

XUÑO 2022. VOLUME 5

MARCO NUMÉRICO

PUBLICACIÓN MATEMÁTICA DO IES MARCO DO CABBALLÓN

XUÑO 2022. VOLUME 5

MARÇO NUMÉRICO

PUBLICACIÓN MATEMÁTICA DO IES MARCO DO CABBALLÓN

MARCO NUMÉRICO

Xuño 2022. Volume 5

Marco Numérico é a publicación matemática do IES Marco do Camballón (Vila de Cruces, Pontevedra). De carácter anual, está aberta a artigos de investigación suxeitos ás esixencias profundidade, rigor e forma (metodolóxica e expositiva) propias dos estándares habituais do ámbito académico.

EDICIÓN	Departamento de Matemáticas
REVISIÓN LINGÜÍSTICA	Departamento de Lingua e Literatura Galega Departamento de Inglés
DESEÑO E IMPRESIÓN	Sacauntos S. Coop. Galega Santiago de Compostela
ISSN	2603-7459
DEP. LEGAL	PO 581-2017
CONTACTO	marconumerico@iesmarcodocamballon.com

Os textos contidos neste volume poden ser difundidos, dados a coñecer e transmitidos en calquera medio ou ámbito comunicativo e da información, sempre que se respecte con veracidade o traballo autorial e non se persigan fins comerciais.

MARCO NUMÉRICO

PUBLICACIÓN MATEMÁTICA DO IES MARCO DO CABBALLÓN

Xuño 2022. Volume 5



Índice

<i>Limiar</i>	7
<i>Contribucións a conxectura de Hardy-Littlewood</i>	9
Verde-Salgado, N.; Ferro-Pérez, J. M.; Álvarez-Lodeiro, I.; Castro-Ramos, I.	
<i>Conxectura de Collatz: abordaxe computacional e estudo de ciclos</i>	21
Fraiz-Casas, A.; López-Calviño, E.; Puga-Silva, L.	
<i>Entre primos. A Conxectura de Brocard</i>	31
Casal-Blanco, R.; González-Castro, M.; Pacheco-Fariña, M.F.; Torres-García, A.	
<i>$n^2 + 1$: unha forma de números primos</i>	43
Castelo-García, S.; López-Castro, M.; Moreira-Pereiro, A.	
<i>Sobre a natureza dos números primos. A conxectura de Legendre</i>	51
Amosa-Martínez, U.; Silva-Lodeiro, X.; Torres-Novo, N.	
<i>O teorema central do límite para variables aleatorias continuas</i>	59
Ferro-Pérez, M.; López-Calviño, E.; Puga-Silva, L.; Verde-Salgado, N.	
<i>Converxencia entre distribucións discretas. Binomial e Poisson</i>	75
González-Castro, M.; Pacheco-Fariña, M.F.; Torres-García, A.	
<i>Aproximación normal da distribución binomial</i>	87
Álvarez-Lodeiro, I.; Amosa-Martínez, U.; Castro-Ramos, I.; Silva-Lodeiro, X.	
<i>Simulación en R do dilema Monty Hall</i>	99
Castelo-García, S.; López-Castro, M.; Torres-Novo, N.	
<i>A distribución de Poisson na desintegración nuclear radiactiva</i>	107
Casal-Blanco, R.; Fraiz-Casas, A.; Moreira-Pereiro, A.	

Limiar

Marco Numérico, nome sutilmente escollido, case seguro, por alguén que sabe xogar coas palabras con elegante mestría, chega aquí á súa quinta edición. Ser convidado a presentar este novo volume representa un recoñecemento que honestamente creo que non chego a merecer, e que vai máis alá das enormes e eternas amizade e respecto mutuos. Porén, asegúrovos que non dubidei o máis mínimo en aceptar a encomenda.

É evidente que *Marco Numérico* xira arredor diso, dos números, das matemáticas. Pero o radio da súa órbita é maior do que pensamos. As matemáticas, e a ciencia en xeral, non só son contar e medir; vai moito máis aló: é puro pensamento! Todos levamos un pequeno ou grande matemático dentro, e levarnos ben con el vai facer que vivamos un pouco máis plenos, se me permitides a expresión. Pero para que serven as matemáticas? Cantas xeracións de novos e vellos levan facéndose a mesma pregunta?

Pois, primeiro, serven para a construción da persoa. As matemáticas son a linguaxe na que nós lemos o mundo, e hai poucas cousas máis humanas cás matemáticas.

E, segundo, son unha ferramenta moi poderosa para exercer dunha forma crítica a nosa liberdade como cidadáns. Isto pode, certamente, soar raro ou pretencioso. Pero canto máis sexamos capaces de analizar con rigor as situacións cotiás, máis difíciles seremos de enganar. Cada día hai alguén que está tratando de disfrazar os datos coa idea de manipularnos, e se temos con nós o rigor das matemáticas e a capacidade de interpretalos seremos cidadáns críticos, libres e máis difíciles de enganar.

Desgraciadamente, hai na sociedade unha especie de teoría que asume que as matemáticas son difíciles, que son soamente para as persoas que dicimos “listas”, e cando cremos que non valem para as matemáticas, non somos “listos”. Se alguén que lea isto tivese ese complexo, por favor, desterrádeo! As matemáticas que se aprenden na escola son soamente un tipo de habilidade e un tipo de ope-

racións. Non somos máis ou menos intelixentes por saber facer este tipo de cousas. As matemáticas escolares tenden tamén a centrarse demasiado na gramática, no vocabulario, na parte técnica, sen preocuparse de contar historias máis grandes.

Precisamente isto último é o que sucede con *Marco Numérico*: a partir dun pensamento crítico, dunha observación da realidade, dunha análise rigurosa de datos, os mozos e mozas están a contar-nos historias marabillosas. Están explicándonos o mundo dende a súa perspectiva. E explicando o que facemos adquirimos coñecementos. Explicar observacións e contar historias son grandes ferramentas de aprendizaxe, que incorporan ademais dous importantes ingredientes de cara á motivación: curiosidade e pracer. E sempre cociñado todo no gran pote da creatividade. E mires cara onde mires, *Marco Numérico* está cheño creatividade, esa fermosa capacidade que tanto e tanto axuda a fomentar o estudo das matemáticas. *Marco Numérico* non deixa de ser tamén unha bela obra de arte.

A álgebra computacional mestúrase coa estatística, coa física nuclear e cos números primos. E eu, que poucos máis primos coñezo que o número primo de Grothendieck, morro de ganas de ter na man o quinto episodio de *Marco Numérico*.

Pablo Cabanelas Eiras
Instituto Galego de Física de Altas Enerxía

Contribucións a conxetura de Hardy-Littlewood

Verde-Salgado, N.; Ferro-Pérez, J. M.;
Álvarez-Lodeiro, I.; Castro-Ramos, I.

IES MARCO DO CAMBALLÓN

Resumo. Con este artigo buscamos validar a coñecida conxetura de Hardy-Littlewood, esta postula que existe un número infinito de primos p tales que $(p+2)$ é un número primo. Para iso fixemos un código en R para poder xustificala.

Abstract. With this article we want to check the famous Hardy-Littlewood law. This conjecture postulates that exists an infinite number of primes p which $(p+2)$ is a prime number. For that we did a code in R to prove it.

1. Introducción

A conxetura de Hardy-Littlewood da que imos tratar é unha teoría formulada por Godfrey Harold Hardy (1877-1947) e John Edensor Littlewood (1885-1977), dous matemáticos británicos. Dita colaboración duraría 35 anos.

Godfrey Harold Hardy foi o principal valedor en Gran Bretaña e director das teses do matemático indio Ramanujan. Foi un matemático moi galardoado, con premios como o Premio Smith, Royal medal ou Medalla de Morgan. O máis importante da súa carreira foi esta conxetura que fixo con Littlewood. Ademais, el mesmo deu a supoñer que era homosexual, cousa que na súa época non se trataba coa liberdade coa que se trata agora.



A esquerda está G .H .Hardy e a dereita J. E. Littlewood

Denomínase *primo* a todo número natural maior que 1 que só ten dous divisores, o 1 e o propio número. Por outro lado denomínase *composto* a todo número natural que ten máis de dous divisores.

A partir destes conceptos, Hardy e Littlewood propuxeron esta conxectura:

Existe un número infinito de primos p tales que $(p + 2)$ tamén é primo.

En matemáticas, máis concretamente na teoría dos números, dous números primos (p, q) son números primos xemelgos se, sendo $q > p$, cúmplese $q - p = 2$. Os únicos números primos consecutivos son o 2 e o 3. A cuestión xorde de atopar dous números primos que sexan impares consecutivos, é dicir, que a diferenza do maior e menor sexa 2.

O primeiro en chamarlles así foi Paul Stäckel. A lei postula que dous números primos denomínanse xemelgos se un deles é igual ao outro máis dúas unidades, así pois, os números primos 3 e 5 forman unha parella de primos xemelgos. Outros exemplos de pares de primos xemelgos son 11 e 13 ou 41 e 43.

2. Verificación computacional da conxectura de Hardy-Littlewood

Para a seguinte toma de datos, empregaremos un código en R co que pretendemos recopilar os primos xemelgos do intervalo que buscamos estudar e representar os valores obtidos. Este algoritmo contará cos seguintes partes:

2.1 Función “primo”

Creamos unha función que analice números individualmente para determinar se son primos. Para isto analizamos os restos do resultado de dividir o número “n” entre os valores do intervalo: [2 , n-n/2]. Se ao comprobalos ningún dos restos é igual a 0 o número declarárase como primo. Isto faise gardando todos os restos nun vector do que se comprobarán todos os valores.

```
Primo <- function(n){
  var <- 0
  for (i in 2:(n-n/2)) {
    result <- n %% i
    if(result==0)
      var <- var+1
    else(
      var <- var+0
    )
  }
  if(var > 0) {
    "No"
  } else {
    "sí"
  }
}
```

2.2 Análise dos primos

Na seguinte acción deberemos determinar o intervalo de números que queremos analizar, ao cal chamaremos “p”. Unha vez feito isto

o programa analizará todos os números pertencentes a este coa función primo. No momento no que encontre un primo (i) analizará o número (i+2) e deste sendo primo declarado como un par de primos xemelgos. Para facilitar os cálculos en lugar de gardarse o intervalo (i , i+2) utilizaremos (i + 1) que representarao como o seu punto medio. Estos valores gardanse nun intervalo ao que chamamos: “Data”

```
data <- c()
for (i in 1:p){
  if(Primo(i) == "sí"){
    if(Primo(i+2) == "sí"){
      data <- append(data, i+1 )
    }
  }
}
```

2.3 Análisis das distancias

Con este programa tiñamos o obxectivo de comprobar se os pares de primos xemelgos eran infinitos. Para isto calculamos a Diferenza absoluta e a Diferenza relativa. O primeiro é o resultado de restar o valor $Data[n+1]-Data[n]$. Desta maneira podemos ver como aumenta ou diminúe a distancia entre primos. Mentres que na Diferenza relativa calculamos $(Data[n+2]-Data[n+1])/(Data[n+1]-Data[n])$, o resultado desta operación permítenos saber como aumenta ou diminúe a velocidade. Tamén crearemos unha variable co número de valores, ou pares de primos, xerados para o seu uso en representacións.

```

x <- length(data)
Dif.abs <- c()
for (z in 1:x) {
  calc <- (data[z+2]-data[z+1])/(data[z+1]-data[z])
  Dif.abs <- append(Dif.abs, calc )
}
Dif.rel <- c()
for (k in 1:x) {
  calc1 <- (data[k+1]-data[k])
  Dif.rel <- append(Dif.rel, calc1 )
}

```

2.4 Representación gráfica

Coas variables xeradas creamos gráficas nas que representan o eixe y. O eixe x é o intervalo xerado polos valores das variables usadas. Utilizamos ademáis o comando “abline” para obter a unha línea recta que describa a pendente resultante da función. O comando “lm” sírvenos para aproximar os valores conseguidos e facer a súa representación máis uniforme.

```

plot(1:x, Dif.abs, type = "l")
abline(lm(Dif.abs ~ 1:x), lwd = 2)

```

3. Análise dos datos

Usando o algoritmo anterior e os datos recopilamos xeramos distintas gráficas cas que estudamos a posición dos primos xemelgos así como as distancias entre eles dunha maneira máis visual. Desta maneira pretendemos obter información da súa tendencia e intentar extrapolala a valores cada vez maiores.

3.1 Intervalo [1 , 10000]

A primeira en ser analizada será unha gráfica que comprende a distancia absoluta dos 10000 primeiros números.

A distancia absoluta é a diferenza entre intervalos. Para determinar esta restamos o número par que se atopa entre os dous primos xemelgos, ou a media de ambos, menos o número par que se atopa no intervalo anterior . Por exemplo:

En $\{ 11, 13 \}$ o número que utilizaríamos sería igual a $(13 + 11)/2$ e no caso de $\{ 5, 7 \}$ o $(7 + 5)/2$. Con isto poderíamos dicir que o valor da distancia absoluta entre os dous intervalos sería igual a $12 - 6$ ou 6 .

No eixo de abscisas estúdase dende 0 até o 200, xa que dende o 0 até o 10000 hai pouco máis de 200 números primos xemelgos. Podemos observar na gráfica que hai unha tendencia de pequenos extremos relativos controlados, pero cando chegamos ao par de números primos xemelgos da posición 50, vemos un máximo relativo. Isto significa que o seguinte par de números primos vai estar moito máis alonxado que os demais (20 números).

Logo, vemos que a gráfica dende o 50 até o 200 replica a do intervalo (0,50), pero non chega aos extremos relativos tan altos como no intervalo anterior. Compréndense desde os 5 números até os 17.

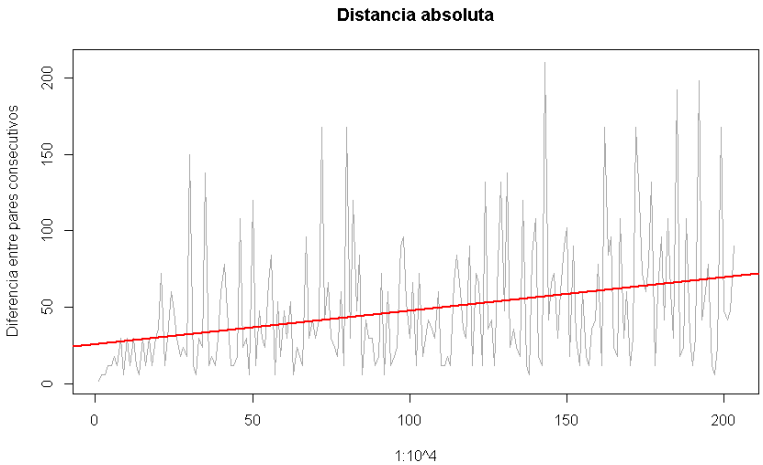


Fig.1. Gráfica desde o 0 ao 10000

A recta que atravesa a función representada na Fig.2 é a pendente media da mesma,e o seu valor é de 48.4, é decir, sube 48.4 unidades no eixe de ordenadas por cada unidade que aumenta no eixe de abscisas. Este valor obtémolo a partir de do cociente de Δy e Δx .

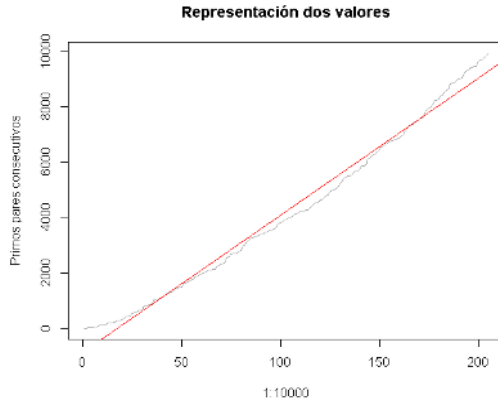


Fig. 2

3.2 Intervalo [1 , 100000]

Agora, contamos coa gráfica que abarca os 100000 primeiros números. Canto máis avanzamos nos números, podemos observar que a función tende a crecer.

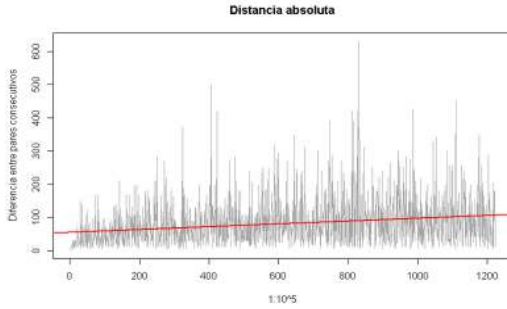


Fig. 3

Volvendo cos máximos e mínimos, obsevamos extremos como na función anterior [1,10000], pero a distancia media e os máximos que describe a función son moito máis altos que na anterior, chegando a máximos relativos de até 600 números. Para poder visualizalo empregáremos a Fig.3.

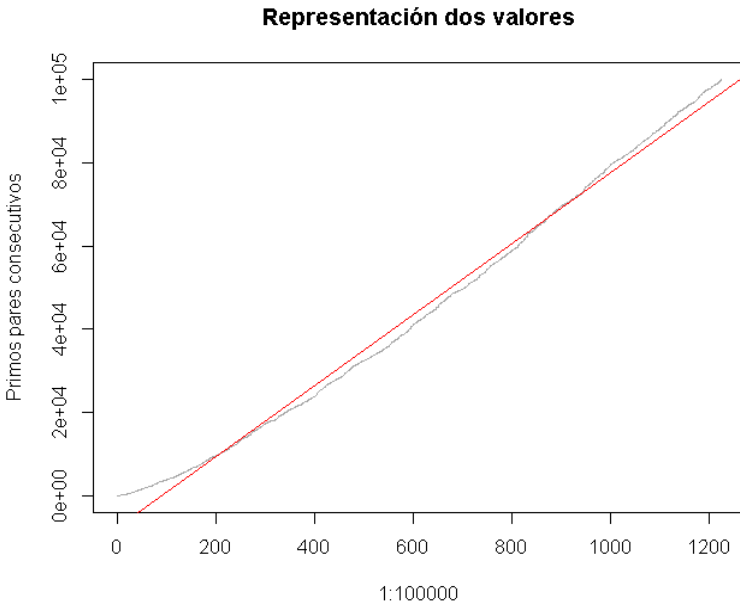


Fig. 4

A pendente xeral, que observamos na Fig.4 aumenta respecto do intervalo anterior sendo igual a 81.7 e 1.69 veces maior.

3.3 Tendencia xeral

Seguimos agora co intervalo comprendido desde o 1 ao $2.26 \cdot 10^6$, utilizamos este intervalo para estudar a tendencia xeral xa que é ao maior intervalo ao que puidemos chegar.

Primeiro temos a Fig.5, gráfica na que vemos os pares de primos xemelgos representados:

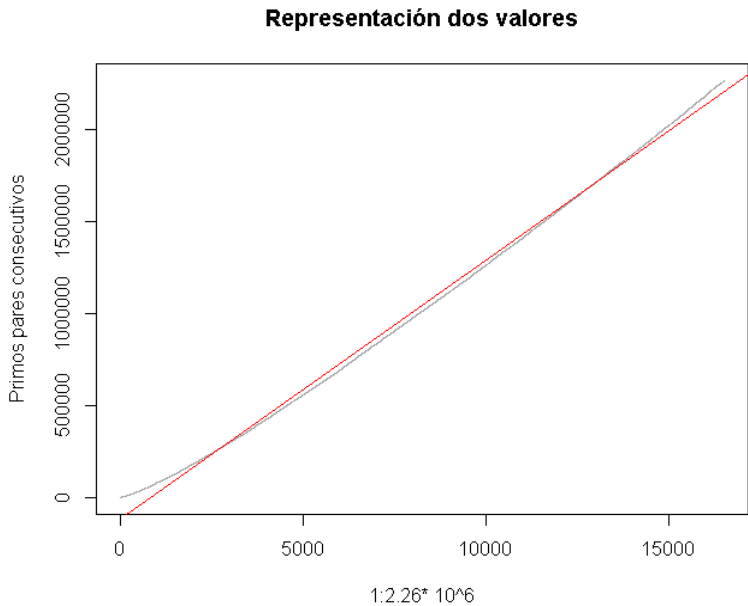


Fig. 5

Vemos que a función toma unha pendente media equivalente a 137. Isto mostranos como en comparación aos resultados anteriores, concretamente dos intervalos $[1, 10000]$ e $[1, 100000]$ vemos que o crecemento é menos pronunciado. Concretamente se comparamos

o intervalo $[1,10000]$ e $[1,100000]$ vemos que o crecemento aumenta 1.69 veces tendo en conta que os valores utilizados son dez veces máis grandes no caso do intervalo $[1, 100000]$ resultando nun aumento por unidade de 0.169. Mentres que se comparamos os intervalos $[1, 100000]$ e $[1, 2.26 \cdot 10^6]$ o crecemento ten unha razón cercana a do caso anterior sendo 1.68 pero cun aumento na magnitude de valores de 22.6 cun aumento de 0.074.

Cabe destacar que no comenzo da función a función adopta unha pendente que non concorda ca recta que representa a pendente media da función. Isto é algo lóxico ,xa que nos primeiros números estudados as distancias entre os números pares primos eran moito menores,esto podémolo ver reflexado no intervalo $[0, 10000]$. Tamén podemos observar que ao final da gráfica a función parece replicar o que pasa ao principio, pero non o podemos asegurar xa que non poidemos estudar por encima do $2.26 \cdot 10^6$.

Na seguinte gráfica Fig.6 analizamos a distancia absoluta. A diferenza entre valores aumenta de maneira constante e cunhas valores que se manteñen aproximadamente por debaixo dos 500 números, non se atopan signos que fagan pensar que estea habendo cambios na mesma. Temos diferentes extremos relativos pero sen ocupar posicións determinantes, é decir, non é ao que tende a gráfica, pero a cada paso que aumentamos de cifra os máximos son casa vez maiores.

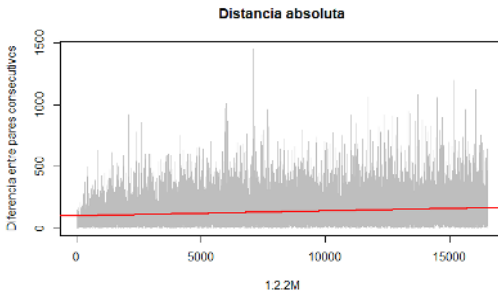


Fig. 6

Na última representación Fig.7 que é a que corresponde ca distancia relativa, é dicir, o cociente entre o calquera valor da distancia absoluta e o seu anterior. Desta maneira obsevamos que a pendente media tende a 1, esto confirmanos que a pesar da existencia de extremos relativos de gran magnitude a tendencia xeral é que os valores non aumentan significativamente entre sí.

