PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ

ESCUELA DE POSGRADO

Modelos de regresión robusta para datos de conteo

Tesis para obtener el grado académico de Maestro en Estadística que presenta:

Christoffer Augusto Villar Naccha

Asesor:

Cristian Luis Bayes Rodriguez

Lima, 2025

Declaración jurada de autenticidad

Yo, Cristian Luis Bayes Rodriguez, docente de la Escuela de Posgrado de la Pontificia Universidad Católica del Perú, asesor de la tesis titulada *Modelos de regresión robusta para datos de conteo*, del autor Christoffer Augusto Villar Naccha, dejo constancia de lo siguiente:
- El mencionado documento tiene un índice de puntuación de similitud de 18 %. Así lo consigna el reporte de similitud emitido por el software Turnitin el 09/06/2025.
- He revisado con detalle dicho reporte y confirmo que cada una de las coincidencias detectadas no constituyen plagio alguno.
- Las citas a otros autores y sus respectivas referencias cumplen con las pautas académicas.

Lima, 09 de junio de 2025

<table>
<tr><td colspan="2">Apellidos y nombres del asesor:
Cristian Luis Bayes Rodriguez</td></tr>
<tr><td>DNI: 40372640</td><td rowspan="2">Firma:  </td></tr>
<tr><td>ORCID: https://orcid.org/0000-0003-0474-7921</td></tr>
</table>

Resumen

En esta tesis se propone un nuevo modelo, denominado Regresión Binomial Negativa con Mixtura en la Dispersión (NB-H), como una alternativa robusta para el análisis de datos de conteo caracterizados por sobredispersión y presencia de valores atípicos. La propuesta se basa en la introducción de una estructura de mixtura en el parámetro de dispersión de la distribución Binomial Negativa, lo que permite que el modelo sea menos sensible a observaciones extremas, preservando así la estructura general de los datos. Se presentan dos formulaciones específicas, denominadas NB-G y NB-IG, que emplean distribuciones Gamma e Inversa Gamma, respectivamente, como componentes de mezcla.

Se adopta un enfoque bayesiano para la estimación de los parámetros, utilizándose simulaciones de cadenas de Markov Monte Carlo (MCMC) implementadas en el lenguaje Stan. Se realiza un estudio de simulación para evaluar la robustez del modelo frente a diferentes escenarios de contaminación, así como dos aplicaciones prácticas con datos reales provenientes del ámbito de salud. Los resultados muestran que las variantes propuestas presentan mejor desempeño respecto al modelo de Regresión Binomial Negativa tradicional en términos de estabilidad y precisión, especialmente en presencia de observaciones atípicas. Esta investigación aporta una estrategia robusta y flexible para el modelado de datos de conteo, capaz de adaptarse a contextos con alta variabilidad y presencia de valores extremos.

Palabras-clave: Regresión Binomial Negativa, Modelos con Mixtura, Inferencia Bayesiana, Datos de Conteo, Valores Atípicos.

Abstract

This thesis proposes a new model, referred to as Negative Binomial Regression with Mixture in the Dispersion (NB-H), as a robust alternative for analyzing count data affected by overdispersion and the presence of outliers. The proposed approach incorporates a mixture structure in the dispersion parameter of the Negative Binomial distribution, making the model less sensitive to extreme observations, thereby preserving the overall data structure. Two specific formulations are developed: NB-G and NB-IG, which use Gamma and Inverse Gamma distributions, respectively, as mixing components.

A Bayesian approach is adopted for parameter estimation, using Markov Chain Monte Carlo (MCMC) simulations implemented in the Stan programming language. A simulation study is performed to evaluate the robustness of the model under various contamination scenarios, alongside two real-world applications involving health-related count data. The results show that the proposed variants outperform the traditional Negative Binomial model in terms of stability and estimation accuracy, particularly in the presence of extreme values. This research introduces a flexible and resilient strategy for modeling count data under challenging conditions of high variability and outliers.

Keywords: Negative Binomial Regression, Mixture Models, Bayesian Inference, Count Data, Outliers.

Índice general

Lista de Tablas

Lista de Figuras

Capítulo 1

Introducción

1.1. Consideraciones preliminares

El estudio de los datos de conteo es importante para comprender una gran variedad de fenómenos en diversas disciplinas y aplicaciones prácticas. Estos datos se obtienen al cuantificar la frecuencia de que ocurra un evento y se pueden modelar mediante el uso de variables de conteo, las cuales solamente pueden tomar valores enteros no negativos. Según Sellers y Shmueli (2010), los modelos de regresión han demostrado ser herramientas cruciales para describir la relación entre una variable dependiente y una o varias variables explicativas.

Entre los modelos de regresión más utilizados para modelar datos de conteo se encuentran los que utilizan a la distribución de Poisson y a la distribución Binomial Negativa. La distribución de Poisson, sin embargo, se basa en la suposición de equidispersión, lo que implica que la media y la varianza son iguales. Esta condición no siempre es aplicable en escenarios reales, lo cual puede ocasionar problemas de sobredispersión. Además, el incumplimiento de la equidispersión basta para invalidar el supuesto de la distribución Poisson (Germán-Soto et al., 2009).

Para abordar el problema de la sobredispersión, una opción adecuada es utilizar la distribución Binomial Negativa, la cual surge como una alternativa más flexible y robusta. Esta distribución resulta particularmente adecuada para modelar datos de conteo cuando la varianza supera a la media, ya que incorpora un parámetro adicional que permite ajustar la varianza de manera independiente (Aeberhard et al., 2014). Debido a esta flexibilidad, la distribución Binomial Negativa se ha consolidado como una opción preferente para modelar datos de conteo con sobredispersión, proporcionando una solución eficaz para capturar las características reales de los datos.

En situaciones reales, los conjuntos de datos suelen contener valores atípicos, lo cual puede limitar la efectividad de la distribución Binomial Negativa. Para afrontar este problema, en el trabajo de Bayes et al. (2025) se propuso un modelo de regresión basado en la distribución Binomial Negativa Gamma. Esta distribución introduce un nuevo parámetro que permite

generar colas más pesadas en comparación con la distribución Binomial Negativa, haciendo al modelo más robusto frente a valores atípicos. Siguiendo una dirección similar a la de dicho trabajo, en esta tesis se busca desarrollar un modelo de regresión con una estructura diferente, que también sea capaz de manejar la sobredispersión y los valores atípicos de manera eficaz.

En este estudio, se introducirá la distribución Binomial Negativa con Mixtura en la Dispersión (en adelante, NB-H), diseñada para abordar efectivamente los problemas de sobredispersión y la presencia de datos atípicos en datos de conteo. Esta formulación se basa en incorporar una mixtura en el parámetro de dispersión dentro del marco de la distribución Binomial Negativa. A partir de esta distribución, se plantea un modelo de regresión que permite establecer la relación entre una variable respuesta y un conjunto de variables explicativas. En el presente trabajo se examinan las principales propiedades del modelo propuesto, sus métodos de estimación, así como un estudio de sensibilidad, y se evalúa su desempeño a través de dos aplicaciones prácticas basadas en datos reales, comparándolo frente a otros modelos de regresión para datos de conteo comúnmente utilizados en la literatura.

Dado que los datos de conteo suelen presentar sobredispersión, valores atípicos y estructuras complejas, esta tesis adopta un enfoque bayesiano para la estimación del modelo propuesto. A diferencia de los métodos clásicos, la inferencia bayesiana permite incorporar información previa mediante distribuciones a priori y evita calcular numéricamente la función de probabilidad completa, lo cual representa una ventaja computacional importante. Estas propiedades hacen que este enfoque sea especialmente adecuado para ajustar modelos flexibles y robustos como el que se analiza en este trabajo.

1.2. Objetivos

El objetivo general de esta investigación es estudiar en profundidad el modelo de regresión NB-H para datos de conteo. Para ello, se desarrollará un enfoque de estimación bayesiana utilizando simulaciones de cadenas de Markov Monte Carlo (MCMC), y se analizará su desempeño tanto en escenarios simulados como en dos aplicaciones con datos reales. Este análisis permitirá evidenciar las ventajas de esta formulación frente a modelos tradicionales en situaciones de sobredispersión y presencia de valores atípicos.

Los objetivos específicos son los siguientes:

- Realizar una revisión crítica de la literatura sobre modelos de regresión aplicados a datos de conteo.
- Presentar formalmente la estructura del modelo de regresión NB-H, resaltando sus

fundamentos conceptuales y propiedades teóricas.

- Proponer un esquema de estimación bayesiana para el modelo NB-H e implementarlo mediante simulaciones de cadenas de Markov Monte Carlo (MCMC), utilizando la librería PyStan en el lenguaje de programación Python.

- Evaluar el desempeño del modelo NB-H mediante un estudio de sensibilidad basado en simulaciones, considerando distintos niveles de severidad en la presencia de valores atípicos.

- Aplicar el modelo NB-H a dos bases de datos reales para evaluar su desempeño, y comparar sus resultados con los obtenidos mediante la regresión Binomial Negativa tradicional, utilizando criterios bayesianos de información como el DIC y el WAIC.

1.3. Organización del trabajo

La presente tesis se estructura en siete capítulos. El Capítulo 2 introduce los conceptos preliminares necesarios para la comprensión de los modelos desarrollados. Se abordan las distribuciones Gamma, Inversa Gamma, Poisson y Binomial Negativa, destacando sus propiedades, parametrizaciones y aplicaciones en el modelado de datos de conteo.

Posteriormente, en el Capítulo 3, se presenta la Distribución Binomial Negativa con Mixtura en la Dispersión (NB-H), la cual permite abordar tanto la sobredispersión como los valores atípicos. Se estudian sus propiedades teóricas y se presentan dos casos particulares: NB-G, que emplea una distribución Gamma como mixtura, y NB-IG, basada en la Inversa Gamma.

El Capítulo 4 se enfoca en la definición del modelo de regresión basado en la distribución NB-H. En esta sección se describe la formulación matemática del modelo, la relación entre la variable de respuesta y las covariables, así como los métodos de estimación bayesiana utilizados, incluyendo la especificación de distribuciones a priori y la implementación del muestreo MCMC.

En el Capítulo 5 se desarrolla un estudio de simulación orientado a evaluar el comportamiento del modelo NB-H en comparación con alternativas tradicionales. Se consideran distintos escenarios con presencia de datos atípicos, lo que permite analizar la robustez de las estimaciones bajo condiciones de perturbación controlada.

En el Capítulo 6 se desarrollan dos aplicaciones prácticas del modelo NB-H sobre bases de datos reales. En ambas aplicaciones se ajusta el modelo propuesto y se evalúa su desempeño en

comparación con el modelo de Regresión Binomial Negativa tradicional, empleando criterios bayesianos de información como el DIC y el WAIC.

Finalmente, el Capítulo 7 expone las conclusiones más relevantes del estudio y propone recomendaciones para futuras líneas de investigación.

Capítulo 2

Conceptos preliminares

En este capítulo se introducirán los conceptos esenciales necesarios que se necesitan para comprender mejor los modelos propuestos en los capítulos subsiguientes. Se describirán diversas distribuciones de probabilidad, junto con sus propiedades, parametrizaciones alternativas y aplicaciones en modelos de regresión.

Primero, se introducirán las distribuciones Gamma e Inversa Gamma, junto con sus funciones de densidad y propiedades, como la media y la varianza. La Distribución Binomial Negativa con Mixtura en la Dispersión se construye haciendo uso de estas distribuciones, aplicándolas de forma individual. Dichas distribuciones continuas resultan fundamentales para modelar la dispersión y asegurar la robustez ante la presencia de valores atípicos.

A continuación, se abordarán las distribuciones Poisson y Binomial Negativa, junto con sus respectivas funciones de masa de probabilidad. Estas distribuciones discretas son ampliamente utilizadas en la literatura para modelar datos de conteo. Luego, se discutirán algunas de sus propiedades como la media y la varianza. Posteriormente, se introducirá una parametrización alternativa para la distribución Binomial Negativa, y se procederá a determinar nuevamente su media y varianza bajo esta parametrización.

Finalmente, se introduce el concepto de modelos de regresión para datos de conteo. Estos modelos son esenciales en situaciones donde la variable de respuesta corresponde a la frecuencia de ocurrencia de un evento, como el número de visitas a un sitio web o el conteo de incidentes en un periodo. Se detallan los modelos de regresión Poisson y Binomial Negativa, dos de las distribuciones más utilizadas para este tipo de datos. Se discutirá cómo la varianza de los datos influye en la selección del modelo adecuado y la necesidad de ajustar la estructura de dispersión en situaciones con alta variabilidad.

2.1. Distribución Gamma

La distribución Gamma es una distribución de probabilidad continua que posee dos parámetros α y β. Esta distribución permite modelar variables aleatorias cuyo rango es el conjunto de los números reales positivos $\mathbb{R}^+$. La función de densidad de probabilidad de una variable $W \sim \text{Gamma}(\alpha, \beta)$ está dada por:

$$g(\omega \mid \alpha, \beta) = \frac{\beta^{\alpha}\omega^{\alpha-1}e^{-\beta\omega}}{\Gamma(\alpha)}, \quad \omega > 0 \ ,$$

donde la función Gamma es representada por $\Gamma(\cdot)$, α es el parámetro de forma y β es el parámetro de tasa. Estos parámetros pueden tomar los valores $\alpha > 0$ y $\beta > 0$. La media y la varianza de esta distribución están dadas por:

$$\mathbb{E}(W) = \frac{\alpha}{\beta} \quad \text{y} \quad \text{Var}(W) = \frac{\alpha}{\beta^2} \ .$$

2.2. Distribución Inversa Gamma

La distribución Inversa Gamma es una distribución de probabilidad continua caracterizada por dos parámetros α y β. Esta distribución puede utilizarse para modelar variables aleatorias que tienen como rango el conjunto de los números reales positivos $\mathbb{R}^+$. La función de densidad de probabilidad de una variable $W \sim \text{IG}(\alpha, \beta)$ se muestra a continuación.

$$ig(\omega \mid \alpha, \beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} \ \omega^{-\alpha-1} \ e^{-\frac{\beta}{\omega}}, \quad \omega > 0 \ ,$$

donde α es el parámetro de forma y β es el parámetro de escala. Estos parámetros deben cumplir que $\alpha > 0$ y $\beta > 0$. La media y la varianza de la distribución Inversa Gamma son respectivamente:

$$\mathbb{E}(W) = \frac{\beta}{\alpha - 1} \quad \text{para } \alpha > 1, \quad \text{y} \quad \text{Var}(W) = \frac{\beta^2}{(\alpha-1)^2(\alpha-2)} \quad \text{para } \alpha > 2 \ .$$

La distribución Inversa Gamma se puede derivar como una transformación de la distribución Gamma. Específicamente, si una variable aleatoria V sigue una distribución Gamma con parámetros α y β, es decir, $V \sim \text{Gamma}(\alpha, \beta)$, entonces la variable $W = \frac{1}{V}$ sigue una distribución Inversa Gamma con los mismos parámetros y es denotada como $W \sim \text{IG}(\alpha, \beta)$.

2.3. Distribución Poisson

La distribución Poisson es una distribución de probabilidad discreta que posee un solo parámetro μ. Una variable aleatoria Y que sigue la distribución Poisson, se representa como $Y \sim P(\mu)$ y su función de masa de probabilidad está dada por:

$$p(y \mid \mu) = \frac{\mu^y e^{-\mu}}{y!}, \quad y = 0, 1, 2, \ldots$$

donde el parámetro puede tomar los valores $\mu > 0$. La media y varianza de esta distribución coinciden y están dadas por:

$$\mathbb{E}(Y) = \mu \qquad \text{y} \quad \text{Var}(Y) = \mu \ .$$

Dado que Y solo puede tomar valores enteros no negativos, la distribución Poisson resulta apropiada para modelar datos de conteo. Además, como señala Berk y MacDonald (2008), el modelo de regresión que utiliza la distribución Poisson es sencillo de interpretar, lo que explica su uso frecuente en aplicaciones prácticas.

2.4. Distribución Binomial Negativa

La distribución Binomial Negativa es una distribución de probabilidad discreta que tiene dos parámetros r y p. A lo largo de este trabajo, nos referiremos a ella indistintamente como distribución Binomial Negativa o distribución Binomial Negativa estándar, con el fin de distinguirla de modelos extendidos que se introducirán más adelante. Una variable aleatoria que presenta esta distribución se representa como $Y_r \sim \text{NB}(r, p)$ y su función de masa de probabilidad se define de la siguiente manera:

$$nb(y \mid r, p) = \binom{y-1}{r-1}(1-p)^{y-r}p^r, \quad y = r, r+1, r+2, \ldots$$

donde los parámetros pueden tomar los valores $r \in \mathbb{Z}^+$ y $p \in (0,1)$. La media y varianza de esta distribución están dadas por:

$$\mathbb{E}(Y_r) = \frac{r}{p} \quad \text{y} \quad \text{Var}(Y_r) = \frac{r(1-p)}{p^2} \ .$$

Johnson et al. (2005) afirma que la distribución Binomial Negativa es un modelo paramétrico que es útil frente a datos con sobredispersión. Por ello, esta distribución es ampliamente utilizada en el estudio de los datos de conteo.

Sin embargo, para el análisis de datos de conteo, se necesita que la variable pueda tomar valores en $\mathbb{Z}_{\geq 0}$. Por ello, se utiliza la distribución Binomial Negativa desplazada, en la cual se sustituye la variable aleatoria Y_r por la variable aleatoria $Y = Y_r - r$, quedando la siguiente función de masa de probabilidad:

$$nb(y \mid r, p) = \binom{y+r-1}{r-1}(1-p)^y p^r, \quad y = 0, 1, 2, \ldots$$

con dos parámetros $r \in \mathbb{Z}^+$ y $p \in (0,1)$. La media y varianza de la distribución Binomial Negativa desplazada están dadas por:

$$\mathbb{E}(Y) = \frac{r(1-p)}{p} \quad \text{y} \quad \text{Var}(Y) = \frac{r(1-p)}{p^2} \ .$$

Los trabajos de Anscombe (1950) y Johnson et al. (2005) emplearon una parametrización alternativa de la distribución Binomial Negativa desplazada, utilizando los parámetros que se describen a continuación:

$$\mu = \frac{r(1-p)}{p} \qquad \phi = \frac{1}{r}$$

Además, su relación inversa está dada por:

$$r = \frac{1}{\phi} \qquad p = \frac{1}{1+\phi\ \mu}$$

donde μ representa la media y ϕ es llamado parámetro de dispersión. Así, al reemplazar estos parámetros en la varianza obtenemos:

$$\text{Var}(Y) = \mu + \phi\ \mu^2.$$

En la parametrización basada en la media μ y la dispersión ϕ, el valor $1/\phi$ puede no ser un número entero. En estos casos, la forma clásica de la función de masa de probabilidad, que se basa en expresiones combinatorias, deja de ser válida. Por ello, se emplea una formulación general utilizando la función Gamma, que permite definir la distribución para cualquier $\phi > 0$ y facilita su aplicación en modelos de regresión. Bajo este enfoque, la función de masa de probabilidad de $Y \sim \text{NB}(\mu, \phi)$ se expresa como:

$$nb(y \mid \mu, \phi) = \frac{\Gamma(y + \frac{1}{\phi})}{\Gamma(\frac{1}{\phi})\,\Gamma(y+1)} \left(\frac{\phi\ \mu}{1+\phi\ \mu}\right)^{y} \left(\frac{1}{1+\phi\ \mu}\right)^{\frac{1}{\phi}}, \quad y = 0, 1, 2, \dots\ .$$

También se pueden derivar expresiones para la asimetría y la curtosis de la distribución Binomial Negativa. Estas medidas se expresan en función de los parámetros μ y ϕ, y se emplearán como referencia en los capítulos posteriores:

$$\text{Skew}(Y_{NB}) = \frac{1 + 2\,\phi\,\mu}{\mu^{1/2}\,(1+\phi\,\mu)^{1/2}}\,, \tag{1}$$

$$\text{Kurt}(Y_{NB}) = \frac{1 + 6\,\phi\,\mu + 6\,\phi^2\,\mu^2}{\mu\,(1+\phi\,\mu)}\,. \tag{2}$$

2.5. Modelos de Regresión para Datos de Conteo

La regresión para datos de conteo es una herramienta muy importante en el análisis estadístico, donde la variable de interés puede representar la cantidad de veces que ocurre un evento en un periodo de tiempo o espacio determinado. En este contexto, el objetivo es describir el comportamiento de la variable de respuesta a partir de las covariables que podrían estar relacionadas con ella. Algunos ejemplos de datos que pueden ser utilizada en este tipo de regresión son las quejas de los clientes, admisiones hospitalarias, casos de enfermedades, número de accidentes de tráfico, entre otros. Este tipo de variable solo permite tomar valores enteros no negativos, por lo que necesita modelos específicos para su análisis. Algunos ejemplos de este tipo de modelos son el modelo de Regresión Poisson y el modelo de Regresión Binomial Negativa.

