UNIVERSIDAD NACIONAL DE INGENIERIA
FACULTAD DE CIENCIAS

Título:

Aproximaciones Gaussianas: Binomial, Poisson y Normal

Curso:

INTRODUCCIÓN A LA ESTADÍSTICA Y PROBABILIDADES

Autores:

BARRIENTOS PORRAS HERLEES BRAYAN

MEDRANO QUISPE WILMER

TAPAHUASCO DE LA CRUZ JORDI EDGAR

BORJAS CÓRDOVA GUILLERMO JOEL

Profesor:

LARA ÁVILA CÉSAR

LIMA – PERÚ

2018

***Resumen* — En este informe se dará una breve explicación de los distintos métodos de aproximación de las distribuciones Gaussianas, se expondrán algunos ejemplos aplicativos de estas aproximaciones y finalmente un algoritmo en lenguaje R para su verificación teórica.**

INTRODUCCIÓN

Conceptos previos.-

i. Una función Gaussiana es una función definida por:

$$f(x) = ae^{-\frac{(x-b)^2}{2c^2}}$$

donde ‘a’, ‘b’ y ‘c’ son constantes reales.

La gŕafica de esta función describe una curva acampanada llamada campana de Gauss.

En el caso en que ‘a’ sea igual a $\frac{1}{c\sqrt{2\pi}}$

esta función correspondería a la función de densidad (PDF) de una variable aleatoria llamada distribución Normal o distribución Gaussiana de media $\mu=b$ y varianza $\sigma^2=c^2$ N(μ, σ), siendo ésta una de las distribuciones de probabilidad de variable continua que con más frecuencia aparece en estadística y en la teoría de probabilidades

ii. La distribución Binomial con parámetros ‘n’ y ‘p’, es una distribución de probabilidad discreta del número de éxitos en una secuencia de ‘n’ ensayos independientes de Bernoullí de un experimento aleatorio. Los éxitos tendrán una probabilidad ‘p’, mientras que los fracasos probabilidad q=1-p. La función de masa de probabilidad (PMF) de esta distribución es:

$$f(k,n,p) = \Pr(k;n,p) = \Pr(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$$

El cual nos daría la probabilidad de conseguir exactamente ‘k’ éxitos en ‘n’ ensayos, cada uno con probabilidad ‘p’ de ser un éxito (n,k ∈ ℕ, p ∈ [0,1].

iii. La distribución de Poisson con parámetros ‘λ’ y ‘k’, es una distribución de probabilidad discreta que expresa la probabilidad de que un número dado de eventos ocurra en un intervalo fijo de tiempo si estos eventos ocuerren con una velocidad constante conocida e independientemente del tiempo transcurrido desde el último evento. Su PMF es dada por:

$$f(k,\lambda) = \frac{e^{-\lambda}\lambda^k}{k!}$$

λ > 0 (real), k ∈ ℕ.

Esto nos da la probabilidad de observar ‘k’ eventos en un intervalo dado que el promedio del número de eventos es ‘λ’ en ese mismo intervalo.

OBJETIVO GENERAL

- El objetivo del presente trabajo es la demostración empírica mediante la implementación de algoritmos en R la aproximación de estas distribuciones a otras. Las aproximaciones a tratar serán: Binomial-Poisson, Binomial-Normal, y Poisson-Normal.

OBJETIVOS ESPECÍFICOS

- Se mostrará gráficamente la aproximación de cada una de las distribuciones y se corroborará mediante la variación de sus parámetros.
- Se expondrán las diferencias entre cada una de las aproximaciones y en qué casos éstas son usadas.

El modelo a utilizar será una composición de otros modelos realizados por distintos autores cuyos trabajos son ejemplos de aproximaciones de estas distribuciones a otras un poco más complejas obtenidas de diferentes artículos que expondremos más adelante.

ESTADO DEL ARTE

A continuación haremos un pequeño análisis de diferentes artículos que emplean las distribuciones Binomial Negativa, Poisson y Normal para hacer sus aproximaciones a otras distribuciones. Esto será una base para el desarrollo de los métodos a emplear en la resolución de nuestros problemas de aproximaciones.

