PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ

ESCUELA DE GRADUADOS

# Modelo de censura intervalar para datos positivos

TESIS PARA OPTAR POR EL GRADO DE MAGISTER EN ESTADÍSTICA

**Presentado por:**

**Justo Andrés Manrique Urbina**

**Asesor: Cristian Luis Bayes Rodríguez**

**Miembros del jurado:**
**Dr. Nombre completo jurado 1**
**Dr. Nombre completo jurado 2**
**Dr. Nombre completo jurado 3**

Lima, Diciembre 2020

# Dedicatoria

Dedicatoria...

# Agradecimentos

Agradecimentos ...

# Resumen

Existen preguntas específicas que, dada las naturaleza de las mismas, los encuestadores temen que los encuestados respondan inadecuadamente. Esto se debe a que dichas preguntas podrían considerarse intrusivas e inclusive vergonzosas. Ello motiva a los encuestados a rehusarse de responderlas, u ofrecer - deliberadamente - respuestas incorrectas. No obstante, esta clase de preguntas crea un dilema para los encuestados: ellos acordaron a los investigadores a proveer información, pero no estarían de acuerdo en brindar completamente la información requerida (2009: 240-241). El encuestador, consciente de esta realidad, busca salvaguardas y equilibrio entre la privacidad del encuestado y la obtención del dato.

Una vez obtenidos los datos, al momento de efectuar el análisis y modelamiento de los mismos, el modelo se debe acoplar a la nueva estructura de los mismos.
**Palabras clave:** censura intervalar, regresión con censura.

# Abstract

Abstract ...

**Keywords:** keyword1, keyword2, keyword3.

# Índice general

# Lista de Abreviaturas

| | |
|---|---|
| fdp | Función de densidad de probabilidad . |
| pBF | Pseudo factor de Bayes(*Pseudo bayes factor*). |

# Lista de Símbolos

$\mu$ Media.

# Índice de figuras

# Índice de cuadros

# Capítulo 1

# Introducción

## 1.1. Consideraciones Preliminares

En muchas situaciones prácticas deseamos investigar como ciertas variables influyen en una variable continua que asume valores en el intervalo $(0, 1)$, tales como, porcentajes, proporciones, tasas, etc. Por ejemplo, la tasa de desnutrición de una cierta región puede ser influenciada por el PBI, o la fracción del gasto de un hogar destinado a alimentos puede ser influenciada por variables como el tamaño de la familia, el ingreso total de la familia, etc. En estos casos, los modelos de regresión pueden no ser apropiados para modelar este tipo de datos porque la variable respuesta sólo toma valores en un rango limitado y los valores estimados pueden caer fuera del rango. Una posible solución es transformar este tipo de datos para que asuman valores en toda la recta y modelarlos mediante una regresión. Este enfoque puede presentar varios problemas, uno de ellos es que los parámetros del modelo no sean fácilmente interpretados en términos de los datos originales. Otro problema es que generalmente las proporciones presentan asimetría, por lo tanto la suposición de normalidad no seria adecuada.[1]

## 1.2. Objetivos

El objetivo general de la tesis es estudiar propiedades, estimar y aplicar a conjuntos de datos reales el modelo de regresión beta desde el punto de vista de la estadística bayesiana. De manera específica:

- Revisar la literatura acerca de los diferentes propuestas de modelos de regresión para proporciones.
- Proponer estudiar, propiedades, e implementar la estimación del modelo de regresión beta desde la perspectiva bayesiana.
- Realizar estudios de simulación de computación intensiva aprovechando el uso de un GRID computacional (proyecto LEGION).
- Aplicar el modelo a conjunto de datos reales.

## 1.3. Organización del Trabajo

En el Capítulo 2, presentamos conceptos previos al desarrollo de ...

[1] Pie de pagina de ejemplo.

Finalmente, en el Capítulo 6 discutimos algunas conclusiones obtenidas en este trabajo. Analizamos la ventajas y desventajas de los métodos propuestos,...

