variables latentes y otras variables aleatorias cuyos rangos de movimiento están definidos en el espacio (0,1), se asume el modelo probabilístico Beta estándar, el cual es un caso especial del modelo Beta generalizado, como muestran [7]. Para esta distribución, es posible encontrar la forma de estimar los parámetros utilizando métodos como el de momentos y el de la máxima verosimilitud. La estimación usando método de máxima verosimilitud, no es sencilla y es necesario el uso de métodos iterativos como el algoritmo Newton Raphson, ya que, cuando se necesita derivar la función de log-verosimilitud con respecto a los parámetros, la expresión se indetermina.

Cuando no es posible hacer la estimación usando métodos del paradigma frecuentista, o se desea contar con una fuente adicional de información para obtener las estimaciones, el paradigma Bayesiano surge como una alternativa importante. Contar con estimaciones obtenidas a partir de distribuciones posteriores es beneficioso, dado que dichas estimaciones consideran el conocimiento o experiencia subjetiva que tienen los expertos en el problema objeto de estudio (Ver [1], [6], [16] y [17]. Cuando a los datos de la muestra se les puede asociar una distribución Beta estándar para modelar su comportamiento (verosimilitud Beta), usar metodología bayesiana puede volverse una tarea bastante compleja, pues no se cuenta con opciones de distribuciones de probabilidad previas (a priori) que permitan obtener formas mínimamente amigables para la distribución posterior. Adicionalmente, se presenta la limitación de que a los dos parámetros de forma no es posible darles algún tipo de interpretación que se adapte a algún contexto real en específico, solo se conocen interpretaciones desde el punto de vista matemático (constantes que permiten que la función tome diferentes formas) o desde el punto de vista probabilístico (parámetros de forma de la distribución de probabilidad).

Algunos autores como [11], realizaron la estimación bayesiana en los Modelos de Mixtura Beta (BMM) proponiendo un método, bajo la inferencia variacional, el cual consiste en una aproximación a la distribución de los parámetros y de esta manera obtener una solución de forma cerrada. Empleando los principios del método de aproximación factorizada. Asumiendo independencia entre los parámetros, les asignan un producto de distribuciones Gamma como distribución previa conjunta. Usando un modelo jerárquico obtienen las estimaciones de los parámetros sujetando la $n-ésima$ observación Beta a una variable indicadora latente, esto genera un conjunto de vectores indicadores que se asumen como independientes, obteniendo una mezcla de proporciones condicionantes de la distribución de la variable indicadora a las que se les asignan distribuciones Dirichlet. Se obtiene entonces, una distribución conjunta y siguiendo el esquema de actualización de la inferencia variacional son obtenidos los hiperparámetros de manera iterativa.

Por su parte, [13] estimó los parámetros de la distribución Beta usando el método de la máxima verosimilitud, el método de los momentos y tres métodos más; La técnica de revisión y evaluación del programa (PERT) en el cual calculan la media y la varianza de la distribución en función de un valor modular y con un procedimiento análogo al método de los momentos, igualan estos resultados a la media y la varianza de la distribución Beta para obtener las estimaciones. El otro método es una modificación de la distribución de potencia de dos colas (TSP) donde los autores plantean que esta distribución tiene un comportamiento similar al de la distribución Beta y calculan su media y su varianza para luego igualar los resultados a la media y la varianza de la distribución Beta y así estimar. El último método consiste en obtener un estimador basado en el primer y tercer cuartil de la distribución Beta, el cual es un método empírico que utiliza el primer y tercer cuartil poblacional de la distribución asumiendo que el valor real del parámetro es conocido y después se generan muestras de la distribución cuyos valores de los parámetros se encuentran en un intervalo construido alrededor de la estimación obtenida con el método de los momentos.

También, [9] realizaron un estudio para estimar los parámetros de la distribución Beta Generalizada (BG)basado en datos agrupados desde una perspectiva Bayesiana utilizando el algoritmo bloques aleatorios a medida Metrópolis-Hastings (TaRBMH), que se basa en el existente algoritmo de caminata aleatoria Metrópolis-Hastings (RWMH).