2.5.1. Modelo de Regresión Poisson

El modelo de Regresión Poisson es ampliamente utilizado en la literatura para datos de conteo. Sea Y_i la variable de respuesta y $X_i \in \mathbb{R}^p$ el vector de covariables para la observación i, con $i = 1, 2, \ldots, n$. Se asume que las respuestas Y_i son condicionalmente independientes dado X_i, y que cada una sigue una distribución Poisson con media μ_i:

$$Y_i \sim P(\mu_i)$$

La relación entre μ_i y las covariables se especifica mediante una función de enlace adecuada. Esta función de enlace permite transformar el dominio de las covariables al dominio de la media del modelo. McGree y Eccleston (2012) utilizan la función de enlace logarítmica para el modelo de regresión Poisson, la cual es la más utilizada en la literatura para este modelo:

$$\log(\mu_i) = X_i^\top \boldsymbol{\beta},$$

donde $\boldsymbol{\beta} \in \mathbb{R}^p$ es el vector de coeficientes. La función de enlace logarítmica garantiza que los valores de μ_i sean positivos. Este modelo de regresión asume que la media y la varianza de Y tienen el mismo valor, lo cual no es adecuado en situaciones reales donde los datos presentan sobredispersión.

2.5.2. Modelo de Regresión Binomial Negativa

El modelo de Regresión Binomial Negativa también es popular en la literatura. Sea Y_i la variable de respuesta y $X_i \in \mathbb{R}^p$ el vector de covariables correspondiente a la observación i, con $i = 1, 2, \ldots, n$. Se asume que las respuestas Y_i son condicionalmente independientes dado X_i, y que cada una sigue una distribución Binomial Negativa con media μ_i y parámetro de dispersión ϕ:

$$Y_i \sim \mathrm{NB}(\mu_i, \phi).$$

Este modelo es especialmente útil cuando los datos exhiben mayor variabilidad que la permitida por el modelo Poisson. En particular, la varianza de Y_i está dada por:

$$\mathrm{Var}(Y_i) = \mu_i + \phi\, \mu_i^2,$$

lo cual permite capturar diferentes niveles de dispersión mediante el parámetro $\phi > 0$, inde-

pendientemente de la media.

Al igual que en el modelo Poisson, se utiliza una función de enlace para relacionar la media μ_i con las covariables. La elección más común es la función logarítmica:

$$\log(\mu_i) = X_i^\top \boldsymbol{\beta},$$

donde $\boldsymbol{\beta} \in \mathbb{R}^p$ representa el vector de coeficientes del modelo. Esta formulación garantiza que los valores de μ_i sean positivos y ofrece una interpretación clara de los coeficientes, que describen los efectos de las covariables.

En este capítulo se han expuesto los modelos y distribuciones fundamentales, resaltando sus principales propiedades y su importancia para modelar fenómenos específicos, como la sobredispersión. Este marco teórico proporciona los fundamentos necesarios para comprender y desarrollar los conceptos que serán explorados en los capítulos siguientes.

Capítulo 3

Distribución Binomial Negativa con Mixtura en la Dispersión

3.1. Introducción

En este capítulo, se presenta la distribución Binomial Negativa con Mixtura en la Dispersión (NB-H), una extensión de la distribución Binomial Negativa que busca abordar los problemas de sobredispersión y valores atípicos presentes en datos de conteo. Esta extensión se logra mediante la inclusión de una variable aleatoria que modula el parámetro de dispersión, permitiendo un ajuste más flexible y preciso.

El capítulo se estructurará detallando la definición formal de la distribución Binomial Negativa con Mixtura en la Dispersión y sus propiedades, lo cual proporcionará una visión más profunda de su comportamiento. Además, se explorarán algunos casos particulares que utilizan distribuciones como la Gamma y la Inversa Gamma para modular la dispersión. Estas combinaciones ofrecen una mayor robustez al modelo frente a la presencia de datos atípicos, haciendo que esta distribución sea una alternativa sólida en situaciones donde la sobredispersión y los valores extremos desafían los enfoques tradicionales.

3.2. Definición

Se define la variable de conteo Y con distribución Binomial Negativa con Mixtura en la Dispersión, la cual se representa como $Y \sim \text{NB-H}(\mu, \phi, \nu)$, cuya representación jerárquica se muestra a continuación:

$$
\begin{aligned}
Y|W = \omega &\sim \text{NB}(\mu, \phi\,\omega) \ , \\
W &\sim H(\nu) \ .
\end{aligned}
$$

donde $H(\nu)$ representa una distribución para una variable aleatoria positiva con parámetro ν y función de densidad $h(\omega \mid \nu)$. Los parámetros del modelo son $\mu > 0$, $\phi > 0$ y $\nu > 0$. Para mantener la estructura de la varianza, se considera que $\mathbb{E}(W) = 1$. Además, la función de masa de probabilidad correspondiente al modelo NB-H se expresa como:

$$nb\text{-}h(y \mid \mu, \phi, \nu) = \int_0^\infty nb(y \mid \mu, \phi\,\omega)\ h(\omega \mid \nu)\ d\omega\ , \qquad y = 0, 1, 2, \ldots$$

que al no tener una forma cerrada, se requiere de la utilización de métodos numéricos para su evaluación.

La distribución Binomial Negativa con Mixtura en la Dispersión tiene una mayor flexibilidad al permitir que el parámetro de dispersión varíe según la distribución $H(\nu)$. De esta forma se logra capturar mejor la sobredispersión y manejar adecuadamente los valores atípicos, ajustando la variabilidad en los datos con mayor precisión que las distribuciones tradicionales utilizadas para datos de conteo. La media de esta distribución se puede calcular mediante la Ley de la Esperanza Total:

$$\mathbb{E}(Y) \ = \ \mathbb{E}(\mathbb{E}(Y|W)) \ = \ \mathbb{E}(\mu) \ = \ \mu$$

De manera similar, la varianza se puede obtener mediante la Ley de la Varianza Total:

$$\begin{aligned}
\mathrm{Var}(Y) &= \mathrm{Var}(\mathbb{E}(Y|W)) + \mathbb{E}(\mathrm{Var}(Y|W)) \\
&= \mathrm{Var}(\mu) + \mathbb{E}(\mu + \phi\ W\mu^2) \\
&= 0 + \mathbb{E}(\mu) + \phi\ \mu^2\ \mathbb{E}(W) \\
&= \mu + \phi\mu^2\ (1) \\
&= \mu + \phi\mu^2
\end{aligned}$$

Como se puede observar, tanto la media como la varianza de la distribución Binomial Negativa se mantienen en el modelo propuesto.

A continuación, se procederá a calcular la asimetría (skewness) asociada a la distribución NB-H. En este trabajo, se adopta la definición de asimetría basada en momentos centrales, la cual se expresa como el cociente entre el tercer momento centrado y la varianza elevada a la potencia 3/2. Esta elección responde a que dicha formulación permite analizar propiedades teóricas de la distribución de forma más estructurada, especialmente en modelos con estructura jerárquica o mezclas como el NB-H. A diferencia de otras medidas de asimetría, como

la asimetría de Bowley (basada en cuartiles) o las definiciones propuestas por Pearson, la medida basada en momentos no depende de la estimación empírica de percentiles, y resulta más adecuada para derivaciones analíticas, ya que se puede expresar en términos de los parámetros del modelo. Esto la convierte en una herramienta más versátil para evaluar la forma y la asimetría de la distribución en contextos inferenciales y de modelamiento probabilístico. Este análisis nos permitirá evaluar en qué medida la distribución se desvía de la simetría:

$$\text{Skew}(Y) = \frac{\mathbb{E}[(Y-\mu)^3]}{Var(Y)^{3/2}}.$$

Primero se calculará $\mathbb{E}[(Y-\mu)^3]$ mediante la Ley de Esperanza Total:

$$\mathbb{E}[(Y-\mu)^3] \;=\; \mathbb{E}[\; \mathbb{E}[(Y-\mu)^3 \mid W] \;] \;.$$

Como $Y|W$ sigue una distribución Binomial Negativa, se puede utilizar la fórmula de la asimetría de esta distribución para hallar $\mathbb{E}[(Y-\mu)^3 \mid W]$:

$$\begin{aligned}
\text{Skew}(Y \mid W) &= \frac{\mathbb{E}[(Y-\mu)^3 \mid W\,]}{Var(Y \mid W)^{3/2}} \\
\mathbb{E}[(Y-\mu)^3 \mid W\,] &= \text{Skew}(Y \mid W)\; Var(Y \mid W)^{3/2} \\
&= \frac{1}{\phi\; W}\;\frac{\phi\; W\; \mu}{1+\phi\; W\; \mu}\;\frac{1+2\;\phi\; W\; \mu}{1+\phi\; W\; \mu}\;(1+\phi\; W\; \mu)^3 \\
&= \mu\;(1+2\;\phi\; W\; \mu)\;(1+\phi\; W\; \mu)
\end{aligned}$$

Luego, reemplazamos en la Ley de Esperanza Total:

$$\begin{aligned}
\mathbb{E}[(Y-\mu)^3] &= \mathbb{E}[\; \mu\;(1+2\;\phi\; W\; \mu)\;(1+\phi\; W\; \mu)\;] \\
&= \mu\;(\;1+3\;\phi\;\mu\;\mathbb{E}[W] \;+\; 2\;\phi^2\;\mu^2\;\mathbb{E}[W^2]\;) \\
&= \mu\;(\;1+3\;\phi\;\mu\;(1) \;+\; 2\;\phi^2\;\mu^2\;\mathbb{E}[W^2]\;) \\
&= \mu\;(\;1+3\;\phi\;\mu \;+\; 2\;\phi^2\;\mu^2\;\mathbb{E}[W^2]\;)
\end{aligned}$$

Entonces, se puede reemplazar el valor obtenido para hallar la asimetría:

$$\begin{aligned}\text{Skew}(Y) &= \frac{\mathbb{E}[(Y-\mu)^3]}{Var(Y)^{3/2}} \\ &= \frac{\mu\ (\ 1+3\ \phi\ \mu\ +\ 2\ \phi^2\ \mu^2\ \mathbb{E}[W^2]\ )}{(\ \mu+\phi\mu^2\ )^{3/2}} \\ &= \frac{1+3\ \phi\ \mu\ +\ 2\ \phi^2\ \mu^2\ \mathbb{E}[W^2]}{\mu^{1/2}\ (\ 1+\phi\mu\ )^{3/2}}\end{aligned}$$

La expresión final queda en función de $\mathbb{E}[W^2]$, cuyo valor será evaluado más adelante para los casos en que W sigue una distribución Gamma o Inversa Gamma.

Luego, se realizará el cálculo de la curtosis asociada a la distribución NB-H. Este análisis nos permitirá evaluar el comportamiento de las colas:

$$\text{Kurt}(Y) = \frac{\mathbb{E}[(Y-\mu)^4]}{Var(Y)^2}.$$

Luego, se procederá a calcular $\mathbb{E}[(Y-\mu)^4]$ mediante el uso de la Ley de Esperanza Total:

$$\mathbb{E}[(Y-\mu)^4)]\ =\ \mathbb{E}[\ \mathbb{E}[(Y-\mu)^4\mid W]\ ]\ .$$

Dado que $Y|W$ sigue una distribución Binomial Negativa, se puede hacer uso de la expresión de la curtosis de esta distribución para hallar $\mathbb{E}[(Y-\mu)^4 \mid W]$:

$$\begin{aligned}\text{Kurt}(Y\mid W) &= \frac{\mathbb{E}[(Y-\mu)^4\mid W\ ]}{Var(Y\mid W)^2} \\ \mathbb{E}[(Y-\mu)^4\mid W\ ] &= \text{Kurt}(Y\mid W)\ Var(Y\mid W)^2 \\ &= \frac{1+6\ \phi\ W\ \mu+6\ \phi^2\ W^2\ \mu^2}{\mu\ (1+\phi\ W\ \mu)}\ \ (\mu+\phi\ W\ \mu^2)^2 \\ &= \frac{1+6\ \phi\ W\ \mu+6\ \phi^2\ W^2\ \mu^2}{\mu\ (1+\phi\ W\ \mu)}\ \ \mu^2\ (1+\phi\ W\ \mu)^2 \\ &= \mu\ (1+\phi\ W\ \mu)\ (1+6\ \phi\ W\ \mu+6\ \phi^2\ W^2\ \mu^2)\end{aligned}$$

Luego, se reemplazará en la Ley de Esperanza Total:

$$\mathbb{E}[(Y-\mu)^4] \;=\; \mathbb{E}[\; \mu\; (1+\phi\; W\; \mu)\; (1+6\; \phi\; W\; \mu+6\; \phi^2\; W^2\; \mu^2)\;]$$

Por último, reemplazamos el valor obtenido para obtener la curtosis:

$$\begin{aligned}
\text{Kurt}(Y) &= \frac{\mathbb{E}[(Y-\mu)^4]}{Var(Y)^2} \\
&= \frac{\mathbb{E}[\; \mu\; (1+\phi\; W\; \mu)\; (1+6\; \phi\; W\; \mu+6\; \phi^2\; W^2\; \mu^2)\;]}{(\; \mu+\phi\mu^2\;)^2} \\
&= \frac{\mu\;\; \mathbb{E}[\; (1+\phi\; W\; \mu)\; (1+6\; \phi\; W\; \mu+6\; \phi^2\; W^2\; \mu^2)\;]}{\mu^2\; (\; 1+\phi\mu\;)^2} \\
&= \frac{\mathbb{E}[\; (1+\phi\; W\; \mu)\; (1+6\; \phi\; W\; \mu+6\; \phi^2\; W^2\; \mu^2)\;]}{\mu\; (\; 1+\phi\mu\;)^2} \\
&= \frac{\mathbb{E}[\; 1+7\; \phi\; W\; \mu+12\; \phi^2\; W^2\; \mu^2\; +6\; \phi^3\; W^3\; \mu^3)\;]}{\mu\; (\; 1+\phi\mu\;)^2} \\
&= \frac{1+7\; \phi\; \mu\; +\; 12\; \phi^2\; \mu^2\; \mathbb{E}[W^2]\; +\; 6\; \phi^3\; \mu^3\; \mathbb{E}[W^3]}{\mu\; (\; 1+\phi\mu\;)^2}
\end{aligned}$$

La fórmula obtenida depende de $\mathbb{E}[W^2]$ y $\mathbb{E}[W^3]$, cuyos valores serán examinados en los casos en que W sigue una distribución Gamma o Inversa Gamma.

3.3. Casos Particulares de la Distribución NB-H

3.3.1. Distribución NB-H utilizando la Distribución Gamma

La distribución Gamma se emplea en este modelo como una distribución latente y su función de densidad de probabilidad se describe a continuación:

$$g(\omega \mid \alpha, \beta) = \frac{\beta^{\alpha}\omega^{\alpha-1}e^{-\beta\omega}}{\Gamma(\alpha)}, \quad \omega > 0 \ ,$$

donde los parámetros α y β determinan directamente la forma de la distribución. Una de las propiedades destacadas de esta distribución es que, cuando $\alpha > 1$, la función de densidad es cóncava hacia abajo y presenta un máximo único en $\omega = \frac{\alpha-1}{\beta}$, garantizando una estructura unimodal. Se puede apreciar en la Figura 1 que, al considerar distintos valores del parámetro ν, la función de masa de probabilidad conserva una única moda claramente identificable.

Luego, se examinará la mixtura de la distribución Binomial Negativa con la distribución Gamma. La distribución Gamma se parametrizará con $\alpha = \nu + 1$ y $\beta = \nu + 1$. Esta elección asegura que la distribución sea unimodal, ya que $\nu > 0$ implica $\alpha > 1$. Además, al establecer ambos parámetros con el mismo valor, se garantizará que la esperanza de la variable latente W se mantenga en el valor de 1 y que la varianza sea positiva. Es así que se define la variable $W \sim \text{Gamma}(\nu + 1, \nu + 1)$, cuya media es la siguiente:

$$\begin{aligned} \mathbb{E}(W) &= \frac{\alpha}{\beta} \\ &= \frac{\nu+1}{\nu+1} \\ &= 1 \end{aligned}$$

además, su varianza se calcula de la siguiente manera:

$$\begin{aligned} \text{Var}(W) &= \frac{\alpha}{\beta^2} \\ &= \frac{\nu+1}{(\nu+1)^2} \\ &= \frac{1}{\nu+1} \end{aligned}$$

Luego, se define la variable de conteo Y con distribución Binomial Negativa con Mixtura en la Dispersión utilizando la Distribución Gamma (en adelante, NB-G), la cual se representa como $Y \sim \text{NB-G}(\mu, \phi, \nu)$. La estructura jerárquica de esta distribución se describe de la siguiente manera:

$$Y|W = \omega \sim \text{NB}(\mu, \phi\ \omega)$$
$$W \sim \text{Gamma}(\nu + 1, \nu + 1)$$

donde los parámetros correspondientes al modelo son $\mu > 0$, $\phi > 0$ y $\nu > 0$. Cabe señalar que la función de masa de probabilidad asociada a esta distribución no posee una forma cerrada, por lo que su evaluación requiere la aplicación de métodos numéricos. Esta función se define como:

$$nb\text{-}g(y \mid \mu, \phi, \nu) = \int_0^\infty nb(y \mid \mu, \phi\ \omega)\ g(\omega \mid \nu, \nu)\ d\omega\ , \quad y = 0, 1, 2, \ldots$$

Con el objetivo de ilustrar el efecto del parámetro ν en la forma de la distribución, a continuación se presenta una comparación gráfica de las funciones de masa de probabilidad del modelo NB-G para diferentes valores de este parámetro. En todos los casos, se mantiene constante la media en $\mu = 50$ y el parámetro de dispersión en $\phi = 0.25$, lo que permite analizar exclusivamente el impacto de ν sobre la forma de la distribución.

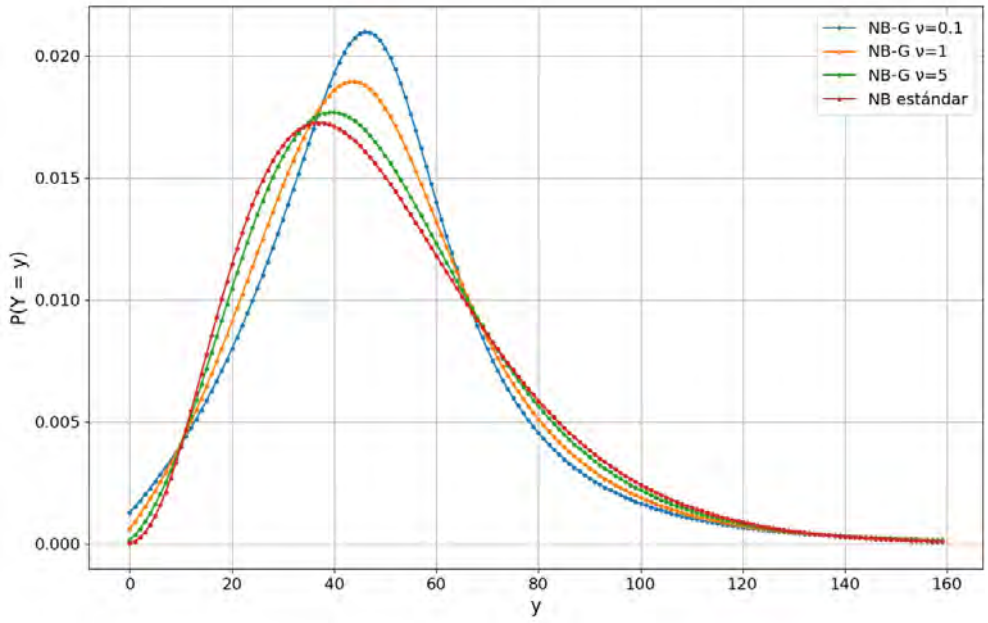

Figura 1: Comparación de las funciones de masa de probabilidad del modelo NB-G para distintos valores del parámetro ν, frente a la distribución Binomial Negativa estándar. Se considera $\mu = 50$ y $\phi = 0.25$.

La Figura 1 permite observar que, a medida que el valor de ν disminuye, la distribución NB-G se vuelve más concentrada en torno a su media, adoptando una forma más aguda en su núcleo. Este patrón se traduce en un incremento de la curtosis con respecto a la Binomial Negativa estándar, lo que implica una mayor propensión a valores extremos, tanto bajos como altos. Esta característica es clave en contextos donde se requiere mayor sensibilidad ante eventos poco frecuentes.

Ahora bien, si bien la Figura 1 permite apreciar la concentración central, no brinda una representación clara del comportamiento en las colas, donde se manifiestan los efectos de la asimetría y de la dispersión extrema. Para abordar este aspecto, se presenta una segunda figura que amplía el rango superior del soporte de la distribución.