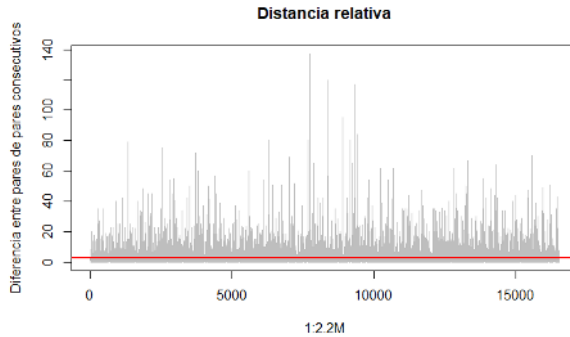


Fig.7. Distancia relativa

4. Conclusións

Os números primos contan cunhas propiedades fascinantes que por agora non podemos analizar de maneira precisa. Polo momento, non podemos asegurar que haxa números primos até o infinito, xa que como vimos no noso análise as distancias entre eles non seguen un patrón definido, non obstante, decatámonos de que estas tenden a aumentar de forma xeral describindo unha pendente que é inversamente proporcional a magnitude do valor que estudemos, é dicir, canto máis grande é o valor a pendente aseméllase máis a unha paralela co eixe de abscisas como nos demostrou o estudo das Fig. 2 , 4 e 5.

En conclusión, non podemos asegurar que se cumpra esta conxectura debido a aparente aletoriedade da distancia entre primos xemelgos, pero non atopamos ningunha proba do contrario, polo que, baseandonos no noso estudo podería ser certa.

5. Bibliografía / Webgrafía

<https://vixra.org/pdf/1406.0025v2.pdf> (Ramón Ruiz) Año de publicación: 2015 (Recuperado o 14/10/2021).

http://wikipedia.es.nina.az/N%C3%BAmeros_primos_gemelos.html
(Recuperado o 28/10/21)

https://es.wikipedia.org/wiki/Godfrey_Harold_Hardy (Recuperado o 28/10/21)

https://es.wikipedia.org/wiki/John_Edensor_Littlewood (Recuperado o 28/10/21)

[https://www.google.es/search?](https://www.google.es/search?q=hardy+y+littlewood&rlz=1C1GCEA_enES843ES843&source=lnms&tbn=isch&sa=X&ved=2ahUKEwj8u_LDx6H0AhXx4uAKHellDQUQ_AUoAXoECAEQAw&biw=1600&bih=789#imgsrc=Tfj_7kFseXcLBM&imgdii=dshz8zDQ233vIM)

[q=hardy+y+littlewood&rlz=1C1GCEA_enES843ES843&source=lnms&tbn=isch&sa=X&ved=2ahUKEwj8u_LDx6H0AhXx4uAKHellDQUQ_AUoAXoECAEQAw&biw=1600&bih=789#imgsrc=Tfj_7kFseXcLBM&imgdii=dshz8zDQ233vIM](https://www.google.es/search?q=hardy+y+littlewood&rlz=1C1GCEA_enES843ES843&source=lnms&tbn=isch&sa=X&ved=2ahUKEwj8u_LDx6H0AhXx4uAKHellDQUQ_AUoAXoECAEQAw&biw=1600&bih=789#imgsrc=Tfj_7kFseXcLBM&imgdii=dshz8zDQ233vIM)
Imaxe dos dous matemáticos(recuperada o 18/11/21)

Conxectura de Collatz: abordaxe computacional e estudo de ciclos

*Fraiz-Casas, A.; López-Calviño, E.;
Puga-Silva, L.*

IES MARCO DO CAMBALLÓN

Resumo. No seguinte estudo analízase a conxectura de Collatz. Para isto, deseñamos varios algoritmos empregando o programa de análise estatística R. Unha vez plantexados, centrámonos no estudo da lonxitude dos ciclos deste problema.

Abstract. The following research analyzes the Collatz conjecture. Therefore, we designe some programs by using the estadistic analysis programme R. When we have done it, we have to focus in the study of cicles length of this problem.

1. Introducción

A conxectura de Collatz, tamén coñecida como $3n+1$, foi formulada polo matemático alemán Lothar Collatz (1910-1990). Traballou en distintas universidades como a Universidade de Siracusa (Nova York). En 1937 propón esta conxectura que, na actualidade, aínda segue sen resolver por completo, aínda que conta con avances parciais e expúxoos ante diferentes seminarios. Lothar compartiu esta cuestión a varios matemáticos como Hasse.

Collatz decatouse da seguinte cuestión: se tomamos un número natural calquera (n), sendo este un número par, divídese entre 2. Mentres que se n é un número impar, multiplícase por 3 e sumáselle 1.

$$\left\{ \begin{array}{l} \frac{n}{2} \text{ se } n \text{ é par} \\ 3*n+1 \text{ se } n \text{ é impar} \end{array} \right.$$

Fig. 1. Cálculo da conxectura

Ao reiterar este proceso, as veces que sexa necesario, decatámonos de que a secuencia de Collatz consiste en que as iteracións sempre rematan en 1. Poñamos un exemplo, se tomamos o número 1221, a secuencia de Collatz é a seguinte:

1221 , 3664 , 1832 , 916 , 458 , 229 , 688 , 344 , 172 , 86 , 43 ,
 130 , 65 , 196 , 98 , 49 , 148 , 74 , 37 , 112 , 56 , 28 , 14 , 22 , 11 ,
 34 , 17 , 52 , 26 , 13 , 40 , 20 , 10 , 5 , 16 , 8 , 4 , 2 , 1.

Poñamos outro caso, esta vez cun número par, o 824 :

824 , 412 , 206 , 103 , 310 , 155 , 466 , 233 , 700 , 350 , 175 ,
 526 , 263 , 790 , 395 , 1186 , 593 , 1780 , 890 , 445 , 1336 , 668 ,
 334 , 167 , 502 , 251 , 754 , 377 , 1132 , 566 , 283 , 850 , 425 , 1276 ,
 638 , 319 , 958 , 479 , 1438 , 719 , 2158 , 1079 , 3238 , 1619 , 4858 ,
 2429 , 7288 , 3644 , 1822 , 911 , 2734 , 1367 , 4102 , 2051 , 6154 ,
 3077 , 9232 , 4616 , 2308 , 1154 , 577 , 1732 , 866 , 433 , 1300 ,
 650 , 325 , 976 , 488 , 244 , 122 , 61 , 184 , 92 , 46 , 23 70 , 35 , 106 ,
 53 , 160 , 80 , 40 , 20 , 10 , 5 , 16 , 8 , 4 , 2 , 1.

Como podemos observar, estas secuencias rematan en 4 2 1, o que confirma o enunciado desta conxectura.

Nos anos sesenta, en plena guerra fría, estendeuse o murmuro de que esta conxectura era parte dun complot soviético para frear outras investigacións de maior importancia dos EE.UU. Tras décadas sen avances, en setembro de 2019, o matemático Terence Tao publicou un novo resultado significativo. Tao estableceu unha serie de conexións entre o problema e a área das ecuacións diferenciais

parciais, chegando así a unha nova mostra das ramificacións desta conxectura.

Os cálculos indican que as secuencias sempre rematan en $4 \cdot 2^k$. Asegura isto que sempre sexa así para calquera número independentemente do número de partida?

2. Verificación computacional da conxectura de Collatz

A partir deste achado, queremos comprobar este fenómeno de xeito descritivo mediante o programa de análise estadístico R, que nos permite elaborar algoritmos para analizar o problema. Para realizar o noso algoritmo que nos permita analizar esta conxectura necesitamos dun algoritmo auxiliar que se encargará de achar a paridade do n .

2.1 Función “par”

En primeiro lugar, creamos a función *par* tal que se o resto da división entre dous é cero, n é par e o programa devolvenos o resultado de $a=1$. Mentres que no caso de que o resto non sexa cero, o resultado de a é igual a 0. Para que o programa realice estas operacións, é necesario empregar a función auxiliar %% que calcula o resto da división para así poder determinar se o número é par ou impar.

```
par<-function(n){  
  if(n%%2==0){a<-1}  
  else{a<-0}  
  return(a)}  
}
```

Fig. 2. Estrutura da función “par”

2.2 Función “co”

A continuación, creamos outro programa que nos dará cantas iteracións que leva a cabo para chegar ao final ao número 1. Para iniciar este código, créanse dúas variables. Un vector a que comeza en n e unha variable b que será n que irá mudando ao longo do programa. Deseguido, esta estrutura entra nun bucle *repeat* onde irá calculando as iteracións. Neste momento, empregamos a función “par” creada anteriormente para poder determinar se n é par ou impar. Se o número é par o programa debeo dividir entre dous, mentres que se é impar debe multiplicar por tres e sumarlle un. Este resultado é introducido no vector a co comando *append* coa finalidade de que ao rematar obteñamos todas as iteracións que se gardarán na variable b como o resultado da iteración. Por último, cando b sexa 1 o bucle detense e o programa devolve o vector a con todas as iteracións.

```
co<-function(n){
  a<-c(n)
  b<-n
  repeat{
    if(b==1){break}
    else if(par(b)==1){
      a<-append(a, (b/2))
      b<-(b/2)}
    else{
      a<-append(a, (3*b+1))
      b<-((3*b)+1) }}
  return(a)
}
```

Fig. 3. Estrutura da función “co”

2.3 Función “co” modificada

Unha vez finalizado este programa, é necesario crear un programa semellante que utilice o comando *length* para precisar

cantas iteracións se levaron a cabo para completar a secuencia.

```

co<-function(n){
  a<-c()
  b<-n
  repeat{
    if(b==1){break}
    else if(par(b)==1){
      a<-append(a,(b/2))
      b<-((b/2))}
    else{
      a<-append(a,(3*b+1))
      b<-((3*b)+1) }}

  f<-length(a)
  return(f)
}

```

Fig. 4. Estrutura da función “co” modificada

2.4 Función “gr”

Para concluir esta análise, creamos o programa que nos vai devolver a gráfica co número e as iteracións. Comezaremos creando dous vectores para as coordenadas x e y , o vector x que gardará os números e o vector y onde se almacenarán o número de iteracións. Despois, entramos nun bucle no que estas coordenadas se irán introducindo co comando *append* no vector x , o número (i). Mentres que no vector y , da mesma maneira, o número de iteracións. Para iso, utilizamos a función anterior que calcula a lonxitude das iteracións. Unha vez computamos as iteracións dos números menores que o introducido no programa, crea a gráfica cos vectores de coordenadas utilizando o comando *plot*. Dentro deste construiremos a nosa gráfica establecendo o título co comando *main*; o que se representa en x co comando *xlab*; o que se representa en y co comando *ylab*; a cor dos puntos da gráfica co comando *col* mentres que o comando *pch* escribiraos un punto e por último o comando *cex* que utilizamos para definir o grosor deste.

```
gr<-function(n){  
  y<-c()  
  x<-c()  
  for(i in 2:n){  
    x<-append(x,i)  
    y<-append(y,co(i))  
  }  
  plot(x, y,main="Collatz",xlab="Numeros",ylab="Iteracions",col="blue",pch=16,cex=0.01)}
```

Fig. 5. Estrutura da función “gr”

3. Estudo da lonxitude dos ciclos de Collatz

Coa a análise de datos feita anteriormente, podemos determinar que para cada número a lonxitude da secuencia varía. Para comprobar este feito, decidimos empregar gráficas que amosen dun xeito máis descritivo o que acontece.

Comezamos facendo unha distinción entre os números pares e impares co obxectivo de observar as diferenzas no número de iteracións. Iniciamos este proceso analizando a gráfica dos números pares, no eixe das X colocamos a variable “números” mentres que no eixe das Y colocamos a variable “iteracións” e tomamos o intervalo de números do 0 ao 10000. Na gráfica podemos observar que a medida que os números medran as iteracións van aumentando ata chegar a números menores que 250, porén os que están máis próximos a 0 procesáanse cun número menor de iteracións.

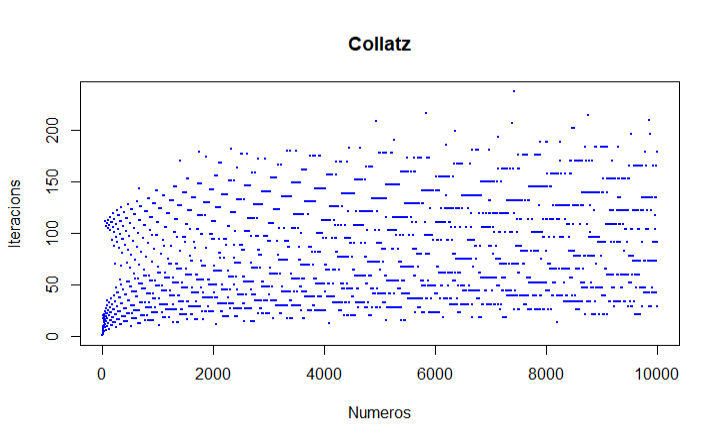


Fig. 6. Gráfica números pares

A seguinte gráfica ilustra o que acontece nos números impares tomando o mesmo intervalo (0,10000) e as mesmas variables. Podemos observar que os números máis próximos a 0 teñen un número semellante de iteracións que van aumentando a medida que os números se incrementan chegando a superar as 250.

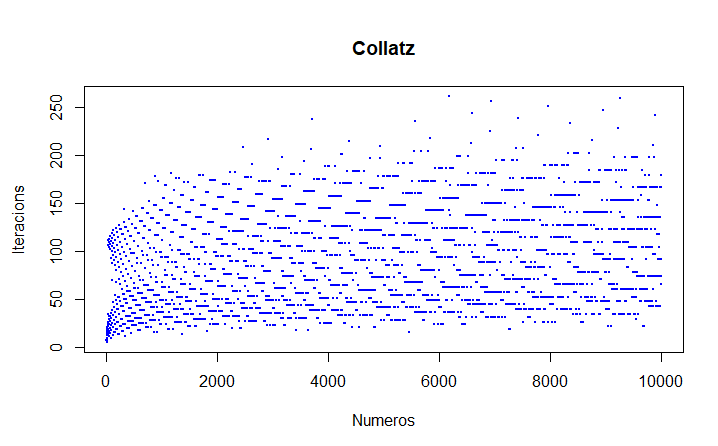


Fig. 7. Gráfica números impares

A diferenza máis notable entre estas dúas gráficas é que nos números máis grandes as iteracións dos números impares son maiores que nos números pares.

Pasamos agora a confrontación entre as gráficas dos números primos e compostos. A primeira amosa o que acontece nos números primos tomando as mesmas variables e o intervalo (0,1000). Podemos comprobar que estes reparten o número de iteracións formando dúas curvas que se distribúen homoxeneamente.

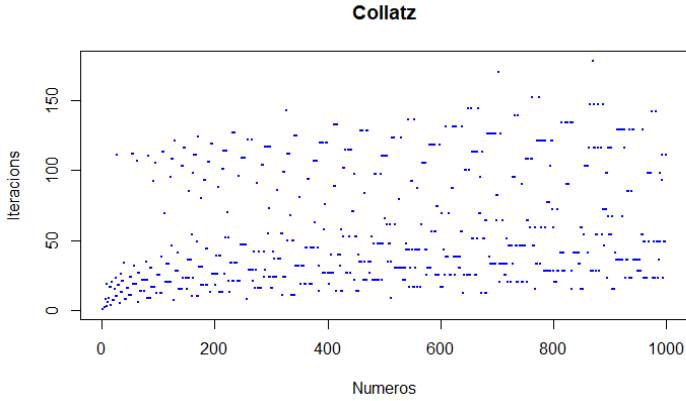


Fig.8. Gráfica de números primos

Na seguinte gráfica, a dos números compostos, tomamos as mesmas variables e o mesmo intervalo. Podemos comprobar que estes reparten o número de iteracións formando dúas curvas que quedan distribuídas de forma desigual.

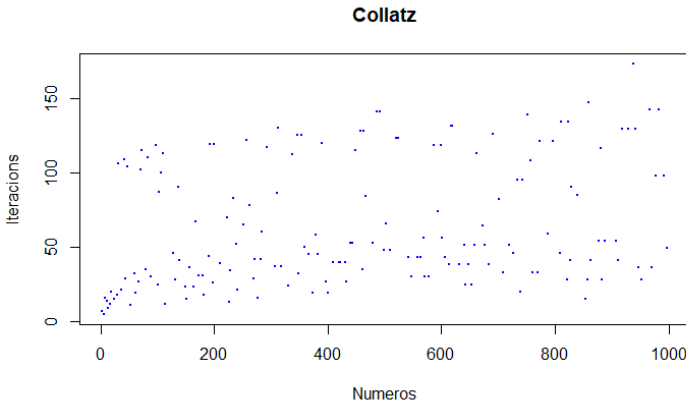


Fig.9. Gráfica de números compostos

Comparando estas dúas gráficas podemos observar que nos números compostos a tendencia é a de ter un número menor de iteracións, porén nos números primos repártese de maneira homoxénea formando dúas curvas ben diferenciadas. Á vista da gráfica, as iteracións nos números compostos son menores.

Para rematar, relacionamos as gráficas con funcións elementais coa fin de establecer unha conexión entre estas. Deste xeito, chegamos a que a función logarítmica ten unha forma similar ao comezo da gráfica polo que, ao representar a función logaritmo aproximamos a base para que a forma sexa o máis semellante posible. Chegamos a conclusión de que a base que mellor se adapta é π/e (1,1557).

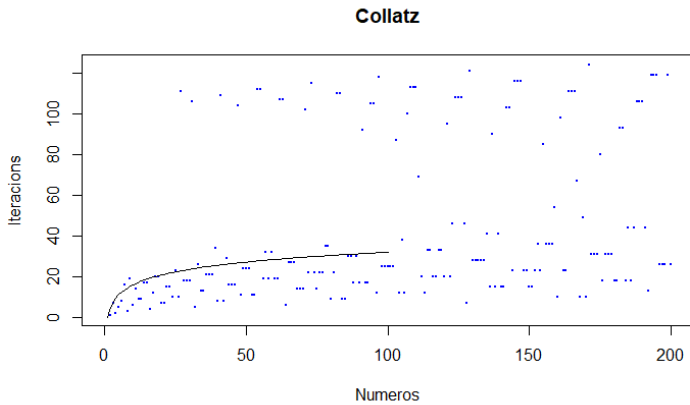


Fig.10.Gráfica de comparación co logaritmo

4. Conclusións

Neste estudo decatámonos do complicadas que poden chegar a ser as matemáticas a través dun problema aparentemente sinxelo pero cun transfondo de grandes dimensións. Queremos destacar que conseguimos facer un avance nesta conxectura, descubriendo a semellanza que as gráficas do número de iteracións teñen en valores próximos a 100 coa función logarítmica. O logaritmo que mellor se adapta a esta forma é o $\log_{1.1557} x$, no que curiosamente a base é

π/e . Este progreso non está evidenciado en ningún outro artigo científico.

5. Bibliografía / Webgrafía

El País (2020). *Primer avance en décadas en un problema aparentemente imposible*. Recuperado o día 17 de novembro de 2021, do sitio Web:

https://elpais.com/elpais/2020/01/08/ciencia/1578499346_232520.html

ABC (2021). *La conjetura de Collatz: un problema «sencillo» que desafiará tu intuición*. Recuperado o día 17 de novembro de 2021, do sitio Web:

https://www.abc.es/ciencia/abci-conjetura-collatz-problema-sencillo-desafiara-intuicion-202101180110_noticia.html

Investigación y ciencia (2019). *Un gran resultado matemático para un «problema peligroso»*. Recuperado o día 17 de novembro de 2021, do sitio Web:

<https://www.investigacionyciencia.es/noticias/un-gran-resultado-matemtico-para-un-problema-peligroso-18125>

La ciencia de la Mula Francis (2020). *Los primos de la Conjetura de Collatz*. Recuperado o día 17 de novembro de 2021, do sitio Web:

<https://francis.naukas.com/2020/02/03/los-primos-de-la-conjetura-de-collatz/>

Gaussianos (2006). *La conjetura de Collatz*. Recuperado o día 17 de novembro de 2021, do sitio Web:

<https://www.gaussianos.com/la-conjetura-de-collatz/c>

Entre primos. A Conxectura de Brocard

*Casal-Blanco, R.; González-Castro, M.;
Pacheco-Fariña, M.F.; Torres-García, A..*

IES MARCO DO CABBALLÓN

Resumo. O obxectivo deste artigo é o estudo da conxectura de Brocard, que asegura un mínimo de números primos entre dous primos consecutivos ao cadrado, mediante a creación de varios códigos en R.

Abstract. The aim of this article is to study Brocard's conjecture, which claims a minimum of prime numbers between two squared consecutive prime numbers, by creating some codes in R program.

1. Introducción

A palabra *conxectura* fai referencia á unha hipótese que xorde de indicios pero que non está probada. A conxectura tratada neste artigo xira arredor dos *números primos*; aqueles que só teñen dous divisores: o 1 e o propio número. Deste xeito, sobre os números primos existen diversas conxecturas, entre as que destacan os problemas de Landau.

A análise da conxectura de Brocard partirá dos problemas de Landau, propostos por Edmund Landau, entre os que podemos distinguir a conxectura de Legendre, que presenta unha estreita relación coa nosa. En primeiro lugar, Edmund Georg Hermann Landau foi un matemático alemán



Fig. 1. Fotografía de Edmund Landau

que ao longo da súa vida mostrou interese na teoría de números (propiedades dos números) e na análise complexa (investigación de funcións analíticas), á vez que propuxo catro problemas sobre números primos coñecidos como os problemas de Landau (1912), entre os que podemos distinguir:

- A conxectura de Brocard.
- A conxectura dos números primos xemelgos.
- A conxectura sobre primos da forma $(n^2 + 1)$.
- A conxectura de Legendre.

Esta última, a conxectura de Legendre, foi enunciada por Adrien-Marie Legendre, físico e matemático francés que ao longo da súa vida contribuíu á estatística, á teoría de números, á alxebra e á análise matemática, na que afirma que dado un número enteiro maior ca cero ($n > 0$), existe entre o seu cadrado n^2 e o cadrado do seu consecutivo, $(n+1)^2$, polo menos un número primo independentemente de cal sexa o número natural proposto:



Fig. 2. Fotografía de Henri Brocard

\exists número primo entre n^2 e $(n+1)^2$

O resultado máis relevante obtido desta conxectura está da man do matemático Chen Jingrun, quen en 1965 demostrou que nestas condicións existe sempre un número primo ou semiprimo, é dicir, un número natural que é produto de dous primos non necesariamente distintos.