En este nuevo algoritmo, los parámetros son codificados y se agrupan en varios bloques distintos aleatoriamente, lo cual evita una mala elección de bloques a priori. Luego, cada bloque de parámetros se actualiza en secuencia por un paso del Metrópolis-Hastings teniendo en cuenta la densidad propuesta, hasta que ésta se adapta tanto en ubicación como en la curvatura de la densidad posterior del bloque, condicionado al valor más actual de los parámetros en los bloques restantes.
Los autores, tienen en cuenta la verosimilitud Beta Generalizada con sus respectivos parámetros, a los cuales les asigno una distribución a priori a cada uno de ellos teniendo en cuenta las restricciones que éstos presentan. Al parámetro de localización se le asignó una distribución a priori Normal, al parámetro de escala y a los dos de forma les asignaron distribuciones a priori Gamma, y cada una de estas distribuciones cuentan con sus correspondientes hiperparámetros. Luego, multiplicando estas distribuciones a previas con la verosimilitud BG obtuvieron la función de densidad de la distribución posterior e implementaron el algoritmo TaRBMH para calcular las estimaciones.
Una de las limitaciones que tiene el método Bayesiano de la inferencia estadística, es aproximar formas distribucionales complejas y con estructura analítica no cerrada. El problema de aproximar integrales y distribuciones ha sido ampliamente trabajado en las matemáticas y la estadística. Algunos autores como [15], realizaron la comparación de métodos de integración pertenecientes a las fórmulas de Newton - Cotes, éstas son un grupo de fórmulas de integración numérica como lo son la regla de Trapecio y la regla de Simpson entre otras, de tipo interpola torio que permiten hallar un valor aproximado de la integral, evaluando la función en puntos equidistantes, entre mayor sea el número de intervalos en que se divida la función, más preciso será el resultado. Esta comparación la llevaron a cabo con la intención de analizar cuál de los métodos utilizados realiza la aproximación más precisa de una integral. Entonces, teniendo en cuenta todas las restricciones teóricas que los métodos utilizan, llevaron a cabo los respectivos procedimientos y al obtener los resultados de las aproximaciones de las integrales de una función, los autores llegaron a la conclusión que los algoritmos que en la práctica demostraron producir resultados más precisos en cuanto a la aproximación al valor de la integral, utilizando pocas iteraciones y con errores pequeños, son los que se basan en el esquema del trapecio.
Por otra parte [4] hicieron uso de algunos métodos de Monte Carlo vía Cadenas de Markov (MCMC) como lo son los algoritmos de Metropolis – Hasting y el Muestreo de Gibbs por bloques, que generalmente son empleados en la estadística Bayesiana. Estos autores, realizaron esto con la intención de obtener valores de una distribución de probabilidad asociada a la arena C4 de la Formación Misoa, en el Lago de Maracaibo, debido a que el muestreo directo resulta complejo. Entonces, con los valores obtenidos los autores generaron pseudo-columnas estratigráficas para caracterizar un yacimiento. Luego, combinaron la información de las columnas reales y las columnas estimadas, y generaron mapas de contenido de arena. Finalmente, después de observar los resultados llegaron a la conclusión que el método de Metrópolis – Hasting distinguió, principalmente, la presencia de arenas a lo largo de las pseudo secuencias generadas, las cuales representan más del 70% del sedimento presente en el área de estudio, también, respecto al algoritmo Muestreo de Gibbs por bloques, cuando se utiliza una longitud apropiada del bloque, es capaz de detectar la presencia de capas delgadas de las otras litologías observadas en el área
Dado el nivel de complejidad tan alto que presenta operar algebraicamente con la función de densidad de la distribución Beta, estimar los parámetros de la misma es una tarea bastante tediosa si, por ejemplo, se considera el método de la máxima verosimilitud. Desde el punto de vista Bayesiano, el problema es aún mayor puesto que ambos parámetros de la distribución Beta son parámetros de forma para los cuales no existe una interpretación sencilla desde el punto de vista de las aplicaciones. El objetivo principal de este artículo es, desarrollar un método que permita obtener formas para las distribuciones a priori en escenarios en los que se requiere estimar los parámetros de la distribución Beta debido a que se asume que dicho modelo es el más adecuado para explicar el comportamiento natural de los datos observados en la muestra. El procedimiento parte del supuesto, de que el proceso de extracción de información externa a la muestra (la cual puede estar en el conocimiento de un especialista en el tema, en publicaciones, datos históricos, etc.) se ha soportado en la media y la varianza de la distribución previa. Adicionalmente, se parte del supuesto de que la muestra de datos de trabajo cumple con la propiedad de ser permutable.
Dado que el proceso de estimación considera una distribución previa biparamétrica, la metodología propuesta considera inicialmente construir una distribución a priori conjunta para los dos parámetros y luego hacer un procedimiento para obtener las distribuciones marginales que serían la de cada uno de los parámetros. Como la forma de la distribución conjunta para los dos parámetros de la distribución Beta previa, es analíticamente intratable, se desarrolló una propuesta que combina el algoritmo de generación de muestras para distribuciones Metropolis Hastings con la regla de los trapecios, método de aproximación numérica para integrales complejas.

2. REFERENTES TEÓRICOS

Para [8], la distribución Kumaraswamy pertenece a una familia de distribuciones de probabilidad para variables aleatorias continuas, definidas en el intervalo $(0,1)$ cuyos parámetros de forma (a_1, b_1) corresponden al conjunto de los reales no negativos. La función de densidad de probabilidad y la distribución de probabilidad acumulada para una variable X que se distribuye Kumaraswamy son:

$$f(x) = a_1\, b_1\, x^{\,a_1-1}\,(1-x^{a_1})^{\,b_1-1};\ 0 < x < 1;\ a_1 > 0;\ b_1 > 0 \qquad \textbf{(1)}$$

$$F(x) = \int_0^x f(t)dt = 1-(1-x^{a_1})^{b_1} \qquad \textbf{(2)}$$

El $n-ésimo$ momento de la distribución puede ser obtenido usando la expresión:

$$m_n = \frac{b_1\Gamma\left(1+\frac{n}{a_1}\right)\Gamma(b_1)}{\Gamma\left(1+b_1+\frac{n}{a_1}\right)} \qquad \textbf{(3)}$$

De modo que, la esperanza matemática y la varianza son:

$$\mu^* = m_1 = \frac{b_1\Gamma\left(1+\frac{1}{a_1}\right)\Gamma(b_1)}{\Gamma\left(1+b_1+\frac{1}{a_1}\right)} \qquad \textbf{(4)}$$

$$\sigma^{2*} = m_2 - m_1^2 \qquad \textbf{(5)}$$

También es posible obtener la mediana usando la expresión:

$$m_e^* = \left(1-2^{-1/b_1}\right)^{1/a_1} \qquad \textbf{(6)}$$

2.1. Algoritmo Metropolis-Hastings (MH)

Se han ideado muchos métodos inteligentes para construir y muestrear distribuciones posteriores arbitrariamente. La simulación de una cadena de Markov Monte Carlo (MCMC) es un método general basado en obtener valores de θ a partir de distribuciones aproximadas y luego corrigiendo dichos valores para aproximar la distribución posterior objetivo, $f(\theta|x)$ [5]. El Algoritmo MH es un algoritmo MCMC, utilizado para obtener una secuencia de muestras aleatorias a partir de una distribución de probabilidad para la cual resulta difícil realizar el muestreo directo [5].
Se define la probabilidad de aceptación de tal manera que cuando se combine con un kernel de transición arbitrario q (distribución propuesta), este define una cadena reversible. La expresión más comúnmente citada para la probabilidad de aceptación es:

$$\alpha(\theta,\phi) = min\left\{1, \frac{\pi(\phi)q(\phi,\theta)}{\pi(\theta)q(\theta,\phi)}\right\} = min\left\{1, \frac{\pi(\phi)w(\phi|\theta)h(\theta)}{\pi(\theta)w(\theta|\phi)h(\phi)}\right\} \qquad \textbf{(7)}$$

El kernel de transición propuesto q es arbitrario y, por lo tanto, es una herramienta flexible para la construcción del algoritmo. Ver [5] para más detalles.