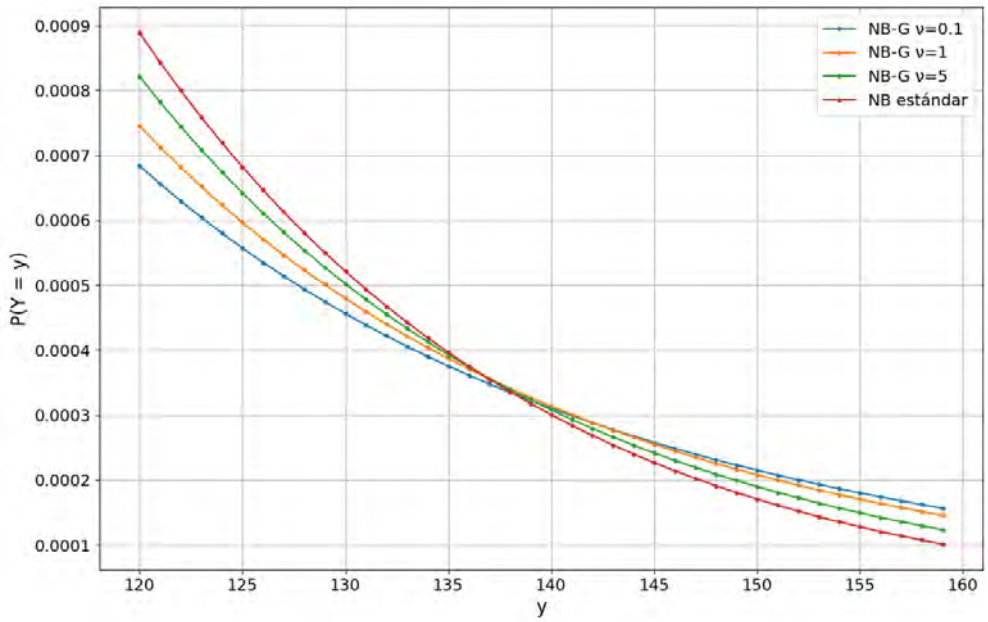

Figura 2: Ampliación de las colas derechas de las funciones de masa de probabilidad del modelo NB-G para distintos valores de ν, en comparación con la distribución Binomial Negativa estándar, en el rango $y \in [120, 160]$.

En la Figura 2 se aprecia con mayor claridad la evolución de la probabilidad en la cola derecha. Específicamente, para valores altos de y, la distribución NB-G con $\nu = 0.1$ decrece a un ritmo más lento que las restantes, incluida la Binomial Negativa estándar. Este comportamiento confirma la presencia de colas más pesadas cuando ν es reducido, lo cual se traduce en una mayor asimetría positiva y en una probabilidad no despreciable de ocurrencia de valores extremos. Dicha propiedad es particularmente útil al modelar fenómenos donde la presencia de datos atípicos y de alta dispersión representa una preocupación relevante. El código empleado para la generación de ambas figuras ha sido incluido en el Apéndice A.1, a fin de garantizar la trazabilidad y replicabilidad de los resultados presentados.

A continuación, se presentan las expresiones correspondientes a la media y varianza de la distribución NB-G:

$$\mathbb{E}(Y) = \mu$$

$$\text{Var}(Y) = \mu + \phi\mu^2$$

Estas expresiones coinciden con las de la distribución Binomial Negativa estándar, lo que resalta que el parámetro ν no modifica los momentos de primer y segundo orden, sino que influye en la forma de la distribución.

Seguidamente, se procederá a calcular la asimetría de esta distribución:

$$\begin{aligned}
\text{Skew}(Y) &= \frac{1 + 3\ \phi\ \mu\ +\ 2\ \phi^2\ \mu^2\ \mathbb{E}[W^2]}{\mu^{1/2}\ (\ 1 + \phi\mu\ )^{3/2}} \\
&= \frac{1 + 3\ \phi\ \mu\ +\ 2\ \phi^2\ \mu^2\ (1 + \frac{1}{\nu+1})}{\mu^{1/2}\ (\ 1 + \phi\mu\ )^{3/2}} \\
&= \frac{1 + 3\ \phi\ \mu\ +\ 2\ \phi^2\ \mu^2\ +\frac{2\ \phi^2\ \mu^2}{\nu+1}}{\mu^{1/2}\ (\ 1 + \phi\mu\ )^{3/2}} \\
&= \frac{(\ 1 +\ \phi\ \mu)(\ 1 + 2\ \phi\ \mu) + \frac{2\ \phi^2\ \mu^2}{\nu+1}}{\mu^{1/2}\ (\ 1 + \phi\mu\ )^{3/2}} \\
&= \frac{1 + 2\ \phi\ \mu}{\mu^{1/2}\ (\ 1 + \phi\mu\ )^{1/2}} + \frac{2\ \phi^2\ \mu^{3/2}}{(\nu + 1)\ (\ 1 + \phi\mu\ )^{3/2}}
\end{aligned}$$

La primera fracción coincide exactamente con la asimetría de la distribución Binomial Negativa, como se indica en la Ecuación (1). Por tanto, se puede reescribir la expresión anterior de forma más concisa como:

$$\text{Skew}(Y)\ =\ \text{Skew}(Y_{NB})\ +\ \frac{2\,\phi^2\,\mu^{3/2}}{(\nu + 1)(1 + \phi\,\mu)^{3/2}}$$

De esta manera, se puede observar que la distribución NB-G siempre tendrá mayor asimetría que la distribución Binomial Negativa estándar, ya que el término adicional $\frac{2\ \phi^2\ \mu^{3/2}}{(\nu+1)\ (\ 1\ +\ \phi\mu\ )^{3/2}}$ es positivo. Asimismo, cuando el valor de ν disminuye, el valor de la asimetría aumenta. La Figura 3 nos permite apreciar cómo cambia la asimetría para diferentes valores de ν:

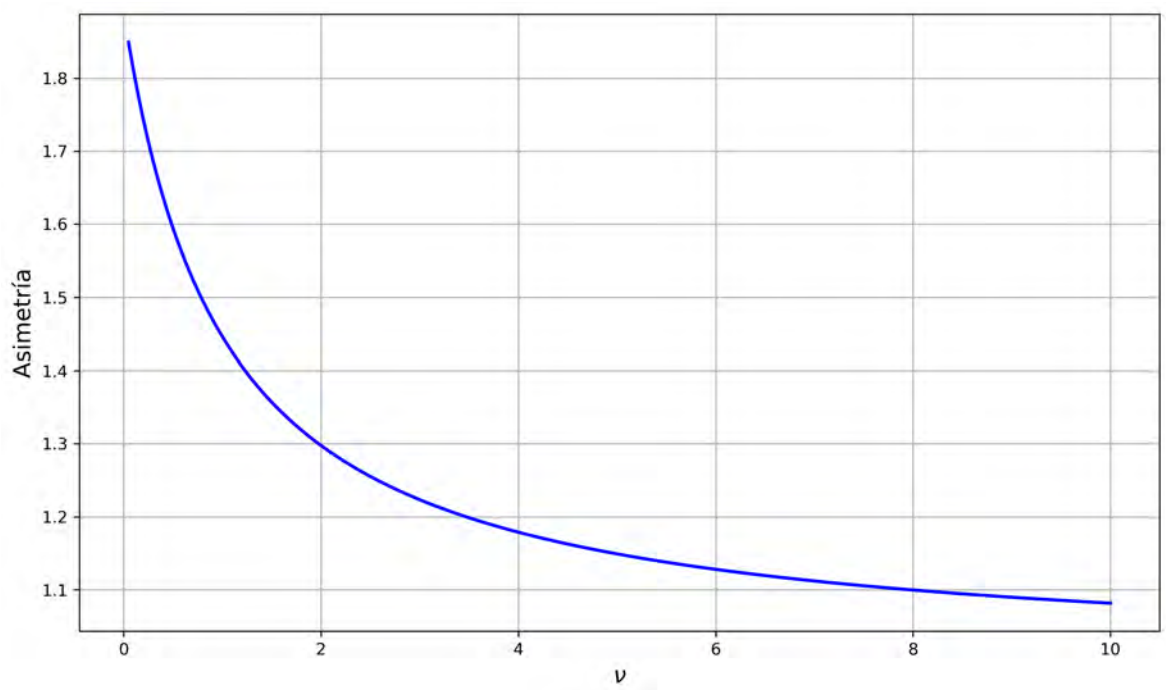

Figura 3: Asimetría de la distribución NB-G en función del parámetro ν, considerando $\mu = 50$ y $\phi = 0.25$.

El código utilizado para calcular y graficar esta figura se encuentra documentado en el Apéndice A.2.

Luego, se calculará la curtosis de esta distribución:

$$\begin{aligned}
\text{Kurt}(Y) &= \frac{1 + 7\ \phi\ \mu\ +\ 12\ \phi^2\ \mu^2\ \mathbb{E}[W^2]\ +\ 6\ \phi^3\ \mu^3\ \mathbb{E}[W^3]}{\mu\ (\ 1 + \phi\mu\ )^2} \\
&= \frac{1 + 7\ \phi\ \mu\ +\ 12\ \phi^2\ \mu^2\ \ (1 + \frac{1}{\nu+1})\ +\ 6\ \phi^3\ \mu^3\ (1 + \frac{3}{\nu+1} + \frac{2}{(\nu+1)^2})}{\mu\ (\ 1 + \phi\mu\ )^2} \\
&= \frac{1 + 7\ \phi\ \mu\ +\ 12\ \phi^2\ \mu^2 +\ 6\ \phi^3\ \mu^3\ +\ \ (\frac{12\ \phi^2\ \mu^2 +\ 18\ \phi^3\ \mu^3}{\nu+1} + \frac{12\ \phi^3\ \mu^3}{(\nu+1)^2})}{\mu\ (\ 1 + \phi\mu\ )^2} \\
&= \frac{(\ 1\ +\ \phi\ \mu\ )\ (\ 1\ +\ 6\ \phi\ \mu\ +\ 6\ \phi^2\ \mu^2) +\ 6\ \phi^2\ \mu^2(\frac{2\ +\ 3\ \phi\ \mu}{\nu+1} + \frac{2\ \phi\ \mu}{(\nu+1)^2})}{\mu\ (\ 1 + \phi\mu\ )^2} \\
&= \frac{1\ +\ 6\ \phi\ \mu\ +\ 6\ \phi^2\ \mu^2}{\mu\ (\ 1 + \phi\mu\ )}\ + \frac{6\ \phi^2\ \mu(\frac{2\ +\ 3\ \phi\ \mu}{\nu+1} + \frac{2\ \phi\ \mu}{(\nu+1)^2})}{(\ 1 + \phi\mu\ )^2}
\end{aligned}$$

La primera fracción coincide exactamente con la curtosis de la distribución Binomial

Negativa, tal como se muestra en la Ecuación (2), por lo que se puede reescribir la expresión de manera más compacta como:

$$\text{Kurt}(Y) \;=\; \text{Kurt}(Y_{NB}) \;+\; \frac{6\,\phi^2\,\mu\left(\frac{2+3\,\phi\,\mu}{\nu+1} + \frac{2\,\phi\,\mu}{(\nu+1)^2}\right)}{(1+\phi\,\mu)^2}$$

Es así, que se puede observar que la distribución NB-G siempre tendrá mayor curtosis que la distribución Binomial Negativa, pues el término adicional $\frac{6\ \phi^2\ \mu(\frac{2\ +\ 3\ \phi\ \mu}{\nu+1}+\frac{2\ \phi\ \mu}{(\nu+1)^2})}{(\ 1\ +\ \phi\mu\ )^2}$ es estrictamente positivo para todo $\nu > 0$. Entonces, se puede concluir que la distribución NB-G presenta colas más pesadas que la distribución NB, ya que su curtosis es mayor. Al presentar colas más pesadas, esta distribución será más robusta frente a los datos atípicos. Además, también se puede deducir que cuando el valor de ν disminuye, la curtosis aumenta, lo que provoca colas más pesadas. La Figura 4 muestra cómo varía la curtosis de la distribución NB-G para diferentes valores del parámetro ν:

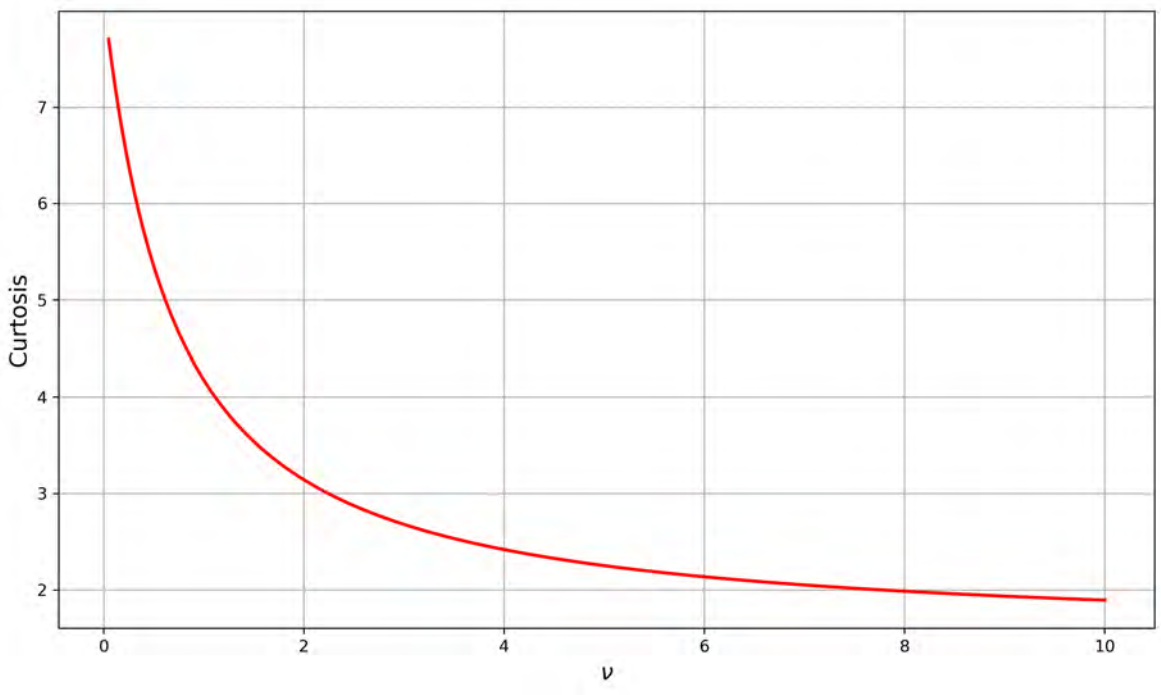

Figura 4: Curtosis de la distribución NB-G en función del parámetro ν, considerando $\mu = 50$ y $\phi = 0.25$.

El código utilizado para calcular y graficar esta figura se encuentra documentado en el Apéndice A.3.

3.3.2. Distribución NB-H utilizando la Distribución Inversa Gamma

La distribución Inversa Gamma se utilizará en este modelo como una distribución latente y a continuación se describe su función de densidad de probabilidad:

$$ig(\omega \mid \alpha, \beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)}\ \omega^{-\alpha-1}\ e^{-\frac{\beta}{\omega}}, \quad \omega > 0\ ,$$

donde los parámetros α y β configuran la forma de la distribución. Una característica de esta distribución es que su función de densidad es cóncava hacia abajo y presenta un máximo único en $\omega = \frac{\beta}{\alpha+1}$, lo que asegura una estructura unimodal. Tal como se aprecia en la Figura 5, la función de masa de probabilidad mantiene una única moda claramente distinguible al variar los valores del parámetro ν. La existencia de su esperanza está garantizada si $\alpha > 1$, mientras que su varianza existe únicamente si $\alpha > 2$.

Luego, se examinará la mixtura de la distribución Binomial Negativa con la distribución Inversa Gamma. La distribución Inversa Gamma se parametrizará con $\alpha = \nu+2$ y $\beta = \nu+1$. Esta elección asegura que la esperanza y la varianza de la distribución existan, ya que $\nu > 0$ implica $\alpha > 2$. Además, el valor dado al parámetro β garantizará que la esperanza de la variable latente W se mantenga en el valor de 1. Es así que se define la variable
$W \sim \text{IG}(\nu + 2, \nu + 1)$, cuya esperanza es la siguiente:

$$\begin{aligned}
\mathbb{E}(W) &= \frac{\beta}{\alpha - 1} \\
&= \frac{\nu + 1}{\nu + 2 - 1} \\
&= 1
\end{aligned}$$

además, su varianza se calcula de la siguiente manera:

$$\begin{aligned}
\text{Var}(W) &= \frac{\beta^2}{(\alpha - 1)^2(\alpha - 2)} \\
&= \frac{(\nu + 1)^2}{(\nu + 2 - 1)^2(\nu + 2 - 2)} \\
&= \frac{1}{\nu}
\end{aligned}$$

Luego, se define la variable de interés Y con distribución Binomial Negativa con Mixtura en la Dispersión utilizando la Distribución Gamma Inversa (en adelante, NB-IG), que tiene

la siguiente representación: $Y \sim \text{NB-IG}(\mu, \phi, \nu)$. Además, su estructura jerárquica es la siguiente:

$$\begin{aligned} Y|W = \omega &\sim \text{NB}(\mu, \phi\ \omega) \\ W &\sim \text{IG}(\nu + 2, \nu + 1) \end{aligned}$$

donde los parámetros del modelo son $\mu > 0$, $\phi > 0$ y $\nu > 0$. La función de masa de probabilidad (PMF) correspondiente al modelo es la siguiente:

$$nb\text{-}ig(y \mid \mu, \phi, \nu) = \int_0^\infty nb(y \mid \mu, \phi\ \omega)\ ig(\omega \mid \nu, \nu) d\omega\ , \quad y = 0, 1, 2, \ldots$$

que, al igual que la distribución NB-G, no posee una expresión cerrada para su función de masa de probabilidad.

Con el propósito de examinar visualmente el efecto del parámetro ν sobre la forma de la distribución NB-IG, se presenta a continuación una comparación de sus funciones de masa de probabilidad para distintos valores de dicho parámetro. En todos los escenarios considerados, se han fijado los valores $\mu = 50$ y $\phi = 0.25$, de manera que las diferencias observadas en las curvas se atribuyen exclusivamente a la variación de ν. Esta visualización resulta útil para identificar cómo la mezcla con una Inversa Gamma afecta tanto la concentración alrededor de la media como el comportamiento en las colas.

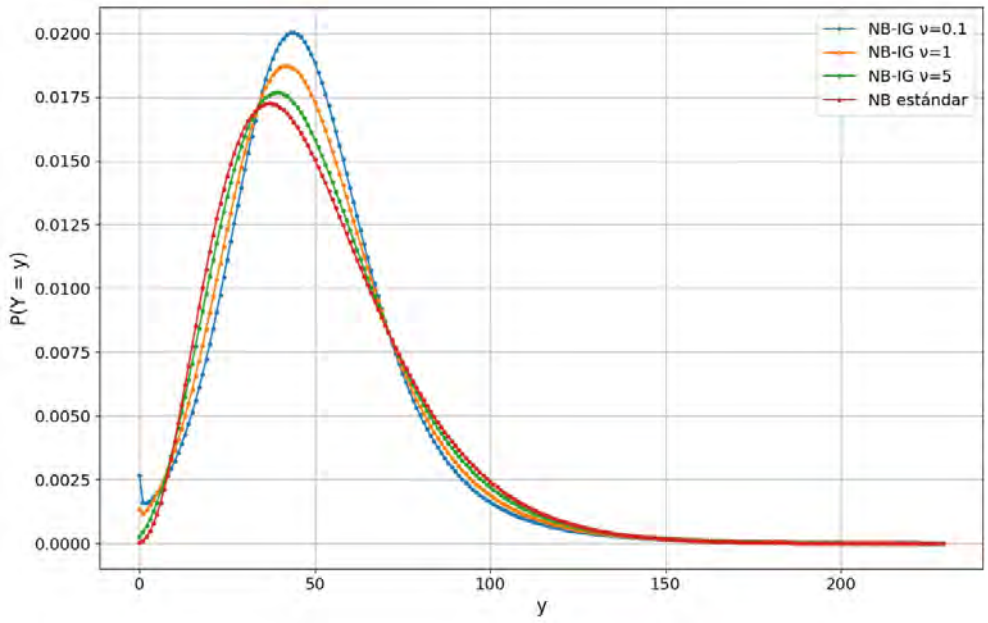

Figura 5: Comparación de las funciones de masa de probabilidad del modelo NB-IG para distintos valores del parámetro ν, frente a la distribución Binomial Negativa estándar. Se considera $\mu = 50$ y $\phi = 0.25$.

La Figura 5 evidencia que la curtosis de la distribución NB-IG se incrementa conforme ν toma valores más pequeños. Este efecto se traduce en una mayor concentración de probabilidad en torno a la media, junto con colas más prolongadas. En contraste con la distribución Binomial Negativa estándar, la NB-IG adquiere una forma más aguda y con peso adicional en los extremos cuando ν es bajo. A medida que ν se incrementa, la distribución pierde esta forma puntiaguda, sus colas se contraen gradualmente, y su perfil se aproxima al de la Binomial Negativa estándar, recuperando una estructura más equilibrada y con menor dispersión.

Para examinar con mayor detalle el comportamiento de la distribución NB-IG en los valores elevados de y, se presenta a continuación una visualización ampliada de la cola derecha. Esta permite comparar la tasa de decaimiento de la función de masa de probabilidad entre las distintas curvas, lo cual resulta fundamental para comprender el comportamiento de las colas bajo distintos valores de ν.