Esta conxectura tamén se demostrou en números grandes, porén o problema segue aberto.

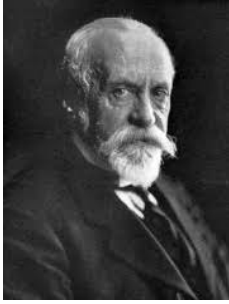


Fig. 3. Fotografía de Henri Brocard

A conxectura de Brocard leva este nome na honra de Henri Brocard; matemático francés que se adicou á xeometría, especialmente ao estudo do triángulo e quen á súa vez, propuxo a hipótese baseada na existencia de polo menos catro números primos entre p_n^2 e p_{n+1}^2 , se $n > 1$; onde p_n é o n -ésimo número primo, isto é, un número primo calquera e onde p_{n+1} é o primo seguinte.

\exists mínimo 4 números primos entre p_n^2 e p_{n+1}^2

Expresada matemáticamente sería:

$\Pi_b \geq 4$ entre n^2 y $(n+2)^2$, donde $\Pi_b(n)$ é a cantidade de números primos entre os dous cadrados, de forma que:

$$\Pi_b(n) = \Pi((n+2)^2) - \Pi(n^2) \geq 4$$

$\Pi((n+2)^2)$ y $\Pi(n^2)$ son a cantidade de números primos entre n^2 y $(n+2)^2$, mentras que $\Pi_b(n)$ é a cantidade de primos entre a diferenza das funcións anteriores.

2. Análise dos datos

Para estudar a conxectura de Brocard, creamos varios programas en R que nos permitiron probar se se cumpría.

En primeiro lugar, creamos un programa que nos permitise detectar números primos. Para iso empregamos a variable “cont” na que imos gardar os divisores do número que estamos analizando. A continuación dividimos de forma enteira

(%%) o número entre tódolos comprendidos entre o 1 ata el mesmo coa función *for*. Por último, se só hai dous divisores gardados na variable “cont” significa que o número é primo e entón pedimos que o devolva coa función *return*.

```
primo<-function(numero){
  cont<-0
  for(i in 1:numero){
    x<-numero%%i
    if(x==0){
      cont<-cont+1
    }
  }
  if(cont==2){
    return(numero)
  }
}
```

Fig. 4. Código para detectar primos

Agora precisamos un programa que busque o número primo seguinte ($p_{n+p_{a1}}$) do número primo que queremos analizar (Fig.5). Para iso creamos en primeiro lugar un vector “v” baleiro onde poder gardar os posibles números primos. Entón pedimos que detecte co programa “primo”, o código que creamos anteriormente, os primos existentes entre n (o número a analizar) e $n+100$ (un valor arbitrario para poñer como límite) co bucle *for*. Por último, pedimos que garde os primos atopados no noso vector “v” coa función *append*, que nos permite engadir novos valores no vector mantendo os xa existentes. Pedímoslle tamén que nos devolva o segundo primo atopado con *return* e “v[2]” para indicar a segunda posición, xa que o primeiro número do vector sería o propio número n .

```

seguinte<-function(n){
  v<-c()
  for(i in n:(n+100)){
    v<-append(v,primo(i))}
  return(v[2])
}

```

Fig. 5. Código para detectar o primo seguinte

Unha vez que xa somos capaces de calcular o primo seguinte ao que estamos analizando faltaríanos calcular os números primos que existen entre os cadrados de ambos (Fig.6). Primeiro creamos de novo un vector “v” onde poder gardar os primos existentes. A continuación, de novo usando o bucle *for*, empregamos o programa “primo” para tódolos números existentes entre o noso número a analizar p_n^2 e o primo seguinte p_{n+1}^2 ; o que nos permitirá obter os números primos que estamos buscando. Para poder engadir valores novos ao vector sen eliminar os xa existentes usamos outra vez a función *append*. Finalmente, pedimos ao programa que nos devolva a lonxitude do vector “v” (é dicir, a cantidade de números que gardou) coa función *length*.

```

m<-function(n){
  v<-c()

  for(i in (n^2):(seguinte(n)^2)){
    v<-append(v,primo(i))}
  return(length(v))
}

```

Fig. 6. Código 3

Para poder traballar con varios números simultaneamente imos implementar un novo programa (Fig.7). Para iso creamos en primeiro lugar un vector ao que chamamos “pr”. Nel introducimos unha lixaxe cos 300 primeiros números primos. O programa daranos a escoller con cantos números primos queremos traballar. Para isto debemos usar a función *readline* e introducir a pregunta entre aspas ao tratarse dunha cadea de texto. Por exemplo, se queremos traballar cos oitenta primeiros números, o programa preguntaranos “Introduce o total de números cos que queres traballar” e deberemos introducir o número 80. Esta resposta o programa debe gardala como un número e non como unha cadea de texto, polo que debemos usar a función *as.integer* antes do *readline*.

Agora imos crear dous novos vectores: o vector “espello”, onde imos gardar o número primo co que está a traballar o programa nese momento; e o vector “resultado”, onde se van almacenar os primos existentes entre o número a analizar ao cadrado e o seguinte primo o cadrado. Para facer isto creamos outro bucle *for* que vaia desde o 1º número do vector “pr” (o primeiro primo) ata o número que nós introducimos previamente no programa (o último número primo que queremos analizar). Dentro deste bucle gardamos na posición “i” (a posición do primo co que estamos a traballar) o número de primos que existen dentro do vector “resultado” e o primo co que estamos a traballar dentro do vector “espello”. Por exemplo, se estamos traballando co número 5, que é o 3º número primo, gardarase o 5 na terceira posición do vector “espello” e o 6, que son os primos obtidos, na terceira posición do vector “resultado”.

Finalmente pedimos que nos amose os vectores “espello” e “resultado” coa función *print*.

```

numero<-as.integer(readline('Introduce o total de números que queres: '))
pr<-c(2,3,5,7,11,13,17,19,23,29,31,37,41,43,47,53,59,61,67,71,73,
79,83,89,97,101,103,107,109,113,127,131,137,139,149,151,157,163,167,173,
179,181,191,193,197,199,211,223,227,229,233,239,241,251,257,263,269,271,
277,281, 283,293,307,311,313,317,331,337,347,349,353,359,367,373,379,383,
389,397,401,409,419,421,431,433,439,443,449,457,461,463,467,479,487,491,
499,503,509,521,523,541,547,557,563,569,571,577,587,593,599,601,607,613,
617,619,631,641,643,647,653,659,661,673,677,683,691,701,709,719,727,733,
739,743,751,757,761,769,773,787,797,809,811,821,823,827,829,839,853,857,
859,863,877,881,883,887,907,911,919,929,937,941,947,953,967,971,977,983,
991,997,1009,1013,1019,1021,1031,1033,1039,1049,1051,1061,1063,1069,1087,
1091,1093,1097,1103,1109,1117,1123,1129,1151,1153,1163,1171,1181,1187,
1193,1201,1213,1217,1223,1229,1231,1237,1249,1259,1277,1279,1283,1289,
1291,1297,1301,1303,1307,1307,1319,1321,1327,1361,1367,1373,1381,1399,
1409,1423,1427,1429,1433,1439,1447,1451,1453,1459,1471,1481,1483,1487,
1489,1493,1499,1511,1523,1531,1543,1549,1553,1559,1567,1571,1579,1583,
1601,1607,1609,1613,1619,1621,1627,1637,1657,1663,1667,1669,1693,1697,
1699,1709,1721,1723,1733,1741,1747,1753,1759,1777,1783,1787,1789,1801,
1811,1823,1831,1847,1861,1867,1871,1873,1877,1879,1889,1901,1907,1913,
1931,1933,1949,1951,1973,1979,1987)
resultado<-c()
espello<-c()
for(i in 1:numero){
  resultado[i]<-m(pr[i])
  espello[i]<-pr[i]
}
print(espello)
print(resultado)

```

Fig. 7. Código 4

Como resultado deste programa temos dúas listas como as que aparecen a continuación. Collimos de exemplo os 16 primeiros números primos e obtivemos os seguintes resultados:

Número estudado : 2 3 5 7 11 13 17 19 23 29 31 37 41 43 47

Cantidade de primos obtidos : 2 5 6 15 9 22 11 27 47 16 57 44 20 46 80

Na primeira liña atopamos os 15 primeiros números primos e debaixo de cada un aparece o total de números primos entre p_n^2 e p_{n+1}^2 correspondente.

Para facer máis visibles os resultados creamos un código (Fig.8) que elaborase unha gráfica de novo co programa *R*. Optamos por realizar gráficas coa función *plot*. No eixo de abscisas representaremos o vector “espello” e no eixo de ordenadas “resultado”. Para poñer título á gráfica usamos a función *main* e para o título de cada eixo as funcións *xlab* e *ylab* respectivamente. Como no eixo de abscisas precisamos poñer unha expresión con subíndices e expoñentes temos que usar a función *expression* combinada con *paste* para poder combinar

estas expresi3ns con texto normal. Para cambiar o tipo de gr3fica usamos *type*. N3s escollemos a opci3n *o*, que permite unir os puntos con li3as. O tipo de punto seleccion3molo con *pch*. Na nosa gr3fica imos usar o tipo *21*. Para cambiar as cores empregamos d3as funci3ns: *bg* para a cor dos puntos (escollemos o vermello) e *col* para a cor da li3a (escollemos o negro). As funci3ns *cex* e *lwd* perm3tennos cambiar o tama3o tanto dos puntos como do grosor das li3as. Por 3ltimo, para reducir o tama3o dos t3tulos dos eixos usamos *cex.lab*.

```
plot(espello, resultado, type = 'o',
     pch = 21, bg= 'red', col = 'black',
     cex = 1.5, lwd = 1.5,
     main = "Comprobaci3n conxectura de Brocard",
     xlab = "n3meros primos",
     ylab = expression(paste("primos entre ", p[n]^2, " e ", p[n+1]^2))
     , cex.lab = 0.75)
```

Fig. 8. C3digo para elaborar as gr3ficas

Para comezar, imos probar os c3digos cos 15 primeiros n3meros primos (Fig.9), para os cales obtivemos os seguintes resultados:

N3mero primo : 2 3 5 7 11 13 17 19 23 29 31 37 41 43 47
 Resultados obtido : 2 5 6 15 9 22 11 27 47 16 57 44 20 46 80

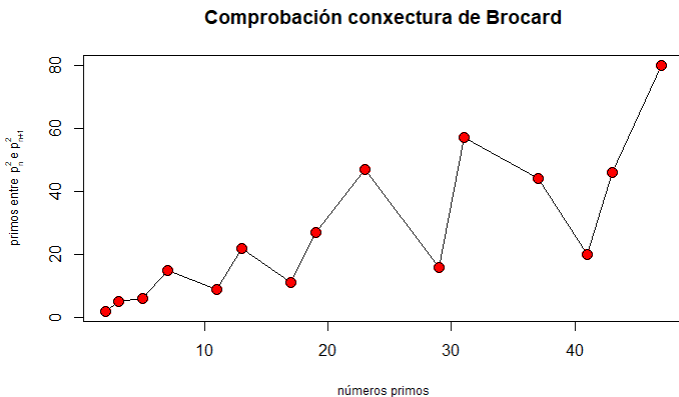


Fig.9. Comprobaci3n cos 15 primeiros n3meros.

Tras os resultados obtidos podemos observar como a conxectura non se cumpre no primeiro número analizado: o 2, que tan só ten dous números primos entre o seu cadrado (4) e o cadrado do primo seguinte (9). Nos números seguintes a conxectura si que se cumpre e cunha tendencia que semella ascendente, é dicir, cada vez existen máis números primos entre p_n^2 e p_{n+1}^2 . Non obstante, é preciso estudar máis números para poder chegar a unha conclusión fiable.

Con esta finalidade elaboramos unha nova gráfica introducindo os 100 primeiros números primos, para os cales obtivemos os seguintes resultados:

Primos estudados:

2 3 5 7 11 13 17 19 23 29 31 37 41 43 47 53 59
61 67 71 73 79 83 89 97 101 103 107 109 113 127
131 137 139 149 151 157 163 167 173 179 181 191
193 197 199 211 223 227 229 233 239 241 251 257
263 269 271 277 281 283 293 307 311 313 317 331
337 347 349 353 359 367 373 379 383 389 397 401
409 419 421 431 433 439 443 449 457 461 463 467
479 487 491 499 503 509 521 523 541

Resultado correspondente:

2 5 6 15 9 22 11 27 47 16 57 44 20 46 80 78 32
90 66 30 106 75 114 163 89 42 87 42 100 354 99 165
49 299 58 182 186 128 198 195 76 . 356 77 144 75
463 479 168 82 166 270 90 438 275 274 292 91 292
199 99 512 735 220 107 215 784 341 579 125 241
363 489 381 380 252 394 530 262 531 670 151 709
151 424 290 430 599 305 145 294 934 612 313 654
318 483 995 165 1513 498

Comprobación conxectura de Brocard

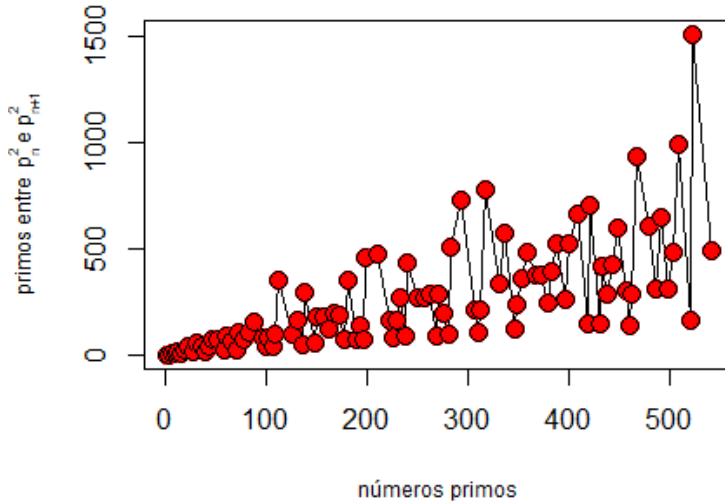


Fig.10. Comprobación cos 100 primeiros números.

Así, podemos observar de novo unha tendencia ascendente na que a cantidade de números primos entre p_n^2 e p_{n+1}^2 é cada vez maior e para todos os números a partir do 2, maior que 4.

3. Conclusións

O estudo da conxectura de Brocard permitiunos obter unha serie de conclusións. Para o primeiro número estudado o 2, observamos que non se cumpre que a cantidade de números primos entre 2^2 e 3^2 , ambos números primos consecutivos, sexa maior que 4 (tan só o 5 e o 7). Por outra parte, malia só poder estudar un máximo de 100 números por limitación computacionais, puidemos observar, tanto a partir da primeira gráfica (Fig. 9) como da segunda (Fig.10), que a cantidade de números primos entre p_n^2 e p_{n+1}^2 , tende a ser maior a medida que incrementamos o número primo estudado. Por iso, a

conxectura semella que se vai seguir cumprindo co resto de números primos.

Ademais, sería interesante poder chegar á expresión matemática que nos permitise saber como se comporta a gráfica para poder así prever o que pode suceder nos números non estudados.

4. Webgrafía

- Roldán Martínez, A. (2017). Conjeturas. Recuperado o 20 de novembro de 2021 do sitio web: <http://hojamat.es/publicaciones/conjeturas.pdf>
- Roldán Martínez, A. (2014). Números y hoja de cálculo. Recuperado o 20 de novembro de 2021 do sitio web: <http://hojaynumeros.blogspot.com/2014/05/>
- Porras Ferreira, J.W. (2011). Una aproximación a la solución de la conjetura de Brocard. Recuperado o 20 de novembro de 2021 do sitio web: <https://historiaybiografias.com/Conjetura%20de%20Brocard.pdf>
- Muñoz Santoja, P. (2020). Algo más que números: El Problema de Brocard. Recuperado o 20 de novembro de 2021 do sitio web: <http://algotomasquenumeros.blogspot.com/2020/04/el-problema-de-brocard.html>
- LinkFang (s.f). Conjetura de Legendre. Recuperado o 20 de novembro de 2021 do sitio web: https://es.linkfang.org/wiki/Conjetura_de_Legendre
- UstaMathematica (2017). Conjetura de Brocard. Recuperado o 20 de novembro de 2021 do sitio web: <https://ustamathematica.blogspot.com/2017/01/conjetura-de-brocard.html>
- Porras Ferreira, J.W. y Alberto Andrade, C. (2015). Solución a los problemas de Landau y otras conjeturas sobre primos. Recuperado o 20 de novembro de 2021 do sitio web: https://www.researchgate.net/profile/Jose-Porras-4/publication/333056758_SOLUCION_A_LOS_PROBLEMAS_DE_LANDAU_Y_OTRAS_CONJETURAS SOBRE PRIMOS/links/5cd99433299bf14d9592e604/SOLUCION-A-LOS-PROBLEMAS-DE-LANDAU-Y-OTRAS-CONJETURAS-SOBRE-PRIMOS.pdf

$n^2 + 1$: unha forma de números primos

Castelo-García, S.; López-Castro, M.;
Moreira-Pereiro, A.

IES MARCO DO CABBALLÓN

Resumo. Con este artigo buscamos responder a pregunta de si existen infinitos números primos da forma n^2+1 . Para iso, recompilamos todos os números ata un valor determinado e, a través do programa R, realizamos a análise de frecuencias correspondentes.

Abstract. With this research, we are trying to find out if there're infinite prime numbers which are n^2+1 . For that, we compile all numbers until a certain number, and thanks to the program R, we carry out the corresponding analysis of frequencies.

1. Introducción

As matemáticas n^2 son unha das ciencias que acompañaron ao ser humano ao longo de toda a súa historia. Unha das primeiras inquedanzas foi a numeración, é dicir, sermos capaces de contar. Inicialmente só se consideraban os números 1, 2, 3... os que actualmente coñecemos como números naturais. Na Antiga Grecia, aínda que se descoñece se antes tamén, xa se percataron de que había números con propiedades especiais respecto á divisibilidade. Estes números son os que actualmente coñecemos como números primos.

Deste xeito, na teoría dos números existen numerosas conxecturas relacionadas cos números primos que parecerían indemostrables, principalmente por non coñecerse a orde na que aparecen estes números.

Edmund Landau foi un matemático alemán de finais do século XIX e principios do XX que traballou, principalmente, en teoría de números e análise complexo. É coñecido, entre outras cousas, pola

súa demostración do teorema dos números primos, que di que se $\pi(x)$ representa a cantidade de números primos que son menores ou iguais que x , entón

$$\pi(x) \approx \frac{x}{\ln(x)}.$$

Os problemas de Landau son catro coñecidos problemas básicos sobre os números primos, que Edmund Landau catalogou como “inabarcables no estado actual da ciencia” durante o Quinto Congreso Internacional de Matemáticos do ano 1912.

Os catro problemas son os seguintes:

- A conxectura de Goldbach establece que todos os números pares maiores que 2 se poden expresar coma a suma de dous números primos
- A conxectura dos números primos xemelgos establece que hai infinitos números primos p tales que $p+2$ tamén é un número primo.
- A conxectura de Legendre establece que sempre existe un número primo entre dous cadrados perfectos.
- Por último a conxectura aquí tratada. De antemán compre destacar que esta é a conxectura das catro anteriores sobre a que menos literatura hai na rede. Expón a existencia de infinitos números primos p tales que $p-1$ é un cadrado perfecto. Dito doutra maneira, hai infinitos números primos da forma n^2+1 .

Relacionado con isto está o Teorema de Friedlander-Iwaniec, probado en 1997, que afirma que hai infinitos números primos da forma a^2+b^4 . Cando b é igual a 1 (a^2+1^4), os números primos de Friedlander-Iwaniec teñen a forma a^2+1 , é dicir, a mesma forma que a do problema de Landau que pretendemos analizar.

A dificultade do problema radica na ínfima cantidade de números da forma a^2+b^4 : o número de enteiros menores que x que

cumpren as condicións é, aproximadamente, da orde $x^{\frac{3}{4}}$.

2. Análise dos datos

Para tratar de demostrar a conxectura da existencia de infinitos números primos da forma $n^2 + 1$, creamos os programas que desenvolveremos máis adiante axudándonos de R.

O programa que averiguará os números da forma $n^2 + 1$ require a creación de outros dous previamente. Estes exercerán como funcións dentro del, e axudarannos a levalo a cabo.

O primeiro que debemos establecer é unha función que confirme se un número é primo ou non. Para iso creamos a función “primo”, á que se lle introduce un número entre paréntese ao que denominamos “numero”. Logo creamos unha estrutura de bucle (coa función for) para os números do 1 a “numero”, aos que chamamos “j”. Introducimos unha estrutura condicional (coa función if) dentro do bucle: se “numero” ao dividilo entre calquera “j” ten resto 0, daquela sumamos 1 a un contador que xa estableceramos inicialmente. Fóra da estrutura de bucle creamos outra condicional: se o resultado do contador é 2, é dicir, se só a división de “numero” entre si mesmo e entre 1 teñen resto 0, entón o programa devolverá “numero”.



```
1- primo = function(numero){
2-   cont = 0
3-   for (j in 1:(numero)){
4-     if (numero%j==0){
5-       cont = cont+1
6-     }
7-     if (cont==2){
8-       return(numero)
9-     }
10- }
11- }

R 4.1.2 > primo(47)
[1] 47
> |
```

Fig.1. Programa da función “primo”.

A función anterior empregarase para realizar a criba de Eratóstenes. Para iso, antes de nada, creamos un vector baleiro “v” (coa función

$c()$, coa que se declaran os vectores en R). Á función “criba” introdúceselle “numero”, que de novo funciona cunha estrutura de bucle. Someterá a cada “i” (de 2 a “numero”) á función “primo”, averiguando así se “i” é ou non primo. Dado que “primo” só devolve “numero” se este é primo, engadimos (coa función *append*) “numero” ao vector “v”. Finalmente, a función “criba” devolverá “v”, isto é, os números primos do 2 (onde comeza o bucle) a “numero” (onde acaba).

	2	3	4	5	6	7	8	9	10
11	12	13	14	15	16	17	18	19	20
21	22	23	24	25	26	27	28	29	30
31	32	33	34	35	36	37	38	39	40
41	42	43	44	45	46	47	48	49	50
51	52	53	54	55	56	57	58	59	60
61	62	63	64	65	66	67	68	69	70
71	72	73	74	75	76	77	78	79	80
81	82	83	84	85	86	87	88	89	90
91	92	93	94	95	96	97	98	99	100

Fig.2. Criba de Eratóstenes ata o número 100.