3. MÉTODO PROPUESTO

Bajo el supuesto de que el conjunto de observaciones obtenidas después de realizar un muestreo para obtener información acerca de un parámetro poblacional, presenta un comportamiento natural que puede ser modelado una distribución Beta estándar, el modelo estadístico asociado a la situación, considera que la función de densidad es la que aparece en la ecuación **(8)**.

$$f_x(x|\alpha,\beta) = \frac{1}{B(\alpha,\beta)}x^{\alpha-1}(1-x)^{\beta-1}, \qquad 0 \le x \le 1;\ \alpha > 0,\ \beta > 0 \qquad \textbf{(8)}$$

Con,

$$E(X) = \mu = \frac{\alpha}{\alpha+\beta} \quad ; \quad Var(X) = \sigma^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \qquad \textbf{(9)}$$

Donde α y β son los parámetros de forma de la distribución a los cuales no es posible darles una interpretación práctica, hecho que complejiza bastante llevar a cabo un proceso de elicitación, procedimiento indispensable cuando se quiere hacer inferencia Bayesiana (Ver [12], para ampliar conceptos de elicitación). Sin embargo, la media y la varianza de esta distribución si tienen fácil interpretación desde cualquier marco contextual pues son indicadores de tendencia central y dispersión de los datos respectivamente. Dado que la media y la varianza son funciones de α y β, entonces, es posible elicitar información para μ y σ^2 como parámetros de interés, asignarles distribuciones previas y después derivar las distribuciones previas para α y β

En este contexto, se hace necesario conocer una serie de requerimientos para crear una idea de cuales distribuciones serán razonables, para asignarlas como distribuciones previas para μ y σ^2 . El primero de estos, es establecer en qué espacio de los números reales toman valores los dos parámetros. Así, teniendo en cuenta que $\beta \in (0, \infty)$ y $\alpha \in (0, \infty)$, se establece que:

$$\lim_{\beta \to \infty} \mu = \lim_{\beta \to \infty} \frac{\alpha}{\alpha+\beta} = \lim_{\alpha \to 0} \mu = \lim_{\alpha \to 0} \frac{\alpha}{\alpha+\beta} = 0 \qquad \textbf{(10)}$$

$$\lim_{\beta \to 0} \mu = \lim_{\beta \to 0} \frac{\alpha}{\alpha+\beta} = \lim_{\alpha \to \infty} \mu = \lim_{\alpha \to \infty} \frac{\alpha}{\alpha+\beta} = 1 \qquad \textbf{(11)}$$

Considerando que los límites anteriores, es evidente que $\mu \in (0,1)$. Procediendo de manera análoga para σ^2 se tiene que:

$$\lim_{\beta \to \infty} \sigma^2 = \lim_{\beta \to \infty} \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} = \lim_{\alpha \to \infty} \sigma^2 = \lim_{\alpha \to \infty} \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} = 0 \qquad \textbf{(12)}$$

$$\lim_{\beta \to 0} \sigma^2 = \lim_{\beta \to 0} \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} = \lim_{\alpha \to 0} \sigma^2 = \lim_{\alpha \to 0} \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} = 0 \qquad \textbf{(13)}$$

Basado en el resultado anterior, si la intención es; establecer un rango de números reales para σ^2, analizar los límites no proporciona suficiente información ya que, es claro que la varianza se aproxima a cero cuando alguno de los parámetros de forma tiende a cero o cuando ambos tienden a infinito.

Se realizó un análisis gráfico al comportamiento de σ^2en el espacio, cuando están variando α y β (Ver **Figura 1)**. Es posible observar que el máximo valor que la varianza puede tomar, está muy cercano a 0.25, lo cual concuerda con lo inferido en el cálculo de sus límites, pues sin importar el valor de α, a medida que los valores del parámetro β se acercan a cero o incrementan después de dos, σ^2 siempre tiende a cero. En consecuencia, es razonable pensar que la variabilidad de la distribución Beta siempre está dentro del intervalo (0, 0.25).

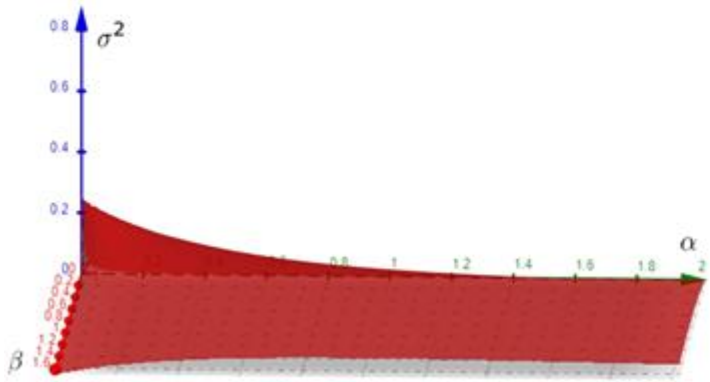

Figura 1: Comportamiento de σ^2 ante diferentes valores de (α, β)

Como μ y σ^2 son funciones de los parámetros de forma, lo que se propone es obtener una distribución a priori conjunta $f(\mu, \sigma^2)$, de modo tal, que utilizando el método de la transformación para distribuciones de variables aleatorias, se pueda derivar una distribución previa conjunta de la forma $f(\alpha, \beta)$, y finalmente marginalizar $f(\alpha)$ y $f(\beta)$. En este contexto, es necesario conocer si la transformación es uno a uno.