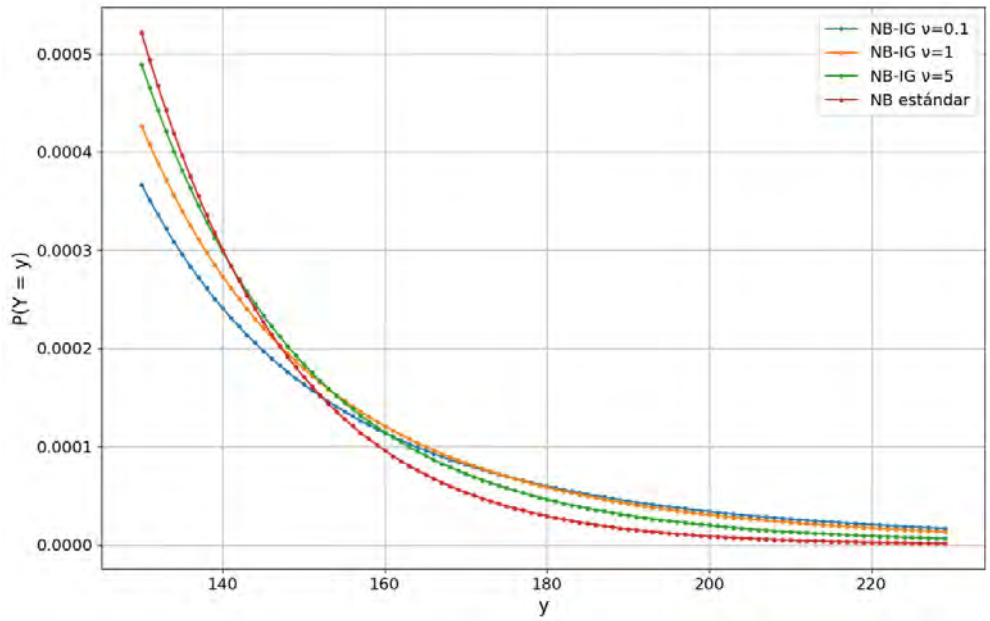

Figura 6: Ampliación de las colas derechas de las funciones de masa de probabilidad del modelo NB-IG para distintos valores de ν, en comparación con la distribución Binomial Negativa estándar, en el rango $y \in [130, 230]$.

La Figura 6 confirma que, al decrecer ν, la distribución NB-IG asigna mayor probabilidad a los valores extremos, lo que refleja la presencia de colas más pesadas. Esta característica implica una mayor propensión a observar valores atípicos. Además, se observa que la asimetría positiva se intensifica a medida que ν disminuye, evidenciando una mayor inclinación de la distribución hacia valores bajos de y.

En conjunto, los efectos observados sobre la curtosis y la asimetría refuerzan la utilidad del modelo NB-IG en situaciones donde los datos presentan datos atípicos y alta sobredispersión. El código empleado para la generación de ambas figuras ha sido incluido en el Apéndice B.1, a fin de garantizar la trazabilidad y replicabilidad de los resultados presentados.

La esperanza y varianza de la distribución NB-IG son las siguientes:

$$\mathbb{E}(Y) = \mu$$

$$\text{Var}(Y) = \mu + \phi\mu^2$$

Luego, se calculará la asimetría de esta distribución:

$$
\begin{aligned}
\text{Skew}(Y) &= \frac{1 + 3\ \phi\ \mu\ +\ 2\ \phi^2\ \mu^2\ \mathbb{E}[W^2]}{\mu^{1/2}\ (\ 1 + \phi\mu\ )^{3/2}} \\
&= \frac{1 + 3\ \phi\ \mu\ +\ 2\ \phi^2\ \mu^2\ (1 + \frac{1}{\nu})}{\mu^{1/2}\ (\ 1 + \phi\mu\ )^{3/2}} \\
&= \frac{1 + 3\ \phi\ \mu\ +\ 2\ \phi^2\ \mu^2\ +\ \frac{2\ \phi^2\ \mu^2}{\nu}}{\mu^{1/2}\ (\ 1 + \phi\mu\ )^{3/2}} \\
&= \frac{(\ 1 +\ \phi\ \mu)(\ 1 + 2\ \phi\ \mu) + \frac{2\ \phi^2\ \mu^2}{\nu}}{\mu^{1/2}\ (\ 1 + \phi\mu\ )^{3/2}} \\
&= \frac{1 + 2\ \phi\ \mu}{\mu^{1/2}\ (\ 1 + \phi\mu\ )^{1/2}} + \frac{2\ \phi^2\ \mu^{3/2}}{\nu\ (\ 1 + \phi\mu\ )^{3/2}}
\end{aligned}
$$

La primera parte de la expresión corresponde exactamente a la asimetría de la distribución Binomial Negativa, la cual se encuentra definida en la Ecuación(1). En consecuencia, es posible reformular la expresión total de manera más simplificada como:

$$
\text{Skew}(Y)\ =\ \text{Skew}(Y_{NB})\ +\ \frac{2\ \phi^2\ \mu^{3/2}}{\nu\ (\ 1 + \phi\mu\ )^{3/2}}
$$

De esta manera, se puede observar que la distribución NB-IG va a tener un valor mayor de asimetría, en comparación con la distribución NB, ya que el término adicional $\frac{2\ \phi^2\ \mu^{3/2}}{\nu\ (\ 1\ +\ \phi\mu\ )^{3/2}}$ siempre es positivo. Adicionalmente, cabe destacar que cuando el valor de ν disminuye, el valor de la asimetría aumenta. La Figura 7 nos permite observar como varía la asimetría para diferentes valores de ν y valores fijos de $\mu = 50$ y $\phi = 0.25$:

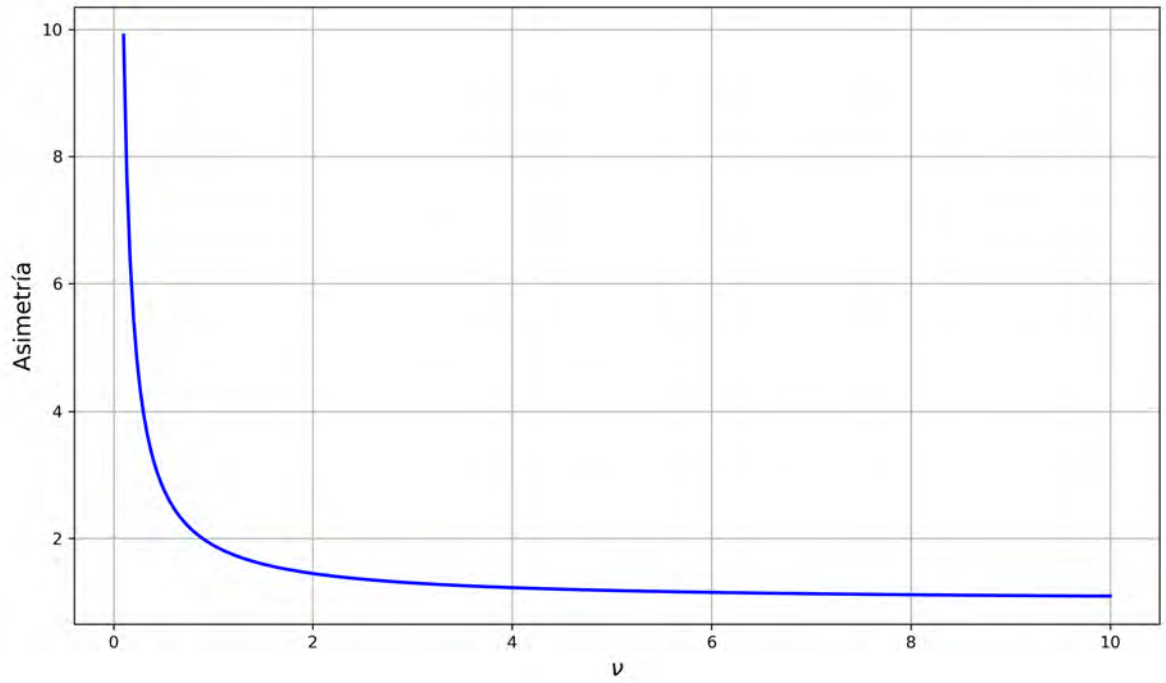

Figura 7: Asimetría de la distribución NB-IG en función del parámetro ν, considerando $\mu = 50$ y $\phi = 0.25$.

El código utilizado para calcular y graficar esta figura se encuentra documentado en el Apéndice B.2.

Luego, se calculó la curtosis de esta distribución, que existe solo para valores de $\nu > 1$:

$$\text{Kurt}(Y) = \frac{1 + 7\ \phi\ \mu\ +\ 12\ \phi^2\ \mu^2\ \mathbb{E}[W^2]\ +\ 6\ \phi^3\ \mu^3\ \mathbb{E}[W^3]}{\mu\ (\ 1 + \phi\mu\ )^2}$$

$$= \frac{1 + 7\ \phi\ \mu\ +\ 12\ \phi^2\ \mu^2\ \ (1 + \frac{1}{\nu})\ +\ 6\ \phi^3\ \mu^3\ (\frac{(\nu+1)^2}{\nu\ (\nu-1)})}{\mu\ (\ 1 + \phi\mu\ )^2}$$

$$= \frac{1 + 7\ \phi\ \mu\ +\ 12\ \phi^2\ \mu^2 +\ 6\ \phi^3\ \mu^3\ + 12\ \phi^2\ \mu^2 \frac{1}{\nu} +\ \ 6\ \phi^3\ \mu^3\ (\frac{3\ \nu\ +\ 1}{\nu\ (\nu-1)})}{\mu\ (\ 1 + \phi\mu\ )^2}$$

$$= \frac{(\ 1\ +\ \phi\ \mu\ )\ (\ 1\ +\ 6\ \phi\ \mu\ +\ 6\ \phi^2\ \mu^2) +\ 6\ \phi^2\ \mu^2\ (\frac{2(\nu-1)\ +\ \phi\ \mu\ (3\nu+1)}{\nu\ (\nu-1)})}{\mu\ (\ 1 + \phi\mu\ )^2}$$

$$= \frac{1\ +\ 6\ \phi\ \mu\ +\ 6\ \phi^2\ \mu^2}{\mu\ (\ 1 + \phi\mu\ )}\ + \frac{6\ \phi^2\ \mu\ (\ 2(\nu - 1)\ +\ \phi\ \mu\ (3\nu + 1)\ )}{\nu\ (\nu - 1)\ (\ 1 + \phi\mu\ )^2}$$

El primer término de la expresión corresponde exactamente a la curtosis de la distribución Binomial Negativa, la cual se encuentra definida en la Ecuación (2). Por ello, es posible reformular la expresión total de forma más resumida como:

$$\text{Kurt}(Y) \;=\; \text{Kurt}(Y_{NB}) \;+\; \frac{6\;\phi^2\;\mu\;(\;2(\nu-1)\;+\;\phi\;\mu\;(3\nu+1)\;)}{\nu\;(\nu-1)\;(\;1+\phi\mu\;)^2}$$

Se puede observar que la distribución NB-IG tiene una mayor curtosis que la distribución NB, ya que el término adicional $\frac{6\;\phi^2\;\mu\;(\;2(\nu\;-\;1)\;+\;\phi\;\mu\;(3\nu\;+\;1)\;)}{\nu\;(\nu\;-\;1)\;(\;1\;+\;\phi\mu\;)^2}$ es positivo para $\nu > 1$. Por tanto, llegamos a la conclusión que la distribución NB-IG va a tener colas más pesadas que la distribución NB, debido a que su curtosis es mayor. Debido a que esta distribución tiene colas más pesadas, se puede afirmar que esta distribución es más robusta ante la presencia de datos atípicos. Adicionalmente, es importante destacar que cuando el valor de ν disminuye, la curtosis incrementa, generando que la distribución tenga colas más pesadas. La Figura 8 nos permite apreciar como varía la curtosis para diferentes valores de ν:

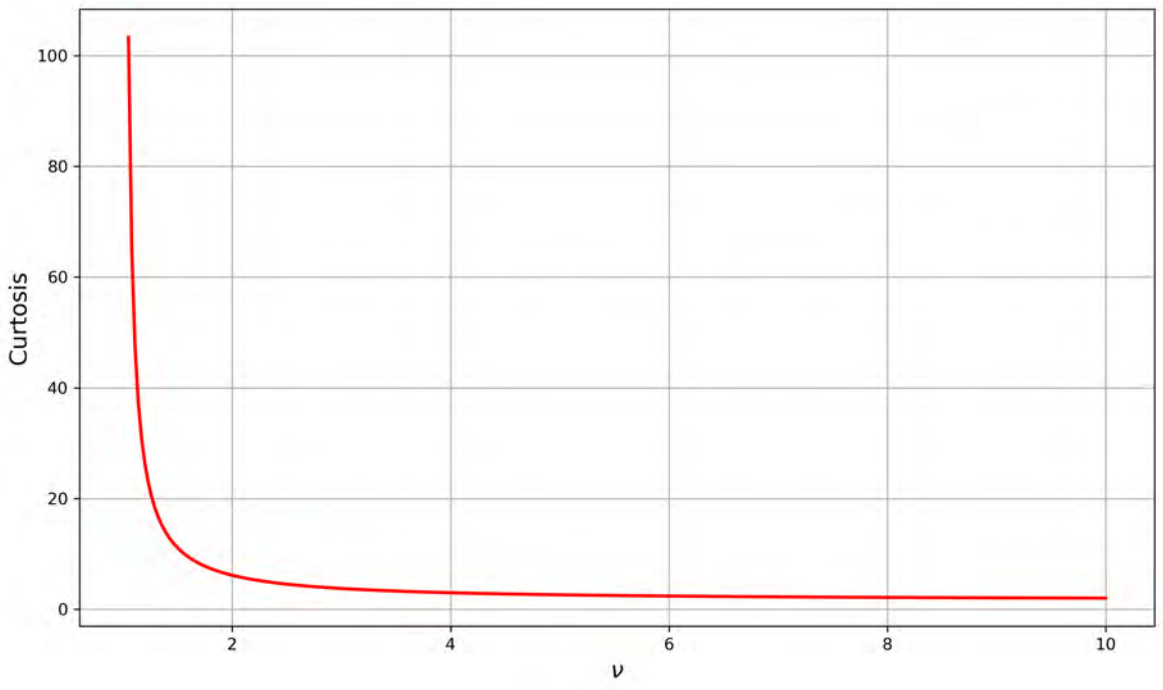

Figura 8: Curtosis de la distribución NB-IG en función del parámetro ν, considerando $\mu = 50$ y $\phi = 0.25$.

El código utilizado para calcular y graficar esta figura se encuentra documentado en el Apéndice B.3.

Finalmente, si bien en este estudio se ha optado por utilizar las distribuciones Gamma e Inversa Gamma para modelar la variabilidad en el parámetro de dispersión, esta elección responde a su flexibilidad, soporte positivo y propiedades analíticas bien conocidas, que permiten un tratamiento computacional eficiente en entornos bayesianos. No obstante, la estructura del modelo propuesto es lo suficientemente general como para incorporar otras familias de distribuciones continuas y positivas, como la distribución Log-Normal o la Inversa Gaussiana. Estas alternativas pueden resultar útiles en contextos donde se sospecha de colas más pesadas o asimetrías específicas en la dispersión latente, constituyendo así una posible extensión del presente trabajo en futuras investigaciones.

Capítulo 4

Modelo de Regresión Binomial Negativa con Mixtura en la Dispersión

En este capítulo, se introduce el modelo de Regresión Binomial Negativa con Mixtura en la Dispersión (NB-H). Este modelo se propone como una herramienta robusta para abordar situaciones donde los datos de conteo presentan sobredispersión o están afectados por valores atípicos.

En este contexto, la propiedad de unimodalidad en la regresión adquiere una relevancia particular. Contar con una única moda favorece la estabilidad de las estimaciones y facilita la identificación clara de los efectos de las covariables, incluso en presencia de ruido o valores atípicos. Esto se traduce en mayor robustez inferencial y mejor capacidad explicativa del modelo.

A lo largo del capítulo, se presentará la formulación matemática del modelo y los métodos utilizados para la estimación de sus parámetros, con un enfoque particular en la inferencia bayesiana. Se detallará cómo las técnicas de muestreo, como el método *No-U-Turn Sampler* (NUTS), permiten estimar eficientemente los parámetros del modelo. Finalmente, se incluirán criterios de comparación de modelos, como el DIC y el WAIC, que permiten evaluar el desempeño del modelo propuesto frente a otros enfoques convencionales.

4.1. Definición del Modelo

Sea Y_i la variable de respuesta y $X_i \in \mathbb{R}^p$ el vector de covariables asociado a la observación i, con $i = 1, 2, \ldots, n$. Se asume que las variables Y_i son condicionalmente independientes dado X_i, y que cada una sigue la distribución:

$$Y_i \sim \text{NB-H}(\mu_i, \phi, \nu),$$

donde $\mu_i > 0$ representa la media de la i-ésima observación, $\phi > 0$ es el parámetro de dispersión y $\nu > 0$ es el parámetro que se encarga de modular la dispersión. La relación entre

la media μ_i y las covariables se establece mediante una función de enlace logarítmica:

$$\log(\mu_i) = X_i^\top \boldsymbol{\beta},$$

donde $\boldsymbol{\beta} \in \mathbb{R}^p$ es el vector de coeficientes del modelo. Esta transformación asegura que las medias predichas sean estrictamente positivas, además de facilitar la interpretación de los efectos de las covariables.

Con el objetivo de proporcionar una comprensión más profunda del modelo propuesto, a continuación se presenta su formulación jerárquica explícita. Este enfoque permite visualizar la estructura latente que da origen a la distribución utilizada:

$$y_i \mid W_i = \omega_i \sim \text{NB}(\mu_i, \phi\omega_i)$$

$$W_i \sim H(\nu)$$

donde la distribución de W_i, denotada como $H(\nu)$, corresponde a una variable aleatoria continua y positiva parametrizada por ν. Esta variable latente introduce una mayor flexibilidad al modelo. Algunas de las distribuciones que cumplen con el requisito son las distribuciones Gamma e Inversa Gamma, entre otras.

Aunque el modelo puede ampliarse para incorporar covariables en el parámetro de dispersión ϕ, en este estudio se decidió enfocar el análisis en la media, con el objetivo de resaltar la relación entre las covariables y la variable de respuesta. La exploración de estructuras adicionales en la dispersión se propone como una línea interesante para trabajos futuros.

La función de verosimilitud aumentada del modelo está dada por:

$$\begin{aligned} L(y, \omega \mid \theta) &= \prod_{i=1}^{n} p(y_i \mid \omega_i)\ p(\omega_i) \\ &= \prod_{i=1}^{n} nb(y_i \mid \mu_i, \phi\ \omega_i)\ h(\omega_i \mid \nu) \end{aligned}$$

donde $\theta = (\boldsymbol{\beta}^\top, \phi, \nu)^\top$.

En este trabajo se utilizará inferencia bayesiana debido a que en escenarios donde existe sobredispersión, datos atípicos y complejidades estructurales, los métodos clásicos suelen presentar limitaciones. A diferencia de métodos clásicos, más restrictivos en sus supuestos, el enfoque bayesiano permite incorporar información previa mediante distribuciones a priori, lo que añade flexibilidad y mayor solidez a las estimaciones. Además, este enfoque evita la

necesidad de evaluar numéricamente la función de probabilidad completa, lo cual representa una ventaja computacional significativa. En conjunto, la aproximación bayesiana se presenta como la alternativa más versátil para ajustar e interpretar modelos NB-H.

4.2. Inferencia Bayesiana

En esta sección se presentará el proceso de inferencia bayesiana para el modelo de Regresión Binomial Negativa con Mixtura en la Dispersión NB-H(μ_i, ϕ, ν). Para la estimación de los parámetros del modelo se utiliza la librería PyStan, el cual implementa el método *No-U-Turn Sampler* (NUTS), que permite la estimación eficiente de los parámetros.

4.2.1. Distribución a posteriori aumentada

La distribución a posteriori aumentada se detalla de la siguiente manera:

$$p(\theta, \omega \mid y) \;\propto\; L(y, \omega \mid \theta)\; p(\theta),$$

que también puede ser expresado como:

$$p(\theta, \omega \mid y) \;\propto\; \prod_{i=1}^{n} nb(y_i \mid \mu_i, \phi\; \omega_i)\; h(\omega_i \mid \nu) \;\; p(\theta),$$

donde $p(\theta)$ es la distribución a priori, $nb(y_i \mid \mu_i, \phi\; \omega_i)$ es la función de masa de probabilidad de la distribución Binomial Negativa para la i-ésima observación, con media μ_i y dispersión modulada por el parámetro $\phi\; \omega_i$, y $h(\omega_i \mid \nu)$ es la función de densidad de la distribución de las variables latentes ω_i, que dependen del parámetro ν.

Dado que se considera que los parámetros $\boldsymbol{\beta}$, ϕ y ν son independientes, la distribución a priori conjunta puede expresarse como:

$$p(\theta) = p(\boldsymbol{\beta})\, p(\phi)\, p(\nu),$$

lo que nos permite reescribir la distribución a posteriori aumentada como:

$$p(\theta, \omega \mid y) \;\propto\; \prod_{i=1}^{n} nb(y_i \mid \mu_i, \phi\; \omega_i)\; h(\omega_i \mid \nu) \;\; p(\boldsymbol{\beta})\; p(\phi)\; p(\nu).$$

Esta expresión refleja la complejidad del modelo NB-H(μ_i, ϕ, ν), ya que la verosimilitud aumentada incorpora tanto las variables latentes como los parámetros de dispersión y mixtura.

4.2.2. Estimación de los parámetros

La estimación de los parámetros $\boldsymbol{\beta}$, ϕ y ν se realizará utilizando el método *No-U-Turn Sampler* (NUTS), implementado en Stan a través de la interfaz de PyStan. Para garantizar una adecuada identificación del modelo, se incorporaron distribuciones a priori debidamente especificadas para cada parámetro.

