

Distribución Discreta Poisson XLindley *

Mariana Blandón Mejía *Universidad Nacional de Colombia - sede Medellín*
Freddy Hernandez Barajas *Docente Universidad Nacional de Colombia - sede Medellín*

Resumen: Se resaltarán las ventajas de la Poisson-XL, su función masa de probabilidad y su función de distribución acumulada, previamente se realizó la creación de la nueva distribución en RStudio utilizando los modelos gamlss, lo que permite elaborar estudios de simulación y modelamiento con datos reales. Se encontró que las propiedades de la PXL se pueden expresar en formas cerradas, lo que la convierte en una herramienta de probabilidad propuesta para establecer modelos de regresión y series de tiempo; la nueva función masa de probabilidad se puede utilizar para modelar datos sesgados positivamente con forma leptocúrtica, además, la nueva distribución PXL, es útil para el modelamiento de fenómenos equidispersos y sobredispersos con una función de tasa de riesgo creciente.

Keywords: gamlss, leptocúrtica, equidisperso, sobredisperso.

Introducción

Se estudiará una nueva distribución de conteo discreta de un parámetro la cual combina las distribuciones Poisson y XLindley, se obtiene como resultado la "Poisson-XL", esta nueva distribución tiene múltiples ventajas como la capacidad de describir datos equidispersos y sobredispersos, además, puede ser utilizada como una herramienta estadística adecuada para modelar datos sesgados positivamente con forma leptocúrtica.

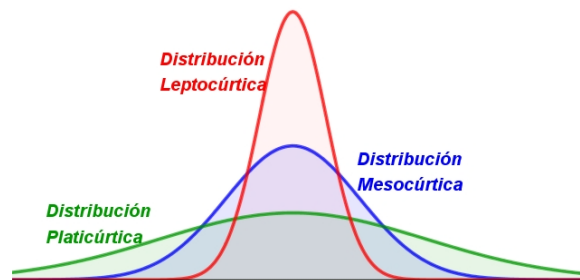


Figure 1: Curtosis.

El coeficiente de curtosis se utiliza para medir la forma de la distribución de los datos. En una distribución leptocúrtica, el coeficiente de curtosis es positivo, indicando colas más pesadas en comparación con una distribución normal. La leptocurtosis es el opuesto de la platicurtosis, que se refiere a colas más ligeras en la distribución. La forma leptocúrtica puede deberse a diversos factores, como la presencia de eventos extremos o la naturaleza de los datos en estudio.

* Autor de contacto: mblandonm@unal.edu.co.

Dado que existe la necesidad de un modelo más flexible para modelar datos estadísticos Mahmoudi y Zakerzadeh, docentes asociados al departamento de estadística de la Universidad de Yazd, ampliaron la distribución de Poisson-Lindley y revelaron que su distribución generalizada es más flexible a la hora de evaluar los datos de recuento, en este estudio se ahondará en una distribución discreta flexible combinando las distribuciones de Poisson y XL, llamada *Poisson-XLindley*.

En este estudio se abordarán las principales características de la distribución discreta de un solo parámetro Poisson XLindley, propiedades estadísticas y la implementación de esta al modelamiento de datos reales y simulados.

Propiedades Estadísticas

La función masa de probabilidad de la distribución Poisson-XL está dada por:

$$f(x) = \frac{\mu^2(2 + \mu + x)e^{-\mu x}}{(1 + \mu)^2}; x > 0, \mu > 0$$

La función de probabilidad conjunta para Poisson-XL está dada por:

$$F(x) = 1 - \frac{(1 + \mu(x + 4 + \mu(3 + \mu)))}{(1 + \mu)^{4+x}}$$

Se presentará un gráfico de la función masa de probabilidad para la distribución PoissonXL variando el valor del parámetro μ

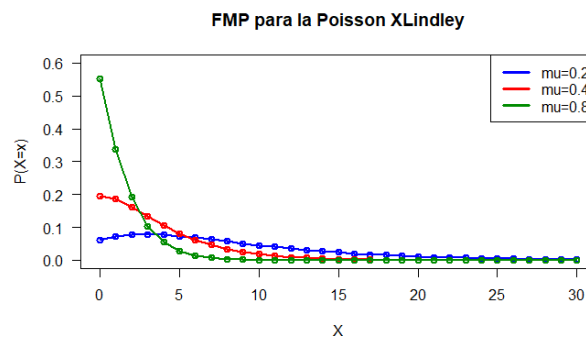


Figure 2: función masa de probabilidad.