Hasta el momento, se sabe que valores positivos de α y β establecen que $\mu \in (0, 1)$ y $\sigma^2 \in (0, 0.25)$. se requiere entonces, conocer si estos últimos derivan en valores positivos para los parámetros de forma al realizar la transformación inversa:

$$\alpha = \mu\left[\frac{\mu(1-\mu)}{\sigma^2} - 1\right], y\ \ \beta = (1-\mu)\left[\frac{\mu(1-\mu)}{\sigma^2} - 1\right] \qquad \textbf{(14)}$$

Analíticamente se observa que para α y β pertenecer al intervalo $(0, \infty)$, restricción de la función de densidad Beta, se debe cumplir que $\frac{\mu(1-\mu)}{\sigma^2} - 1 > 0$ y $\sigma^2 < \mu(1-\mu)$. Entonces, para que α y β sean positivos,μ puede tomar cualquier valor dentro del intervalo (0, 1), sin embargo, aunque σ^2 siempre toma valores en (0, 0.25), para este parámetro no es posible tomar cualquier valor en ese intervalo, pues si no se cumple la condición impuesta por la ecuación, α y β pueden presentar valores negativos. Esto lleva a concluir que $\sigma^2 \in (0, \mu(1-\mu))$ lo que implica que, σ^2 depende de los valores que pueda tomar μ, comportamiento representado gráficamente en la **Figura 2**, en la que el área sombreada representa el rango donde la varianza de la distribución Beta tiene sentido, bajo la curva de la parábola $\mu(1-\mu)$, y es donde la transformación en cuestión presenta la característica de ser uno a uno, y cumplir con la restricción de una densidad Beta, de que los parámetros de forma siempre deben ser positivos.

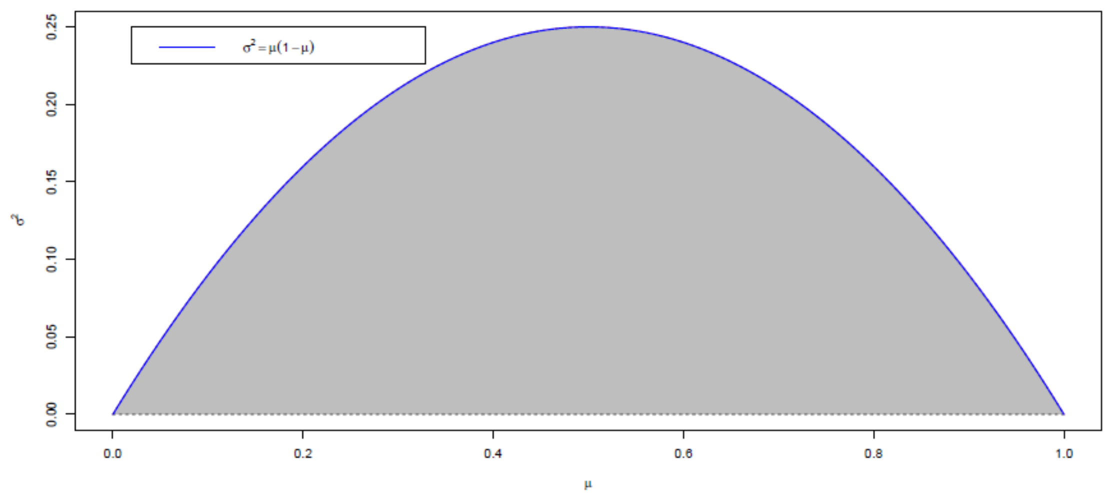

Figura 2: Gráfica del comportamiento de σ^2 frente a diferentes valores de μ

Los resultados anteriores, permiten concluir que es posible considerar una distribución a priori para σ^2 condicional en μ, de la manera $f(\sigma^2|\mu)$. También, una distribución a priori para μ, $f(\mu)$, tal que se puede expresar una distribución a priori conjunta, teniendo en cuenta la definición de probabilidad condicional, de la forma:

$$f(\mu,\sigma^2) = f(\mu)f(\sigma^2|\mu)$$

Dado que la media toma valores en el intervalo (0,1), considerar una distribución Kumaraswamy con hiperparámetros (a_1, b_1), como previa para μ es, una buena alternativa, pues su función de densidad es bastante flexible para modelar variables aleatorias continuas, restringidas en el intervalo (0, 1), y su integral es de forma cerrada.

Entonces, $\mu \sim Kumaraswamy(a_1, b_1)$, con función de densidad:

$$f(\mu) = a_1 b_1 \mu^{a_1-1}(1-\mu^{a_1})^{b_1-1};\ 0<\mu<1;\ a_1>0,\ b_1>0$$

Si se le asigna la misma importancia, o la misma probabilidad de ocurrencia a cualquier valor de σ^2 dentro de la región sombreada en la **Figura 2**. Podemos establecer como distribución previa condicional, una distribución uniforme continua, de modo que $\sigma^2|\mu \sim Uniforme\ (0, \mu(1-\mu))$, con función de densidad:

$$f(\sigma^2|\mu) = \frac{1}{\mu(1-\mu)};\ 0<\sigma^2<\mu(1-\mu)$$

Se cuenta con la estructura suficiente para establecer una a priori conjunta para la media y la varianza de la distribución Beta:

$$f(\mu,\sigma^2|a_1,b_1) = \frac{a_1 b_1}{\mu(1-\mu)}\mu^{a_1-1}(1-\mu^{a_1})^{b_1-1} \quad \textbf{(15)}$$

Este escenario plantea que el proceso de elicitación se concentrará solo en obtener los valores de los hiperparámetros (a_1, b_1), lo cual sugiere que solo se requiere tener conocimiento previo del valor esperado de la variable aleatoria de interés.

Una vez obtenida la forma de la distribución de densidad previa conjunta para μ y σ^2 es necesario encontrar la forma de la distribución de densidad previa conjunta para los parámetros α y β y posteriormente obtener sus marginales. Empleando el método de la transformación para pares de variables aleatorias, explicado en [2] y [3], se tiene que: $f(\alpha,\beta) = f(\mu,\sigma^2)|J|$ donde $J = \frac{-\alpha\beta}{(\alpha+\beta)^3(\alpha+\beta+1)^2}$

Al realizar los respectivos reemplazos, se obtiene la forma analítica de la distribución a priori conjunta $f(\alpha,\beta)$:

$$f(\alpha,\beta) = \frac{a_1 b_1}{\left(\frac{\alpha}{\alpha+\beta}\right)\left(1-\frac{\alpha}{\alpha+\beta}\right)} \cdot \left(\frac{\alpha}{\alpha+\beta}\right)^{a_1-1}\left(1-\left(\frac{\alpha}{\alpha+\beta}\right)^{a_1}\right)^{b_1-1}\left[\frac{\alpha\beta}{(\alpha+\beta)^3(\alpha+\beta+1)^2}\right] \quad \textbf{(16)}$$

