

# Modelación Bayesiana de la distribución del Carbón de la Espiga del Maíz (*Ustilago Maydis*) con Modelos Inflados en Cero

Videla María Eugenia<sup>1,2,4</sup>, Kistner Belén<sup>1,3</sup>, Iglesias Juliana<sup>1,3</sup> y Bruno Cecilia<sup>1,2</sup>

<sup>1</sup> Consejo Nacional de Investigaciones Científicas y Tecnológicas (CONICET)

<sup>2</sup> Cátedra de Estadística y Biometría. Facultad de Ciencias Agropecuarias (FCA). Universidad Nacional de Córdoba. Grupo vinculado a UFyMA (Unidad de Fitopatología y Modelización Agrícola INTA -CONICET).

<sup>3</sup> Mejoramiento Genético de Maíz. EEA INTA Pergamino.

<sup>4</sup> Universidad Nacional de Villa María (UNVM)

eugeniavidela12@gmail.com

## Resumen

*Algunas variables epidemiológicas en cultivos agrícolas se miden como cantidad de plantas enfermas en una parcela (conteos) y tienen la particularidad de contener una gran proporción de valores iguales a cero. El objetivo de este trabajo es comparar el desempeño de diferentes modelos bayesianos ajustados sobre variables de conteo utilizando el método de estimación por aproximación de Laplace anidada integrada (INLA- Integrated Nested Laplace Approximation). Se ilustra sobre un conjunto de 79 líneas estables de maíz con carbón de la espiga. Los modelos evaluados fueron MLGM Poisson, MLGM Binomial Negativo, Modelo Inflado en Cero Poisson y Modelo Inflado en Cero Binomial Negativo.*

**Palabras Clave: Modelos Lineales Generalizados Mixtos – Variables Discretas – Conteos – INLA**

## Introducción

Las variables epidemiológicas que afectan a cultivos agrícolas suelen registrarse con variables discretas (plantas enfermas de una parcela) y modelarse a través de Modelos Lineales Generalizados (MLG) (Diaz-Quijano, 2016). En estudios para identificar genotipos resistentes al carbón de la espiga del maíz (*Ustilago Maydis*) la opción clásica es ajustar Modelos Lineales Generalizados Mixtos (MLGM) con distribución Poisson. Sin embargo, estos modelos son vulnerables a la sobredispersión debido a la relación de igualdad que existe entre media y varianza. En este contexto, los MLGM basados en la distribución Binomial Negativa incorporan un componente de variabilidad para datos con sobredispersión (Ismail y Jemain, 2007). Otro desafío en este tipo de variables es que contienen una gran proporción de valores iguales a cero (plantas sin síntoma en toda la parcela de estudio), lo cual también genera sobredispersión (Phang y Loh, 2013). Estos ceros pueden clasificarse en “falsos ceros”, causados por parcelas que no estén expuestas a la enfermedad o por errores del observador y en “ceros auténticos” que son aquellos generados por parcelas que han estado expuesta al hongo pero por diversos motivos (resistencia del genotipo, condiciones ambientales, etc) ninguna planta ha sido infectada. Distinguir unos ceros de otros es fundamental para identificar genotipos resistentes. Los MLGM de tipo “Inflados en Cero” (MIC), surgen como una alternativa para ajustar el exceso de ceros y evitar la sobredispersión. Para cada observación se producen dos procesos, el primer proceso genera el dato para cada observación como resultado de un ensayo Bernoulli con probabilidad, asumiendo que existen dos posibles resultados para cada observación (sana/enferma), mientras que el segundo

proceso genera los resultados de una distribución paramétrica específica, como Poisson (Modelo MICP) o binomial negativo (Modelo MICBN) con probabilidad (Lambert, 1992). Recientemente, se ha desarrollado un método de estimación para ajustar modelos bayesianos bajo la aproximación de Laplace, INLA por sus siglas en inglés Integrated Nested Laplace Approximation (Rue et al., 2009) como una alternativa computacionalmente eficiente para el método de cadenas de Markov Monte Carlo (MCMC) (Blangiardo et al., 2013).

El objetivo de este trabajo es comparar el desempeño de modelos bayesianos, ajustados sobre variables de conteo, que no contemplan los excesos de ceros (MLGM Poisson (P) y MLGM Binomial Negativa (BN)) respecto a modelos de conteo para datos inflados en cero (MICP y MICBN) para modelar la distribución del carbón de la espiga del maíz con la finalidad de seleccionar genotipos resistentes.

## Material y Métodos

Se evaluaron 79 genotipos de maíz pertenecientes al grupo de mejoramiento de maíz de INTA Pergamino, Argentina en dos localidades (Tucumán y Pergamino). Se contabilizaron la cantidad de plantas enfermas con carbón de la espiga del maíz (*Ustilago maydis*) sobre parcelas con 30 plantas. El 50% de los datos fueron 0. Se ajustaron cuatro modelos bayesianos para datos de conteo: MLGM Poisson (P), MLGM Binomial Negativa (BN), Modelo Poisson Inflado en Cero (MICP) y Modelo Binomial Negativa Inflado en Cero (MICBN). Los modelos se ajustaron utilizando el método INLA implementado en la función `inla()` de la librería INLA de R (Martino y Rue, 2010). Los modelos contemplaron el efecto de localidad y de repetición dentro de localidad como fijo y el efecto de genotipo como aleatorio para estimar el mejor predictor lineal insesgado (BLUP) y posteriormente realizar un ranking de los genotipos de acuerdo a su comportamiento de resistencia al carbón de la espiga. El desempeño de los modelos se comparó con el criterio de información de deviance (DIC), número de parámetros efectivos ( $p$ ) y marginal log-Likelihood (MlogLink).