Fig.3. Programa da función “criba”.

O último programa será o que realmente nos aproxime á solución

do problema tratando de atopar, dentro dos primos achegados por “criba”, os que sexan da forma $n^2 + 1$. Devolverá o vector “w”, que conterá 1 ou 0 nas posicións dos números de “criba” segundo sexan primos da forma $n^2 + 1$ ou non. Para iso creamos un novo vector valeiro “w” e, a continuación, un bucle para as posicións no vector “criba(numero)”, ás que denominamos “i”. Dentro del establecemos que o vector “w” debe incluír en cada “i”, isto é, en cada posición de “criba(numero)”, un 0, pero con excepcións. Dentro do bucle establecemos un novo que afectará aos números do 1 ao que se atope na posición [i] do vector “criba(numero)”, incluíndo tamén os que non pertencen ao conxunto “criba(numero)”, a estes chamarémoslle “j”. Se existe algún “j” tal que ao elevalo ao cadrado e sumarlle 1 sexa igual a algún número do vector “criba(numero)” ($j^2 + 1 = \text{criba}(\text{numero})[i]$), entón o vector w engadirá un 1 na posición dese número ([i]).



```
1 numero<-as.integer(readline(prompt="Introduce un número: "))
2 w<-C()
3
4 for (i in 1:length(criba(numero))) {
5   w[i]<-0
6   for (j in 1:criba(numero)[i]){
7     if ((j^2)+1==criba(numero)[i]){
8       w[i]<-1
9     }
10  }
11  print(w)
```

Console Terminal Jobs

R 4.1.2 ~> v>
Introduce un número: 60
[1] 1 0 1 0 0 0 1 0 0 0 0 1 0 0 0 0 0 0
>

Fig.4. Programa que analiza se un número é primo da forma $n^2 + 1$.

Coa secuencia de uns e ceros que nos devolve o último programa podemos crear gráficas que permiten visualizar os resultados. No eixe de abscisas atópanse representados todos os números primos, é dicir, os números achegados pola función “criba”. No eixe de ordenadas o vector “w”, que contén ceros e uns.

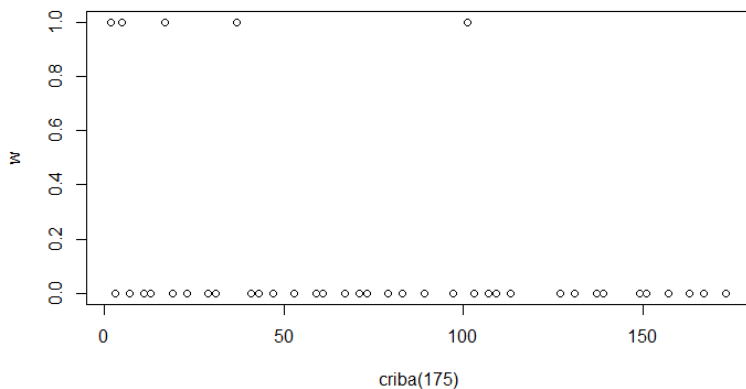


Fig.5. Gráfica que mostra os números primos da forma n^2+1 ata 75.

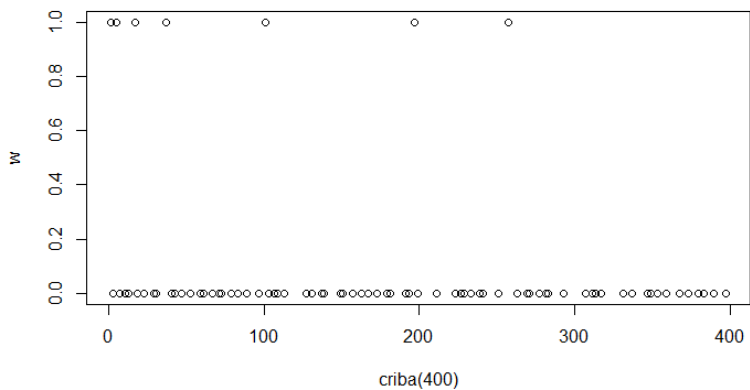


Fig.6. Gráfica que mostra os números primos da forma n^2+1 ata 175.

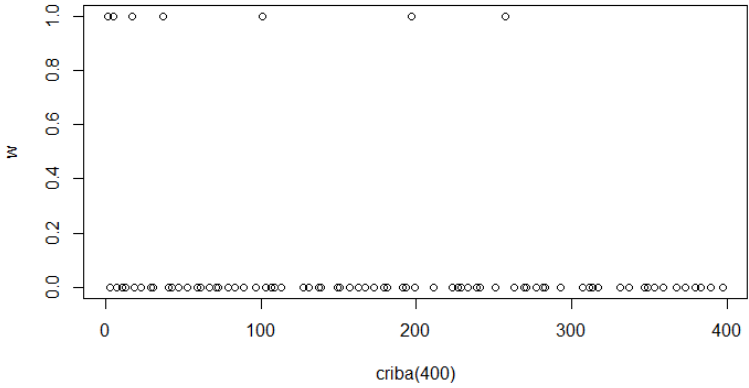


Fig.7. Gráfica que mostra os números primos da forma n^2+1 ata 400.

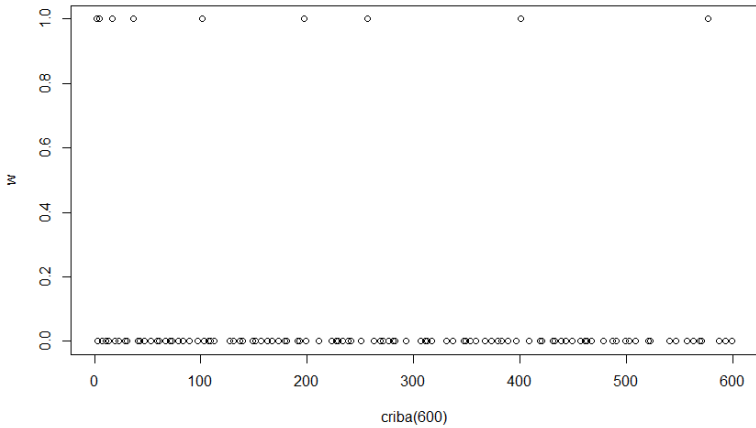


Fig.8. Gráfica que mostra os números primos da forma n^2+1 ata 600.

3. Conclusións

Tras a análise dos resultados dos gráficos creados cos datos do programa, chegamos a unha serie de conclusións. Como se observa claramente nas gráficas do apartado anterior, os primos da forma

aparecen cunha taxa de incremento menor conforme avanzamos cara o infinito, aínda así, continuámoslos atopando. Porén, non fomos capaces de determinar a existencia destes números máis alá do 600, ata onde a computadora foi capaz de procesar. Por iso, non podemos responder con certeza á cuestión exposta máis alá de dito número, mais podemos prever que a frecuencia coa que surxen segundo nos achegamos ao infinito se reduce considerablemente. Podemos intuír, axudándonos do Teorema de Friedlander-Iwaniec, que esta secuencia pode ser infinita. Os pasos a seguir para logralo serían continuar introducíndolle ao programa números cada vez máis grandes, e observar os resultados do vector “w”, é dicir, se devolve uns (primos da forma $n^2 + 1$) ou só obtemos ceros.

4. Webgrafía

- Andrade, Carlos (s.f.). Solución a los problemas de Landau y otras conjeturas sobre primos. Recuperado o 17 de novembro de 2021, do sitio Web: https://www.researchgate.net/publication/333056758_SOLUCION_A_LOS_PROBLEMAS_DE_LANDAU_Y_OTRAS_CONJETURAS_SOBRE_PRIMOS
- Roldán, Antonio (2014). Números y hoja de cálculo. Recuperado o 17 de novembro de 2021, do sitio Web: <http://hojaynumeros.blogspot.com/2014/11/conjeturas-con-hoja-de-calculo-primos.html>
- Gaussianos (2012). Los problemas de Landau, después de 100 años “nada nuevo bajo el sol”. Recuperado o 17 de novembro de 2021, do sitio Web: <https://www.gaussianos.com/los-problemas-de-landau-despues-de-100-anos-nada-nuevo-bajo-el-sol/>
- Forišek, Michal (s.f.). ¿Son realmente los “problemas de Landau” un problema? Recuperado o 17 de novembro de 2021, do sitio Web: <https://edu.sacolife.com/62767/son-realmente-los-problemas-de-landau-un-problema.html>

Sobre a natureza dos números primos. A conxectura de Legendre

*Amosa-Martínez, U.; Silva-Lodeiro, X.;
Torres-Novo, N.*

IES MARCO DO CABBALLÓN

Resumo. Con este traballo buscamos realizar un estudo da conxectura de Legendre e o posterior análise dos datos para procurar a verificación e analizar como varia o comportamento da dita conxectura nos diferentes números.

Abstract. With this essay we aim to carry out a study of the Legendre conjecture and the following analysis of the data to try to verify and analyze how the behavior of the conjecture varies in the different numbers.

1. Introducción

A conxectura de Legendre leva máis de dous siglos sen resolverse dende que se propuxo por primeira vez por Adrien-Maire Legendre en 1808. Nesta conxectura Legendre afirma que entre dous cadrados de números consecutivos hai polo menos un número primo é dicir, entre n^2 e $(n+1)^2$ sempre hai polo menos un número primo. (Se por exemplo collemos un número primo $52 = 25$ e outro $62 = 36$ encontraríamos os números primos 29 e 31).

Asemade, o matemático xaponés Chen Jing-Run demostrou que entre estes cadrados ademais de existir un número primo habería tamén polo menos un semiprimo (produto de dous primos). Ademais, está comprobado que hai un número primo no intervalo dun número “ n ” e o seu dobre “ $2n$ ”. Ao mesmo tempo tamén se quixo demostrar que a conxectura era aplicable para números grandes mais isto non implica que dita conxectura estea demostrada para todos os núme-

ros. Por outra banda, esta conxectura forma parte dos problemas de Landau que serán explicados nos seguintes apartados.

Os problemas de Landau son 4 problemas básicos sobre os números primos enunciados por Edmund Landau no Congreso Internacional de Matemáticas en Cambridge no 1912.

Estes catro problemas son a conxectura de Goldbach (que afirma que todo número primo maior que 2 pode ser escrito como a suma de outros dous números primos), a conxectura de Legendre (para todos os n naturais existe un n^o primo entre n^2 e $(n+1)^2$), a conxectura dos primos xemelgos (existen infinitas parellas $(p,p+2)$ nas que ambos números son primos) e primos da forma n^2+1 (cuestiona se existen infinitos números primos p tales que $p-1$ sexa un cadrado perfecto). Ningún dos problemas de Landau está resolto hoxe en día, aínda que algúns están demostrados parcialmente.

2. Análise dos datos

En este apartado, procedemos a explicar como diseñamos un código en R que nos permitira traballar coa conxectura de Legendre. Unha vez elaborado o código, procedemos á creación de gráficas de diferentes números para a súa posterior análise. A partires de estas gráficas imos tentar tirar conclusións sobre como funciona a conxectura de Legendre e intentar demostrala. Para iso, explicaremos primeiro, como elaboramos os programas que intentan verificar dita conxectura.

2.1. Elaboración do código en R

```
primo<-function(n){  
  cont<-0  
  for(i in 1:n){  
    resto<-n%i  
    if(resto==0){  
      cont<-cont+1  
    }  
  }  
  if(cont==2){  
    return(n)}  
}
```

Fig.1. Programa “primo”

```
llegendre<-function(n){  
  v<-c()  
  for(i in n^2:(n+1)^2){  
    v<-append(v,primo(i))  
  }  
  print(v)  
  print(paste("lonxitude do vector=",length(v)))  
}
```

Fig.2. Programa “Conxectura de Legendre”

Como xa explicamos anteriormente, a conxectura de Legendre afirma que entre un número “n” elevado ao cadrado e o seguinte elevado ao cadrado hai sempre un número primo, como mínimo.

Para comprobar se hai ou non un número primo entre estes creamos unha función no programa R que calculara os números primos comprendidos entre o 1 e o número introducido. Este programa emprega un contador, que nun comezo é cero. A este contador súmase-lle 1 cada vez que o resultado da división entre o número introducido “n” e un número do intervalo (1,n). O programa devolve como primos os números cuxo contador é igual a 2, xa que os números primos só son divisibles entre 1 e entre eles mesmos, polo que só teñen 2 números no contador. Un número non primo ten máis de dous unidades no seu contador, porque é divisible entre máis de 2 números.

Con isto, buscamos unha función que calcule todos os números primos encontrados no intervalo (n², (n+1)²). Este programa emprega a función “primo” programada previamente ao intervalo que abarca (n², (n+1)²). Tras facer este cálculo, o programa devólvenos a lonxitude do vector, no que foi gardando todos os números primos atopados.

```
vectorr<-function(n){  
  v<-c()  
  for(i in 1:n){  
    v<-append(v,lonxitude(i))  
  }  
  print(v)  
}
```

Fig.3. Programa “vector lonxitude”

Logo, creamos unha función á que lle chamaremos “vectorr” na cal vanse gardando as lonxitudes de cada de cada intervalo.

Por último, utilizamos a función hist que fai un diagrama de barras a partir dos datos do vector lonxitude.

Engadindo á función main, e entre aspas o que queremos escribir como título da gráfica, dos eixes vertical e horizontal , podemos personalizar o noso histograma.

```
hist(vectorr(442),main="Histograma legendre frecuencia",  
      xlab="Número de primos existentes no vector", ylab="Frecuencia" )|
```

Fig.4. Programa “histograma”

2.2. Análise das gráficas

Agora procedemos a analizar as gráficas da conxectura de Legendre para ver o seu comportamento en función do número proposto:

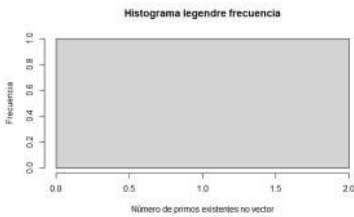


Fig.4. Histograma do número 1

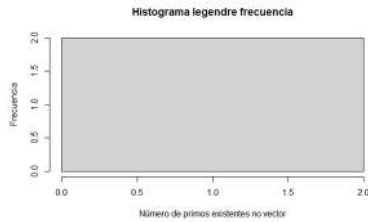


Fig. 5. Histograma do número 2

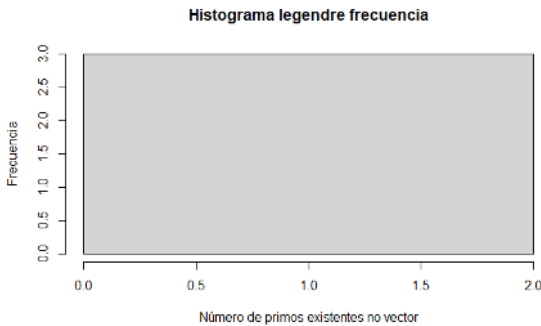


Fig. 6. Histograma do número 3

Utilizando legendre, (fig. 4, 5 e 6) as gráficas resultan, a primeira vista, iguais. Isto débese a que ao facer o cadrado de números pequenos, a cantidade de números primos comprendidos entre estes, é a mesma.

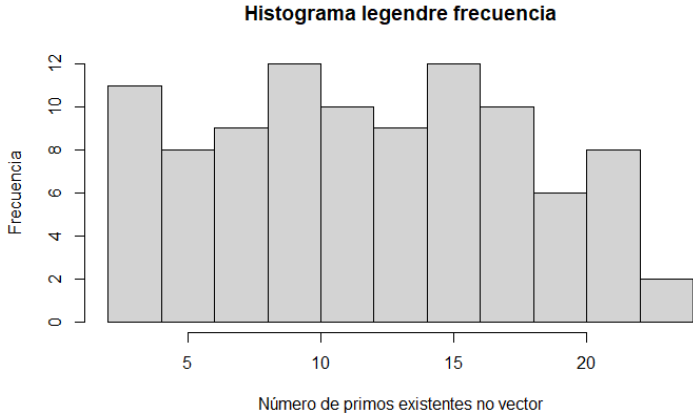


Fig.7. Histograma do número 9

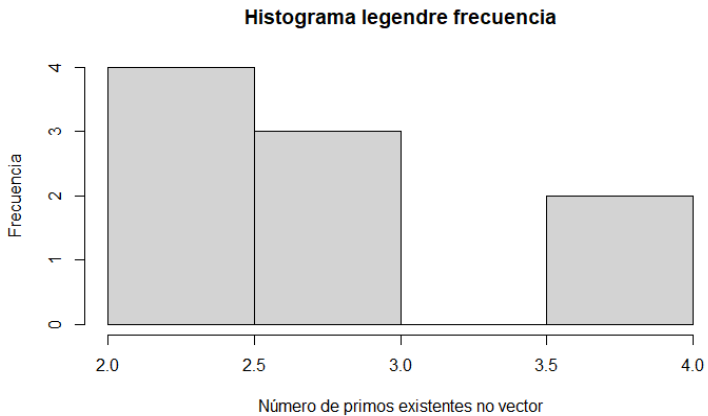


Fig.8. Histograma do número 97

Pola contra, nas anteriores gráficas pódese ver que, calculando os histogramas de números maiores, o número de primos comprendidos entre estes son cada vez máis, e polo tanto os histogramas tamén son

máis grandes (fig.7 e 8), comparados coas gráficas analizadas anteriormente (fig. 4, 5 e 6) e, obsérvase que en números máis grandes a semellanza entre estas é menor. Polo tanto, canto máis grande sexa un número, máis números primos hai entre estes ao aplicar Legendre

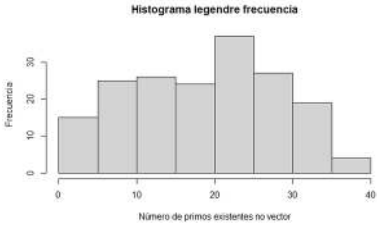


Fig.9. Histograma do número 177

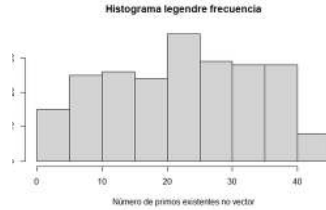


Fig. 10. Histograma do número 220

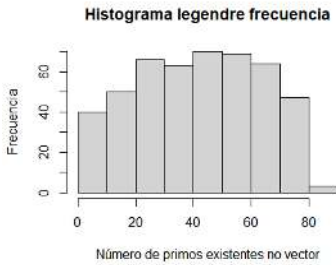


Fig. 11. Histograma do número 300

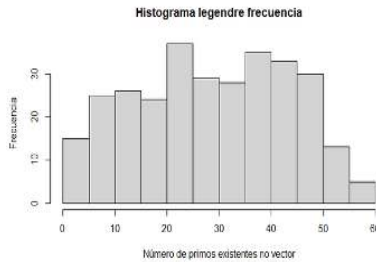


Fig. 12. Histograma do número 472

Asemade, os primos que máis se repiten son os que están entre o 20 e o 40 (fig. 9, 10, 11 e 12). Logo, se nos centramos en números máis grandes (fig. 11) contemplamos que o número de primos contidos no vector, é menor que en números máis altos (fig. 12).

3. Conclusións

Con este artigo elaboramos unha análise dos histogramas de números cada vez máis grandes, e concluímos que o número de primos

atopados no intervalo $(n^2, (n+1)^2)$, é cada vez maior.

Non podemos asegurar que a conxectura de Legendre sexa aplicable a todos os números, pero podemos asegurar que nos números que nos estudamos, entre o 1 e o 472, se cumpría en todos e que cada vez se encontraba máis números primos no intervalo $(n^2, (n+1)^2)$.

Tendo en conta as conclusións que sacamos/tiramos dos números que nós estudamos podemos supoñer que canto máis grandes sexan os números estudados maior cantidade de primos apareceran ao aplicar legendre, pero esto non podemos aseguralo, debido a que esta conxectura non está demostrada para todos os números e pode ser que en números moi grandes sucedan cousas diferentes. Isto tería que estudarse nun futuro traballo.

4. Bibliografía / Webgrafía

- Wikipedia (2021). Conjetura de Legendre. Recuperado o 17 de novembro do 2021, do sitio Web: http://wikipedia.es.nina.az/Conjetura_de_Legendre.html
- Teinteresa.es (2013). Tres conjeturas con máis de 200 anos... y solo una ha sido resuelta. Recuperado o 17 de novembro do 2021, do sitio Web: https://www.teinteresa.es/ciencia/conjeturas-anos_0_937706662.html
- ^DiAmOnD^ . Los problemas de Landau, después de 100 años “nada nuevo bajo el sol”. Recuperado o 17 de novembro do 2021, do sitio web: <https://www.gaussianos.com/los-problemas-de-landau-despues-de-100-anos-nada-nuevo-bajo-el-sol/>
- Amp (2020). Problemas de Landau. Recuperado o 17 de novembro do 2021, do sitio Web: <https://amp.es.what-this.com/2449336/1/problemas-de-landau.html>
- ResearchGate (2013). Solución a los problemas de Landau y otras conjeturas sobre primos. Recuperado o 17 de novembro do 2021, do sitio Web: https://www.researchgate.net/publication/333056758_SOLUCION_A_LOS_PROBLEMAS_DE_LANDAU_Y_OTRAS_CONJETURAS_SOBRE_PRIMOS
- Ustamathematica (2016). Conjetura de Legendre. Recuperado o 17 de novembro do 2021, do sitio Web: <https://ustamathematica.blogspot.com/2016/10/conjetura-de-legendre.html>
- Leebran (2018). Demostración de la conjetura de Legendre. Recuperado o 17 de novembro do 2021, do sitio Web: <https://leebran.wordpress.com/2018/03/26/conjetura-de-legendre/>
- Xataka (2021). ¿Por qué seguimos buscando números primos más allá de los 22 millones de dígitos? . Recuperado o 17 de novembro do 2021, do sitio Web: ¿Por qué seguimos buscando números primos más allá de los 22 millones de dígitos? (xataka.com)

O teorema central do límite para variables aleatorias continuas

*Ferro-Pérez, M.; López-Calviño, E.;
Puga-Silva, L.; Verde-Salgado, N.*

IES MARCO DO CABBALLÓN

Resumo. No seguinte estudo analízase o teorema central do límite para variables aleatorias continuas, concretamente, a uniforme e a exponencial. Para isto, deseñamos algoritmos empregando o programa de análise estatístico R. Unha vez expostos, centrámonos no estudo gráfico de ambas distribucións.