Para obtener las distribuciones previas marginales, es necesario integrar la distribución a priori conjunta **(16)** con respecto a β, con el propósito de obtener la distribución a priori de α. De este modo, se tiene que $f(\alpha) = \int_0^\infty f(\alpha,\beta)d\beta$**.** Integrando por partes se tiene que:

$f(\alpha) = [uv - \int vdu]_0^\infty$ con $dv = \left(\frac{\alpha}{\alpha+\beta}\right)^{a_1-1}\left(1-\left(\frac{\alpha}{\alpha+\beta}\right)^{a_1}\right)^{b_1-1}\left(\frac{\alpha}{(\alpha+\beta)^2}\right)d\beta$ y

$u = \frac{\alpha+\beta}{\alpha(\alpha+\beta+1)^2}$

Entonces; $v = \left(1 - \left(\frac{\alpha}{\alpha+\beta}\right)^{a_1}\right)^{b_1} - 1$, y $du = -\left[\frac{\alpha+\beta-1}{\alpha(\alpha+\beta+1)^3}\right] d\beta$. Conocidas las formas analíticas de u y v, la ecuación **(16)** tiene solución al resolver la integral:

$$\int vdu = -\int \left(1 - \left(\frac{\alpha}{\alpha+\beta}\right)^{a_1}\right)^{b_1} \frac{\alpha+\beta-1}{\alpha(\alpha+\beta+1)^3} d\beta + \int \frac{\alpha+\beta-1}{\alpha(\alpha+\beta+1)^3} d\beta$$

Si se escribe, $\int vdu = \Omega - \int du = \Omega - u$ donde $\Omega = -\int \left(1 - \left(\frac{\alpha}{\alpha+\beta}\right)^{a_1}\right)^{b_1} \frac{\alpha+\beta-1}{\alpha(\alpha+\beta+1)^3} d\beta$ y sustituyendo, se tiene la distribución a priori de α:
$f(\alpha) = u(1+v)|_0^\infty - \Omega|_0^\infty$, reemplazando los valores de u, v y Ω en términos de los parámetros de la distribución Beta α y β, entonces:

$$f(\alpha|a_1, b_1) = \int_0^\infty \left(1 - \left(\frac{\alpha}{\alpha+\beta}\right)^{a_1}\right)^{b_1} \frac{\alpha+\beta-1}{\alpha(\alpha+\beta+1)^3} d\beta \qquad \textbf{(17)}$$

Ésta última integral no tiene forma cerrada, lo que motivó a desarrollar el siguiente algoritmo de aproximación empleando la regla de los trapecios, la cual explican [10].

1. Obtener los valores de los hiperparámetros (a_1, b_1), elicitando μ.
2. Definir valores para el parámetro α, de acuerdo a su espacio paramétrico. Por ejemplo, si α esta restringido a al rango comprendido por los números reales positivos $(0, \infty)$, la idea es tomar desde un valor muy cercano a cero α_0 hasta uno muy grande α_M, es decir, que se tendrán M valores $\alpha_0, \alpha_1, \alpha_2, \ldots, \alpha_M$.
3. Ahora bien, para emplear la regla de los trapecios, tener en cuenta Θ_β el espacio paramétrico de β, donde se deben definir los límites de integración. En ese contexto, determinar el número de trapecios n y el número máximo b donde se definirá el límite superior de la integral marginal y a el límite inferior de la misma integral. Esto significa, que se está integrando la marginal de α, en el espacio definido para β en (a, b).
4. Inicializar un bucle que recorrerá la posición de los valores de α previamente definidos en (α_0, α_M), el cual será recorrido por el contador j, desde $j = 1$., generando el estado α^j.
5. Inicializar un bucle que recorrerá la posición de los trapecios previamente definidos en (a, b), el cual será recorrido por el contador i, desde $i = 1$.
6. Especificar un valor para el parámetro β, el cual es igual a la posición $i-ésima$ de los trapecios, el cual será el estado β^i.
7. Evaluar la función $f(\alpha^j, \beta^i|a_1, b_1)$. Donde, el valor de α^j es igual a la posición $j-ésima$ del intervalo (α_0, α_M).
8. Si i es igual a 1 o b, multiplicar la expresión anterior por la constante $\frac{b}{2n}$ de lo contrario, multiplicarla por $\frac{b}{n}$.
9. Almacenar el resultado preliminar en la posición $i-ésima$ de un vector de densidades D (este vector se debe crear previamente, y tiene la función de almacenar las densidades encontradas en el paso anterior).
10. Aumentar el contador tal que $i = i + 1$, y repetir los pasos desde el 6 hasta el 9, b veces.
11. Sumar todas las densidades almacenadas en el vector de densidades D, y almacenarlas en la posición $j-ésima$ de un vector de áreas A(esta será la aproximación por la regla de los trapecios para el valor $j-ésimo$ de α en el intervalo (α_0, α_M)).
12. Aumentar el contador tal que $j = j + 1$, y repetir los pasos desde el 5 al 11, tantas veces como valores de α se definan en el intervalo (α_0, α_M).
13. Por último, generar un número aleatorio para σ^2 a partir de la distribución $\sigma^2 | \mu \sim Uniforme\ (0, \mu(1-\mu))$ y calcular la constante $k = \mu\left[\frac{\mu(1-\mu)}{\sigma^2} - 1\right]$ para corregir los valores para α definidos en el intervalo (α_0, α_M) de la manera $(\alpha_0 + k, \alpha_1 + k, \ldots, \alpha_M + k)$.