En el caso de los coeficientes $\boldsymbol{\beta}$, se asumió una distribución normal con media cero y desviación estándar 10, lo que permite capturar una amplia gama de posibles efectos de las covariables sobre la media, sin imponer restricciones excesivas. El parámetro ϕ, asociado a la dispersión de la distribución Binomial Negativa, fue modelado mediante una distribución gamma con parámetros de forma y escala iguales a 2, asegurando su positividad y controlando posibles sobreestimaciones. Por su parte, el parámetro ν, que modula la forma de las colas, fue asignado a una distribución exponencial con media igual a 0.5 (equivalente a una tasa de 2), favoreciendo valores acotados sin excluir casos extremos plausibles.

Estas elecciones reflejan un compromiso entre flexibilidad y control, permitiendo que la información contenida en los datos sea la principal guía del proceso inferencial, al mismo tiempo que se estabilizan los componentes más sensibles del modelo. Así, se garantiza una estimación más robusta y confiable, asegurando que los resultados reflejen adecuadamente los patrones observados en los datos.

4.2.3. Criterios de Comparación de Modelos

Para evaluar y comparar el desempeño de la estimación del modelo mediante el enfoque bayesiano, existen varios criterios. En esta tesis, utilizaremos dos criterios de comparación ampliamente reconocidos: Deviance Information Criterion (DIC) y Widely Applicable Information Criterion (WAIC).

El criterio DIC es utilizado para comparar modelos bayesianos y combina la devianza con una penalización por la complejidad del modelo. Este criterio es útil para la selección del modelo porque optimiza el equilibrio entre la exactitud de los datos y la simplicidad del modelo. La devianza se define de esta manera:

$$D(\theta) = -2\log(L(y, \omega \mid \theta)) \ ,$$

donde $\theta = (\boldsymbol{\beta}^\top, \phi, \nu)^\top$ y $L(y, \omega \mid \theta)$ es la función de verosimilitud aumentada descrita anteriormente. Luego, definimos el criterio DIC:

$$\text{DIC} = D(\bar{\theta}) + 2\ p_D\ ,$$

donde $p_D = \overline{D(\theta)} - D(\bar{\theta})$ y $\bar{\theta}$ representa la esperanza de θ. Teniendo un tamaño de muestra M para las simulaciones de Montecarlo para la distribución a posteriori aumentada, los términos $\overline{D(\theta)}$ y $\bar{\theta}$ se calculan de la siguiente manera:

$$\overline{D(\theta)} = \sum_{j=1}^{M} \frac{D(\theta_j)}{M}\ ,$$

$$\bar{\theta} = \sum_{j=1}^{M} \frac{\theta_j}{M}\ .$$

El criterio WAIC también estima el ajuste y la capacidad predictiva del modelo mientras ajusta por la complejidad del mismo. Se calcula similarmente, pero, a diferencia del DIC, el WAIC incorpora la varianza del logaritmo de la probabilidad Su fórmula es la siguiente:

$$\text{WAIC} = D(\bar{\theta}) + 2\ p_{\text{WAIC}}\ ,$$

donde $p_{\text{WAIC}} = \sum_{i=1}^{n} \text{Var}(\log(p((Y_i \mid \theta)))$.

Ambos criterios expuestos permitirán seleccionar el modelo más adecuado para los datos, teniendo en cuenta el ajuste y la simplicidad del modelo. Estos criterios proporcionan una base sólida para la comparación de modelos en el ámbito de la inferencia bayesiana. Adicionalmente, es importante mencionar que mientras menor sea el valor del DIC y el WAIC, mejor será el ajuste y la parsimonia del modelo. Por tanto, en esta tesis se considerará como el mejor modelo al que tenga el menor valor de estos indicadores, pues será el que presente el mejor balance entre ajuste y simplicidad.

Capítulo 5

Estudio de Simulación

En este capítulo se presenta un estudio de simulación orientado a evaluar el desempeño de la Regresión Binomial Negativa con Mixtura en la Dispersión, la cual fue presentada en el Capítulo 4. El análisis se centra en estudiar la robustez de las estimaciones de los parámetros ante distintos patrones de contaminación en los datos, entendida esta como la introducción artificial de valores atípicos en las respuestas simuladas. Para ello, se compararán sus resultados con los del modelo Binomial Negativa estándar.

5.1. Generación de los datos

Como punto de partida para la evaluación de los modelos propuestos, se elaboró un conjunto de datos simulados. El conjunto simulado consta de 100 observaciones y se construyó a partir de una covariable continua x y una variable de respuesta y.

Para representar los efectos fijos, se construyó una matriz de diseño basada en la variable $x = [x_1, \ldots, x_{100}]^\top$, cuyos valores fueron generados de forma independiente según una distribución uniforme sobre el intervalo $[0, 1]$. Esta elección busca garantizar una cobertura equilibrada del dominio de la covariable.

La variable de respuesta y_i fue generada en función de una media μ_i, especificada mediante la relación $\mu_i = \exp(\beta_0 + \beta_1 x_i)$, donde los coeficientes se fijaron en $\beta_0 = 1$ y $\beta_1 = 3$. Este vínculo asegura que los valores esperados de y aumenten conforme lo hace x.

A partir de los valores de μ_i, se generaron las respuestas y_i de acuerdo con una distribución Binomial Negativa parametrizada por la media μ_i y un parámetro de dispersión $\phi = 0.25$. De este modo, se tiene:

$$y_i \sim \text{NB}(\mu_i, \phi).$$

En la Figura 9 se muestra la nube de puntos del escenario sin contaminación y las curvas ajustadas correspondientes a los modelos NB, NB-G y NB-IG. La superposición de las tres curvas sugiere que todos los modelos ofrecen estimaciones consistentes de la relación entre x y y en ausencia de datos atípicos.

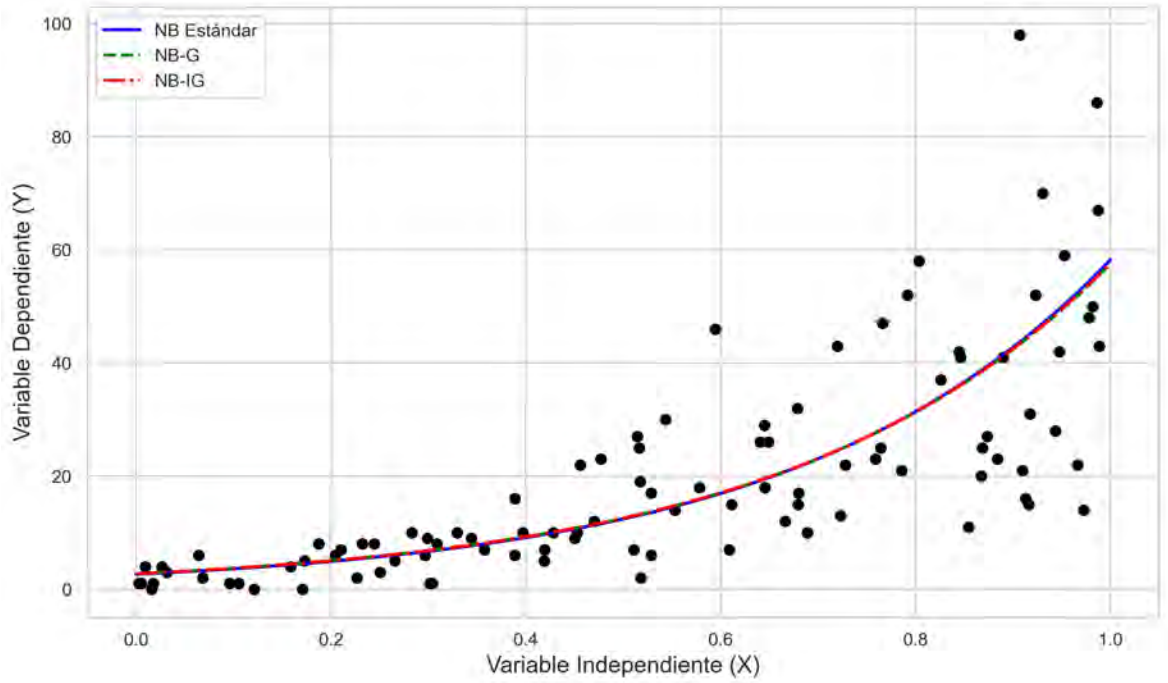

Figura 9: Curvas ajustadas por los modelos NB, NB-G y NB-IG sobre los datos simulados sin contaminación.

A continuación se presentan las estimaciones posteriores de los parámetros correspondientes a los modelos NB, NB-G y NB-IG. Tal como se muestra en la Tabla 1, los valores obtenidos para β_0, β_1 y ϕ resultan prácticamente iguales entre los tres enfoques. Esta similitud en los valores estimados es consistente con lo observado en la representación gráfica previa, en la que las curvas ajustadas por cada modelo coincidían visualmente a lo largo del dominio de la covariable. En conjunto, los resultados confirman que, en ausencia de valores atípicos, los tres enfoques generan estimaciones muy similares para los parámetros del modelo.

Tabla 1: Media posterior e intervalo de credibilidad bayesiano al 95 % para los modelos NB, NB-G y NB-IG.

	NB			**NB-G**			**NB-IG**		
Parámetro	**Media**	**2.5 %**	**97.5 %**	**Media**	**2.5 %**	**97.5 %**	**Media**	**2.5 %**	**97.5 %**
β_0	0.98	0.70	1.26	1.01	0.73	1.29	1.01	0.73	1.29
β_1	3.09	2.67	3.51	3.04	2.63	3.47	3.05	2.62	3.47
ϕ	0.26	0.17	0.37	0.30	0.19	0.46	0.32	0.20	0.49
ν	–	–	–	0.74	0.03	2.35	0.63	0.02	2.19

5.2. Escenarios de Contaminación

Con el fin de estudiar la robustez de los modelos frente a valores atípicos, se definieron tres escenarios de contaminación. En cada uno, se seleccionaron cuatro observaciones y se les incrementó de forma artificial su respuesta y_i, sin alterar los valores valores originales de x_i. Cada uno de los escenarios se distinguen según la ubicación de las observaciones modificadas:

- **Escenario 1**: se alteran las observaciones con los valores más bajos de x.
- **Escenario 2**: se perturban aquellas con valores cercanos a la mediana de x.
- **Escenario 3**: se modifican las observaciones con los valores más altos de x.

La Figura 10 muestra los Escenarios 1 y 2, en los que se modificaron cuatro observaciones con valores bajos y centrales de la variable independiente x, respectivamente. Estas alteraciones generan puntos que se apartan del comportamiento general de los datos, y han sido señalados con flechas para facilitar su identificación. La Figura 11 corresponde al Escenario 3, en el que las observaciones intervenidas se ubican en el extremo superior de x. Estas visualizaciones permiten establecer un punto de partida claro para evaluar cómo responde cada modelo ante distintos patrones de contaminación.

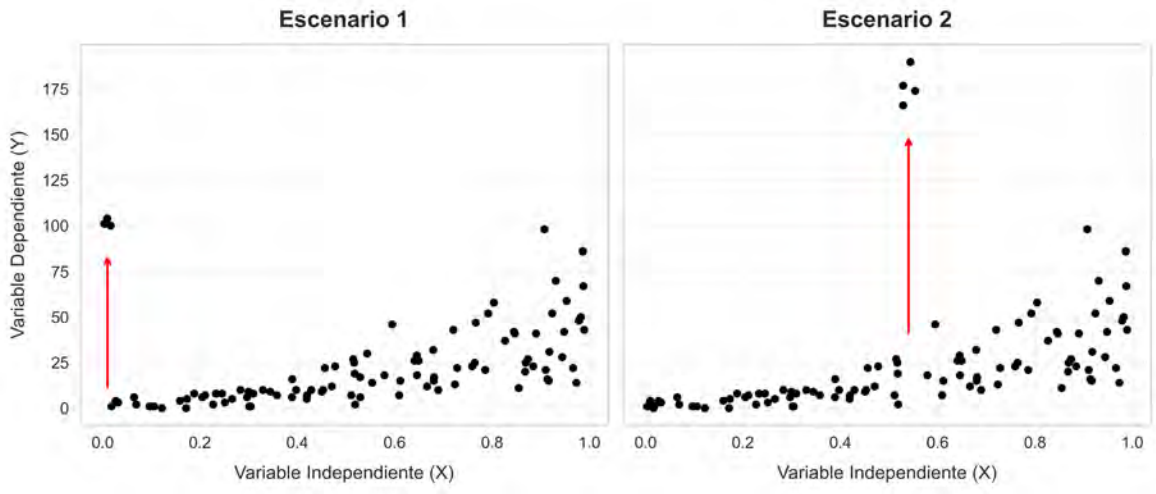

Figura 10: Escenarios 1 y 2: Visualización de la contaminación tras modificar observaciones con valores bajos y centrales de x.

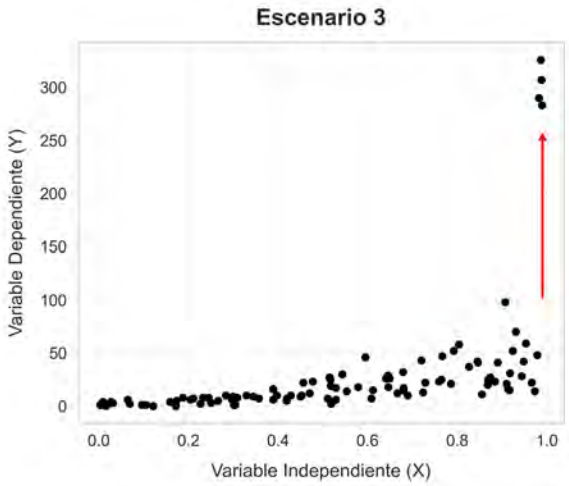

Figura 11: Escenario 3: Visualización de la contaminación tras modificar observaciones con valores altos de x.

5.3. Evaluación del Desempeño de los Modelos

Para cada escenario de contaminación y nivel de perturbación considerado, se ajustaron los tres modelos de regresión: NB, NB-G y NB-IG. El objetivo fue examinar cómo varían las estimaciones de los parámetros β_0, β_1 y ϕ a medida que se incrementa gradualmente la severidad de la contaminación.

Las Figuras 12, 13 y 14 muestran la evolución de las estimaciones puntuales y sus intervalos de credibilidad al 95 % para cada parámetro. En cada figura, las filas representan los modelos considerados (NB, NB-G y NB-IG), mientras que las columnas reflejan los tres escenarios de contaminación: (i) Escenario 1, (ii) Escenario 2, y (iii) Escenario 3. El eje horizontal representa el nivel de perturbación aplicado (Δ), mientras que el eje vertical muestra el valor estimado del parámetro correspondiente.

En los resultados obtenidos para los tres escenarios de contaminación, se evidencia que los modelos NB-G y NB-IG preservan una mayor estabilidad en sus estimaciones conforme aumenta el nivel de perturbación aplicado (Δ). En contraste, el modelo NB presenta una mayor sensibilidad frente a las observaciones alteradas. Esta diferencia resalta la ventaja de las distribuciones NB-G y NB-IG para realizar inferencias más consistentes ante la presencia de valores atípicos.

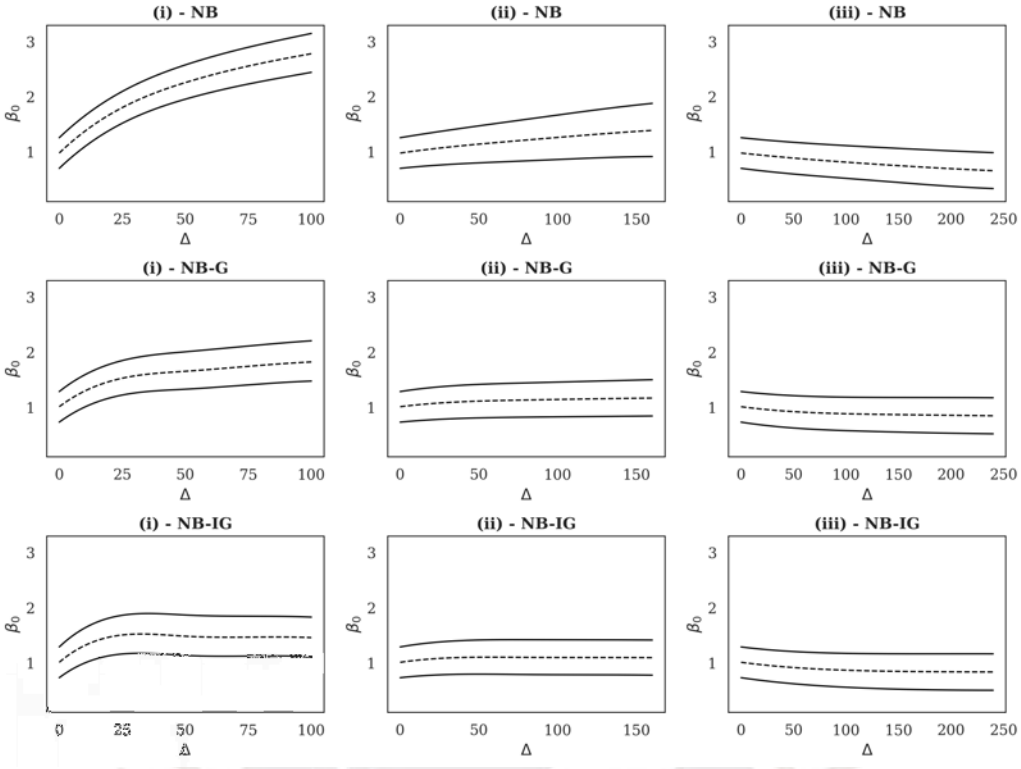

Figura 12: Estimaciones de β_0 para los modelos NB, NB-G y NB-IG bajo distintos niveles de contaminación.

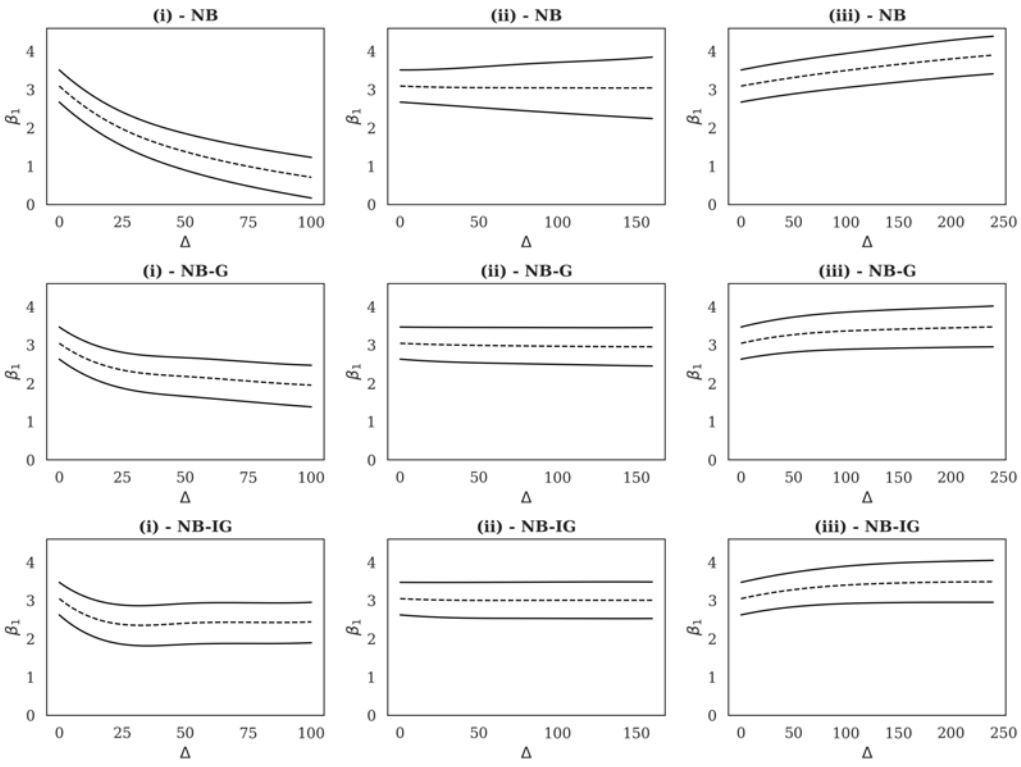

Figura 13: Estimaciones de β_1 para los modelos NB, NB-G y NB-IG bajo distintos niveles de contaminación.