Abstract. The following research analyzes the central limit theorem for continuous random variables, specifically the uniform and exponential. Therefore, we design two programs by using the estadistic analysis programme R. When we have done it, we have to focus on the study of both distributions.

1. Introducción

O teorema central do límite (TCL) é un dos resultados fundamentais de estadística, porén, non é un único teorema, senón que consiste nun conxunto de resultados sobre o comportamento das distribucións da suma de variables aleatorias. Garante que, baixo certas hipóteses, a distribución da suma dun número moi grande de variables aleatorias aproxímase a unha distribución normal. O termo “central” foi acuñado por George Pólya en 1920 e adquire o significado de fundamental ou “importancia central”. A súa relevancia débese a que este conxunto de teoremas desvela as razóns nas que, en moitos campos de aplicación, danse en todo momento distribucións normais ou case normais.

Dende o punto de vista histórico, tras ser estudado rigorosamente por diversos matemáticos destacados, foi establecido por primeira vez en 1738 por Abraham De Moivre, baixo condicións moi res-

trinxidas. A comezos do século XIX, Laplace propuxo unha hipótese que parecía ser verosímil. Así por exemplo, as observacións astronómicas, pequenas variacións de temperatura, correntes irregulares de aire e vibracións de edificios poden considerarse como algúns de ditos ámbitos.

Sexa X_n unhas variables aleatorias independentes dunha poboación de media finita (μ) e desviación típica finita (σ), entón, a distribución da media mostral de tamaño n ten de media e desviación típica n , e apróximase a unha distribución normal a medida que medra o tamaño da mostra. Se a poboación de partida é normal, a distribución das medias mostrais é tamén normal mentres que se a poboación de partida non é normal, a distribución da media mostral aproximarase a unha normal cando o tamaño da mostra sexa suficientemente grande e as variables aleatorias sexan independentes. Na práctica, considérase que mostras con $n > 30$ son suficientemente grandes como para que a distribución da suma sexa normal. O Teorema Central do Límite réxese pola seguinte expresión:

$$\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

Para comprender con exactitude o anterior exposto, é necesario aclarar en que consisten as variables aleatorias. Denóminase así a toda lei que asocia a cada elemento do espazo mostral un número real. O conxunto de valores asignados chámase percorrido da variable. Segundo como sexa este, distínguense as variables discretas e continuas.

Unha variable aleatoria é discreta cando só pode tomar un número finito ou infinito numerable, é dicir, equivalente a tantos valores como elementos do conxunto dos números naturais. Isto ocorre, por exemplo, cando tiramos dous dados. O espazo mostral deste experimento aleatorio é:

$$E = \{ (1,1), (1,2), \dots, (1,6), (2,1), \dots, (6,1), \dots, (6,6) \}$$

Esta lei asocia a cada resultado a suma dos puntos obtidos en cada dado e toma os valores 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12. A representación gráfica deste experimento é a seguinte:

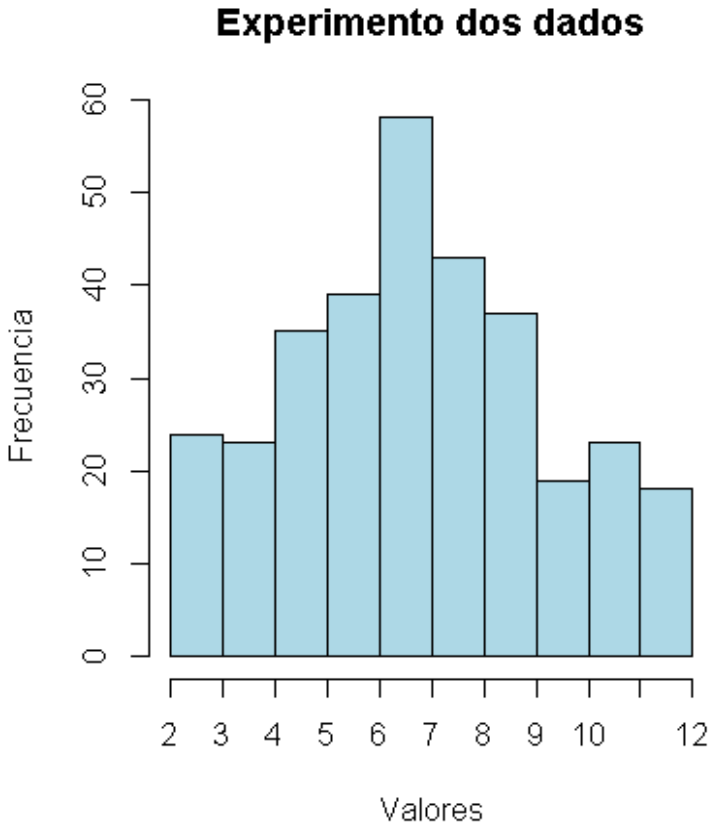


Fig. 1. Gráfica experimento dados

Mentres que unha variable aleatoria é continua cando pode tomar todos os valores posibles dentro dun certo intervalo da recta real. Un exemplo desta variable consiste en realizar o experimento de elixir ao chou cen xudías verdes dunha plantación e medir a súa lonxitude. Na seguinte gráfica amósase a súa representación:

Experimento das Xudías

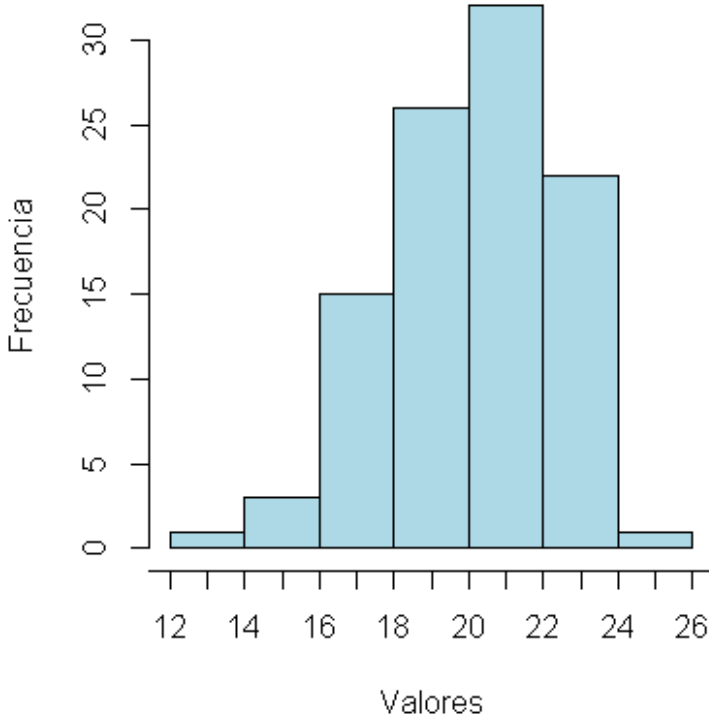


Fig. 2. Gráfica experimento xudías

2. As distribucións uniforme e exponencial. Caracterizacións e aplicacións.

Por un lado, a distribución uniforme, tamén coñecida como rectangular polo aspecto da súa función de densidade, é aquela que pode tomar calquera valor dentro dun intervalo, todos eles coa mesma probabilidade. Esta distribución é o modelo continuo máis simple, corresponde ao caso dunha variable aleatoria que só pode tomar valores comprendidos entre dous extremos a e b , de xeito que todos os

intervalos dunha mesma lonxitude, dentro desas variables, teñen a mesma probabilidade. Unha peculiaridade desta distribución é que a probabilidade dun suceso depende exclusivamente da amplitude do intervalo considerado e non da súa posición no campo de variación da variable. Empregaríase no caso de que queiramos determinar a probabilidade de que, ao lanzar un moeda, todos os resultados, cara e cruz, teñan a mesma probabilidade, é dicir, $\frac{1}{2}$.

Caractérezase por unha función de densidade que é plana, e polo tanto, a probabilidade é uniforme nun intervalo pechado que se define da seguinte maneira:

$$f(x, A, B) = \begin{cases} \frac{1}{B-A} & A \leq x \leq B \\ 0 & \text{outro caso} \end{cases}$$

A media e a varianza da distribución uniforme réxese polas seguintes expresións:

$$\mu = \frac{A+B}{2} \quad \sigma^2 = \frac{(B-A)^2}{12}$$

A continuación, amósase un exemplo gráfico da devandita distribución:

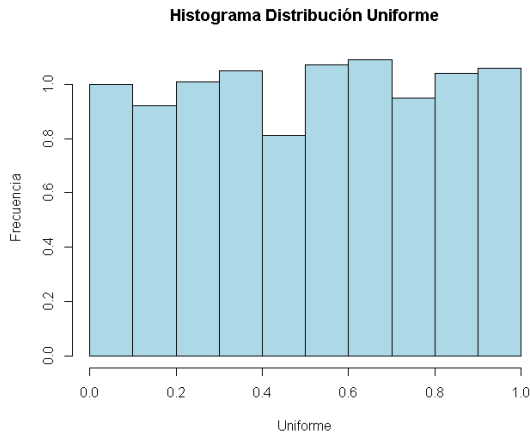


Fig. 3. Gráfica distribución uniforme

Por outro, a distribución exponencial describe procesos nos que interesa saber o tempo ata que ocorre un determinado evento, en concreto, emprégase para modelar tempos de supervivencia como o tempo que tarda unha partícula radiactiva en desintegrarse. O coñecemento da lei que segue este evento emprégase para a datación de fósiles ou calquera materia orgánica mediante a técnica do carbono 14, C14. Esta distribución é continua porque o tempo entre os sucesos non ter que ser un número enteiro. Ademais, unha característica importante desta é a propiedade coñecida como “falta de memoria”. Isto quere dicir que o tempo transcurrido dende calquera instante dado (t) ata que ocorre o evento, non depende do que ocorrira antes do instante t . O uso da distribución exponencial é limitado en bioestadística, debido a esta peculiaridade que a fai demasiado restritiva para a maioría dos problemas.

Pódese considerar como a distribución de tempo entre sucesos consecutivos xerados por un proceso de Poisson, por exemplo, o tempo que trascurre entre dúas feridas graves sufridas por unha persoa. A media da distribución de Poisson, λ , que representa a taxa de ocorrencia do evento por unidade de tempo, é o parámetro da distribución exponencial e a súa inversa o valor medio da distribución.

A súa función de densidade réxese pola seguinte expresión: sexa X unha variable aleatoria continua que pode tomar valores maiores que 0 ($x \geq 0$). Afírmase que X segue unha distribución exponencial de parámetro λ :

$$f(x) = \begin{cases} \lambda \cdot e^{-\lambda x} & \text{se } x \geq 0 \\ 0 & \text{no outro caso} \end{cases}$$

Tamén se pode demostrar que a media e varianza dunha variable exponencial de parámetro λ son iguais a:

$$\mu = \frac{1}{\lambda} \quad \sigma^2 = \frac{1}{\lambda^2}$$

Polo tanto, a desviación típica dunha distribución exponencial é igual a súa media:

$$\sigma = \frac{1}{\lambda}$$

A continuación, amósase un exemplo gráfico da distribución exponencial:

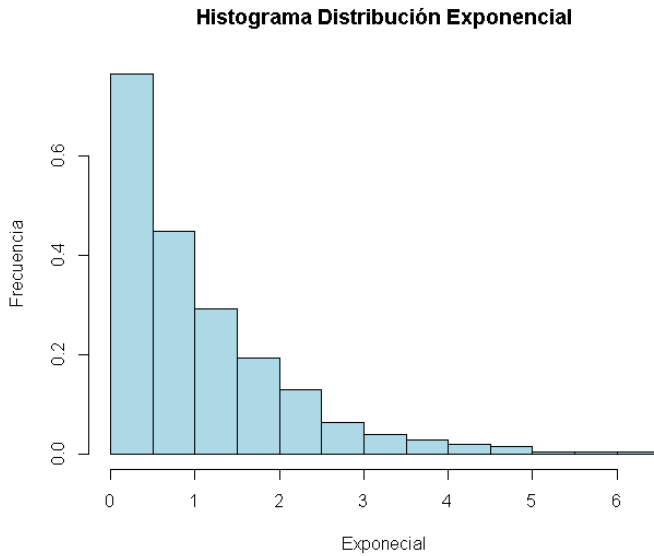


Fig. 4. Gráfica distribución exponencial

3. Análise dos datos

A partir deste análise das distribucións, queremos demostrar o Teorema Central de Límite (TCL) de xeito descritivo partindo das distribucións uniforme e exponencial mediante o programa de análise estadístico R , que nos permite elaborar algoritmos para estudar o teorema. Este teorema tamén estipula que os resultados varían dependendo do tamaño da mostra estudada, por isto crearemos dous algoritmos directamente relacionados, o primeiro usarémolo para comprobar resultados concretos, mentres que no segundo empregaremos intervalos máis amplos coa finalidade de examinar o seu comportamento. Para realizar os nosos algoritmos que nos permita analizar esta proposición necesitamos crear tres variables distintas.

3.1. Estudo de valores concretos

3.1.1. Condicións do estudo

En primeiro lugar, xeramos esas tres variables que teñen como finalidade permitir levar a cabo o programa. A primeira delas désígnase como “Valores” e nela, como o seu nome indica, almacenaránse os valores xerados. Esta variable ten como obxectivo realizar experimentos independentes, e polo tanto, darémoslle un valor nulo que se representa a través da expresión NULL.

A seguinte variable recibe o nome de “n” e corresponde ao número de valores que extraeremos con cada execución da función uniforme ou exponencial que estará determinada por un número aleatorio. A representación vai a depender do número escollido.

Por último, empregamos a variable “m” que corresponde ao número de veces que realizaremos a función uniforme ou exponencial e que tamén está determinada por un número dado.

```
Valores<- NULL
n <- 1000
m <- 100
```

Fig. 5. Parámetros do estudo

3.1.2. Recollida de datos

A continuación, creamos un bloque de duración equivalente á variable “m” no que se realizará a función que nos interese, neste caso *Uniforme*, que xerará “n” valores, empregando o comando `runif` que xenerará números racionais aleatoriamente comprendidos no intervalo $(0,1)$. Despois, a través do comando *append*, almacenaremos a media dos valores xerados no vector *Valores*.

```
for (i in 1:m) {
  uniforme <- runif(n,0,1)
  valores <- append(valores, mean(uniforme))
}
```

Fig. 6. Estrutura bloque de duración uniforme

A diferenza da función uniforme, na exponencial empregaremos o comando `rexp(n,1)` que de igual xeito xerará “n” valores determinados polo segundo valor, neste caso 1 e chamado λ ; este representa o número de sucesos esperados por unidade, por isto, é coñecida como a tasa de ocorrencia.

```
for (i in 1:m) {  
  Exponencial <- rexp(n, 1)  
  valores <- append(valores, mean(Exponencial))  
}
```

Fig. 7. Estructura bloque de duración exponencial

3.1.3.Representación gráfica

Para poder ter unha representación visual dos resultados obtidos polo algoritmo crearemos un histograma no que representaremos a variable “Valores” e usaremos o comando `prob` co obxectivo de cuantificar a probabilidade sobre 1. Ao mesmo tempo, realizaremos graficamente unha curva normal, para ver como os valores representados se asemellan a unha distribución normal. Para isto determinaremos dous vectores “x” e “y”. O eixe “x” virá definido por unha secuencia de “m” valores repartidos de maneira equidistante entre o valor mínimo e máximo; o eixe “y” será determinado polo cálculo da densidade, co comando `dnorm`, a partir do vector “x” e a media e a desviación típica dos datos extraídos previamente. Por último, definimos a curva co comando `lines`. Ademais, coa finalidade de comparar esta representación teórica representaremos a verdadeira distribución da mostra a partir do comando `density` e, de novo, o comando `lines`.

```
hist(valores, prob = TRUE, main, ylab, col, ylim, xlim)  
x <- seq(min(valores), max(valores), length = m)  
y <- dnorm(x, mean = mean(valores), sd = sd(valores))  
lines(x, y, col, lwd)  
z <- density(valores, n = m)  
lines(x, z[["y"]], col, lwd, lty)
```

Fig. 8. Estructura histograma

3.2 Estudio de intervalos de valores

3.2.1 Estudio dos diferentes valores de “m”

Deseguido, buscaremos con esta modificación do programa anterior repetir o proceso para cada valor de “m” repetidas veces co que pretendemos estudar e calcular o valor da correlación lineal, entre a curva normal, representativa da distribución teórica da mostra e a distribución mostral calculada a partir da densidade, para posteriormente calcular a media das mesmas. A correlación indicanos hasta que punto están relacionadas dúas variables linealmente, esto quere decir que busca unha tasa constante de cambio entre elas, o seu valor absoluto, polo tanto, non pode ser maior que 1. Esta medida estadística é calculada a partir da seguinte fórmula:


$$r = \cos(\alpha) = \frac{\sum_{i=1}^N (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^N (y_i - \bar{y})^2}}$$

Para determinar o rango do estudo utilizamos a variable “Nº_de_Experimentos” que está asociada a un intervalo e tamén crearemos unha variable “o” que almacenará o valor da correlación, esta última recollerá os datos dentro do programa anterior, concretamente, para cada experimento de cada valor de “m”.

```
o <- NULL
Nº_de_Experimentos <- 10:1000
o <- append(o, cor(y, z[["y"]]))
```

Fig. 9. Variables “o” e “Nº_de_Experimentos”

Cada valor de “m” é asociado a outro que é o resultado da media de todos os valores da correlación obtidos con ese parámetro. Despois, será representada nunha gráfica cuxos eixes x e y virán definidos polo número de experimentos e o valor calculado, ao que chamaremos “Correlación”, respectivamente.

```
for (m in N°_de_Experimentos) {
  for (j in 1:n) {} #Programa anterior

  Correlación <- append(Correlación, mean(o))
}

plot(N°_de_Experimentos,Correlación, type , col , ylim , lwd , main)
```

Fig. 10. Estructura da variable “N°_de_Experimentos”

4. Estudo gráfico das funcións

A partir da análise de datos feita anteriormente, decidimos empregar gráficas que amosen dun xeito máis descritivo o que acontece.

4.1 Distribución Uniforme

Para realizar este estudo partimos de distribucións uniformes de rango (0,1) xerando para cada unha delas 100 valores. Neste caso, repetimos o proceso 30 veces, que é o umbral estipulado polo teorema central do límite, e imos gardando a media aritmética de cada unha das representacións da distribución uniforme. A continuación, realizamos unha representación gráfica nun histograma coa finalidade de amosar os resultados obtidos. No eixe das X colocamos as medias acadadas e no eixe das Y a densidade. Agora, representamos a densidade da mostra (línea discontinua) e a distribución uniforme teórica da mesma a través da curva normal, todo co obxectivo de comparar. Observamos que as representacións resultantes son semellantes á distribución normal esperada, deste xeito, nunha maioría dos casos cúmprese o determinado no teorema central do límite.

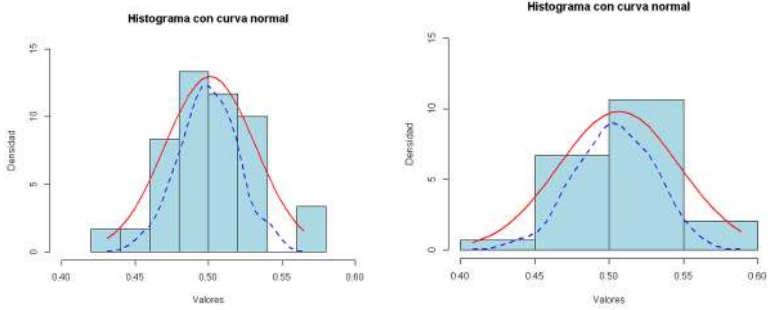


Fig. 11. Gráficas distribución uniforme

Despóis, realizamos a seguinte gráfica coa finalidade de demostrar que a maior número de repeticións a curva dunha distribución normal aseméllase cada vez máis á uniforme. Neste caso, ao igual que o anterior, no eixe das X colocamos as medias acadadas e no eixe das Y a densidade. En definitiva, podemos ver que ao aumentar o número de valores confírmase o estipulado polo teorema central do límite

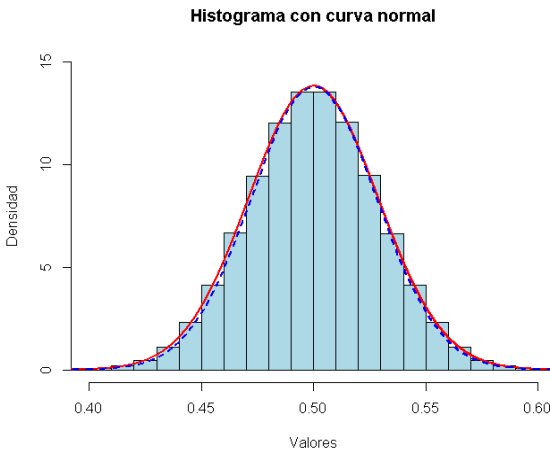


Fig. 12. Gráfica distribución uniforme

Coa intención de comprobar a evolución da correlación entre a distribución mostral e teórica para cada número de repetición, ás que chamamos “m”, recorremos a un algoritmo que calcula o valor medio para cada un delas e as representa. Os resultados móstrannos unha correlación bastante alta incluso para un valor baixo de “m” como, por exemplo, 10. A correlación, que non pode ser maior que 1, aumenta progresivamente, a maior número de experimentos realizados no eixe das X, menos avanzamos no eixe das Y. Cabe destacar que nun principio o crecemento é considerable. Con estes resultados podemos concretar que ao aumentar o valor de “m” obtendremos unha maior “precisión” e, por tanto acercámonos máis a distribución uniforme.