Finalmente, el vector de áreas obtenido es una aproximación empleando la metodología de la regla de los trapecios, para valores definidos de α dentro de su espacio paramétrico, de la distribución a priori marginal de dicho parámetro.
Otro concepto importante es la generación de números aleatorias de esta función marginal. Por ello, después de aproximar la densidad con el algoritmo anterior, es necesario graficar dicha aproximación versus valores de α en $(\alpha_0 + k, \alpha_1 + k, \ldots, \alpha_M + k)$ con el objetivo de que, a través del comportamiento gráfico, se pueda perfilar una distribución propuesta, para integrar este algoritmo con el Metropolis- Hasting, y finalmente

obtener un nuevo algoritmo que genere números aleatorios en este caso. A continuación, se expone el algoritmo de muestreo, que integra el MH (con función propuesta $h(\phi)$) y el algoritmo de aproximación.

Algoritmo de Muestreo

1. Obtener los valores de los hiperparámetros $(a_{1,}b_1)$, elicitando μ.

2. Definir un valor inicial arbitrario para el parámetro de forma α^0, el cual debe cumplir la condición de pertenecer al espacio paramétrico Θ_α.
3. Para emplear la regla de los trapecios, tener en cuenta Θ_β el espacio paramétrico de β, donde se deben definir los límites de integración. En ese contexto, determinar el número de trapecios n y el número máximo b donde se definirá el límite superior de la integral marginal, y a el límite inferior de la misma integral. Esto significa, que se está integrando la marginal de α, en el espacio definido para β en (a, b).
4. Inicializar un bucle que definirá la cantidad de iteraciones que tendrá el algoritmo Metropolis-Hasting, el cual será recorrido por el contador j, desde $j = 1$.
5. Generar un valor aleatorio de ϕ, que será candidato para el estado α^j, a partir de la distribución propuesta $h(\phi)$.
6. Inicializar un bucle que recorrerá la posición de los trapecios previamente definidos en (a, b), el cual será recorrido por el contador i, desde $i = 1$.
7. Especificar un valor para el parámetro β, el cual es igual a la posición $i - ésima$ de los trapecios, el cual será el estado β^i.
8. Evaluar las funciones $f(\alpha^{j-1}, \beta^i|a_1, b_1)$ y $f(\phi, \beta^i|a_1, b_1)$ que es la a priori marginal de .
9. Si i es igual a 1 o b, multiplicar las expresiones anteriores por la constante $\frac{b}{2n}$. De lo contrario, multiplicarlas por $\frac{b}{n}$.
10. Almacenar los resultados preliminares en la posición $i - ésima$ de dos vectores de densidades X y W, respectivamente para la expresión evaluada con α^{j-1}, y la expresión evaluada con ϕ (Estos vectores se deben crear previamente, y tienen la función de almacenar cada una de las densidades encontradas en el paso anterior).
11. Aumentar el contador tal que $i = i + 1$, y repetir los pasos desde el 7 hasta el 10, b veces.
12. Obtener las densidades en $h(\phi)$ y $h(\alpha^{j-1})$.
13. Evaluar la probabilidad de aceptación, de manera tal que esta es igual a $min\left\{1, \frac{h(\alpha^{j-1})\sum_{i=1}^{b} W_i}{h(\phi)\sum_{i=1}^{b} X_i}\right\}$. Si el movimiento es aceptado, entonces $\alpha^j = \phi$ de lo contrario $\alpha^j = \alpha^{j-1}$
14. Aumentar el contador tal que $j = j + 1$, y repetir los pasos desde el 5 al 13, tantas veces como sea el total de iteraciones que haya definido (Se recomienda un número suficientemente grande)
15. Por último, generar un número aleatorio para σ^2 a partir de la distribución $\sigma^2 |\mu \sim Uniforme\ (0, \mu(1 - \mu))$ y calcular constante $k = \mu\left[\frac{\mu(1-\mu)}{\sigma^2} - 1\right]$ para corregir los valores para α de la manera $\alpha *= \alpha + k$. Así, $\alpha *$ serán los números aleatorios para la distribución a priori de α.

Para obtener la función de densidad marginal para el parámetroβ, se integró la distribución a priori conjunta $f(\alpha, \beta)$ con respecto al parámetro de forma α, es decir se resolvió la integral $f(\beta) = \int_0^\infty f(\alpha, \beta)d\alpha$.

La solución de esta integral, exige de nuevo aplicar la integración por partes y los conceptos matemáticos usados para obtener $f(\alpha)$. Así se llegó a la siguiente expresión:

$$f(\beta) = \frac{1}{(\beta + 1)^2} - \int_0^\infty \left[1 - \left(\frac{\alpha}{\alpha + \beta}\right)^{a_1}\right]^{b_1} \left[\frac{\alpha + \beta - 1}{\beta(\alpha + \beta - 1)^3}\right] d\alpha$$

La solución a la integral obtenida para $f(\beta)$ fue obtenida usando de nuevo el algoritmo de aproximación que utiliza la regla de los trapecios. Al realizar el procedimiento de aproximación a la solución numérica de la integral, se observó que dicha solución converge a cero (valores alrededor de 1.7 e^{-7}). Se asumió esta cantidad como un residuo despreciable de modo que, la distribución marginal del parámetro β es

$$f(\beta) = \frac{1}{(\beta+1)^2} \quad \textbf{(18)}$$

Para probar que efectivamente se trata de una función de densidad, se integró en todo su espacio paramétrico, llegándose a la solución $\int_0^\infty f(\beta)d\beta = \left[-\frac{1}{\beta+1}\right]_0^\infty = 1$ Adicionalmente, la forma de la función de distribución acumulada es $F(\beta) = 1 - \frac{1}{\beta+1}$.
Para tener una aproximación de la distribución marginal de β, se empleó el método de la transformada inversa que permite obtener números aleatorios de esta distribución. Sea una variable aleatoria u, tal que $u \sim Uniforme(0, 1)$, entonces $F(\beta) = u$ y $\beta = F^{-1}(u)$. Entonces; $u = 1 - \frac{1}{\beta+1}$ y $\beta = \frac{u}{1-u}$, de esa forma, todos los números generados usando la última ecuación serán originados a partir de la distribución a priori marginal de β.

