

# Análisis Estadístico Lineal Avanzado en R (parte II)

Curso Colaborativo IIAP - UNAMAD

Irwing S. Saldaña

Instituto de Ciencias Antonio Brack

Departamento de Ecoinformática y Biogeografía

Perú, 2021



# Blgo. Irwing S. Saldaña

## Instructor

Dpto. de Ecoinformática y Biogeografía,  
Instituto de Ciencias Antonio Brack, Perú

[Website](#) | [ResearchGate](#) | [Linkedin](#) | [R Latam Blog](#) | [Github](#)



# Hurdle Models

[ Lidiando con exceso de ceros en conteos ]

## RECORDEMOS

En los modelos de conteos (Poisson o Binomial Negativa), se asume que los ceros y los valores no ceros (números enteros positivos) provienen de los mismos procesos de generación de datos.



# Nociones de modelos para lidar con Excesos de Ceros

## Model de Hurdle (Zero-altered models, ZAP/ZANB)

Los ceros son modelados con **Reg. Binomial**; los conteos con **Reg. de Poisson Zero-Truncada**. Todas las razones de que un conteo sea cero se modelan como probabilidad de ausencia.

"Conteo >0" --- ¡Aquí estoy!

"Conteo = 0" --- No estoy:

- Porque el hábitat no es bueno.
- Porque no, aunque el hábitat es bueno.
- Porque no pudiste verme.
- Pensaste que era otra especie.



© Asociación Armonía  
*Carl Christian Tofte*



## Model Zero-Inflados (Zero-altered models, ZAP/ZANB)

Un segundo modelo para lidiar con exceso de ceros son los ZAP/ZANB. El modelamiento es igual a modelos Hurdle, excepto que las razones de que un conteo sea cero se modelan como (1) probabilidad de ausencia y como (2) problemas de detectabilidad (dos procesos).

### Procesos de conteo (P/BN)

"Conteo >0" --- ¡Aquí estoy!

"Conteo = 0 como ausencia" --- No estoy:

- Porque el hábitat no es bueno.

### Procesos de masa (B)

"Conteo = 0 como prob. detect." --- No estoy:

- Porque no pudiste verme.
- Porque no, aunque el hábitat es bueno.
- Pensaste que era otra especie.



© Asociación Armonía

Carl Christian Tofte



# Desarrollemos la práctica

30 : 00



# Modelos Generalizados Lineales de Efectos Mixtos

[ GLMM ]



# GLMM

- Son modelos que involucran la **fusión de las dos grandes ventajas en modelamiento lineal** aprendidas en este curso: los efectos mixtos (proveniente de los LMM) y el relajar las asunciones teóricas de normalidad (proveniente de los GLM).
- Tienen la siguiente representación:

$$\begin{cases} Y_i \mid b_i \sim \mathcal{F}_\psi, \\ b_i \sim \mathcal{N}(0, D), \end{cases}$$

Donde  $Y_i$  es la respuesta de la  $i$ ésima muestra, y  $b_i$  es un vector de efectos aleatorios para esta unidad. La respuesta condicionada por los efectos aleatorios tiene distribución  $\mathcal{F}$  parametrizada por el vector  $\psi$ . Se supone que los efectos aleatorios siguen una distribución normal multivariante con media 0 y matriz de varianza-covarianza  $D$ .



# El problema de la integral cerrada

La **función de verosimilitud** de estos modelos tiene la siguiente forma general:

$$L(\theta) = \prod_{i=1}^n \int p(y_i | b_i; \psi) p(b_i; D) db_i,$$

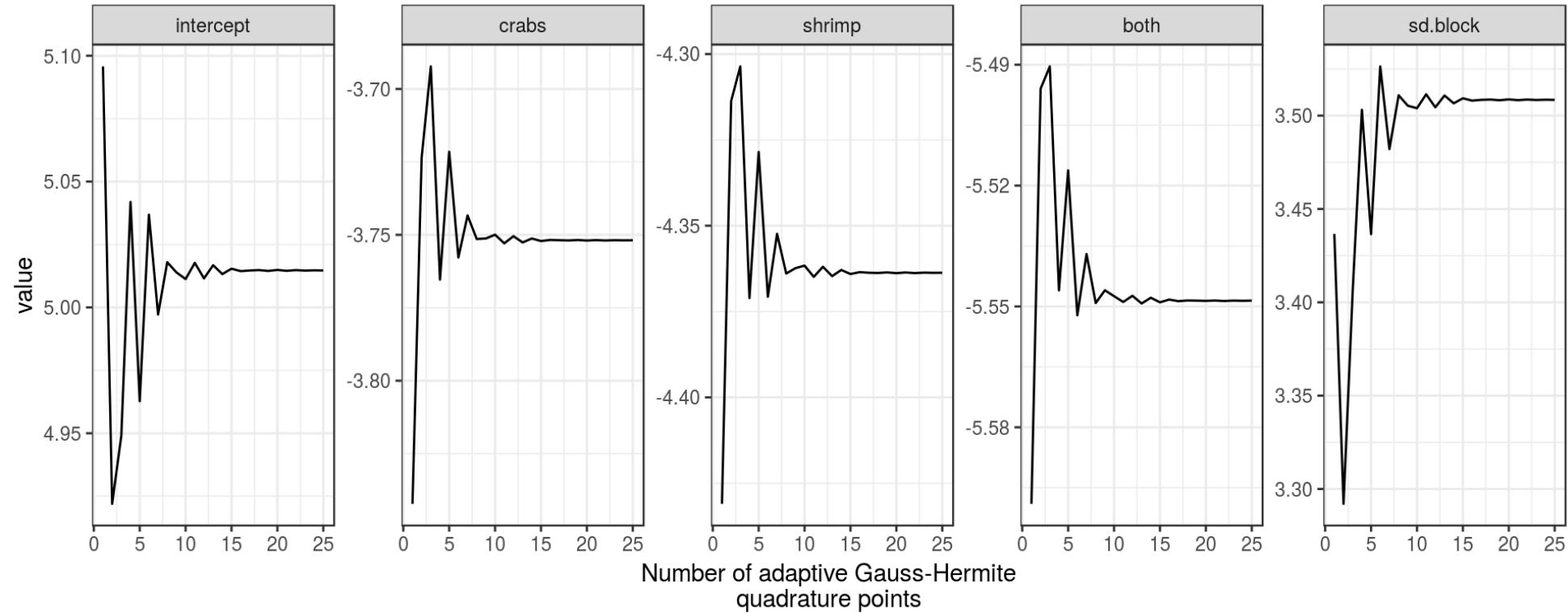
El problema es que la integral aquí no tiene una solución de forma cerrada. Por lo tanto, para estimar los parámetros en estos modelos con máxima verosimilitud, es necesario aproximar de alguna manera esta integral. Se conoce en la literatura dos soluciones:

- **Aproximación del integrando:** PQL y Aproximación de Laplace. Trabaja mejor con modelos:
  - Binomiales con gran N en el divisor (n/N).
  - Poisson con grandes recuentos esperados.
- **Aproximación de la integral:** Monte Carlo y Cuadratura Adaptativa Gaussiana. Trabaja mejor con modelos:
  - Binomiales de bernulli ("Y" binaria).
  - Poisson con bajos recuentos esperados.



```
# Permite Cuadratura GH con solo un efecto aleatorio. Para más efectos  
# aleatorios se aplica automáticamente la Aproximación de Laplace
```

```
library(lme4)  
glmer(..., nAGQ = 25)
```



# Topología de modelos GLMM en R

Modelos de efectos mixtos GLMM usando `lme4::glmer`

```
# Ejemplo GLMM
glmer(Var_respuesta ~ Efec_fijo + (1|Efec_Aleatorio),
      data=DF, family="FAMILIA", nAGQ = 25)
```

La interpretación de los modelos, sea cual sea el método que se aplique, sigue lo aprendido en el curso sobre interpretación de los efectos aleatorios, así como la correcta interpretación de los efectos fijos y su transformación necesaria como en los GLM previamente vistos.



# Desarrollemos la práctica

60 : 00



# Gracias por su atención