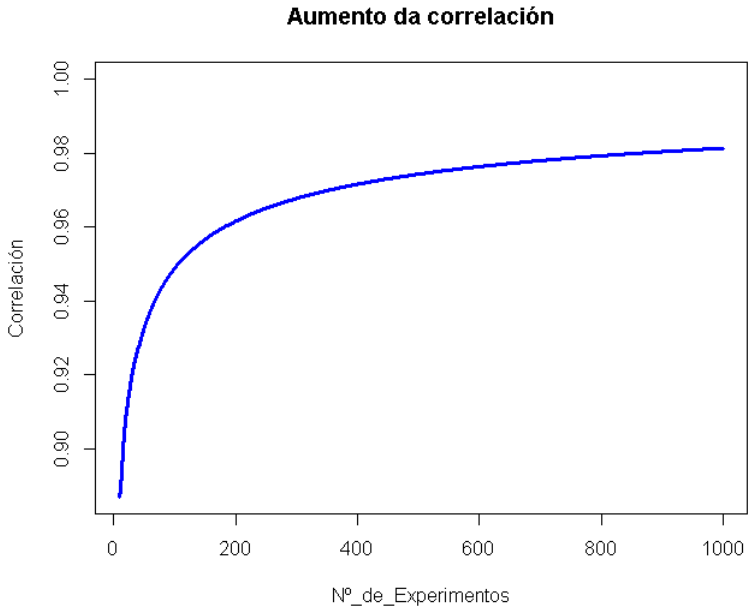


Fig. 13. Gráfica aumento da correlación

4.2 Distribución Exponencial

Deseguido, repetimos un proceso similar ao realizado anteriormente, xerando 30 distribucións, neste caso exponenciais, facendo as medias de cada unha delas para acabar representándoas xunto coa súa curva normal e a densidade. Do mesmo xeito que antes, no eixe das X colocamos as medias aritméticas acadadas e no eixe das Y a probabilidade sobre 100. Agora, facemos unha comparación coa curva normal do que sería unha distribución normal (liña discontinua). Observamos uns resultados similares aos da distribución uniforme, pois as representacións aseméllanse á distribución normal esperada, polo que, de novo podemos concluír que se cumpre o teorema central do límite.

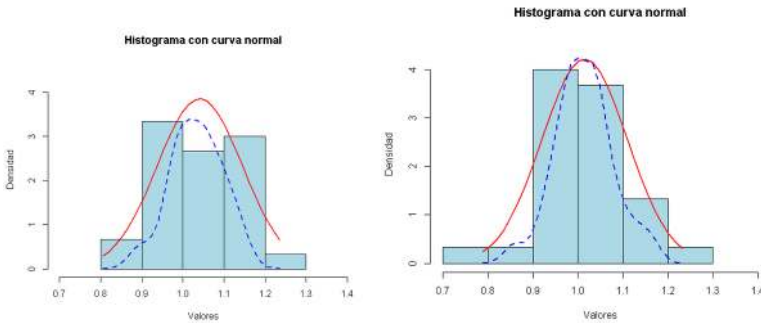


Fig. 14. Gráficas distribución exponencial

Co propósito de evidenciar, ao igual que na distribución anterior, que a maior número de repeticións a curva dunha distribución normal párese cada vez máis á exponencial, realizamos a seguinte gráfica na cal o eixe das X está determinado polas medias acadadas e o eixe das Y pola densidade. Polo que, podemos ver que ao aumentar o número de valores confírmase o que formaliza este teorema.

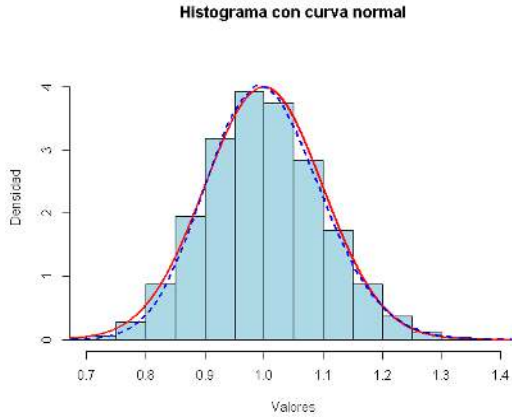


Fig. 15. Gráfica cun “m” moi grande

Finalmente, para comprobar a evolución da correlación entre a distribución mostral e teórica para cada número de repetición, ás que chamamos “m”, faremos o mesmo algoritmo empregado na distribución uniforme. Os resultados que se obteñen son practicamente iguais a outra distribución, por tanto, queda evidenciado que ao aumentar o valor de “m” obteremos unha maior “precisión” e, por tanto acercámonos máis a distribución exponencial.

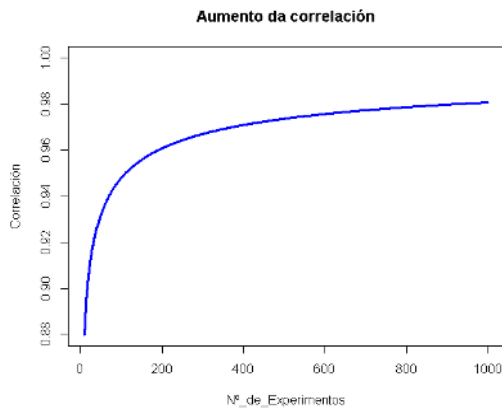


Fig. 16. Gráfica aumento da correlación

5. Conclusións

Neste estudo decatámonos de que o Teorema Central do Límite cúmprese con moita precisión nas distribucións uniforme e exponencial. Ademais, chegamos a conclusión de que a partir dun certo punto, o aumento do número de valores practicamente non afecta a precisión. Cabe destacar que neste proxecto non só acabamos por demostrar este teorema senón que tamén comprobamos que, no caso das distribucións uniformes e exponenciais, son válidos valores menores de 30, como sería o caso de 15, para obter unha similitude coa distribución normal moi elevada neste caso maior a un 88%.

6. Bibliografía / Webgrafía

- Colegio Inmaculada Niña de Granada (S.D.). Teorema Central del Límite. Recuperado o día 3 de marzo de 2022, do sitio Web: http://www.martadcm.villatic.org/IESALMUDE-NA/MATEMATICAS_CCSSII/APUNTES_MATCCSSII/TEOREMA_CENTRAL_DEL_LIMITE.pdf
- Universidad Católica de la Santísima Concepción (S.D.). El Significado del teorema central del límite; evolución histórica a partir de sus campos de problemas. Recuperado o día 3 de marzo de 2022, do sitio Web: <https://www.ugr.es/~batanero/pages/ARTICULOS/Articulo%20TFS.pdf>
- Neus Canal Díaz (2006). Distribuciones de probabilidad. El teorema central del límite. Recuperado o día 3 de marzo de 2022, do sitio Web: <https://www.revistaseden.org/files/8-CAP%208.pdf>
- Escuela Politécnica Superior de Linares Universidad de Jaén (2007). Apuntes de Métodos Estadísticos de la Ingeniería Técnica Industrial Mecánica y Química Industrial. Recuperado o día 3 de marzo de 2022, do sitio Web: <http://www4.ujaen.es/~svilchez/metodos/Apuntes.pdf>
- Universidad Ecotec (S.D.). Distribución de probabilidad continua uniforme. Recuperado o día 3 de marzo de 2022, do sitio Web: https://www.ecotec.edu.ec/material/material_2017F1_MAT260_13_80490.pdf
- IES Pino Manso (S.D.). Distribuciones discretas, distribución binomial. Recuperado o día 3 de marzo de 2022, do sitio Web: https://www.edu.xunta.gal/centros/iespinomanso/aulavirtual/pluginfile.php/12565/mod_resource/content/6/DISTRIBUCIONES%20DISCRETAS.pdf

Converxencia entre distribucións discretas. Binomial e Poisson

González-Castro, M.; Pacheco-
Fariña, M.F.; Torres-García, A.

IES MARCO DO CAMBALLÓN

Resumo. O obxectivo deste artigo é a análise das condicións precisas para que entre dúas distribucións de probabilidade discretas se produza unha converxencia grazas á relación entre os seus parámetros característicos.

Abstract. The objective of this article is to analyze the precise conditions of convergence between two discrete probability distributions due to the relationship between their characteristic parameters.

1. Introducción

As distribucións de probabilidade determinan a probabilidade de que teña lugar un suceso ou fenómeno dunha variable aleatoria, isto é, móstrannos a distribución da probabilidade entre os sucesos que poden ocorrer. Dentro destas distínguense distribucións de probabilidade discretas, que son aquelas que toman un conxunto finito ou infinito numerable de valores, ou distribucións de probabilidade continuas, que toman valores infinitos que non se poden contar.

Neste artigo analizarase a converxencia ou aproximación entre dúas distribucións discretas: a binomial e a de Poisson.

En primeiro lugar, a distribución bino-



Figura 1. Imaxe de Jakob Bernoulli.

mial ten a súa orixe na figura de Jakob Bernoulli, un teólogo, astrónomo e matemático suízo que desenvolveu a considerada como principal distribución de probabilidade discreta para variables que só poden tomar dous posibles resultados, é dicir, variables dicotómicas.

Este proceso consiste en realizar experimentos aleatorios independentes entre sí e observar se teñen lugar ou non, obténdose así resultados excluíntes e exhaustivos, isto é, os resultados só poden ser éxito ou fracaso e sempre ten que ocorrer un dos dous sucesos.

Así, o resultado que se espera obter é o que se denomina como éxito, que ocorre cunha probabilidade p e cuxo valor é 1, mentres o fracaso ten lugar coa probabilidade $q=1-p$, e o seu valor é 0.

Deste xeito, a probabilidade de éxito e fracaso debe ser constante. Por exemplo, realizando o lanzamento dunha moeda, a probabilidade de sacar cara (éxito) é de 0,5 e esta non cambia porque a moeda é a mesma en cada experimento, polo tanto a probabilidade de sacar cruz (fracaso) é a mesma.

Cabe ter en conta que esta distribución é unha xeneralización da distribución de Bernoulli; probabilidade discreta que tamén da lugar a dous eventos mutuamente excluíntes.

É necesario saber que a súa variable é aleatoria, e para representar que X segue unha distribución binomial de parámetros n e p escríbese:

$$X \sim B(n, p)$$

Polo tanto, os parámetros fundamentais desta distribución son n , isto é o número de experimentos, p a probabilidade de éxito, q a probabilidade de fracaso e X o número de éxitos.

A probabilidade pódese calcular grazas á función de probabilidade da distribución binomial, que especifica o número de veces que pode ocorrer un evento nun número independente de lanzamentos n e onde p é a probabilidade de que o evento ocorra nun único lanzamento.

$$f(x) = \binom{n}{x} p^x (1 - p)^{n-x}$$

Así, n debe ser un enteiro positivo e p debe pertencer ao intervalo $0 \leq p \leq 1$.

En canto ás súas aplicacións, emprégase en todos os eventos nos que hai dous resultados, por exemplo, a definición do sexo dun bebé; o feito de que gañe ou perda un partido un equipo, o feito de superar ou non un exame, a efectividade ou inefectividade dun tratamento, o hábito de fumar (si/non). Tamén se usa con frecuencia no control de calidade. É un modelo probabilístico axeitado cando se toman mostras sobre unha poboación que pode considerarse grande, p representa a fracción de elementos defectuosos en dita poboación. Nestas aplicacións X representa o número de elementos atopados ao tomar unha mostra aleatoria de tamaño n .

Por outra parte, a distribución de Poisson debe o seu nome ao matemático francés Siméon Denis Poisson, quen se deu a coñecer grazas á súa obra *Recherches sur la probabilité des jugements en matières criminelles et matière civile* (Investigación sobre a probabilidade de dous procesos xudiciais en materia penal e civil). Esta distribución, ao igual que a binomial é unha distribución discreta que determina o número de sucesos que teñen lugar nun determinado período de tempo. Para isto, debe cumprir as seguintes características: os sucesos teñen unha probabilidade de éxito moi baixa, son independentes e non poden ocorrer dous sucesos ao mesmo tempo. Ademais, estes sucesos dependen do tempo.



Figura 2. Fotografía de Siméon Denis Poisson.

Para poder desenvolver esta distribución é necesario saber o valor de λ . Este parámetro representa o número promedio de eventos esperados por unidade de tempo e espazo, é dicir, é a taxa de ocorrencia do fenómeno que se observa, e o cal se obtén do produto de $n \cdot p$, sendo n o número de veces que se presenta un suceso e p a probabilidade de que ocorran estes sucesos, e isto represéntase do se-

guinte xeito:

$$X, Ps(\lambda)$$

Unha variable segue unha distribución de Poisson se os datos non son negativos, os sucesos teñen unha probabilidade de éxito moi baixa, son independentes e dependen do tempo. Ademais, non poden ocorrer dous sucesos ao mesmo tempo en ningún caso.

A función de probabilidade desta distribución defínese por:

$$F(X=k) = \frac{e^{-\lambda} * \lambda^k}{k!}$$

Nesta, $F(X=k)$ representa a probabilidade de ocorrencia cando a variable discreta X toma un valor finito k , onde k é o número de ocorrencias do suceso, e debe ser un número enteiro positivo.

En canto aos seus usos, emprégase no ámbito do risco operativo co obxectivo de modelar as situacións nas que se produce unha perda operacional. Tamén pode calcular por exemplo: o número de coches que pasan por un determinado punto dunha ruta, o número de erros ortográficos que se comete ao escribir unha única páxina...

As distribucións Binomial e Poisson manteñen unha relación de aproximación cando se satisfan certas condicións nos parámetros característicos de ambas. Así, a aproximación entre a distribución de Poisson, representada por $Ps(\lambda)$, e a distribución Binomial, representada por $B(n,p)$ funciona cando o número de probas é elevado, é dicir, n é un valor grande ($n \rightarrow \infty$) e a probabilidade de éxito é pequena, isto é, o valor de p é moi baixo ($p \rightarrow 0$).

En xeral, a aproximación funciona de maneira axeitada cando:

$$\begin{array}{ccc} n \geq 20 & \text{ou} & n \geq 100 \\ p \leq & & p \leq 0,10 \\ 0,050 & & \end{array}$$

Ao cumprirse estas condicións obtense o parámetro lamba (λ), cuxo valor promedio é $n \cdot p$:

$$\lambda = n \cdot p$$

Deste xeito, para lograr a aproximación entre ambas distribucións discretas e tendo en conta as condicións e parámetros descritos, é preciso obter un valor de λ que sexa menor ou igual ca dez:

$$\lambda \leq 10$$

2. Análise dos datos

A análise da converxencia de variables aleatorias partirá da creación de programas en R que permitan validar computacionalmente o seu funcionamento e efectividade.

En primeiro lugar, creamos un programa que permitise desenvolver a distribución binomial. Para iso, empregamos a función `rbinom`, almacenada no vector `v`, para obter un número determinado de observacións aleatorias que seguen este tipo de distribución.

```
v<-rbinom(size,n,prob)
hist(v)
```

Figura 3. Código distribución binomial

Nesta función `size` é o número de ensaios, `n` é o número de observacións aleatorias a ser xeradas e `prob` actúa como a probabilidade de éxito en cada un dos ensaios. A representación desta distribución conséguese a través da función `hist`.

O programa que permite desenvolver a distribución de Poisson emprega a función `rpois`, almacenada no vector `g`, onde `n` é o número de observacións a ser xeradas e `lambda` é a media, que se obtén do produto $n \cdot p$. Do mesmo xeito, a función `hist` permite representar graficamente esta distribución.


```
g<-rpois(n,lambda)
hist(g)
```

Unha vez definidas ambas distribucións de xeito independente a través do programa R, é preciso realizar un programa que permita a súa análise conxunta.

Para iso, e tendo en conta os parámetros característicos destas variables aleatorias, créanse os vectores n (número de observacións xeradas), p (probabilidade), j (probabilidade de que certo suceso teña lugar unha vez) e $lambda$ para poder almacenar os valores que caracterizan estas distribucións e á vez van determinar as súas aproximacións.

Agora, introdúcese os programas de cada distribución anteriormente expostos e a continuación, o cálculo numérico das probabilidades é posible grazas ás seguintes funcións, almacenadas como *Pbinom* e *Ppois*.

```
n<-
p<-
j<-
lambda<-n*p

g<-rpois(n,lambda)
v<-rbinom(size,n,p)

Pbin<-choose(n,j)*p^j*(1-p)^(n-j)
Ppois<-(exp(-lambda))*(lambda^j)/factorial(j)
```

Figura 4. Código 1 converxencia.

Primeiro, a función de probabilidade da distribución binomial, é calculada a través da función *choose*, que permite o cálculo do número combinatorio $\binom{n}{j}$ para logo ser multiplicado por p^j e finalmente por $(1-p)^{n-j}$.

Deste mesmo xeito, a expresión da función de probabilidade da distribución de Poisson vai permitir o cálculo da probabilidade de cada observación.

Por último, a representación conxunta de ambas distribucións conséguese a través da función *hist*. Os histogramas que se obteñen noméanse como ‘Binomial’ e ‘Poisson’ (*main*) e as escalas de ambos son axustadas empregando *xlim* para o eixo de abscisas e *ylim* para o eixo de ordenadas. Así mesmo, a través do comando *par* e a opción *mfrow* créase unha matriz de gráficos 1x2, isto é, os histogramas distribúense formando unha fila e dúas columnas.

```
n<-
p<-
j<-
lambda<-n*p

g<-rpois(n,lambda)
v<-rbinom(size,n,p)

Pbin<-choose(n,j)*p^j*(1-p)^(n-j)
Ppois<-(exp(-lambda))*(lambda^j)/factorial(j)

hist(v,main="Binomial",xlim = c(0,10), ylim = c(0,10))
hist(g,main="Poisson", xlim = c(0,10), ylim = c(0,60))

par(mfrow=c(1,2))
```

Figura 5. Código 2 converxencia.

Agora, tendo en conta as condicións precisas para que teña lugar a aproximación entre ambas distribucións, é dicir, cun valor de n elevado e un valor de p moi baixo, e obtendo un valor de λ menor ou igual ca dez, procuramos analizar a efectividade da aproximación.

En primeiro lugar, a nosa análise comeza cun valor de λ pequeno.

Pbin	9.9990100485084e-05
Ppois	9.99900004999833e-05

$n=100$
 $p = 0.000001$
 $\lambda = 0.0001$

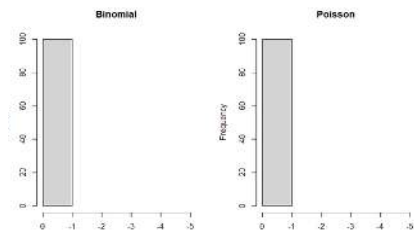


Figura 6. Gráfica 1.

Pbin	0.304407254517954
Ppois	0.303265329856317

$n=100$
 $p = 0.005$
 $\lambda = 0.5$

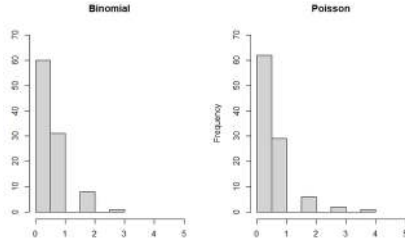


Figura 7. Gráfica 2.

Pbin	0.369729637649726
Ppois	0.367879441171442

$n=1000$
 $p = 0.01$
 $\lambda = 1$

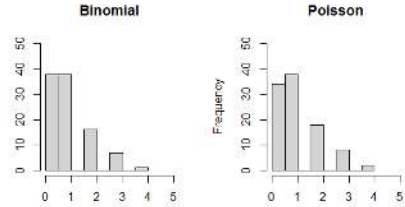


Figura 8. Gráfica 3.

Pbin	0.0904923385897146
Ppois	0.090483741803596

$n=100$
 $p=0.0001$
 $\lambda = 1$

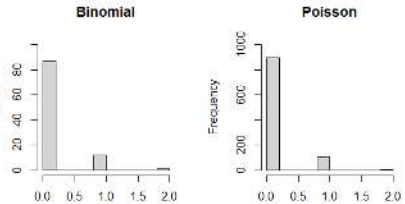


Figura 9. Gráfica 4.

Pbin	0.00261969341852239
Ppois	0.00268370102322009

$n=1000$
 $p=0.008$
 $\lambda = 8$

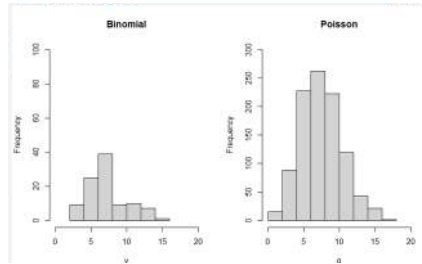


Figura 10. Gráfica 5.

Pbin	0.00268048158396687
Ppois	0.00268370102322009

$n=20000$
 $p=0.0004$
 $\lambda = 8$

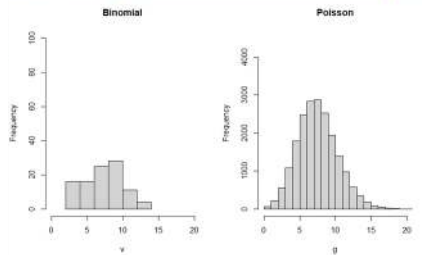


Figura 11. Gráfica 6.

Finalmente, probamos con valores de λ superiores a dez para comprobar a validez das condicións que permiten que se cumpla a converxencia.

Pbin	0.000166142684784144
Ppois	0.000183718708692702

$n=500$
 $p = 0.022$
 $\lambda = 11$

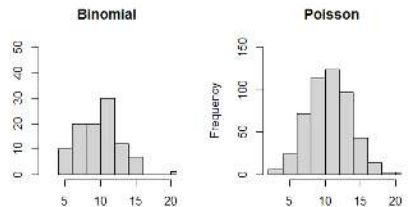


Figura 12. Gráfica 7.

Pbin	5.09258994083624e-09
Ppois	4.12230724487712e-08

$n=100$
 $p = 0.2$
 $\lambda = 20$

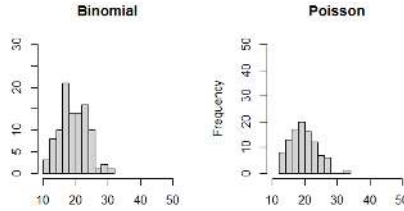


Figura 13. Gráfica 8.

3. Conclusións

O valor de λ determina que se produza ou non a converxencia entre a distribución binomial e a distribución de Poisson. Así, queda probado que cando este posúe valores menores ou iguais ca dez, se produce a aproximación entre ambas distribucións. Isto, ademais, trae consigo unha relación entre os valores n e p , de xeito que, cumpríndose as hipóteses formuladas para o grao de converxencia, canto maior é o valor de n e á vez, menor é o valor de p , as gráficas e a probabilidade presentan unha aproximación maior.

Así mesmo, resultaría interesante continuar a análise para lograr unha relación entre os parámetros que permitise concretar as condicións imprescindibles para a aproximación entre as distribucións.

Malia as limitacións que se nos presentaron ao longo desta investigación, chegamos á conclusión de que o valor de λ desempeña un papel fundamental na relación entre a binomial e a Poisson, que permite establecer unha relación de proporcionalidade inversa entre os parámetros n e p .