4. ESTUDIO DE SIMULACIÓN

4.1. Distribución A Priori para α

Creando un escenario de simulación en el software estadístico R, se fijaron los valores de los parámetros de forma de un conjunto de números aleatorios simulados para la distribución Beta, tal que $\alpha = 2$ y $\beta = 2$. Lo anterior, implica que $\mu = 0.5$. Adicionalmente, se simuló el proceso de elicitación, donde se supone que entrevistando un experto se estableció un valor a priori para $\mu^* = 0.47$, donde los hiperparámetros quedan determinados por $b_1 = 10$ y $a_1 = 3.36$. La densidad aproximada para la distribución a priori obtenida ($\mu \sim Kuma(a_1 = 3.36,\ b_1 = 10)$) aparece en la **Figura 3.**

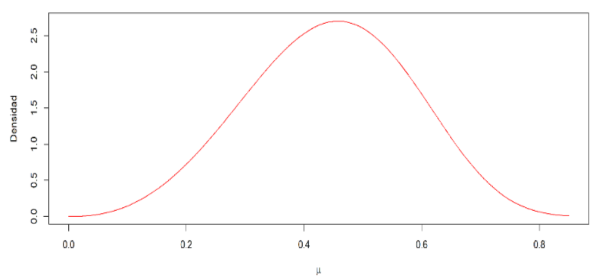

Figura 3 Curva de densidad aproximada para la distribución previa de μ
Al hacer la aproximación de la función de densidad a priori de α, usando el algoritmo propuesto, se obtuvo la forma que aparece en la **Figura 4**

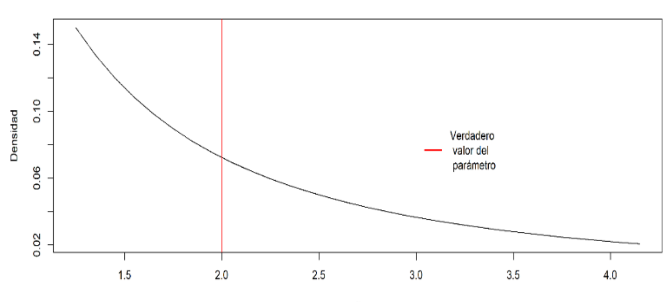

Figura 4: Curva de densidades distribución a priori para α aproximada con el algoritmo

La figura anterior, refleja que partiendo de una buena elicitación para la construcción de la a priori de μ, a través del proceso y algoritmo propuesto se obtiene una distribución a priori para α bastante informativa, ya que muestra como la densidad de esta distribución se agrupa en torno al verdadero valor del parámetro que es $\alpha = 2$.
Ahora bien, para que se pueda obtener una estimación bayesiana a priori de este parámetro, se generaron números aleatorios por medio del algoritmo de muestreo expuesto en la parte metodológica de este

documento. Allí, teniendo en cuenta el comportamiento de la curva en la figura anterior, mediante un análisis de sensibilidad emplearon una distribución propuesta independiente tal que $\phi \sim exp(\lambda = 0.5)$, con 10.000 iteraciones, obteniendo los resultados evidenciados en la **Figura 5**.

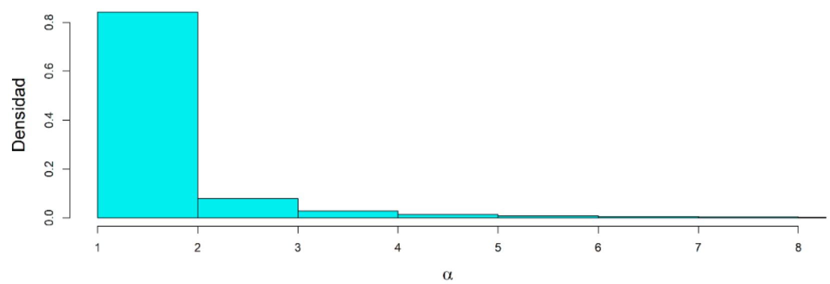

Figura 5: Densidad aproximada para α usando una combinación del algoritmo de muestreo propuesto anteriormente

De la figura anterior, se evidencia que el histograma de los números aleatorios generados, concuerda con el comportamiento visto en la curva de densidad a priori marginal de α en la **Figura 4**, indicando lo acertada que es la metodología en situaciones donde es matemáticamente inviable obtener resultados exactos. Así, con los números aleatorios obtenidos se aproximó la esperanza a priori como $E[\alpha] = 1.76$, el cual no es muy diferente al verdadero valor de $\alpha = 2$.

4.2. Distribución A Priori para β

Después de aplicar el método de la transformada inversa, en la Figura **6**, aparece la forma suavizada de la densidad obtenida a partir de los vectores de números aleatorios generados para la distribución a priori de β a partir de una muestra de 10000 valores de β comprendidos dentro del intervalo $(0, 2000)$.

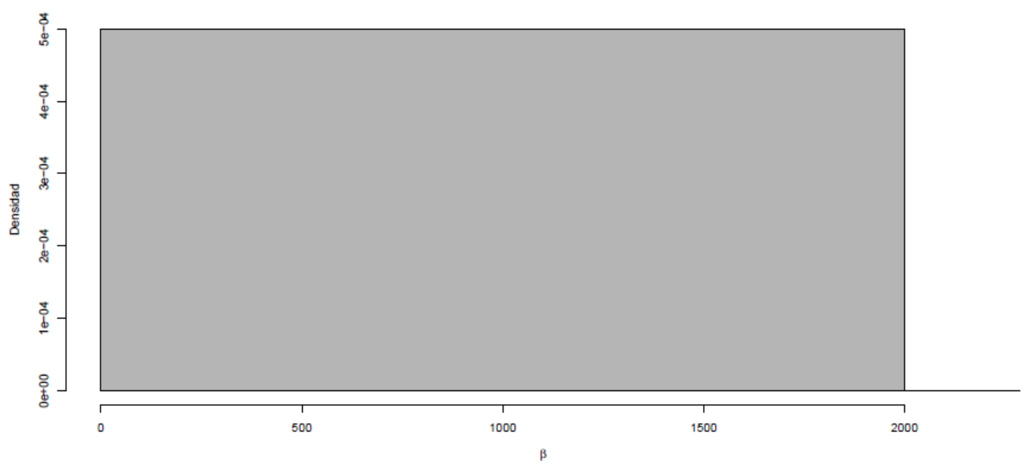

Figura 6: Densidad suavizada para la distribución a priori de β

Esta distribución a priori podría considerarse poco informativa, pues refleja como en el espacio paramétrico de β, se asigna la misma probabilidad de ocurrencia a cualquier valor de esta variable.