A. APROXIMACIÓN DE LA DISTRIBUCIÓN BINOMIAL NEGATIVA A UNA DE POISSON

A. Navarro , F. Utzet , P. Puig , J. Caminal , M. Martin. "La Distribución Binomial Negativa frente a la de Poisson en el análisis de fenómenos recurrentes", Laboratorio de Bioestadística y Epidemiología. Facultad de Medicina. Universidad Autónoma de Barcelona Gac Sanit 2001; 15 (5): 447 – 452.

Objetivos:

- Exponer la posible problemática en el cálculo de riesgos en bases de datos agregados cuando el fenómeno estudiado es recurrente y presentar la distribución binomial negativa como una alternativa válida y sencilla para analizar este tipo de fenómeno.

Métodos:

- En el contexto de los fenómenos recurrentes, el análisis mediante la regresión de Poisson puede provocar sobredispersión o variancia extra-poisson, lo cual conduce a la subestimación de los errores estándares de los coeficientes, pudiendo derivar en la significación estadística de factores que realmente no estén asociados con el fenómeno. La binomial negativa puede captar parte de la variancia que no identifica la regresión de Poisson. Para comprobarlo se comparó ambas distribuciones sobre el número de hospitalizaciones que presentaron individuos entre 65 y 69 años de edad durante el año 1996. Esta comparación fue realizada en dos bases de datos agregados distintas: por individuo y según las variables de interés.

Resultados:

- El ajuste mediante ambas distribuciones presenta diferencias en las dos bases de datos. En ambos casos, la regresión de poisson estima significativamente cuatro de las seis variables estudiadas. Para la binomial negativa son dos en la base por individuo y una en la base por variables.

Conclusiones:

- La existencia de sobredispersión es frecuente en fenómenos recurrentes. Cuando esto sucede, el uso de la binomial negativa es más apropiado que el de la regresión de Poisson.

B. APROXIMACIÓN BINOMIAL A NORMAL

Ortíz Pinilla, Jorge; Castro, Amparo; Neira, Tito; Torres, Pedro; Castañeda, Javier "Criterios sobre el uso de la distribución normal para aproximar la distribución binomial" Revista Colombiana de Estadística, vol. 23, núm. 1, 2000, pp. 65-70 Universidad Nacional de Colombia Bogotá, Colombia

Objetivos:

Presentar propuestas para controlar el error cometido al utilizar la aproximación normal cerca de algunos valores escogidos para el error máximo. En el cual se presentan un criterio y algunas fórmulas.

Resultados:

Es posible encontrar algunas simplificaciones a los modelos encontrados sin sacrificar grandemente las estimaciones de n.

Conclusiones:

El propósito es controlar la máxima diferencia entre la función distribución binomial exacta y la aproximación normal con corrección de Yates, las reglas recomendadas tradicionalmente para el valor mínimo del parámetro "n" presentan inexactitudes importantes, generadas por la falta de regularidad en el control del error máximo.

C. APROXIMACIÓN DE LA DISTRIBUCIÓN POISSON A UNA DISTRIBUCIÓN NORMAL.

I.Arroyo, L.Bravo, Dr. Ret. Nat. Humberto, Msc.Fabián L., "Distribuciones Poisson y Gamma: Una Discreta y Continua Relación", Universidad del Norte, Barranquilla, Colombia, Diciembre 2013 pp 100-102.

Objetivos:

- Verificar que una distribución Poisson p (λ, k) a medida que λ (frecuencia con la que se espera se realice un suceso en un intervalo de tiempo) aumenta, se aproxima a una distribución normal.

Resultados:

- En una distribución Poisson, la varianza y esperanza son iguales a λ.

- La distribución poisson p(λ,k) para $\lambda \geq 9$, converge a una distribución normal N(λ, $\sqrt{\lambda}$).
- Por el teorema central de límite conforme aumenta λ las variables de Poisson se aproximan a una distribución normal.

Conclusiones:

- A medida q λ crece la distribución de Poisson va adquiriendo forma de la campana de Gauss.

Diseño del Experimento

Para la implementación del código en R se usarán las funciones de ploteo 'plot()' para el gráfico y de distribuciones tales como 'dbinom()', 'dpois()' y 'dnorm()'; así como otras funciones auxiliares para cálculos matemáticos.
La técnica a utilizar será la de comparación entre las gráficas de las funciones de distribución, de esta manera al variar el valor de sus parámetros se podrá observar el cambio en las gráficas y cómo se se van asemejando unas con otras.
A continuación se expondrán las teorías de cómo suceden las aproximaciones correspondientes:

Binomial – Normal:
Una distribución binomial con número de ensayos n y probabilidad de éxito p
se puede aproximar a una distribución normal si:

- si $n \to \infty$.
- o si n no es muy grande pero se cumple que $n \cdot p > 5$ y que $n \cdot (1-p) > 5$.