En el anexo presentamos algunas pruebas de resultados en más detalles (Apéndice A) y también los programas utilizadas en las aplicaciones a conjuntos de datos reales.

# Capítulo 2

# Distribución Beta

Introducción al capítulo

## 2.1. Distribución Beta

Una clase de distribuciones que permite modelar variables continuas limitadas al intervalo $(0, 1)$ es la distribución beta.

### 2.1.1. Función de densidad de probabilidad

la función de densidad de probabilidad de una variable aleatória $Y$ que sigue una distribución beta es dada por

$$f_Y(y \mid \alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} y^{\alpha-1}(1-y)^{\beta-1}, \quad 0 < y < 1. \tag{2.1}$$

Esta distribución es bastante flexible para modelar este tipo de datos debido a que su función de densidad de probabilidad puede tomar diversas formas dependiendo del valor de los parámetros que caracterizan a esta distribución (ver figura 2.1).

### 2.1.2. Propiedades

La media y la variancia de una distribución son expresadas por

$$E(y) = \frac{\alpha}{\alpha+\beta} \quad \text{y} \quad Var(y) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}. \tag{2.2}$$

## 2.2. Parametrización alternativa

Por este motivo, recientemente en la literatura han sido propuestos modelos de regresión basados en esta distribución, por ejemplo ver Ferrari y Cribari-Neto (2004). El modelo propuesto por estos autores permite la interpretación de los parámetros en términos de la respuesta en su escala original (variable respuesta sin transformar). Ellos consideran una reparametrización del modelo donde $\mu = \alpha/(\alpha+\beta)$ y $\phi = \alpha + \beta$, y de (2.2) obtenemos que

$$E(y) = \mu \quad \text{y} \quad Var(y) = \frac{V(\mu)}{1+\phi} \tag{2.3}$$

donde $V(\mu) = \mu(1-\mu)$, así tenemos que $\mu$ es la media de la variable respuesta y $\phi$ puede ser interpretado como un parámetro de precisión. Luego, en la nueva parametrización la densidad de la distribución beta puede ser escrita como