4. Bibliografía / Webgrafía

- Minitab.com. (2022). Recuperado do sitio web: <https://support.minitab.com/es-mx/minitab/18/help-and-how-to/probability-distributions-and-random-data/supporting-topics/distributions/poisson-distribution/>
- Sergas. (2014). Distribuciones de Probabilidad. Recuperado o 3 de marzo de 2022 , do sitio web: <https://www.sergas.es/Saude-publica/Documents/1927/4-Ayuda%20Distribuciones%20de%20probabilidad.pdf>
- Matemáticas Visuales. (2013). Distribución Binomial. Recuperado do sitio web: <http://www.matematicasvisuales.com/html/probabilidad/varaleat/binomial.html>
- Matemáticas Visuales. (2013). Distribución de Poisson. Recuperado do sitio web: <http://www.matematicasvisuales.com/html/probabilidad/varaleat/poisson.html>
- Moreno, V. (2021). Biografía de Jakob Bernoulli. Recuperado o 3 de marzo de 2022, do sitio web: <https://www.buscabiografias.com/biografia/verDetalle/8929/Jacob%20Bernoulli%20o%20Jacques%20Bernoulli>
- Morettini, M. (2013). Aproximaciones de distribuciones de probabilidad: Enfoque Empírico. Recuperado o 3 de marzo de 2022, do sitio web: <http://nulan.mdp.edu.ar/2040/1/morettini.2013.pdf>
- Dmuba. (2004). Aproximación Poisson a la distribución binomial. Recuperado do sitio web: https://www.dm.uba.ar/materias/optativas/procesos_estocasticos/2004/2/Proceso-Poisson.doc
- Martínez Gómez, M. y Marí Benlloch, M. (2010). La distribución binomial. Recuperado do sitio web: <https://riunet.upv.es/bitstream/handle/10251/7936/Distribucion%20binomial.pdf>
- Martínez Gómez, M. y Marí Benlloch, M. (2010). La distribución Poisson. Recuperado do sitio web: <https://riunet.upv.es/bitstream/handle/10251/7937/Distribucion%20Poisson.pdf>

Aproximación normal da distribución binomial

Álvarez-Lodeiro, I; Amosa-Martínez, U;
Castro-Ramos, I; Silva-Lodeiro, X.

IES MARCO DO CAMBALLÓN

Resumo. Con este traballo buscamos realizar un código que permita demostrar converxencias de variábeis aleatorias para obter as probabilidades daquelas situacións nas que só se poden obter dous resultados.

Abstract. With this essay we want to create a code that allows us to demonstrate convergences of random variables to obtain the probabilities of those situations in which only two results can be obtained.

1. Introducción

Unha distribución binomial $Bi(n,p)$ pódese aproximar por unha distribución normal, sempre que n sexa maior e p non esté moi próxima a 0 e 1. Dita aproximación consiste en utilizar unha distribución normal coa mesma media e a desviación típica da distribución binomial.

Distribución de probabilidade $Bi(n,p)$. Orixe, caracterización e aplicacións prácticas

Unha distribución de probabilidade é un modelo teórico que trata de explicar o comportamento dun fenómeno real. Actúa como unha función que asigna a cada suceso, cuantificado mediante unha variable aleatoria, a probabilidade correspondente. A historia da probabilidade comeza no S.XVII cando Pierre de Fermat e Blaise Pas-

cal tratan de resolver algúns problemas relacionados cos xogos do azar. Christian Huygens coñeceu a correspondencia entre Blaise Pascal e Pierre de Fermat suscitada polo cabaleiro De Méré, formulouse o como determinar a probabilidade de gañar unha partida e publicou en 1657 o primeiro libro sobre a probabilidade (*Liber de Lulo Alae*). Durante o S.XVIII, debido a popularidade dos xogos de azar, o cálculo de probabilidades tivo un notable desenvolvemento sobre a base da anterior definición de probabilidade. Destacan o teorema de Bernoulli e a distribución binomial, ademais é o primeiro caso particular estudado por De Moivre, do teorema central do límite. En 1809 Gauss iniciou o estudo da teoría de erros e en 1810 Laplace, completou o desenvolvemento desta teoría.

A obra de Gregor Mendel, *A matemática da Herencia*, foi unha das primeiras aplicacións importantes da teoría da probabilidade ás ciencias naturais. Dende as orixes, a principal dificultade para poder considerar a probabilidade como unha rama da matemática foi a elaboración dunha teoría suficientemente precisa como para que fose aceptada como unha forma da matemática. A principios do S.XX o matemático ruso Andrei Kolmogorov definiuna de forma axiomática e estableceu as bases para a moderna teoría da probabilidade que na actualidade é parte dunha teoría máis ampla como é a teoría da medida.

As características máis importantes a considerar sobre a distribución da probabilidade son: a probabilidade dun resultado específico está entre cero e un e que a suma das probabilidades de todos os resultados mutuamente excluíntes é 1. Ademais, toda distribución de probabilidade xérase por unha variable aleatoria x , e pode ser de dous tipos: Variable aleatoria discreta (x), cando unicamente pode tomar valores representados por números enteiros e un número finito de eles.

(Por exemplo: X é a variable que nos define o número de alumnos aprobados na asignatura de matemáticas nun grupo de 30 alumnos. (1,2,3 e así sucesivamente ou os 30)).

Por outra banda, a variable aleatoria continua (x) pode tomar valores expresados en números enteiros ou fraccionados.

(Por exemplo: X é a variable que nos define a concentración en gramos de ouro dalgunhas mostras de mineral (7.4 gr, 6.1, 1.9, 23.3, 12.7, 8.1, 9.5, 11.8, ... n)).

Utilizamos a distribución binomial en todos os eventos nos que só hai dous resultados, por exemplo suspender ou aprobar un exame ou tamén, como o que imos ver a continuación:

Caso 1:

Lánzase unha moeda 4 veces, cal sería a probabilidade de que saian máis caras que cruces? Para isto, utilizaremos os parámetros $n:1$ e $p:0,5$.

Solución:

$$\begin{aligned} B(4, 0.5) \quad p = 0.5 \quad q = 0.5 \\ p(x \geq 3) &= p(x = 3) + p(x = 4) \\ &= \binom{4}{3} \cdot (0.5)^3 \cdot 0.5 + \binom{4}{4} \cdot (0.5)^4 = 0.3125 \end{aligned}$$

Caso 2:

A probabilidade de que un home acerte no blanco é $\frac{1}{4}$. Se dispara 10 veces, cal sería a probabilidade de que acerte 3 veces? Para isto utilizamos o seguinte código:

Solución:

$$\begin{aligned} B(10, \frac{1}{4}) \quad p = \frac{1}{4} \quad q = \frac{3}{4} \\ p(x = 3) &= \binom{10}{3} \left(\frac{1}{4}\right)^3 \left(\frac{3}{4}\right)^7 = 0.25 \end{aligned}$$

Distribución de probabilidade normal. Orixe, caracterización e aplicacións prácticas

A distribución normal é un modelo teórico capaz de aproximar satisfactoriamente un valor dunha variable aleatoria a unha situación ideal, esta adapta unha variable aleatoria a unha función que depende da media e a desviación típica, é dicir, a variable aleatoria terá a

mesma representación pero con lixeiras diferenzas.

A distribución normal coñécese como a curva de Gauss ou a campá de Gauss. A distribución normal foi presentada por primeira vez por Abraham de Moivre nun artigo do ano 1733, que foi reimpresso na segunda edición de *The Doctrine of Changes* de 1738, onde afirma que dita distribución se asemella á distribución binomial, logo foi ampliado por Laplace no seu libro *Teoría Analítica de las Probabilidades* en 1812, e na actualidade coñécese como o Teorema de De-Moivre-Laplace que empregou a distribución na análise dos erros dos experimentos.

Posteriormente, Gauss afirmou usar este método en 1794 e xustificou en 1808 asumíndoa como unha distribución normal dos erros. Como consecuencia esta distribución asociouse a el debido a que a empregou con seguridade na análise dos datos astronómicos. Asemade, o nome de “distribución normal” foi otorgado por Charles S. Pierce, Francis Galton e Wilhelm Lexis. A distribución normal é a máis importante de todas as distribucións de probabilidade. Consiste nunha distribución variable continua, cun campo de variación $[-\infty, \infty]$. Foi recoñecida por Abraham de Moivre. Máis tarde, Gauss elaborou un desenvolvemento máis fondo e formulou a ecuación da curva. De aí o nome da “campá de Gauss”.

Dita distribución pódese afirmar que é asintótica ao eixo de de abscisas, é dicir, os seus extremos nunca tocan o eixo horizontal, cuxos valores tenden ao infinito; é simétrica respecto a súa media; no centro da curva encóntrase a media, mediana e moda; a distancia entre a liña trazada na media e o punto de inflexión é igual a desviación típica (σ), é dicir, canto maior sexa a desviación máis aplanada é a curva de densidade. Ademais, a forma da campá de Gauss depende dos parámetros μ e σ . A media indica a posición da campá, de modo que para diferentes valores de μ a gráfica é desplazada ao longo do eixo horizontal. Por outra banda, a desviación estándar determina o grado de apuntamento da curva, é dicir, canto maior sexa o valor da σ , máis se dispersarán os datos entornando a media, polo que a curva será máis plana.

Converxencia normal dunha $Bi(n,p)$. Valores teóricos de media e desviación típica. Explicación, historia e aplicacións

En primeiro lugar, cando a cantidade de experimentos que facemos tende a infinito, unha binomial acaba asemellándose a unha normal. Por exemplo, se nós realizamos repetidas veces o mesmo proceso, cunha probabilidade de 0,5 (onde hai as mesmas posibilidades de que ocorra unha das opcións); cun total de catro valores posibles[0,1,2,3]; obtemos unha representación dunha binomial que se asemella moito a unha representación dunha normal. Pero se nós collemos calquera número de posibilidades cunha probabilidade calquera obteríamos unha representación semellante a do caso anterior.

Ademais, a distribución binomial é unha distribución de probabilidade discreta que conta o número de éxitos para variables dicotómicas, é dicir, aquelas nas que só existen dúas opcións. Podemos utilizar a distribución binomial para determinar o sexo dun bebé, saber se o noso equipo vai a gañar, saber se imos aprobar ou suspender o exame ou se vai caer unha moeda amosando cara ou cruz. Imos fixarnos no exemplo do sexo do bebé, utilizando os valores de probabilidade $n=1$, é dicir, os resultados poden ser neno ou nena. Segundo numerosos estudos, de cada 20 bebés 12 son nenas (0,51 aproximadamente) e 8 son nenos, pero todo é probabilidade.

En canto aos valores teóricos de media e desviación típica, sempre que representamos unha distribución binomial temos que proporcionar uns parámetros n e p que serán os que determinen a forma da curva. Para calcular os valores teóricos de media ($n \cdot p$) e desviación típica ($\sqrt{n \cdot p \cdot (1-p)}$) temos a seguinte fórmula:

2. Análise dos datos

Neste apartado, procedemos a explicar como deseñamos un código en R que nos permitira traballar coa distribución binomial. Unha vez elaborado o código, procedemos á creación de histogramas de diferentes números para a súa posterior análise. A partir destes his-

togramas imos tentar tirar conclusións sobre como funciona a aproximación normal da distribución binomial. Para iso, explicaremos primeiro como elaboramos os programas que demostren dita converxencia.

2.1. Elaboración do código en R

Como xa se explicou anteriormente, unha distribución binomial $Bi(n,p)$ pódese aproximar por unha distribución normal, sempre que n sexa grande e p non esté moi próxima a 0 e 1. Para iso, creouse unha mostra de 200 valores dunha binomial $Bi(n,p)$ sendo n o número de éxitos nunha mostra, e p a probabilidade de que un valor estadístico calculado sexa posible.

Este programa, ademais de calcular probabilidades, proporciona os valores característicos da distribución (media, varianza ou desviación típica) como información complementaria. Para calcular a media que nos indica o valor medio dun fenómeno aleatorio nunha distribución binomial empregamos a función “*mean*”, e para o cálculo da desviación típica que fai referencia a unha medida que se utiliza para cuantificar a variación ou a dispersión de datos numéricos nunha variable aleatoria, utilizamos a función “*sd*”. Logo, aplicamos tanto na media como na desviación a función “*print*” para que imprima na consola os seus valores correspondentes.

```
m<-200
n<-20
p<-0.4
v<-rbinom(m,n,p)
media<-mean(v)
print(media)
destip<-sd(v)
print(destip)
```

Fig.3. Funcións “*mean*”, “*sd*” e “*print*”.

A continuación, calculamos os valores teóricos da media, tendo como fórmula: $n \cdot p$; e desviación típica $n \cdot p \cdot (1-p) / m^{0.15}$, sendo “ m ”

o número de observacións aleatorias xeradas. Tras facer estes cálculos, o programa devólvenos os valores correspondes da media e desviación típica teóricas tras o emprego da función “*print*”.

Para finalizar, utilizamos a función “*hist*” para obter unha representación gráfica da distribución binomial; a curva normal coa media e desviación típica mostrais e a curva normal coa media e desviación típica teóricas calculadas anteriormente mediante a función “*curve*”. Para distinguir a curva normal da curva teórica, empregamos a función “*main*”, “*col*” e “*lty*” para personalizar o seu nome, a súa cor e o seu trazado respectivamente.

```
hist(v, freq=FALSE, main = "Distribución binomial" , col = "white",
     xlab = "", ylab = "", ylim = c(0,0.26))
curve(dnorm(x, mean=mean(v), sd=sd(v)), add=TRUE, col="red", lwd = 5, lty= 3)
```

Fig.4. Funcións “*hist*” e “*curve*”.

2.2. Análise das gráficas.

Agora procedemos a analizar as gráficas da distribución binomial para ver o seu comportamento en función dos números propostos. Para isto, xeramos unha distribución binomial con parámetros $n=20$, $p=0.5$ e repetimos o proceso 200 veces, obtendo o seguinte histograma:

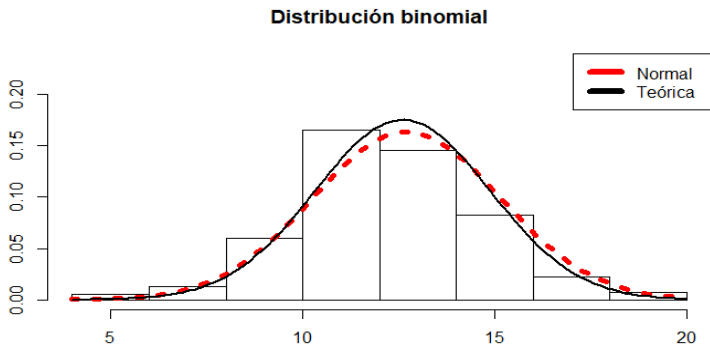


Fig.5. Bi(20,0.5).

Vemos representado o histograma, xunto con dúas funcións: unha discontinua, que representa os valores da media e desviación normal; e outra continua que pertence á desviación e media teóricas.

Observamos que as funcións adquiren unha forma semellante, isto débese a que cantos máis posibles valores obtidos haxa, maior vai ser o parecido entre elas. Por exemplo, se utilizáramos exactamente os mesmos parámetros salvo n , a gráfica de datos amosaría claras diferenzas coa anterior dada.

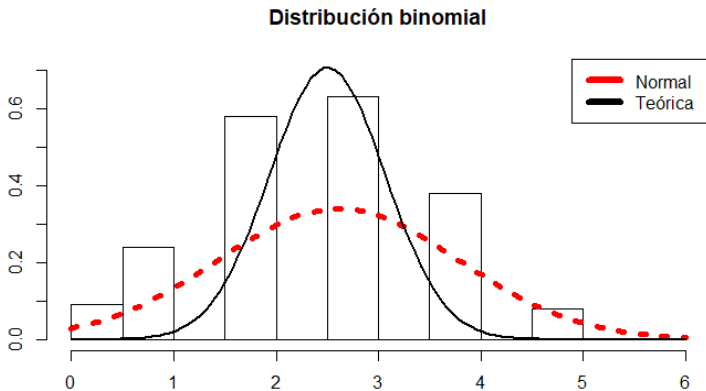


Fig.6. Bi(5,0.5).

Esta gráfica ten $n=5$, $p=0.5$, e repetimos o proceso 200 veces. A diferenza entre funcións é moito máis notable agora. Mentres que a normal tende a valores non moi distanciados entre eles, a teórica ten un crecemento moito máis pronunciado e contrastando co caso anterior no que se asemellaban máis.

Agora imos facer o mesmo, pero variando o valor de p , no que traballaremos con 0.2, e voltaremos ao valor orixinal de $n = 20$.

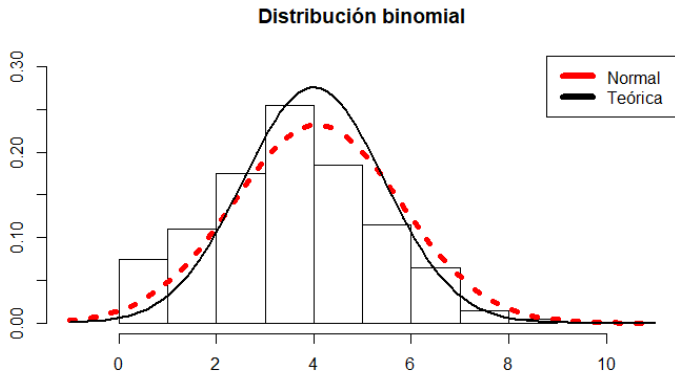


Fig.7. Bi(20,0.2).

Vemos como o histograma tende a agrupar os seus valores nos primeiros números. En canto aos resultados obtidos, temos un caso intermedio entre o primeiro gráfico e o segundo. Crecen máis ou menos igual, pero os seus máximos son diferentes, a ídem da teórica acada un máximo máis alto con respecto da normal.

Por último, para determinar con máis exactitude como pode influír o valor p na representación gráfica. Para este último caso, utilizaremos un valor de p maior que os anteriores utilizados, $p= 0.8$.

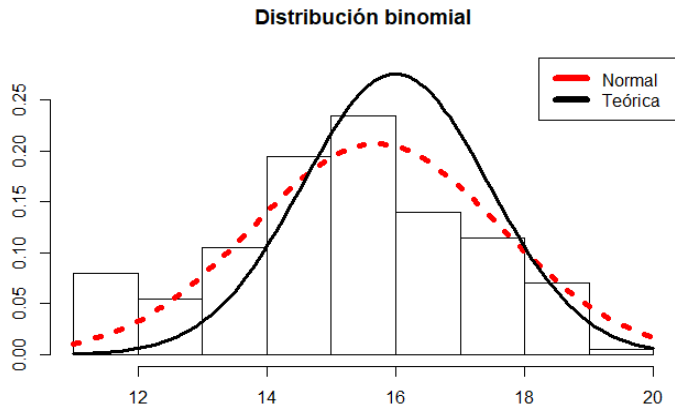


Fig.8. Bi(20,0.8).

Neste caso, os histograma agrúpanse nos valores medios, diferenciándose da anterior, no que se agrupaban nos primeiros valores.

3. Conclusións

Con este artigo elaboramos unha análise dos histogramas donde se pode apreciar nas gráficas anteriores como a medida que aumenta n mellora o parecido das gráficas de barras das distribucións binomiais (discretas) á gráfica da distribución normal estándar (continua).

Tendo en conta isto, podemos supoñer que, dende un punto de vista empírico pódese aproximar a distribución binomial á distribución normal, é dicir, en lugar de realizar os cálculos de probabilidades coa distribución binomial, pódese utilizar a distribución normal. E esta aproximación será o suficientemente boa se $n \geq 30$ e se $0,1 \leq p \leq 0,9$. Con todo, debe quedar claro que se n é infinito, non hai coincidencia entre a distribución binomial e a normal, xa que tamén depende de p .

4. Bibliografía / Webgrafía

- Martínez Mediano José María (s.d.). *Distribuciones de probabilidad*. Recuperado o 3 de marzo do 2022, do sitio Web: <https://matematicasiesoja.files.wordpress.com/2018/12/Funciones-de-probabilidad-2%C2%BA%Bachillrato.pdf>
- Estadística para todos (2008). *Historia de la probabilidad*. Recuperado o 3 de marzo do 2022, do sitio Web: https://www.estadisticaparatodos.es/historia/histo_proba.htm
- Esan graduate school of business (2021). *Conoce las principales distribuciones de probabilidad*. Recuperado o 3 de marzo do 2022, do sitio Web: <https://www.esan.edu.pe/apuntes-empresariales/2016/10/conoce-las-principales-distribuciones-de-probabilidad/>
- Ies Castelar (s.d.). *Aproximación de la distribución binomial por la normal*. Recuperado o 3 de marzo do 2022, do sitio web:<https://iescastelar.educarex.es/web/departamentos/matematicas/matematicascss2ba/matematicas2ccss/tabinomialpornormal.htm>
- Economipedia (2022). *Distribución normal*. Recuperado o 3 de marzo de 2022, do sitio Web: <https://economipedia.com/definiciones/distribucion-normal.html>
- Kunin Daniel (s.d.). *Distribuciones de probabilidad*. Recuperado o 2 de marzo de 2022, do sitio Web: <https://seeing-theory.brown.edu/probability-distributions/es.html>
- Fisterra (2001). *La distribución normal*. Recuperado o 3 de marzo de 2022, do sitio Web: <https://www.fisterra.com/formacion/metodologia-investigacion/la-distribucion-normal/>
- Universidad Politecnica de Valencia (s.d.). *La distribución binomial*. Recuperado o 3 de marzo de 2022, do sitio Web: <https://riunet.upv.es/bitstream/handle/10251/7936/Distribucion%20binomial.pdf>
- Shinigamr13 en Blog educativo (2013). *Historia de la Distribución normal*. Recuperado o 3 de marzo de 2022, do sitio Web: <https://rksbet.wordpress.com/2013/02/26/historia-de-la-distribucion-normal/>
- Warbletoncouncil (s.d.). *Distribución normal: Fórmula, características, ejercicio*. Recuperado o 3 de marzo de 2022, do sitio Web: <https://gl.warbletoncouncil.org/distribucion-normal-10549>

Simulación en R do dilema Monty Hall

Castelo-García, S.; López-Castro, M.;
Torres-Novo, N.

IES MARCO DO CABBALLÓN

Resumo. Con este artigo buscamos simular o dilema de Monty Hall, demostrando a probabilidade de éxito así como de fracaso asociada a cada unha das escollas. Para iso, realizamos unha verificación computacional do problema e, a través diste, analizamos o conxunto de resultados.

Abstract. With this research, we are trying to simulate the Monty Hall dilemma, proving the probability of success or failure associated with each choice. For that, we created a programme in R to verify the problem and we analyzed the results.