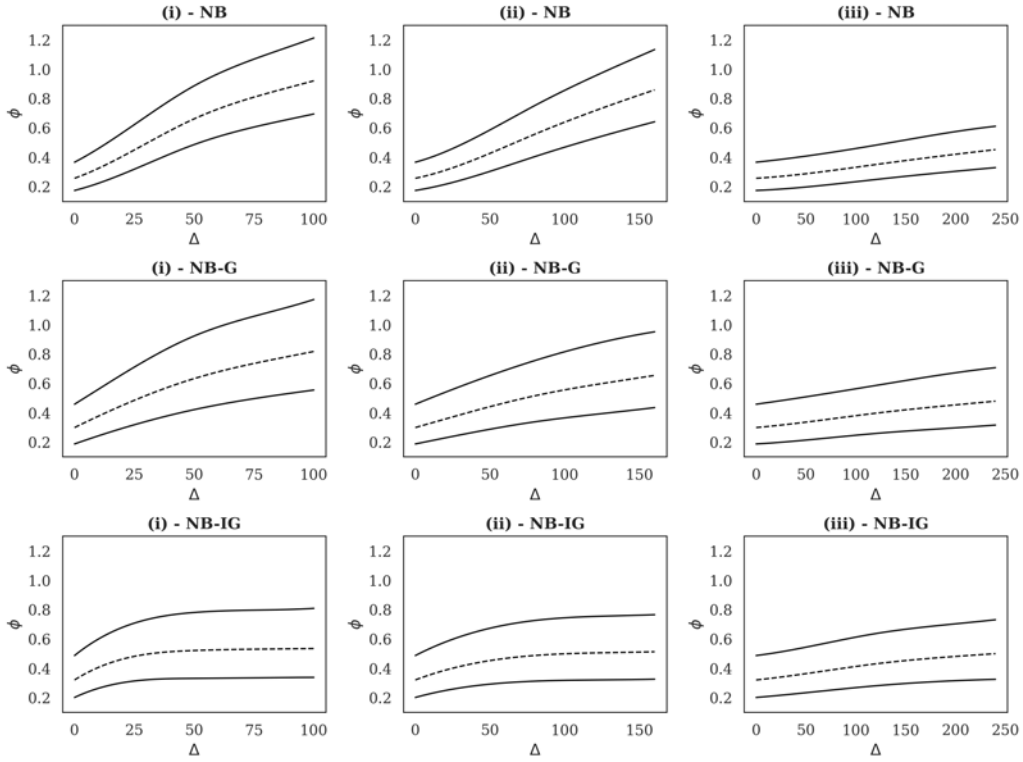

Figura 14: Estimaciones de ϕ para los modelos NB, NB-G y NB-IG bajo distintos niveles de contaminación.

En las Figuras 15 y 17, se presentan las trazas MCMC de los parámetros del modelo NB-G y NB-IG, respectivamente, bajo el Escenario 1, donde los datos han sido completamente contaminados mediante la modificación de las observaciones con los valores más bajos de la covariable x. Se observa una adecuada mezcla y estabilidad de las cadenas en todos los casos, lo cual sugiere una buena convergencia del algoritmo implementado en PyStan. Además, se observó un comportamiento similar en los demás escenarios analizados, lo que indica que el algoritmo mostró una buena convergencia de manera consistente en todos los casos considerados.

Asimismo, las Figuras 16 y 18 muestran las densidades posteriores que se obtuvieron para los parámetros de los modelos NB-G y NB-IG en este mismo escenario. Estas distribuciones son unimodales, con concentraciones razonables en torno a sus respectivas medias posteriores. Estos resultados reafirman la estabilidad de las estimaciones y la capacidad de los modelos NB-G y NB-IG para proporcionar inferencias consistentes incluso en contextos marcados por la presencia de valores atípicos.

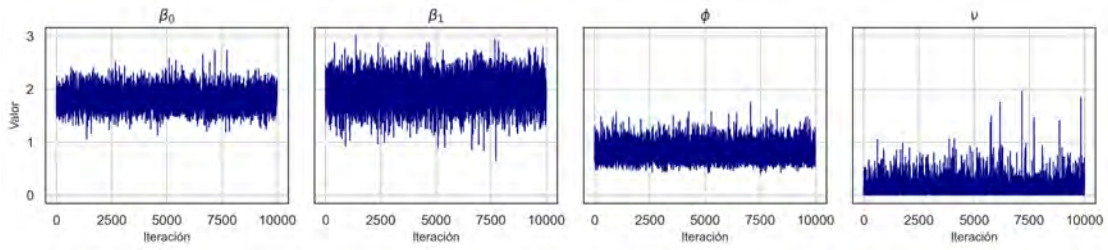

Figura 15: Trazas MCMC de los parámetros del modelo NB-G en el Escenario 1.

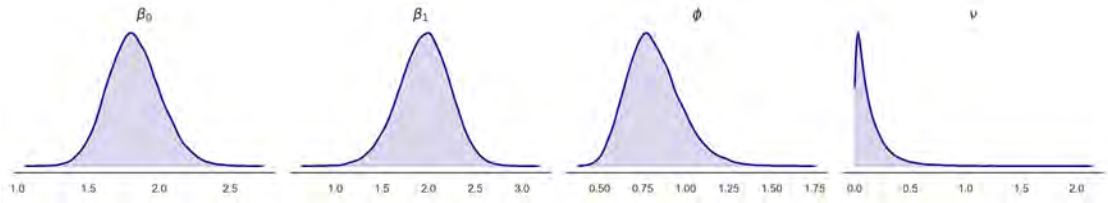

Figura 16: Densidades posteriores de los parámetros del modelo NB-G en el Escenario 1.

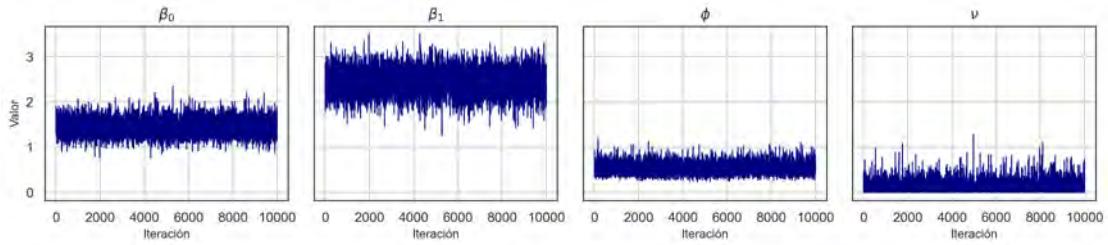

Figura 17: Trazas MCMC de los parámetros del modelo NB-IG en el Escenario 1.

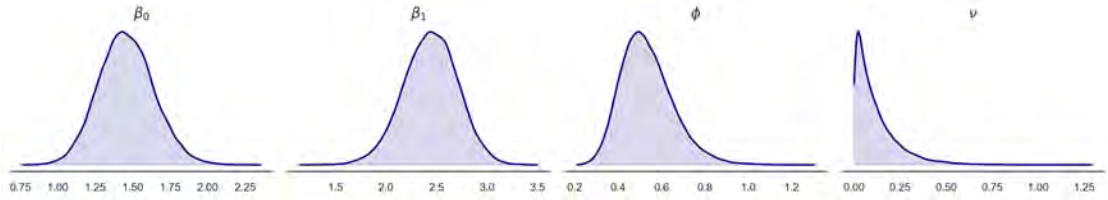

Figura 18: Densidades posteriores de los parámetros del modelo NB-IG en el Escenario 1.

Capítulo 6

Aplicaciones

6.1. Aplicación 1 - Número de visitas a consultorios médicos

6.1.1. Descripción de la base de datos

La primera aplicación utiliza la base de datos *NMES1988*, incluida en la librería *AER* de *R*, la cual ha sido empleada en numerosas investigaciones enfocadas en la demanda de servicios médicos. Este conjunto de datos proporciona información detallada sobre características sociodemográficas, condiciones de salud y uso de servicios médicos. Un análisis previo de estos datos fue realizado por Stasinopoulos et al. (2018), quienes estudiaron la relación entre diversas variables y el uso de servicios médicos. Siguiendo esta línea de investigación, se empleará un subconjunto específico de 343 registros correspondientes a individuos que declararon gozar de un estado de salud excelente. Este grupo presenta observaciones atípicas, lo que permite evaluar con mayor precisión la robustez de los modelos propuestos.

6.1.2. Especificación del modelo

El objetivo de este análisis es modelar la variable de respuesta y (*visits*), que representa el número de visitas a consultorios médicos utilizando modelos de regresión para datos de conteo. Las variables explicativas consideradas son las siguientes:

- **Condiciones crónicas** (*chronic*): Número de afecciones crónicas diagnosticadas.
- **Hospitalizaciones** (*hospital*): Frecuencia de hospitalizaciones.
- **Seguro privado** (*insurance*): Indica si el individuo cuenta con cobertura privada.
- **Género** (*gender*): Masculino o femenino.
- **Años de educación** (*school*): Años de educación formal recibidos.

En el análisis, se emplearán los modelos de regresión NB, NB-G y NB-IG, prestando atención a su desempeño frente a la presencia de datos atípicos. La estimación de parámetros

se realizará mediante PyStan, y la comparación de modelos se llevará a cabo utilizando los criterios DIC y WAIC.

El vector de coeficientes $\boldsymbol{\beta}$ estimado para el modelo puede expresarse como:

$$\boldsymbol{\beta}^\top = \begin{pmatrix} \beta_0 & \beta_1 & \beta_2 & \beta_3 & \beta_4 & \beta_5 \end{pmatrix}$$

donde cada parámetro β_i representa el efecto asociado a las siguientes variables explicativas:

- β_0: Intercepto.
- β_1: Condiciones crónicas.
- β_2: Hospitalizaciones.
- β_3: Seguro privado.
- β_4: Género.
- β_5: Años de educación.

6.1.3. Resultados de la Aplicación 1

A continuación, se presentan los resultados de los parámetros estimados para los tres modelos: NB, NB-G y NB-IG. En todos los casos, se aplicaron simulaciones MCMC utilizando la librería PyStan. Para cada modelo, se ejecutaron 4 cadenas independientes con 11000 iteraciones cada una, de las cuales las primeras 1000 se consideraron como periodo de calentamiento (warm-up). Esto resulta en un total de 40000 muestras posteriores por parámetro, tras descartar la fase inicial de adaptación.

La Tabla 2 resume las estimaciones puntuales y los intervalos de credibilidad al 95 % para los parámetros. Asimismo, se reportan los valores de los criterios de información DIC y WAIC, los cuales permiten comparar el ajuste y la capacidad predictiva de los modelos considerados.

Además, se presentan las gráficas de trazas y de distribuciones posteriores correspondientes a los modelos NB-G y NB-IG (Figuras 19–22). Estas representaciones facilitan la verificación visual de la convergencia de las cadenas MCMC y permiten explorar la estructura de las distribuciones posteriores, aportando evidencia adicional sobre la estabilidad y precisión de las estimaciones obtenidas.

Tabla 2: Resumen de parámetros y criterios de información (DIC y WAIC) para los modelos NB, NB-G y NB-IG en la Aplicación 1.

	NB			**NB-G**			**NB-IG**		
Parámetro	**Media**	**2.5 %**	**97.5 %**	**Media**	**2.5 %**	**97.5 %**	**Media**	**2.5 %**	**97.5 %**
β_0	-0.055	-0.499	0.389	-0.152	-0.588	0.289	-0.145	-0.582	0.294
β_1	0.307	0.175	0.442	0.279	0.155	0.409	0.288	0.164	0.413
β_2	0.353	0.025	0.703	0.407	0.102	0.722	0.419	0.117	0.735
β_3	0.497	0.148	0.842	0.474	0.124	0.820	0.446	0.098	0.789
β_4	0.077	-0.164	0.320	0.043	-0.193	0.279	0.011	-0.225	0.245
β_5	0.043	0.011	0.075	0.052	0.021	0.083	0.053	0.023	0.084
ϕ	0.900	0.712	1.121	0.951	0.702	1.253	0.945	0.696	1.247
ν	–	–	–	0.268	0.006	1.002	0.321	0.008	1.182
DIC	1589.50	–	–	1549.94	–	–	1553.73	–	–
WAIC	1591.34	–	–	1560.64	–	–	1561.21	–	–

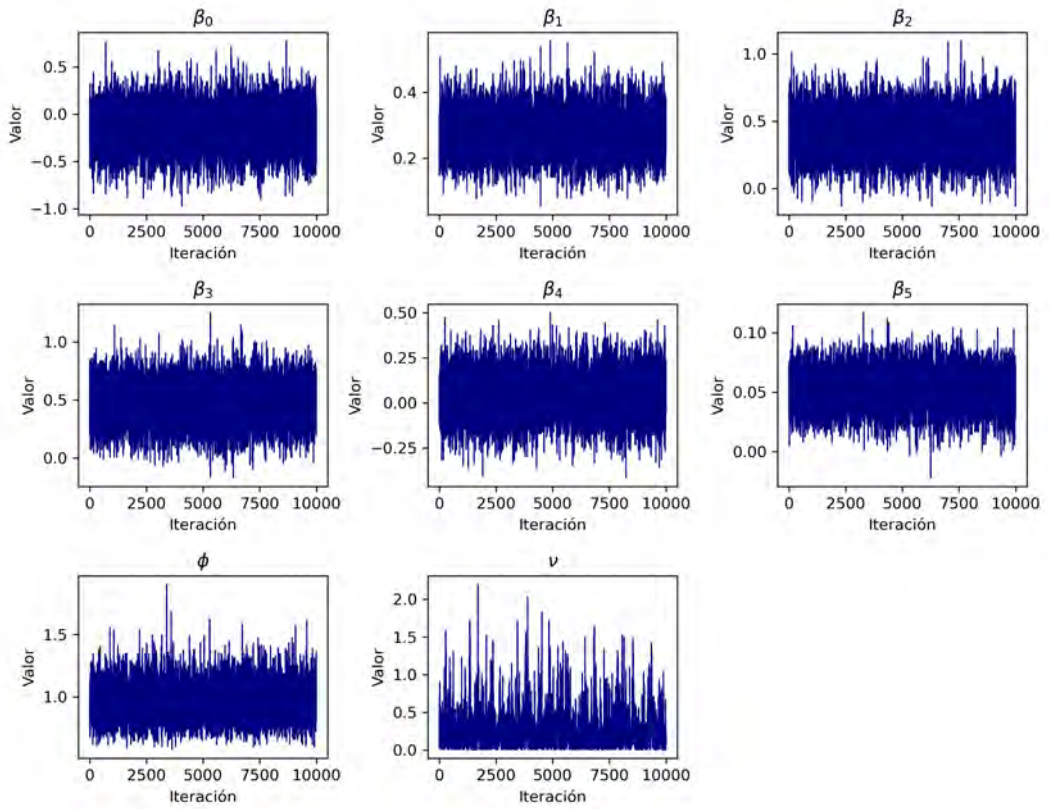

Figura 19: Trazas MCMC para los parámetros del modelo NB-G en la Aplicación 1.

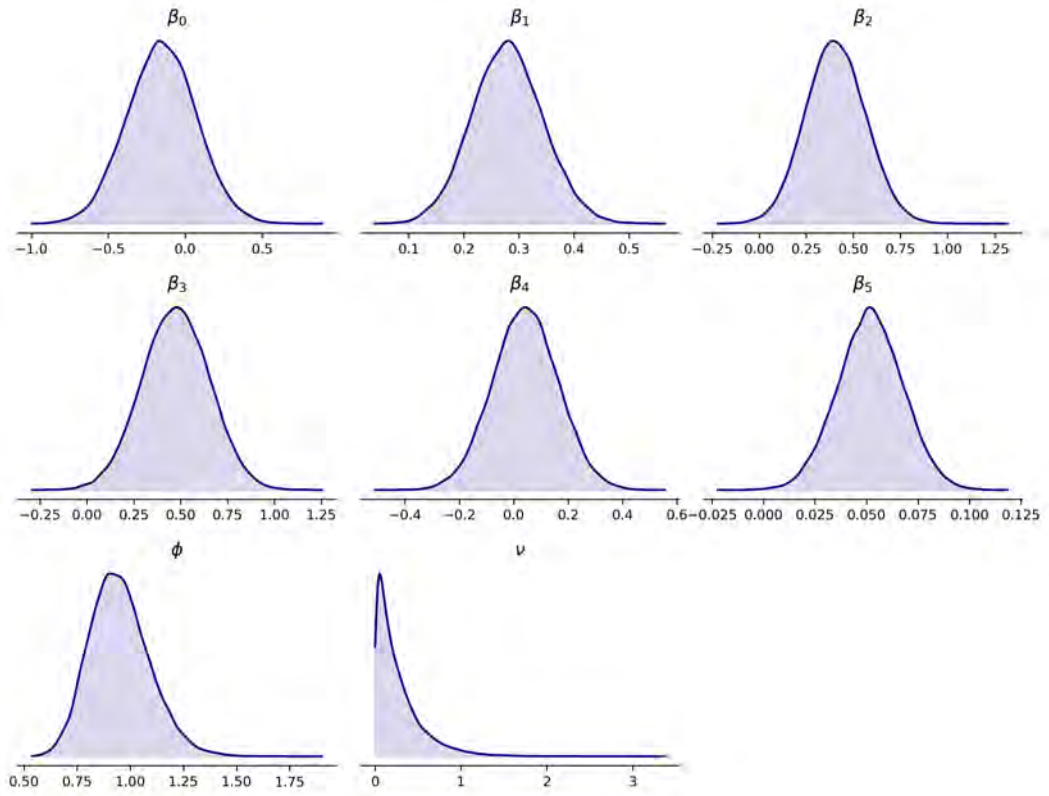

Figura 20: Densidades posteriores de los parámetros del modelo NB-G en la Aplicación 1.

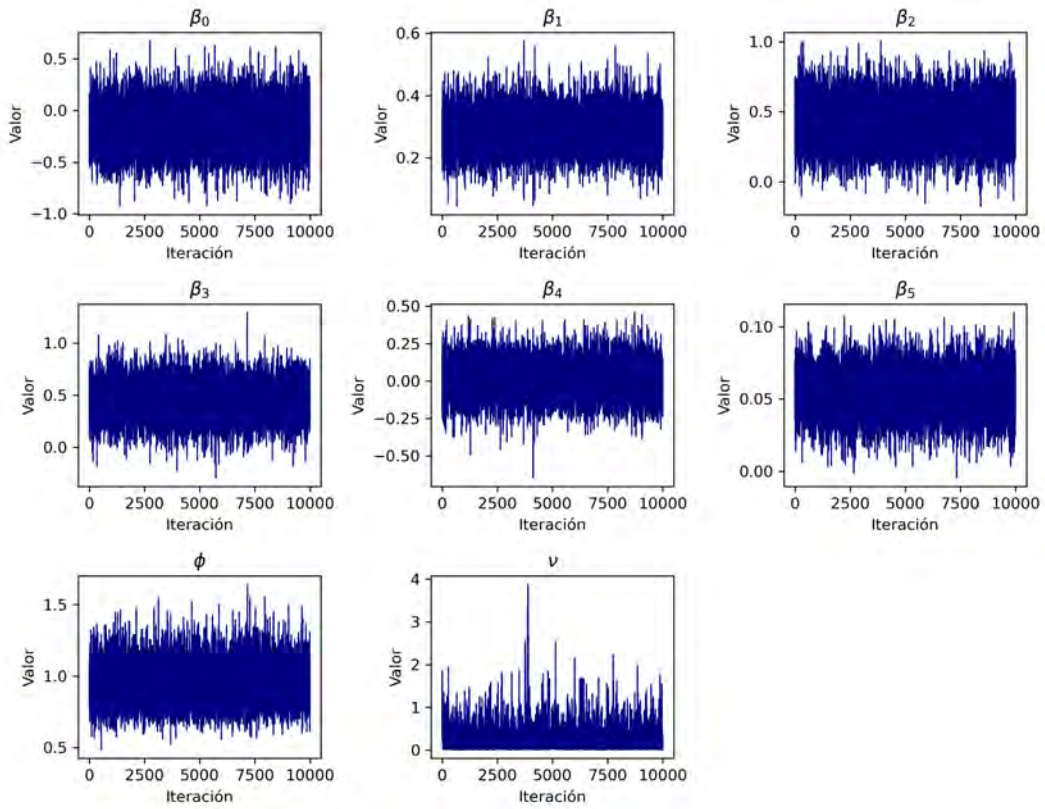

Figura 21: Trazas MCMC para los parámetros del modelo NB-IG en la Aplicación 1.

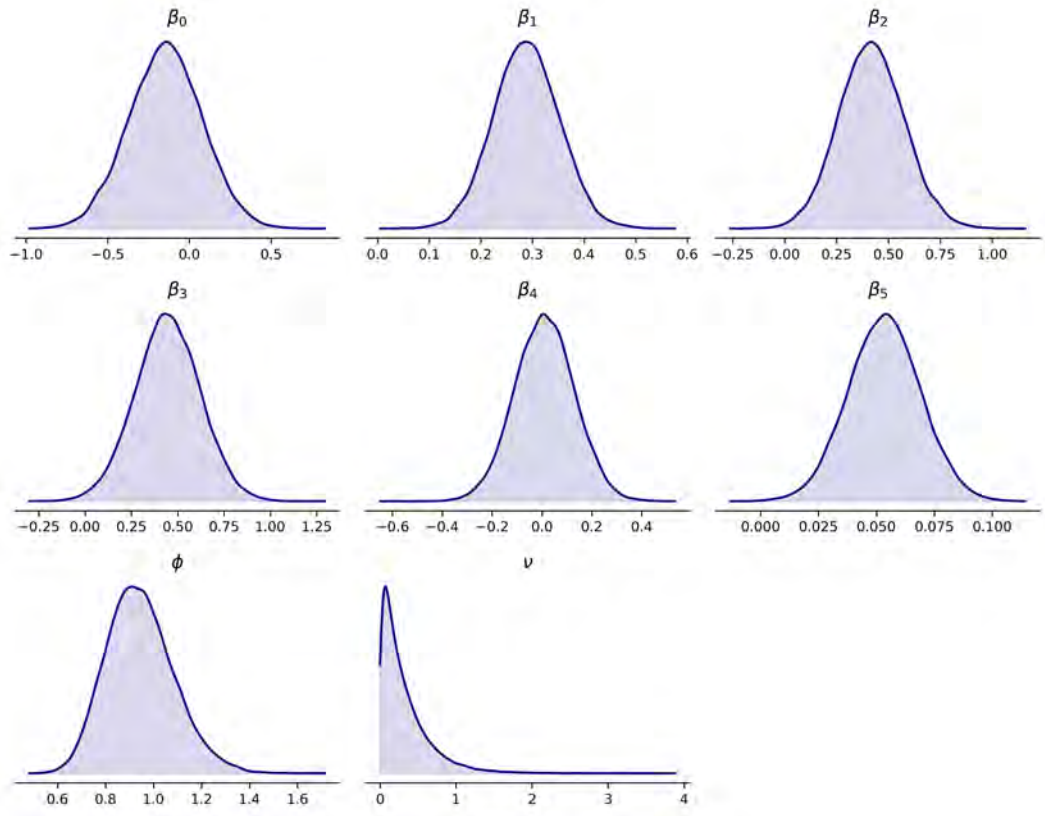

Figura 22: Densidades posteriores de los parámetros del modelo NB-IG en la Aplicación 1.