Del gráfico anterior se puede analizar que a medida que se incrementa el valor del parámetro μ la función masa presenta una probabilidad más alta para valores de muestra pequeños, luego de una muestra de 5 unidades esta se estabiliza y comporta de manera muy similar a valores más pequeños de μ como lo son 0.2 y 0.4.

Implementación de la distribución Poisson XL

Ejemplo con datos Reales.

Los datos con los que se trabajará fueron tomados del estudio del cual fue tomada esta distribución (I-Haq, Al-Bossly.2022), este conjunto de datos contiene el número de lesiones dosimétricas citogenéticas de mamíferos producidas por exposición a estreptogramina.

Las estreptograminas son conocidas por su actividad contra ciertos tipos de bacterias, especialmente aquellas resistentes a otros antibióticos más comunes. Por lo general, se utilizan en combinación para aumentar su eficacia y reducir el riesgo de resistencia bacteriana, el objetivo de este estudio es analizar la cantidad de lesiones presentes en el ADN de los conejos luego de aplicar diferentes dosis de estreptogramina.

Con el objetivo de analizar la robustez de la distribución Poisson-XL y como esta se adapta a estos datos de tipo científico se realizó una prueba en la que se comparará el comportamiento de esta distribución con respecto a los datos reales.

Table 1: Ajuste PXL vs datos reales.

| X | Frecuencia Observada | Frecuencia(promedio) Esperada |
|----------|----------------------|-------------------------------|
| 1 | 200 | 194.6 |
| 2 | 57 | 68.5 |
| 3 | 30 | 24.0 |
| 4 \geq | 13 | 12.9 |
| Total | 300 | 300 |

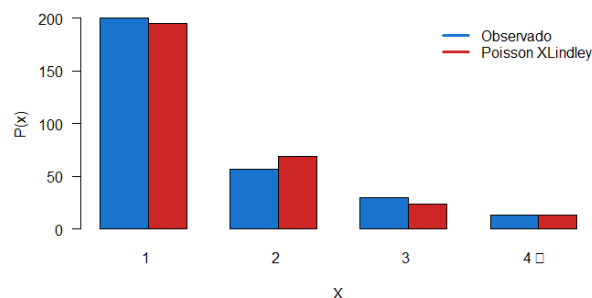


Figure 3: Resultados ajuste a datos reales.

Con base en los resultados de la tabla se puede observar que en este tipo de situaciones la distribución Poisson XL tiene un muy buen rendimiento en el ajuste para este tipo de datos..

Generación de algunos valores aleatorios con μ conocido

GAMLSS (Generalized Additive Models for Location, Scale, and Shape) es un enfoque estadístico que proporciona una metodología flexible para modelar diferentes aspectos de una distribución de probabilidad, incluyendo la ubicación, la escala y la forma. Los modelos GAMLSS son una extensión de los modelos de regresión generalizados (GLM) y los modelos lineales generalizados aditivos (GAM).

En el siguiente código se generará una muestra con $n = 1000$ valores que siguen una distribución Poisson XLindley con μ conocido, en este caso 0.5, luego de generar los valores se ajustará un modelo de la familia gamlss el cual tiene como objetivo acercarse lo más posible al verdadero valor de μ

```
y <- rPOISXL(n=1000, mu=0.5)
library(gamlss)
mod1 <- gamlss(y~1, family=POISXL,
               control=gamlss.control(n.cyc=500, trace=FALSE))

exp(coef(mod1, what='mu'))
(Intercept)
0.4820785
```

El resultado del modelo arroja que $\mu = 0.48$ lo cual es bastante cercano al μ que se estableció inicialmente, así que este modelo cumple el objetivo.

Conclusiones

En general, se encontró que las propiedades de PXL se pueden expresar en formas cerradas, lo que convierte a esta distribución en una herramienta de probabilidad propuesta para establecer modelos de regresión, series de tiempo y es ideal para modelamiento de datos de tipo científico; esto hace de la distribución PXL un elemento muy versátil para el análisis de datos en diversos campos.

Referencias

I-Haq, Al-Bossly. 2022. Poisson XLindley Distribution for Count Data: Statistical and Reliability Properties with Estimation Techniques and Inference
URL: <https://www.hindawi.com/journals/cin/2022/6503670/>

Hassan, Shalbaf, Bilal y Rashid. 2020. Una nueva distribución discreta flexible con aplicaciones para contar datos. Journal of Statistical Theory and Applications, vol. 19
URL: <https://www.atlantis-press.com/journals/jsta/125935241/view>

Grine y Zeghdoudi. 2017. Sobre la distribución cuasi-lindley de Poisson y sus aplicaciones. Journal of Modern Applied Statistical Methods.
URL: <https://www.semanticscholar.org/paper>