5. CONCLUSIONES Y RECOMENDACIONES

Obtener distribuciones previas para los parámetros de la distribución Beta es una terea que puede llegar a ser bastante difícil. Una de las mayores dificultades para lograr este objetivo es el hecho de que ambos parámetros son de forma lo que hace que no sea posible darles una interpretación práctica. En este artículo, se desarrolla y describe un procedimiento que parte de contar con información previa sobre los parámetros μ, y σ^2 los cuales son reparametrizaciones de los parámetros de forma, para obtener la forma de las distribuciones previas para $\alpha\ y\ \beta$ y los respectivos valores para los hiperparámetros.

De acuerdo con los resultados, existe dependencia entre μ y σ^2, pues el parámetro de escala siempre es menor que la curva $\mu(1-\mu)$, y nunca tomará un valor superior a 0.25, ya que $\mu \in (0,1)$. Es posible entonces, obtener una distribución a priori conjunta para los dos primeros momentos de la distribución y a partir de esta, utilizar el método de la transformación y Jacobiano para funciones matemáticas para obtener la distribución de probabilidad conjunta para los parámetros de forma α y β. La ventaja de contar con la distribución conjunta, es que la misma permite derivar las distribuciones marginales para cada uno de los parámetros. La metodología propuesta permite obtener una distribución para el parámetro β con forma analítica cerrada, de la cual se pueden tener números aleatorios sin mucho trabajo.

En el caso del parámetro α, no fue posible obtener una distribución con forma analítica cerrada, sin embargo, se logró aproximar numéricamente la densidad a través de un algoritmo que emplea la regla de los trapecios combinado con el algoritmo de Metropolis Hastings. La distribución obtenida se puede considerar una distribución previa informativa, pues parte de información indirecta obtenida para el primer momento de la distribución previa.

Se considera importante, dar continuidad en trabajos posteriores al planteamiento expuesto, para validar si efectivamente las distribuciones a priori halladas generan resultados óptimos en la estimación de α y β. En consecuencia, una primera opción es suponer uno de los parámetros conocido y así trabajar con la distribución a priori marginal correspondiente al otro y analizar los resultados para cada uno. Otra alternativa, puede ser, utilizar el método propuesto, pero partiendo de una distribución previa para μ diferente a la $Kumaraswamy$, que teóricamente permita modelar el comportamiento natural de variables cuyo rango sea el intervalo $(0,1)$. Una última estrategia podría ser, emplear las distribuciones marginales y condicionales a priori para obtener una distribución posterior conjunta, de la cual sea posible generar números aleatorios por métodos iguales o análogos a los expuestos en este trabajo.

RECEIVED: APRIL, 2020.
REVISED: MAY , 2021.

REFERENCIAS

[1] BOX, G. & TIAO, G. (1992): **Bayesian Inference in Statistical Analysis**, Wiley, EE.UU.
[2] CANAVOS, G. (1984): **Probabilidad y estadística: aplicaciones y métodos**, Limusa, México.
[3] CASELLA, G. & BERGER, R. (2002): **Statistical Inference**, Thomson Learning, EE.UU.
[4] DURÁN, E., SOTO, M. & ALDANA, M. (2014): Metrópolis – Hasting y Gibbs por bloques: utilización de algoritmos de Monte Carlo vía cadenas de Markov en el modelado de columnas estratigráficas, campo lama, lago de Maracaibo, Revista de la Facultad de Ingeniería U.C.V., Vol. 29, No 4, pp. 41-54.
[5] GAMERMAN, D. & LOPES, H. (2006): **Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference**, Second Edition, CRC Press, EE.UU.
[6] GELMAN, A., CARLIN, J., STERN, H., DUNSON, D., VEHTARI, A. & RUBIN, D. (2014): **Bayesian Data Analysis**, Third Edition, Taylor & Francis.
[7] JOHNSON, N., KOTZ, S. & BALAKRISHNAN, N. (1995): **Continuous univariate distributions**, number v.2, Wiley & Sons, New York, EE.UU.
[8] JONES, M.C. (2009): Kumaraswamy's distribution: A beta-type distribution with some tractability advantages, **Journal Statistical Methodology**, 6, 70- 81.
[9] KAKAMU, K. & NISHINO, H. (2018), `Bayesian estimation of beta-type distribution parameters based on grouped data', Computational Economics 54, 625{645.
[10] LARSON, R., HOSTETLER, R. & EDWARDS, B. (2006): **Cálculo con geometría analítica**, McGraw-Hill, México.
[11] MA, Z. & LEIJON, A. (2011): Bayesian Estimation of Beta Mixture Models with Variational Inference, **Journal IEEE Transactions on Pattern Analysis and Machine Intelligence,** 33, 2160-2173.
[12] O'HAGAN, A., GARTHWAITE, P. & KADANE, J. (2005): Statistical Methods for Eliciting Probability Distributions, **Journal of the American Statistical Association**, 100, 680-700.
[13] OWEN, C. (2008): Parameter Estimation For The Beta Distribution, (Thesis Master of Science, Brigham Young University, Utah, EE.UU), Disponible en **https://scholarsarchive.byu.edu/etd/1614. Consultado** 15-9,2018. [13]

[14] R CORE TEAM (2012): R: A Language and Environment for Statistical Computing**, R Foundation for Statistical Computing**, Vienna, Austria.
[15] RAFFO, E., MAYTA, R. & PEREZ, V. (2007): Algoritmos de integración numérica, Departamento de Ingeniería de Sistemas e Informática. **UNMSM**, 10(2).
[16] TOVAR, J. (2012): Eliciting Beta prior distributions for binomial sampling**. Revista Brasileira de Biometria,** 30, 159-172.
[17] TOVAR, J. (2015): Inferencia Bayesiana e Investigación en salud: un caso de aplicación en diagnóstico clínico, **Revista Médica de Risaralda**, 21, 9-16.