La Tabla 2 presenta un resumen detallado de las estimaciones obtenidas para los modelos NB, NB-G y NB-IG. En todos los casos, se aprecia que los coeficientes β_1 (Condiciones crónicas) β_2 (Hospitalizaciones) β_3 (Seguro privado) y β_5 (Años de educación) presentan intervalos de credibilidad al 95 % que no contienen al valor cero, lo cual indica evidencia suficiente para considerar que estas variables explicativas influyen de forma significativa en la cantidad de visitas a consultorios médicos. Por otro lado, se observa que los parámetros β_0 (intercepto) y β_4 (género) no muestran evidencia suficiente de significancia estadística, dado que sus respectivos intervalos de credibilidad al 95 % incluyen al valor cero. Esto sugiere que, en el contexto de los modelos considerados, no se puede afirmar con suficiente respaldo estadístico que estas covariables tengan un efecto relevante sobre el número de visitas a consultorios médicos.

En las Figuras 19 y 21, que muestran las cadenas MCMC de los modelos NB-G y NB-IG, se aprecia un comportamiento estable y sin tendencias. Esta dinámica sugiere una adecuada convergencia de las cadenas, lo cual respalda la validez y confiabilidad de las simulaciones obtenidas. Las densidades posteriores, mostradas en las Figuras 20 y 22, permiten visualizar la distribución de probabilidad de cada parámetro, así como cuantificar de manera precisa la incertidumbre asociada a las estimaciones obtenidas.

En relación con los criterios de evaluación, los resultados presentados en la Tabla 2 evidencian que tanto el DIC como el WAIC favorecen el desempeño de los modelos con mixtura en la dispersión, en comparación con el modelo de Regresión Binomial Negativa estándar. El modelo NB-G exhibe el valor más bajo de DIC, seguido muy de cerca por el NB-IG, mientras que el modelo NB presenta una penalización considerablemente mayor, lo que sugiere un ajuste menos adecuado. Este patrón también se verifica en los valores del WAIC, donde los modelos NB-G y NB-IG obtienen puntuaciones inferiores, lo que respalda su capacidad para modelar datos con posibles valores atípicos de manera más robusta.

En conjunto, los resultados obtenidos respaldan la utilidad de los modelos de regresión con mixtura en la dispersión como enfoques sólidos y adaptables para el análisis de datos de conteo. Su capacidad para capturar adecuadamente la variabilidad inherente a los datos, incluso en presencia de observaciones extremas, los convierte en herramientas especialmente valiosas en aplicaciones reales donde tales escenarios deben ser explícitamente considerados.

6.2. Aplicación 2 - Tiempo de hospitalización en días

6.2.1. Descripción de la base de datos

En esta segunda aplicación, se empleará la base de datos *azpro*, disponible en la librería *COUNT* de *R*. Este conjunto de datos recoge información clínica y demográfica de 3589 hospitalizaciones por afecciones cardíacas. Particularmente, se incluyen datos sobre el tipo de procedimiento realizado, el sexo del paciente, la modalidad de admisión hospitalaria y un indicador de edad avanzada. El análisis se centrará en la variable *los* (length of stay), que mide el número de días de hospitalización de cada paciente. Debido a la posible presencia de alta dispersión y valores extremos, esta base de datos representa un escenario propicio para evaluar el desempeño de los modelos de regresión propuestos.

6.2.2. Especificación del modelo

El objetivo de este análisis es modelar la variable respuesta y (*los*), correspondiente al número de días que un paciente permanece hospitalizado. Dado que se trata de una variable de conteo, se considerarán modelos de regresión para datos discretos, incorporando también la posibilidad de sobredispersión y observaciones atípicas.

Cabe señalar que, para adecuar el soporte de la variable respuesta a las propiedades de la distribución Binomial Negativa, se realizará una transformación previa consistente en restar una unidad al número original de días de hospitalización. De este modo, el rango de y comienza en cero, lo cual es coherente con las características del modelo empleado.

Las variables explicativas consideradas en el modelo son:

- **Sexo** (*sex*): Hombre o mujer.
- **Tipo de admisión hospitalaria** (*admit*): Ingreso programado o de emergencia.
- **Procedimiento médico recibido** (*procedure*): Angioplastia o cirugía de revascularización coronaria.
- **Edad mayor a 75 años** (*age75*): Indicador de si el paciente supera los 75 años.

La estimación de los parámetros del modelo será realizada a través de simulaciones MCMC utilizando *PyStan*. Para evaluar el ajuste y la capacidad predictiva de los modelos considerados, se emplearán los criterios de información DIC y WAIC.

El vector de parámetros $\boldsymbol{\beta}$ se define como:

$$\boldsymbol{\beta}^\top = \begin{pmatrix} \beta_0 & \beta_1 & \beta_2 & \beta_3 & \beta_4 \end{pmatrix}$$

donde cada componente del vector representa el efecto de las covariables sobre el tiempo de hospitalización:

- β_0: Intercepto.
- β_1: Efecto del sexo del paciente.
- β_2: Efecto del tipo de admisión hospitalaria.
- β_3: Efecto del tipo de procedimiento médico realizado.
- β_4: Efecto de tener más de 75 años.

6.2.3. Resultados de la Aplicación 2

En esta sección se presentan los resultados de la estimación de parámetros para los modelos NB, NB-G y NB-IG aplicados al número de días de hospitalización. Para cada modelo, se realizaron simulaciones MCMC con cuatro cadenas independientes, cada una con 11000 iteraciones, de las cuales las primeras 1000 fueron descartadas como periodo de calentamiento.

La Tabla 3 resume las estimaciones puntuales obtenidas para cada modelo, junto con los intervalos de credibilidad al 95 %. Además, se incluyen los valores de los criterios de información DIC y WAIC, los cuales permiten comparar el ajuste y la capacidad predictiva de los modelos.

Asimismo, desde la Figura 23 hasta la Figura 26 se presentan las gráficas de trazas y de distribuciones posteriores correspondientes a los modelos NB-G y NB-IG. Estas representaciones gráficas resultan útiles para inspeccionar la estabilidad de las cadenas MCMC, verificar su convergencia y explorar la forma de las distribuciones posteriores de los parámetros.

Tabla 3: Resumen de parámetros y criterios de información (DIC y WAIC) para los modelos NB, NB-G y NB-IG en la Aplicación 2.

	NB			**NB-G**			**NB-IG**		
Parámetro	**Media**	**2.5 %**	**97.5 %**	**Media**	**2.5 %**	**97.5 %**	**Media**	**2.5 %**	**97.5 %**
β_0	1.145	1.090	1.201	1.116	1.063	1.169	1.107	1.054	1.160
β_1	-0.147	-0.191	-0.103	-0.141	-0.181	-0.100	-0.145	-0.185	-0.104
β_2	0.447	0.402	0.491	0.423	0.382	0.463	0.430	0.389	0.471
β_3	1.139	1.096	1.181	1.120	1.080	1.161	1.133	1.092	1.174
β_4	0.137	0.091	0.184	0.142	0.099	0.186	0.146	0.102	0.190
ϕ	0.243	0.224	0.262	0.232	0.210	0.256	0.230	0.207	0.255
ν	–	–	–	0.033	0.001	0.121	0.049	0.001	0.180
DIC	19852.36	–	–	19433.70	–	–	19435.29	–	–
WAIC	19852.47	–	–	19528.73	–	–	19522.07	–	–

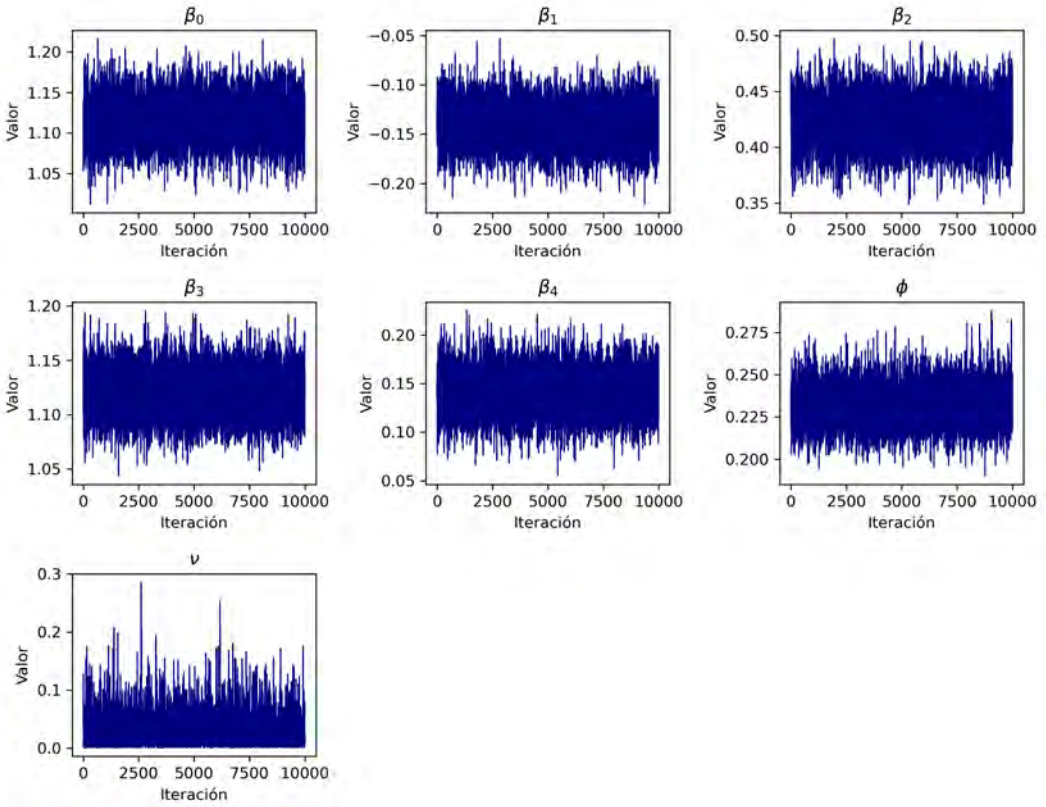

Figura 23: Trazas MCMC para los parámetros del modelo NB-G en la Aplicación 2.

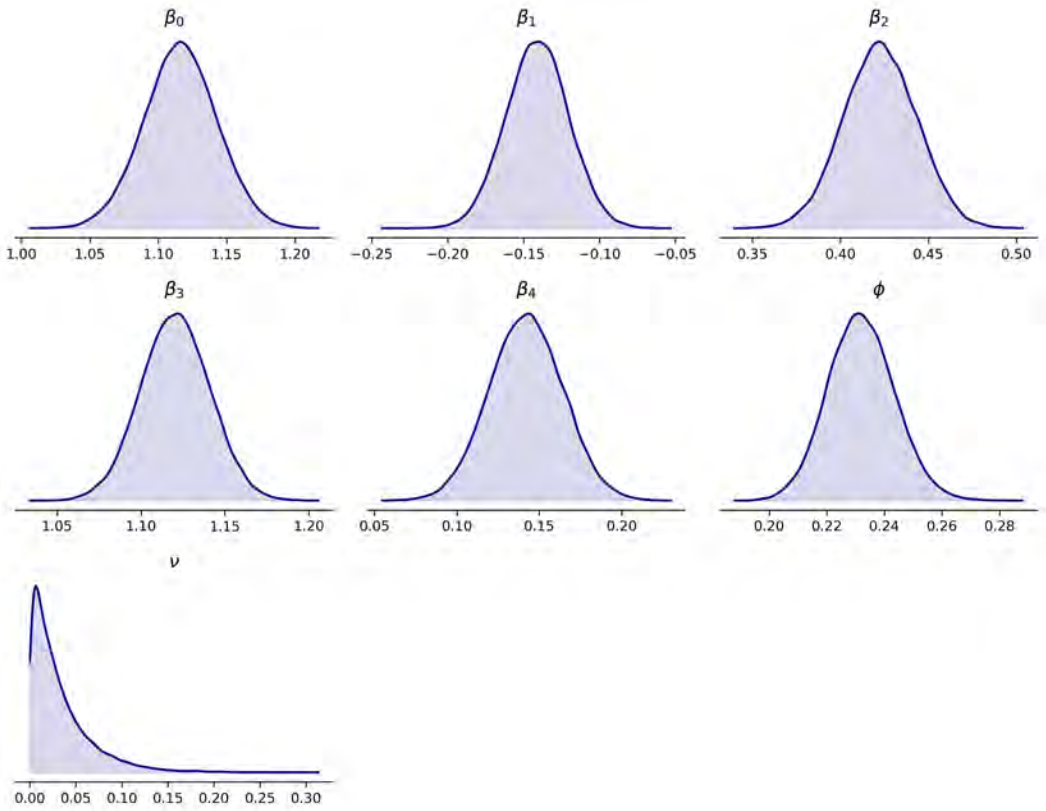

Figura 24: Densidades posteriores de los parámetros del modelo NB-G en la Aplicación 2.

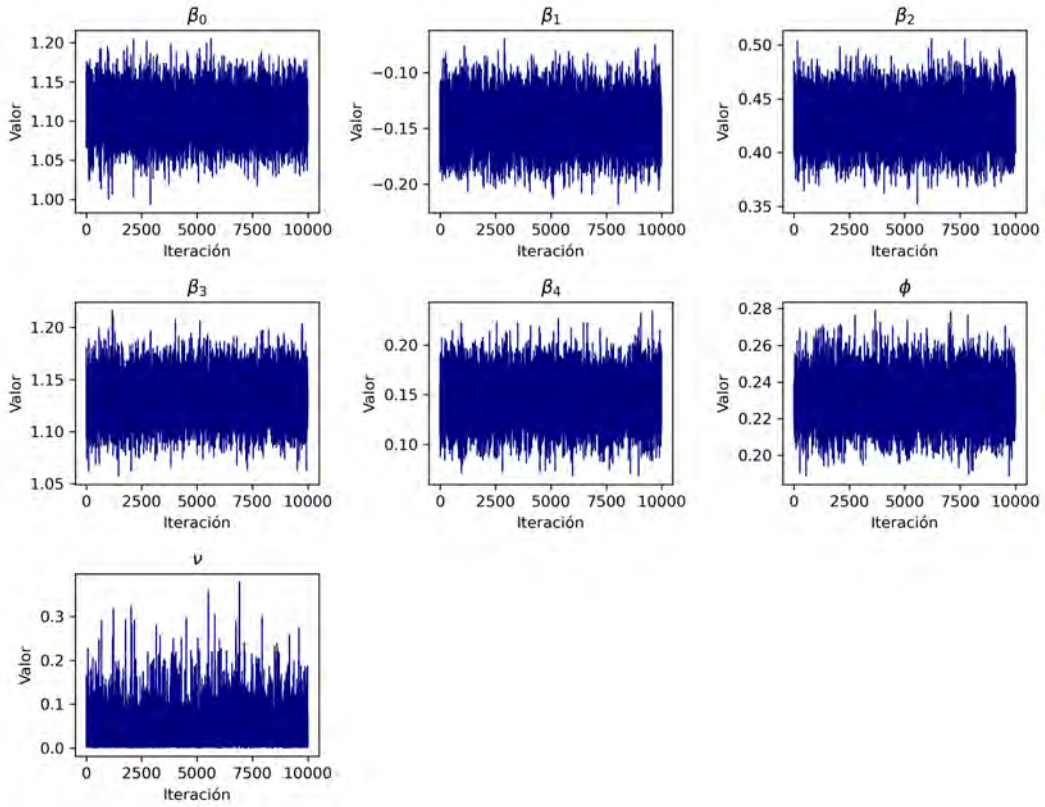

Figura 25: Trazas MCMC para los parámetros del modelo NB-IG en la Aplicación 2.

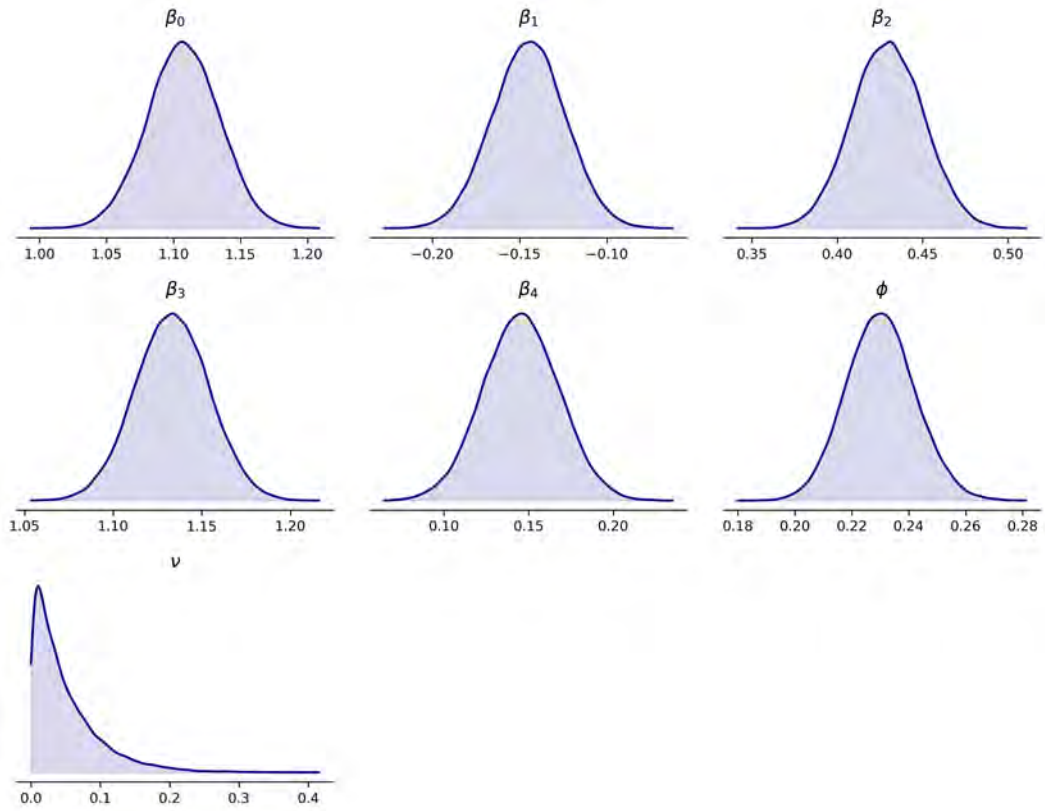

Figura 26: Densidades posteriores de los parámetros del modelo NB-IG en la Aplicación 2.

La Tabla 3 presenta un resumen de los parámetros estimados para los tres modelos en el análisis del tiempo de hospitalización. Se observa que todos los coeficientes β asociados a las variables explicativas muestran intervalos de credibilidad al 95 % que no incluyen al valor cero, lo que indica evidencia suficiente para afirmar que estas covariables tienen un efecto significativo sobre la cantidad de días de hospitalización de los pacientes.

Las Figuras 23 y 25, correspondientes a las cadenas MCMC de los modelos NB-G y NB-IG, muestran un comportamiento estable y sin tendencias aparentes, lo que sugiere que las cadenas han alcanzado convergencia satisfactoria. Asimismo, las densidades posteriores presentadas en las Figuras 24 y 26 permiten apreciar la forma de las distribuciones posteriores de los parámetros, reflejando niveles adecuados de precisión en las estimaciones.

En relación con los criterios de evaluación del ajuste, los resultados de la Tabla 3 indican que los modelos NB-G y NB-IG logran un ajuste más satisfactorio en comparación con el modelo de Regresión Binomial Negativa, tanto en DIC como en WAIC. El modelo NB-G alcanza el menor valor de DIC, seguido de cerca por el NB-IG, mientras que el modelo NB presenta un valor considerablemente más alto, lo cual sugiere un ajuste menos favorable. Una tendencia similar se observa al comparar los valores del WAIC, donde los modelos NB-G y NB-IG muestran un mejor desempeño.

En conjunto, los resultados obtenidos respaldan que los modelos con estructura de mixtura en la dispersión representan alternativas flexibles y eficaces para el análisis de datos de conteo con alta variabilidad y presencia de valores atípicos. Su capacidad de adaptarse a diversas condiciones los convierte en herramientas valiosas para aplicaciones reales donde este tipo de escenarios debe ser considerado.