## Resultados

*Tabla 1: Criterio de información de Deviance (DIC), número de parámetros efectivos ( $p$ ) y marginal log-Likelihood (MlogLink) obtenidos al ajustar 4 modelos.*

Modelo	DIC	$p$	MlogLink
P	809.706	62.16	-464.40
BN	789.692	58.34	-445.65
MICP	794.713	60.27	-455.96
MICBN	788.346	59.34	-446.24

*P: MLGM Poisson, NB: MLGM Binomial Negativa, MICP: Modelo Inflado en Cero Poisson, MICBN: Modelo Inflado en Cero Binomial Negativa.*

Tabla 2: Valores promedios de los efectos aleatorios de cada genotipo para los más susceptibles (valores mayores a 2) y los menos susceptibles (valores menores a -1) obtenidos con el Modelo MICBN.

Genotipo	BLUP
L4674	3.290987
LP605	2.882328
LP915	2.734550
LP579	2.407617
LP1512	-1.142281

Los modelos con distribución Binomial Negativa (BN y MICBN) tuvieron menos cantidad de parámetros efectivos ( $p$ ) que los Poisson (P y MICP) y mayor probabilidad marginal (MlogLink). El modelo inflado en cero Binomial Negativo (MICBN) tuvo mejor ajuste (menor DIC) que el modelo inflado en cero Poisson (MICP) y que los modelos de conteo tradicionales (P y BN) (Tabla 1).

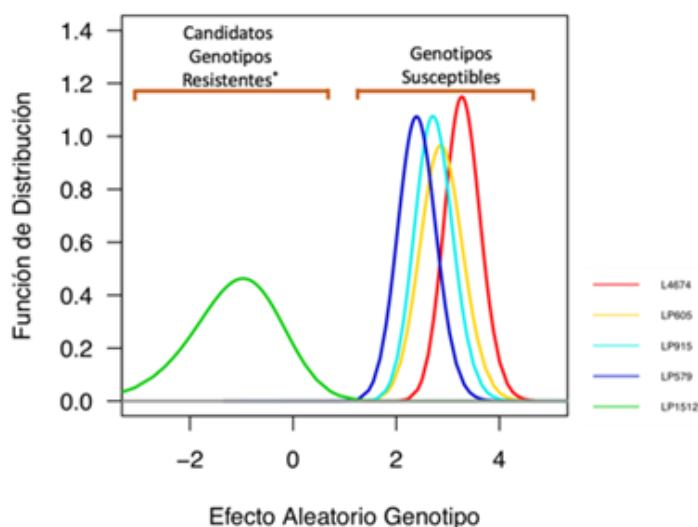


Figura 1: Densidades de los efectos aleatorios de los genotipos más susceptible y más resistentes al carbón del maíz.

\*La densidad de los 18 genotipos más resistentes fueron idénticas. Se muestra sólo para el LP1512.

Con el MICBN se realizó un ranking del efecto aleatorio del genotipo. Se destacaron cuatro genotipos con valores promedios de efecto aleatorio de genotipo mayor a 2 (genotipos más susceptibles) y 18 genotipos con valores promedios de efecto aleatorio menores a -1 (más resistentes) según el modelo seleccionado. Los genotipos seleccionados como más resistentes presentaron idéntico valor medio y sus densidades se superpusieron, por tal motivo solo se

muestra la función de distribución del genotipo LP1512 como representante de este grupo de genotipos (Tabla 2 – Figura 1).

### Conclusión

El método INLA permitió obtener predictores del merito genético en ensayos a campo para selección de genotipos con excesos de cero para una variable discreta. Los modelos Inflados en Cero con distribución Binomial Negativo resultaron más precisos con respecto a los modelos tradicionales como MLGM Poisson y Binomial Negativo o el modelo de Poisson Inflado en Cero.

### Referencias

- Blangiardo, M., Cameletti, M., Baio, G., y Rue, H. (2013). Spatial and spatio-temporal models with R-INLA. *Spatial and spatio-temporal epidemiology*, 4, 33-49.
- Diaz-Quijano, F. (2016). Regressions applied to the study of discrete events in epidemiology. *Revista de la Universidad Industrial de Santander. Salud*, 48(1), 9-15.
- Ismail, N., y Jemain, A. (2007). Handling overdispersion with negative binomial and generalized Poisson regression models. *In Casualty Actuarial Society Forum* (pp. 103-158).
- Lambert, D. (1992). Zero-inflated Poisson regression models with an application to defects in manufacturing. *Technometrics* 34(1), 1 – 14.
- Martino, S. y Rue, H. (2010). Implementing approximate bayesian inference using integrated nested laplace approximation: a manual for the inla program. Available from: <http://www.math.ntnu.no/hrue/>
- Phang, Y. y Loh, E. (2013). Zero inflated models for overdispersed count data. *Int. J. Math. Comput. Nat. Phys. Eng*, 7, 817-819.
- Rue, H., Martino, S. y Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *J R Stat Soc B*;71(2):1-35.