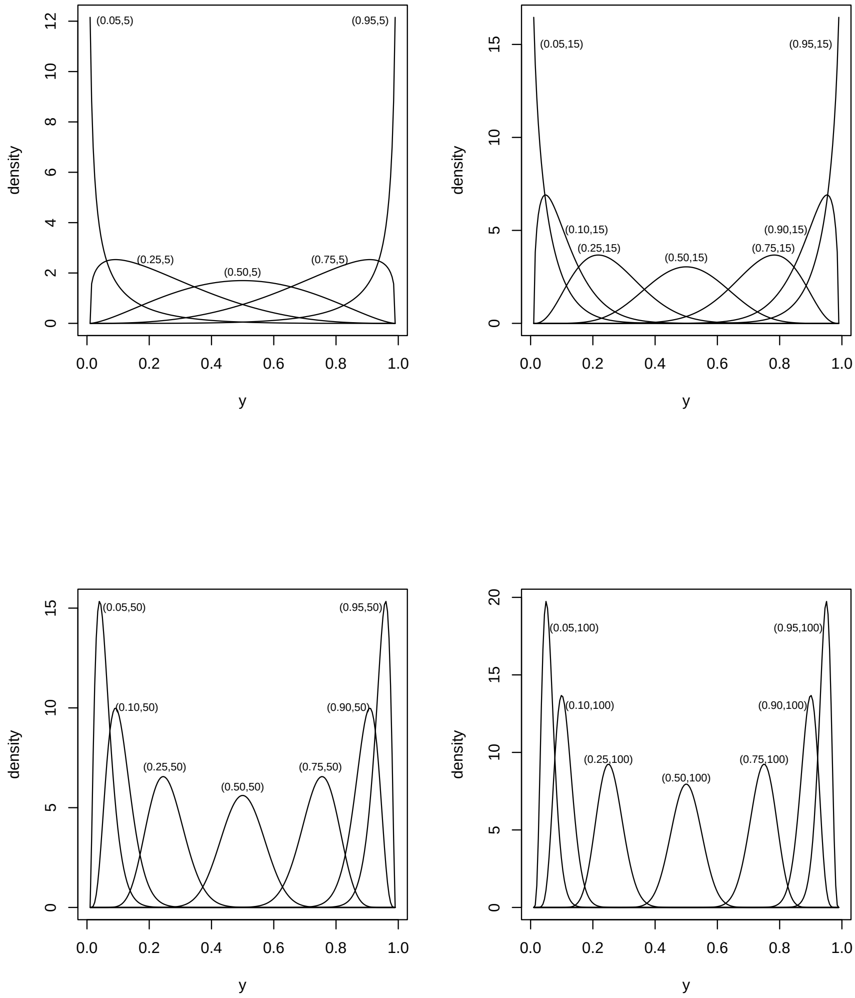


Figura 2.1: Densidades de la distribución beta para diferentes valores de los parámetros

$$f_Y(y \mid \mu, \phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} y^{\mu\phi - 1}(1-y)^{(1-\mu)\phi - 1}, \quad 0 < y < 1. \tag{2.4}$$

donde $0 < \mu < 1$ y $\phi > 0$.

# Capítulo 3

# Modelo de Regresión Beta

El modelo de regresión para $Y_1, ..., Y_n$ variables aleatorias que siguen una distribución beta dada en (2.4) com parámetros $\mu_i$ y $\phi$ será obtenido cuando consideremos

$$g(\mu_i) = \sum_{i=1}^{k} x_{ij}\beta_j \tag{3.1}$$

donde $\beta = (\beta_1, ..., \beta_k)$ es el vector de parámetros de regresión, $x_{i1}, ..., x_{ik}$ son valores de $k$ covariables y $g(.)$ es una función de enlace estrictamente monótona y dos veces diferenciable que va de $(0, 1)$ a $\Re$.

Desde un punto de vista clásico, Ferrari y Cribari-Neto (2004) consideran la estimación de los parámetros del modelo por el método de máxima verosimilitud. De este modo la inferencia en el modelo está basada en propiedades asintóticas de estos estimadores considerando una muestra suficientemente grande.

Sin embargo, en muchas situaciones prácticas el tamaño de muestra puede ser pequeño, o no siéndolo suele existir alguna información preliminar acerca de los parámetros de interés. Estas situaciones son perfectamente abordados desde la perspectiva bayesiana (ver Congdon, 2003) Para la regresión Beta este enfoque fue considerado por Branscum et al. (2007). Por este motivo, en la presente tesis se desarrollará el análisis bayesiano del modelo de regresión beta. Entre los tópicos a ser estudiados adicionalmente consideramos: la estimación del modelo utilizando MCMC, selección de modelos utilizando diferente tipos de criterios y aplicaciones a conjuntos de datos reales.

Capítulo 4

# Estudio de Simulación

## Capítulo 5

# Aplicaciones

# Capítulo 6

# Conclusiones

## 6.1. Conclusiones

Conclusiones

## 6.2. Sugerencias para investigaciones futuras

- Item 1.
- Item 2.
- Item 3.
- ...

**Apéndice A**

# Resultados teóricos

# Bibliografía


Branscum, A. J., Johnson, W. O. y Thurmond, M. C. (2007). Bayesian beta regression; application to househol data and genetic distance between foot-and-mouth disease viruses, *Australian & New Zealand Journal of Statistics* **49**(3): 287–301.

Congdon, P. (2003). *Applied-Bayesian-Modelling* , John Wiley & Sons.

Ferrari, S. y Cribari-Neto, F. (2004). Beta regression for modelling rates and proportions, *Journal of Applied Statistics* **31**: 799–815.