Capítulo 7

Conclusiones y Recomendaciones

7.1. Conclusiones

En la presente tesis se presentó el Modelo de Regresión NB-H como una alternativa robusta para el análisis de datos de conteo con sobredispersión y presencia de valores atípicos. Con el fin de mejorar su capacidad de adaptación frente a este tipo de escenarios, se incorporó una mezcla en el parámetro de dispersión, lo que permite ajustar la forma de las colas ante observaciones extremas.

A lo largo del trabajo, se estudiaron sus principales propiedades y se realizó una evaluación exhaustiva de su desempeño. Como casos particulares de esta formulación general, se desarrollaron dos variantes: el Modelo de Regresión NB-G y el Modelo de Regresión NB-IG. El análisis de la asimetría y curtosis reveló que las distribuciones NB-G y NB-IG presentan colas más pesadas que la distribución Binomial Negativa estándar, lo que respalda su mayor capacidad para manejar observaciones atípicas.

Como parte de este análisis, se incorporó un estudio de sensibilidad que consistió en introducir alteraciones controladas en las observaciones, lo que permitió evaluar la estabilidad de las estimaciones frente a distintos niveles de perturbación. Además, se comparó el comportamiento del modelo NB frente a los modelos NB-G y NB-IG, empleando técnicas de inferencia bayesiana para evaluar su precisión y confiabilidad, tanto en escenarios simulados como ante datos reales.

Los resultados de las simulaciones mostraron que los modelos NB-G y NB-IG producen estimaciones más estables y precisas en contextos con sobredispersión y presencia de valores extremos. Estas formulaciones evidenciaron una mayor resistencia frente a la distorsión generada por observaciones atípicas, lo que se tradujo en ajustes más coherentes con la estructura subyacente de los datos.

En las aplicaciones prácticas desarrolladas utilizando las bases de datos *NMES1988* y *azpro*, los resultados obtenidos evidencian que los modelos NB-G y NB-IG presentan un mejor desempeño frente a la presencia de valores atípicos en comparación con el modelo de

Regresión Binomial Negativa. Este resultado se reflejó en valores inferiores de los criterios DIC y WAIC, lo que refuerza la utilidad de los modelos con mixtura en la dispersión para mejorar la calidad de las inferencias y la capacidad predictiva en situaciones reales donde existen observaciones extremas.

En conclusión, la presente investigación ha demostrado la utilidad de los modelos NB-H como herramientas flexibles y robustas para el análisis de datos de conteo caracterizados por sobredispersión y presencia de valores atípicos. En comparación con la regresión Binomial Negativa convencional, estas formulaciones muestran un mejor rendimiento y una mayor capacidad de adaptación frente a este tipo de escenarios. Los resultados obtenidos abren nuevas posibilidades para la incorporación de estos modelos en contextos prácticos donde los datos de conteo son prevalentes.

7.2. Recomendaciones para Estudios Futuros

A partir de los hallazgos de la presente investigación, se identifican diversas líneas que podrían abordarse en estudios futuros. En primer lugar, se sugiere explorar otras distribuciones para modelar la mixtura en el parámetro de dispersión dentro de los modelos de regresión NB-H, lo que podría aportar una mayor flexibilidad y mejorar el desempeño en distintos contextos.

Asimismo, sería valioso extender el análisis a conjuntos de datos con características distintas, con el fin de evaluar la robustez de los modelos propuestos en una gama más amplia de aplicaciones prácticas.

Finalmente, se recomienda profundizar en los aspectos computacionales relacionados con la estimación bayesiana en modelos con componentes de mezcla. En este trabajo, los tiempos de cómputo para cada aplicación fueron razonables, alrededor de cinco minutos utilizando cuatro cadenas y 11,000 iteraciones. Sin embargo, en estudios con volúmenes de datos más grandes o modelos de mayor complejidad, estos tiempos podrían incrementarse considerablemente. Por ello, la incorporación de técnicas de optimización adicionales o métodos de muestreo más eficientes constituye una línea de investigación prometedora para estudios futuros.

Bibliografía

Aeberhard, W. H., Cantoni, E., & Heritier, S. (2014). Robust Inference in the Negative Binomial Regression Model with an Application to Falls Data. *Biometrics*, *70*(4), 920-931.

Anscombe, F. J. (1950). Sampling Theory of the Negative Binomial and Logarithmic Series Distributions. *Biometrika*, *37*(3-4), 358-382. https://doi.org/10.1093/biomet/37.3-4.358

Bayes, C., Bazán, J., & Valdivieso, L. (2025). A Robust Regression Model for Count Data in Medical Research [Manuscrito sometido para publicación].

Berk, R., & MacDonald, J. M. (2008). Overdispersion and Poisson Regression. *Journal of Quantitative Criminology*, *24*(3), 269-284.

Germán-Soto, V., Gutiérrez Flores, L., & Tovar Montiel, S. H. (2009). Factores y Relevancia Geográfica del Proceso de Innovación Regional en México, 1994-2006. *Estudios Económicos*, *24*(2 (48)), 225-248.

Johnson, N., Kemp, A., & Kotz, S. (2005). Univariate Discrete Distributions (Third Edition). John Wiley & Sons.

McGree, J. M., & Eccleston, J. A. (2012). Robust Designs for Poisson Regression Models. *Technometrics*, *54*(1), 64-72.

Sellers, K. F., & Shmueli, G. (2010). A Flexible Regression Model for Count Data. *The Annals of Applied Statistics*, *4*(2), 943-961.

Stasinopoulos, M. D., Rigby, R. A., & Bastiani, F. D. (2018). GAMLSS: A Distributional Regression Approach. *Statistical Modelling*, *18*(3-4), 248-273.

Apéndice A

Códigos para la exploración de propiedades de la distribución NB-G

A.1. Generación de la función de masa de probabilidad (PMF)

El siguiente código, desarrollado en lenguaje Python, tiene como objetivo calcular de manera númerica la función de masa de probabilidad asociada a la distribución NB-G, considerando diversos valores del parámetro ν.

```
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import gamma, nbinom
from scipy.integrate import quad

# PMF de la distribución NB-G
def nb_g_pmf(y, mu, phi, nu):
    def integrand(w):
        r = 1 / (phi * w)
        p = 1 / (1 + phi * w * mu)
        return nbinom.pmf(y, r, p) * gamma.pdf(w, a=nu + 1, scale=1 / (nu + 1))

    result, _ = quad(integrand, 1e-6, 100, epsabs=1e-10)
    return result

# PMF de la Binomial Negativa Estándar
def nb_standard_pmf(y_vals, mu, phi):
    r = 1 / phi
    p = 1 / (1 + phi * mu)
    return nbinom.pmf(y_vals, r, p)
```

```
# Parámetros
mu = 50
phi = 0.25
y_vals = np.arange(0, 160)
nu_vals = [0.1, 1, 5]

# PMF - NBG - Diversos valores de nu
pmf_dict = {}
for nu in nu_vals:
    pmf_dict[f"NB-G  ={nu}"] = [nb_g_pmf(y, mu, phi, nu) for y in y_vals]

# PMF - NB - Estándar
pmf_dict["NB estándar"] = nb_standard_pmf(y_vals, mu, phi)

# Gráfico
plt.figure(figsize=(12, 8))
for label, probs in pmf_dict.items():
    plt.plot(y_vals, probs, marker='o', linestyle='-', label=label, markersize=3)

plt.xlabel("y", fontsize=16)
plt.ylabel("P(Y = y)", fontsize=16)
plt.xticks(fontsize=14)
plt.yticks(fontsize=14)
plt.legend(fontsize=14)
plt.grid(True)
plt.tight_layout()
plt.show()
```

A.2. Visualización de la asimetría de la distribución NB-G

El siguiente código, desarrollado en Python, permite calcular y graficar la asimetría (skewness) teórica de la distribución NB-G en función del parámetro ν, manteniendo fijos los valores de $\mu = 50$ y $\phi = 0.25$. La visualización resultante es la Figura 3 del documento principal.

```
import numpy as np
import matplotlib.pyplot as plt

# Parámetros fijos
mu = 50
phi = 0.25

# Rango de nu
nu_vals = np.linspace(0.05, 10, 300)
skew_vals = []

# Cálculo de la asimetría
for nu in nu_vals:
    term1 = (1 + 2 * phi * mu) / (mu**0.5 * (1 + phi * mu)**0.5)
    term2 = (2 * phi**2 * mu**1.5) / ((nu + 1) * (1 + phi * mu)**1.5)
    skewness = term1 + term2
    skew_vals.append(skewness)

# Gráfico
plt.figure(figsize=(10, 6))
plt.plot(nu_vals, skew_vals, color='blue', linewidth=2)
plt.xlabel("$\\nu$", fontsize=14)
plt.ylabel("Asimetría", fontsize=14)
plt.grid(True)
plt.tight_layout()
plt.show()
```

A.3. Visualización de la curtosis de la distribución NB-G

El siguiente código, escrito en Python, permite calcular y graficar la curtosis teórica de la distribución NB-G en función del parámetro ν, manteniendo constantes los valores de $\mu = 50$ y $\phi = 0.25$. La visualización resultante es la Figura 4 del documento principal.

```
import numpy as np
import matplotlib.pyplot as plt

# Parámetros
mu = 50
phi = 0.25

# Rango de nu
nu_vals = np.linspace(0.05, 10, 300)
kurt_vals = []

# Cálculo de la curtosis
for nu in nu_vals:
    term1 = (1 + 6 * phi * mu + 6 * (phi**2) * (mu**2)) / (mu * (1 + phi * mu))
    term2 = 6 * (phi**2) * mu * ((2 + 3 * phi * mu) / (nu + 1) + (2 * phi * mu) /
        (nu + 1)**2) / (1 + phi * mu)**2
    kurtosis = term1 + term2
    kurt_vals.append(kurtosis)

# Gráfico
plt.figure(figsize=(10, 6))
plt.plot(nu_vals, kurt_vals, color='red', linewidth=2)
plt.xlabel("$\\nu$", fontsize=14)
plt.ylabel("Curtosis", fontsize=14)
plt.grid(True)
plt.tight_layout()
plt.show()
```

Apéndice B

Códigos para la exploración de propiedades de la distribución NB-IG

B.1. Generación de la función de masa de probabilidad (PMF)

El siguiente código, implementado en `Python`, permite calcular numéricamente la función de masa de probabilidad de la distribución NB-IG para distintos valores del parámetro ν, manteniendo constantes los valores $\mu = 50$ y $\phi = 0.25$.

```python
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import invgamma, nbinom
from scipy.integrate import quad

# PMF de la distribución NB-IG
def nb_ig_pmf(y, mu, phi, nu):
    def integrand(w):
        r = 1 / (phi * w)
        p = 1 / (1 + phi * w * mu)
        return nbinom.pmf(y, r, p) * invgamma.pdf(w, a=nu + 2, scale=nu + 1)
    result, _ = quad(integrand, 1e-6, 100, epsabs=1e-10)
    return result

# PMF de la Binomial Negativa Estándar
def nb_standard_pmf(y_vals, mu, phi):
    r = 1 / phi
    p = 1 / (1 + phi * mu)
    return nbinom.pmf(y_vals, r, p)
```

```
# Parámetros
mu = 50
phi = 0.25
y_vals = np.arange(0, 230)
nu_vals = [0.1, 1, 5]

# PMF - NB-IG para distintos valores de nu
pmf_dict = {}
for nu in nu_vals:
    pmf_dict[f"NB-IG  ={nu}"] = [nb_ig_pmf(y, mu, phi, nu) for y in y_vals]

# PMF - NB estándar
pmf_dict["NB estándar"] = nb_standard_pmf(y_vals, mu, phi)

# Gráfico
plt.figure(figsize=(12, 8))
for label, probs in pmf_dict.items():
    plt.plot(y_vals, probs, marker='o', linestyle='-', label=label, markersize=3)

plt.xlabel("y", fontsize=16)
plt.ylabel("P(Y = y)", fontsize=16)
plt.xticks(fontsize=14)
plt.yticks(fontsize=14)
plt.legend(fontsize=14)
plt.grid(True)
plt.tight_layout()
plt.show()
```

B.2. Visualización de la asimetría de la distribución NB-IG

El siguiente código, implementado en Python, permite calcular la asimetría teórica de la distribución NB-IG en función del parámetro ν, manteniendo fijos los valores $\mu = 50$ y $\phi = 0.25$. La figura generada con este código corresponde a la Figura 7 del cuerpo principal del documento.

```
import numpy as np
import matplotlib.pyplot as plt

# Parámetros fijos
mu = 50
phi = 0.25

# Valores de nu
nu_vals = np.linspace(0.1, 10, 300)  # evitar división por cero
skew_vals = []

# Cálculo usando la fórmula reducida
for nu in nu_vals:
    term1 = (1 + 2 * phi * mu) / (mu**0.5 * (1 + phi * mu)**0.5)
    term2 = (2 * phi**2 * mu**1.5) / (nu * (1 + phi * mu)**1.5)
    skewness = term1 + term2
    skew_vals.append(skewness)

# Gráfico
plt.figure(figsize=(10, 6))
plt.plot(nu_vals, skew_vals, color='blue', linewidth=2)
plt.xlabel("$\\nu$", fontsize=14)
plt.ylabel("Asimetría", fontsize=14)
plt.grid(True)
plt.tight_layout()
plt.show()
```

B.3. Visualización de la curtosis de la distribución NB-IG

El siguiente código, implementado en Python, permite calcular la curtosis de la distribución NB-IG para distintos valores del parámetro ν, manteniendo constantes los valores $\mu = 50$ y $\phi = 0.25$. El gráfico generado con este código corresponde a la Figura 8 en el cuerpo principal del documento.

```
import numpy as np
import matplotlib.pyplot as plt

# Parámetros
mu = 50
phi = 0.25

# Rango de nu
nu_vals = np.linspace(0.05, 10, 300)
kurtosis_vals = []  # almacenará la curtosis calculada

# Calculo de la curtosis
for nu in nu_vals:
    num = 1 + 6*phi*mu + 6*(phi**2)*(mu**2)
    den = mu * (1 + phi*mu)
    term1 = num / den
    term2 = 6*(phi**2)*mu * ( (2 + 3*phi*mu)/(nu+1) + (2*phi*mu)/((nu+1)**2) ) /
        ((1 + phi*mu)**2)
    kurtosis_vals.append(term1 + term2)

# Gráfico
plt.figure(figsize=(10, 6))
plt.plot(nu_vals, kurtosis_vals, color='red', linewidth=2)
plt.xlabel("$\\nu$", fontsize=14)
plt.ylabel("Curtosis", fontsize=14)
plt.grid(True)
plt.tight_layout()
plt.show()
```

Apéndice C

Códigos de Stan para la Simulación

C.1. Código para el Modelo de Regresión NB

El siguiente código en Stan implementa la regresión NB:

```
data {
  int<lower=0> N;
  vector[N] x;
  int<lower=0> y[N];
}

parameters {
  real beta_0;
  real beta_1;
  real<lower=0> phi;
}

model {
  vector[N] mu;
  for (n in 1:N)
    mu[n] = exp(beta_0 + beta_1 * x[n]);
  y ~ neg_binomial_2(mu, 1/phi);
}
```

C.2. Código para el Modelo de Regresión NB-G

El siguiente código en Stan implementa la regresión NB-G:

```
data {
  int<lower=0> N;
  vector[N] x;
  int<lower=0> y[N];
}

parameters {
  real beta_0;
  real beta_1;
  real<lower=0> phi;
  real<lower=0> nu;
  vector<lower=0>[N] omega;
}

transformed parameters {
  vector[N] mu;
  mu = exp(beta_0 + beta_1 * x);
}

model {
  beta_0 ~ normal(0, 10);
  beta_1 ~ normal(0, 10);
  phi ~ gamma(2, 2);
  nu ~ exponential(2);
  omega ~ gamma(nu + 1, nu + 1);

  for (i in 1:N) {
    y[i] ~ neg_binomial_2(mu[i], 1.0 / (phi * omega[i]));
  }
}
```

C.3. Código para el Modelo de Regresión NB-IG

El siguiente código en Stan implementa la regresión NB-IG:

```
data {
  int<lower=0> N;
  vector[N] x;
  int<lower=0> y[N];
}

parameters {
  real beta_0;
  real beta_1;
  real<lower=0> phi;
  real<lower=0> nu;
  vector<lower=0>[N] omega;
}

transformed parameters {
  vector[N] mu;
  mu = exp(beta_0 + beta_1 * x);
}

model {
  beta_0 ~ normal(0, 10);
  beta_1 ~ normal(0, 10);
  phi ~ gamma(2, 2);
  nu ~ exponential(2);
  omega ~ inv_gamma(nu + 2, nu + 1);

  for (i in 1:N) {
    y[i] ~ neg_binomial_2(mu[i], 1.0 / (phi * omega[i]));
  }
}
```

Apéndice D

Códigos de Stan para las Aplicaciones

D.1. Código para el Modelo de Regresión NB

El siguiente código en Stan implementa la regresión NB aplicada con múltiples covariables:

```
data {
  int<lower=0> N;
  vector[N] x1;
  vector[N] x2;
  vector[N] x3;
  vector[N] x4;
  vector[N] x5;
  int<lower=0> y[N];
}

parameters {
  real beta_0;
  real beta_1;
  real beta_2;
  real beta_3;
  real beta_4;
  real beta_5;
  real<lower=0> phi;
}
```

```
transformed parameters {
  vector[N] mu;
  mu = exp(beta_0 +
           beta_1 * x1 +
           beta_2 * x2 +
           beta_3 * x3 +
           beta_4 * x4 +
           beta_5 * x5);
}

model {
  y ~ neg_binomial_2(mu, 1 / phi);
}

generated quantities {
  vector[N] log_lik;
  real deviance;

  deviance = 0;
  for (n in 1:N) {
    real mu_safe = fmin(fmax(mu[n], 1e-8), 1e8);
    log_lik[n] = neg_binomial_2_lpmf(y[n] | mu_safe, 1 / phi);
    deviance += -2 * log_lik[n];
  }
}
```

D.2. Código para el Modelo de Regresión NB-G

El siguiente código Stan implementa el Modelo de Regresión NB-G aplicada con múltiples covariables:

```
data {
  int<lower=0> N;
  vector[N] x1;
  vector[N] x2;
  vector[N] x3;
  vector[N] x4;
  vector[N] x5;
  int<lower=0> y[N];
}

parameters {
  real beta_0;
  real beta_1;
  real beta_2;
  real beta_3;
  real beta_4;
  real beta_5;
  real<lower=0> phi;
  real<lower=0> nu;
  vector<lower=0>[N] omega;
}

transformed parameters {
  vector[N] mu;
  mu = exp(beta_0
           + beta_1 * x1
           + beta_2 * x2
           + beta_3 * x3
           + beta_4 * x4
           + beta_5 * x5);
}
```

```
model {
  // Priors
  beta_0 ~ normal(0, 10);
  beta_1 ~ normal(0, 10);
  beta_2 ~ normal(0, 10);
  beta_3 ~ normal(0, 10);
  beta_4 ~ normal(0, 10);
  beta_5 ~ normal(0, 10);
  phi ~ gamma(2, 2);
  nu ~ exponential(2);
  omega ~ gamma(nu + 1, nu + 1);

  for (i in 1:N)
    y[i] ~ neg_binomial_2(mu[i], 1.0 / (phi * omega[i]));
}

generated quantities {
  vector[N] log_lik;
  real deviance;

  deviance = 0;
  for (n in 1:N) {
    real mu_safe = fmin(fmax(mu[n], 1e-8), 1e8);
    log_lik[n] = neg_binomial_2_lpmf(y[n] | mu_safe, 1.0 / (phi * omega[n]));
    deviance += -2 * log_lik[n];
  }
}
```

D.3. Código para el Modelo de Regresión NB-IG

El siguiente código en Stan implementa el Modelo de Regresión NB-IG aplicada con múltiples covariables:

```
data {
  int<lower=0> N;
  vector[N] x1;
  vector[N] x2;
  vector[N] x3;
  vector[N] x4;
  vector[N] x5;
  int<lower=0> y[N];
}

parameters {
  real beta_0;
  real beta_1;
  real beta_2;
  real beta_3;
  real beta_4;
  real beta_5;
  real<lower=0> phi;
  real<lower=0> nu;
  vector<lower=0>[N] omega;
}

transformed parameters {
  vector[N] mu;
  mu = exp(beta_0
           + beta_1 * x1
           + beta_2 * x2
           + beta_3 * x3
           + beta_4 * x4
           + beta_5 * x5);
}
```

```
model {
  beta_0 ~ normal(0, 10);
  beta_1 ~ normal(0, 10);
  beta_2 ~ normal(0, 10);
  beta_3 ~ normal(0, 10);
  beta_4 ~ normal(0, 10);
  beta_5 ~ normal(0, 10);
  phi ~ gamma(2, 2);
  nu ~ exponential(2);
  omega ~ inv_gamma(nu + 2, nu + 1);

  for (i in 1:N)
    y[i] ~ neg_binomial_2(mu[i], 1.0 / (phi * omega[i]));
}

generated quantities {
  vector[N] log_lik;
  real deviance;

  deviance = 0;
  for (n in 1:N) {
    real mu_safe = fmin(fmax(mu[n], 1e-8), 1e8);
    log_lik[n] = neg_binomial_2_lpmf(y[n] | mu_safe, 1.0 / (phi * omega[n]));
    deviance += -2 * log_lik[n];
  }
}
```