La distribución normal que se usa para la aproximación tiene media $\mu = n \cdot p$ y varianza $\sigma^2 = n \cdot p \cdot (1-p)$.

Binomial – Poisson:

Una distribución binomial con número de ensayos n muy grande y probabilidad de éxito p cercana a 0 (suceso raro), se puede aproximar a una distribución Poisson con $\lambda = np$.
Una regla general aceptable es emplear esta aproximación si:

- $n > 50$ y $n \cdot p < 5$.

Por esta razón se considera a veces al modelo de Poisson como una forma límite de la distribución Binomial y se le utiliza para aproximar probabilidades en ésta.

Poisson – Normal:

La distribución Poisson con parámetro λ se puede aproximar por medio de la distribución normal si $\lambda \to \infty$.
La distribución normal que se usa para la aproximación tiene media $\mu = \lambda$ y varianza $\sigma^2 = \lambda$.

Implementación

Código del script:
El código se encuentra en la siguiente url:
https://github.com/Guillermo-Borjas/Proyecto-Aproximaciones-Gaussianas
La carpeta Aproximaciones-Gaussianas, contendrá 2 archivos: Server.R y ui.R.
Seguir las instrucciones del README.md para ejecutar el programa.

Descripción del código:
Se elaboró una aplicación web que corre en un servidor local. Para ello se usó la librería "Shiny", cuyo paradigma es la programación reactiva, esto es, la interactividad entre la interfaz de usuario y la parte del servidor es casi en tiempo real, pudiendo así reflejar nuevos resultados instantáneamente se modifiquen ciertos valores que el programa le permite al usuario alterar.

En principio se usaron 3 funciones para cada tipo de plot: Binomial - Normal, Poisson – Normal y Binomial – Poisson. Cada función declarada en el lado servidor (Server.R), recoge los valores de los parámetros que son varialbes de tipo input, esto es, variables que son mostradas en la interfaz de usuario (ui.R), y cada vez que se modifican sus valores, vuelven a evaluarse en las funciones, arrojando un nuevo resultado del plot el cual es guardado por otra variable de tipo output, esta variable es recogida por la interfaz mostrando así el nuevo plot con los valores modificados.

La primera función (f1), recoge los valores n (número total de ensayos) y p (probabilidad de éxito), éstos valores se usarán para aproximar la función de masa de probabilidad de la distribución Binomial a la función de densidad de la distribución Normal, usando el teorema de De Moivre – Laplace. Esto es, para el valor de la desviación estándar de la distribución Normal se tomará el valor de la raíz cuadrada de la multiplicación de n, p y (1 – p) y media la multiplicación n · p.

La segunda función (f2), recoge el valor de λ , y mediante el teorema central de límite, la PMF de Poisson se aproxima a la PDF de la Normal cuando la media de la distribución Nromal es λ , y su desviación estándar raíz cuadrada de λ.

La tercera función (f3), recoge también los valores n y p como en la función "f1", y para aproximar el PMF de la Binomial a la de Poisson se dará a λ el valor de n · p.

Resultados y Discudiones

Efectivamente, las gráficas nos muestran una gran similitud en sus aproximaciones, sólo tenemos una discrepancia en el caso de la aproximación de Binomial a Poisson; cuando la probabilidad de éxito p se hace cada vez mayor vemos que el resultado de la probabilidad de Poisson se hace cada vez menor en comparación a la Binomial.

Conclusiones

Es posible usar la distribución Normal como una buena aproximación tanto para las distribuciones Binomial y Poisson, ahorrándonos cálculos tediosos propios de las distribuciones de variables discretas.

Para la aproximación de Binomial a Poisson se ha de tener en cuenta que el valor de p ha de ser pequeño, preferiblemente < 0.2.

Bibliografía

Universidad Nacional de Mar del Plata: http://nulan.mdp.edu.ar/2040/1/morettini.2013.pdf

http://joseluislorente.es/estadistica/Tema7_aproxima_binomial_normal.pdf