1. Introducción

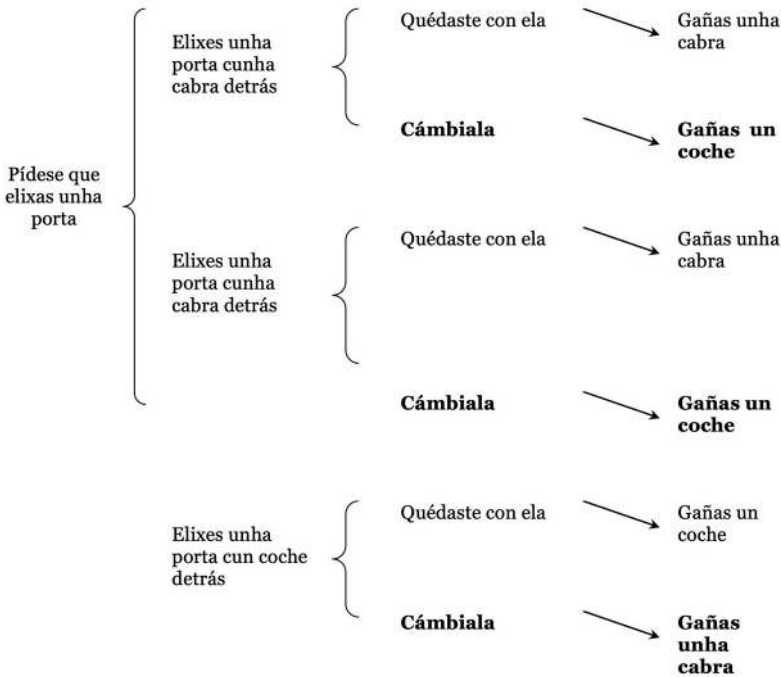
Este dilema está inspirado no concurso de televisión *Let's Make a Deal* emitido entre 1963 e 1986 na televisión estadounidense e o seu nome provén do presentador do concurso, Monty Hall.

Neste concurso, o xogador atópase con 3 portas, tras as que se agochan 3 premios, entre as cales deberá escoller unha. O seu premio será aquilo que se oculte detrás da porta escollida. Nunha delas hai un coche, mentres que nas dúas restantes hai unha cabra. Posteriormente a que o participante realice a súa elección, o presentador abrirá unha das outras dúas portas restantes. Dado que este coñece o que hai detrás de cada porta, abre unha que esconde unha cabra e móstralla ao participante. Nese intre o concursante terá unha última oportunidade de cambiar a súa elección e facerse co automóbil.

Deste xeito, o concursante atoparase ante a seguinte disxuntiva: Debe manter a súa elección orixinal ou pola contra será óptimo es-

coller a outra porta? E neste caso, hai algunha diferenza?

A primeira vista parece obvio que é igual. A intuición dinos que agora, quitando unha porta sen premio, a porta que nós escollimos ten un 50% de posibilidades de ter unha cabra e polo tanto o 50% restante de agochar o coche, deste xeito da igual cambiar que non facelo. Pero en realidade, isto non é correcto. Para comprendelo imos apoiarnos en primeiro lugar nunha explicación gráfica, xa que esta forma visual é a máis sinxela de comprender.



Velaquí un exemplo gráfico:

Cada unha das portas ten $\frac{1}{3}$ de probabilidade de agochar o coche detrás. O xogador escolle a porta 1, que ten $\frac{1}{3}$ desta probabilidade, polo tanto as portas 2 e 3 teñen os $\frac{2}{3}$ restantes (a suma de todas as probabilidade de eventos posibles debe ser 1).

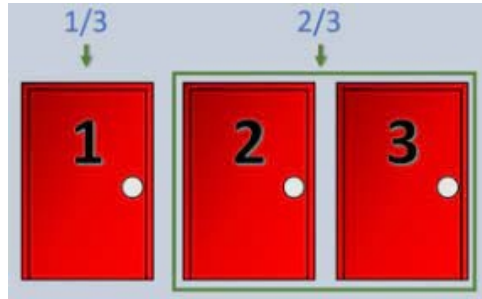


Fig.1. Exemplificación do concurso.

Cando o presentador abre unha porta, neste caso a 3, amosanos que a probabilidade de que nesa estivese o premio era realmente 0. Poderíase pensar que agora, con só 2 portas, cada unha delas terá $\frac{1}{2}$ de probabilidade de agochar o coche detrás, pero en realidade, os $\frac{2}{3}$ seguen estando entre as portas 2 e 3, pero como coñecemos que a porta 3 non agocha o premio, a súa probabilidade pasa á porta 2, polo que esta conta con $\frac{2}{3}$, fronte ao $\frac{1}{3}$ da porta 1.

Isto sucede en cada xogada, polo que a probabilidade de éxito é sempre maior cando o xogador cambia de porta despois de que o presentador abra unha que agochaba unha cabra.

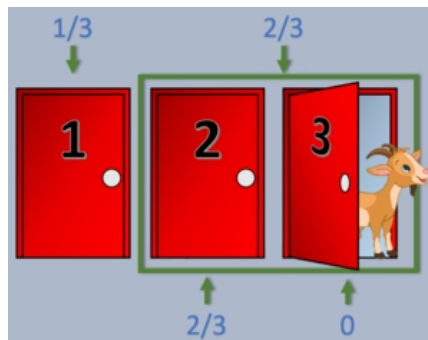


Fig. 2. Exemplificación de Monty Hall.

Como conclusión, a mellor decisión e estratexia será cambiar sempre, pois a probabilidade efectiva de gañar é o dobre da correspondente ao xogador que non cambia nunca.

Unha forma máis clara de observalo é reformular o problema. Se en lugar de haber só tres portas, o xogo estivese formado por 100, e tras a elección orixinal o presentador abraza 98 das restantes para mostrar que tras elas hai cabras, existen dúas posibilidades: si o xogador decide non cambiar a súa elección gañaría o coche só si foi a súa elección orixinal (1 de cada 100 veces), mentres que si a cambia, gañaría o vehículo si non o escolleu previamente (99 de cada 100 veces).

2. Análise dos datos

Para levar a cabo a verificación destas probabilidades elaboramos un programa en R.

```
montyhall<-function(n){
  contador<-0
  contador2<-0
  for (i in 1:n) {
    portas<-c("a","b","c")
    premio<-sample(portas)[1]
    eleccion<-sample(portas)[1]
    aberta<-sample(portas[which((portas!=premio)&(portas!=eleccion))])[1]
    eleccion2<-sample(portas[which((portas!=eleccion)&(portas!=aberta))])
    if(premio==eleccion){
      contador<-contador+1}
    if(premio==eleccion2){
      contador2<-contador2+1}
  }
  print(paste("contador=",contador,"contador2=",contador2))
}
```

Fig.3. Programa da función “montyhall”

No programa, creamos un vector, ao que chamamos “portas”, con 3 variables(“a”, “b” e “c”). A función *sample*, coloca estas tres portas nunha orde aleatoria, cada vez dunha maneira distinta. Creamos “premio”, onde fixemos un *sample* de portas e escollemos a primei-

ra, nesa porta está o noso coche. Da mesma maneira, creamos “eleccion”, que escolle outra porta, que pode ser a mesma ou non, esta é a porta escollida polo xogador. En “aberta”, facemos algo semellante, pero empregando a función *which* especificamos que a porta escollida ten que ser distinta da que esconde o premio e distinta da que abriu o presentador (para isto empregamos o comando `!=`, para impoñer que sexa distinto de “premio” ou de “aberta”, e o comando `&` para que sexa distinta de ambos á vez), esta é a porta que abriu o presentador, que sempre vai agochar unha cabra. Posteriormente creamos a opción de “eleccion2”, onde, empregando os mesmos comandos facemos que sexa distinta da primeira elección e distinta da porta aberta polo presentador, sería a porta escollida se o xogador decidise cambiar de porta.

A continuación creamos dous contadores, que comezan estando en 0. En “contador” súmase 1 cada vez que o premio coincide coa primeira elección do xogador, e en “contador2” súmase 1 cada vez que o premio coincide coa segunda elección do xogador, cando este gaña o premio ao cambiar de porta. Como xa explicamos anteriormente, é máis probable que o xogador gañe ao cambiar de porta, polo que o “contador2” estará formado por valores maiores que o “contador”.

Creamos unha función, chamada “montyhall”, onde facemos que se repita todo o proceso as veces establecidas.

```
> montyhall(100)
[1] "contador= 31 contador2= 69"
> v<-c(31,69)
> barplot(v, main = "Gráfico de frecuencias", xlab = "Acertos 1ª elección
  Acertos cambiando a elección ",ylab = "Frecuencia", col = rainbow(2))
```

Fig.4. Creación dun diagrama de barras

Poñendo na liña de entrada “montyhall” dun número, o programa devolve a lonxitude dos contadores, que gardamos no vector “v”. Facemos un *barplot* (función que fai gráficos de barras) de “v”. En *barplot* modificamos o título empregando *main*, e os nomes dos distintos eixes empregando *xlab* e *ylab*, ademais usamos `col=rainbow(2)` para cambiar as cores da gráfica. O programa de-

volvemos unha gráfica de barras, onde vemos as veces que se gaña mantendo a primeira elección de porta (“contador”) ou cambiando a nosa elección (“contador2”).

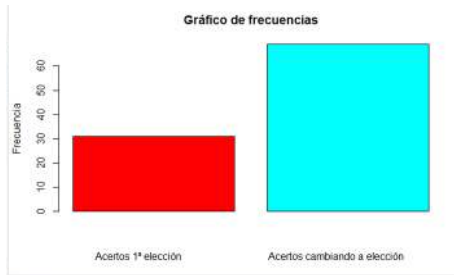


Fig.5. Diagrama de barras do número 100



Fig.6. Diagrama de barras do número 10



Fig.7. Diagrama de barras do número 30

A continuación procedemos a elaborar gráficas de varios números, para ver se verifican as probabilidades demostradas matematicamente. Comezamos facendo “montyhall” de números máis pequenos, como o 10, na figura 6, o ou 30, figura 7. En números pequenos, non é tan visible nas gráficas a proporción de gañar se quedamos coa primeira porta escollida ($1/3$) ou se cambiamos de porta ($2/3$). Conforme imos facendo gráficas con números máis grandes, as veces que o premio está tras a porta escollida na primeira elección e as veces que toca o premio cando cambiamos de porta, vanse achegando a estas proporcións, como vemos nas gráficas do número 1000, figura 8, e na do número 100000, figura 9.

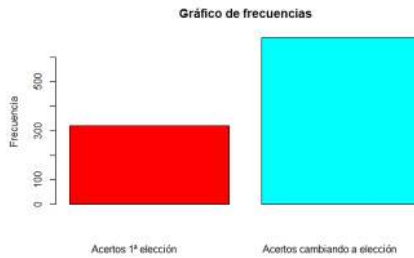


Fig.8. Diagramas de barras do número 1000



Fig.9. Diagrama de barras do número 100000

3. Conclusións

A elaboración do programa en R e os cálculos matemáticos demostran que a mellor estratexia é a do cambio de porta, debido a que a probabilidade de que o premio estea na porta escollida na primeira elección é $\frac{1}{3}$ mentres que a probabilidade de que estea nalgunha das outras é $\frac{2}{3}$. As gráficas demostran isto, e delas podemos concluir que cantas máis veces repetimos o proceso, máis se achegan os resultados ao estimado, é dicir, gañar o premio aparece con maior frecuencia asociado á estratexia de cambio de porta.

4. Webgrafía

- Gabriel Páez (2020). Problema de Monty Hall. Recuperado o 3 de febreiro de 2022, do sitio web: <https://economipedia.com/definiciones/problema-de-monty-hall.html>
- Dina Mohamed (s.f.). El Problema de Monty Hall. Recuperado o 3 de febreiro de 2022, do sitio web: <http://estadis.net/recursosestadisticos/index.php/recursos-virtuales/174-el-problema-de-monty-hall>
- Statologos (s.f.). Una guía para `dnorm`, `pnorm`, `qnorm` y `mnorm` en R. Recuperado o 3 de febreiro de 2022, do sitio web: <https://statologos.jaol.net/dnorm-pnorm-rnorm-qnorm-en-r/>
- Statologos (s.f.). Una guía para `dbinom`, `pbinom`, `qbinom` y `rbinom` en R. Recuperado o 3 de febreiro de 2022, do sitio web: <https://statologos.jaol.net/dbinom-pbinom-qbinom-rbinom-en-r/>
- Statologos (s.f.). Una guía para `dpois`, `ppois`, `qpois` y `rpois` en R. Recuperado o 3 de febreiro de 2022, do sitio web: <https://statologos.jaol.net/dpois-ppois-qpois-rpois-r/>
- R Coder (s.f.). Gráfico de barras en R. Recuperado o 17 de febreiro de 2022, do sitio web: <https://r-coder.com/grafico-barras-r/>

A distribución de Poisson na desintegración nuclear radiactiva

*Casal-Blanco, R.; Fraiz-Casas, A.;
Moreira-Pereiro, A.*

IES MARCO DO CAMBALLÓN

Resumo. Con este traballo trataremos de averiguar se os datos das partículas irradiadas pola desintegración nuclear radiactiva do Americio seguen a distribución de Poisson, e se esta tende a unha distribución normal conforme incrementamos o intervalo temporal da mostra. Para iso almacenamos os datos en LibreOffice Calc para, posteriormente, crear as gráficas correspondentes co programa R.

Abstract. With this project we will try to find out if the data from radiated particles by Americium' radioactive decay takes the shape of Poisson's distribution, and also if it tends to a normal distribution when the time interval and the sample is raised. To archive that, we store up the data in LibreOffice Calc. Afterwards, we create the graphics with the help of R program.

1. Introducción

Ata finais do século XIX dábase por sentado que as compoñentes da materia, os átomos, eran estables, iguais e inmutables. Só as moléculas sufrían transformacións despois da recombinación dunhas con outras, os átomos que as formaban non.

Antoine Henri Becquerel foi quen observou un suceso que parecía contradicilo. Unhas placas fotográficas, rodeadas dun papel negro, ennegrecíase tras estar preto da peclenda durante aproximadamente un día. A lóxica tradicional explicaría coa interacción cun axente externo, porén as placas non foran sometidas a ningún axente químico, a luz nin ao aumento de temperatura. Por esa razón, o científico chegou a conclusión de que tiña que haber algo producido pola peclenda que atravesaba as placas fotográficas conseguindo

velar as placas.

Máis adiante averiguaría que os raios producidos polo uranio, un dos compoñentes da peclenda, son capaces de atravesar láminas finas de metal e, incluso, de ionizar o aire.

Pouco despois, Marie e Pierre Curie conseguiron aillar 0,1 g de cloruro de uranio, que é o material radiactivo da peclenda, para determinar as propiedades do elemento uranio. A continuación, descubriron que o torio se comportaba de maneira similar. Posteriormente, aillaron o Polonio e o Radio, dous novos elementos químicos que eran incluso máis radioactivos que o uranio. Á vez déronlle o nome ao fenómeno: radioactividade.

Ernest Rutherford investigou os raios emitidos polo uranio, concluíndo que eran de dous tipos, alfa e beta, e que logo de emitilos se transformaba noutro elemento. Paul Villard descubre outro tipo de raios, os gamma, similares aos raios X.

Actualmente sabemos que os núcleos se manteñen unidos, a pesar de estar formados por protóns, que se repelen entre si, e neutróns, grazas á forza nuclear forte que os atrae. Porén, existen núcleos incapaces de manterse unidos, é dicir, esa forza nuclear forte é superada pola forza electromagnética, e comezan a emitir partículas e radiacións.

Estes átomos cumpren tres condicións principais: teñen máis protóns que neutróns, son moi pesados e contan cun número impar de partículas no núcleo.

As radiacións que emiten estes átomos son de tres tipos principalmente:

- Alfa: son núcleos de helio, isto é, dous protóns e dous neutróns sen os correspondentes electróns.
- Beta: son electróns ou positróns, partículas cunha masa ínfima e carga negativa e positiva respectivamente.
- Gamma: consiste en luz, é dicir, fotóns, pero cunha enerxía 1 millón de veces superior á luz que nos observamos, polo tanto moi penetrante.

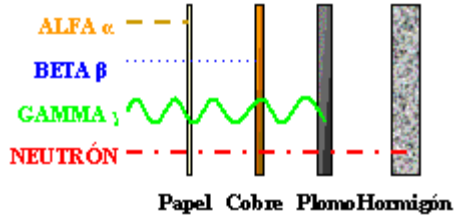


Fig.1. Capacidade de penetración dos distintos tipos de radiación electromagnética.

Esta radiación pódese clasificar en función da súa acción sobre a materia en ionizante e non ionizante. A radiación ionizante, á que pertencen os raios X, alfa, beta, gamma, caracterízase por ser moi enerxética e, polo tanto, capaz de desprender electróns dos átomos. A non ionizante é menos enerxética e non é quen de producir este efecto.

Os periodos de desintegración dos átomos radiactivos son constantes, pero moi distintos entre os elementos. A vida media é o tempo que se precisa para que se desintegre metade dunha cantidade determinada dun elemento radiactivo. Vai desde segundos, por exemplo a do C 15 é de 24 s, a miles de anos, o U 238 tarda 4 500 000 000 anos.

Por último, a desintegración nuclear radiactiva é un proceso que se produce non só nas centrais nucleares, senon que tamén se leva a cabo naturalmente na Terra, nas rochas, no aire, nos seres vivos, etc.

A distribución de Poisson é unha distribución de variable discreta, é dicir, só adquire algúns valores de y . Esta é a distribución aleatoria indicada se queremos determinar o número de sucesos que teñen lugar nun intervalo de tempo determinado. Ademais diso, debe cumprir unhas características para que se poida aplicar:

- Os sucesos deben ter unha probabilidade de éxito moi baixa, isto é, son sucesos raros.
- Non veñen determinados por ningunha condición, polo tanto son

independentes uns doutros.

- Dous sucesos non poden ocorrer á vez en ningún caso.

- A taxa de aparición é máis ou menos constante e independente do número de sucesos que teñen lugar. Porén, si que depende o intervalo de tempo seleccionado.

A expresión desta distribución de probabilidade é a seguinte:

$$P(x) = \frac{e^{-\lambda} \cdot \lambda^x}{x!}$$

x refírese aos sucesos que teñen lugar nun intervalo determinado.

O parámetro λ ten que ver coa frecuencia media coa que ocorren os sucesos nun intervalo de tempo determinado, é dicir, son os sucesos que cabe esperar que teñan lugar nese intervalo. O valor de λ é igual ao da varianza (σ^2), polo tanto, tamén ao cadrado da desviación típica.

Outra propiedade desta distribución é que a suma de n variables independentes de Poisson é, á súa vez, outra distribución de Poisson con λ igual á suma dos λ das variables que as forman.

A distribución Normal é unha variable aleatoria continua que depende da media (μ) e da desviación típica (σ), polo que se designa $N(\mu, \sigma)$.

A media (μ) defínese como o promedio do conxunto de datos. A desviación típica (σ) é a dispersión dos datos con respecto á media; graficamente ao aumentar a desviación típica a gráfica tende a aplanarse, isto é, os valores atopanse máis alonxados da media. Asemade, nos puntos $\mu - \sigma$ e $\mu + \sigma$ presenta puntos de inflexión.

Esta distribución pode adoptar todos valores no intervalo $(-\infty, +\infty)$, polo tanto, o seu dominio son os números reais. Por ese motivo conta cunha asíntota no eixe OX cando x tende a $+\infty$ e $-\infty$.

Unha das súas propiedades é a simetría respecto ao valor da media, onde adopta o seu máximo, por iso ese valor sempre coincide coa mediana (valor central) e coa moda (valor máis repetido).

Ademais, a área comprendida entre a curva e o eixe OX é sempre igual a 1. Isto indica que nesta área se atopan o total de probabilidades. Da mesma forma, podemos averiguar que a área entre $\mu - \sigma$ e $\mu + \sigma$ é de 0.68, é dicir, esa é a probabilidade de que o resultado se atope entre eses dous valores de x . Igualmente, a área comprendida entre $\mu - 2\sigma$ e $\mu + 2\sigma$ é de 0.95, co que podemos concluír que é entre estes dous valores onde se atopa unha maior probabilidade. Isto ten sentido tendo en conta que a curva tende a achegarse ao eixe OX conforme se aproxima ao $+\infty$ e $-\infty$.

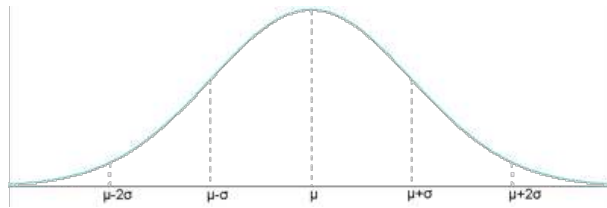


Fig.2. Distribución normal.

Relacionado con ambas distribucións aleatorias está o teorema central do límite, que ten múltiples aplicacións no campo da estatística. O motivo é que certifica que calquera distribución da que partamos se vai aproximar á normal sempre que a mostra sexa suficientemente ampla. Isto permítenos inferir hipóteses ou resultados nun intervalo, simplificando en gran medida problemas estatísticos.

O teorema do límite central certifica que a media dun espazo mostral, con variables aleatorias independentes, aproxímase a unha distribución normal, e que a converxencia é maior canto maior sexa o espazo mostral e o número de medias. Isto quere dicir que se distribúe con forma de campá de Gauss.

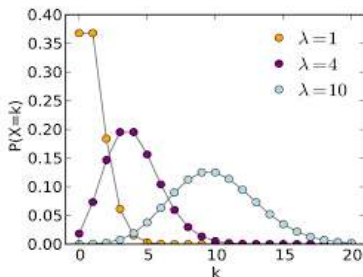


Fig.3. Distribución de Poisson variando o parámetro λ .

2. Análise de datos

Para obter os datos das partículas desprendidas nun intervalo de tempo por un certo material, realizamos o seguinte experimento. Para desenvolvelo empregamos un detector Geiger, que é capaz de detectar as partículas que emite a fonte radiactiva de Americio, o outro material que foi requirido. O proceso foi sinxelo, antes de nada encendemos o detector Geiger e observamos que detecta raios cósmicos. Tras iso, achegamos a fonte de Americio, un elemento radiactivo, a unha distancia de 2cm aprox. co que conseguimos obter a frecuencia aproximada de partículas detectadas por segundo que buscabamos. A continuación iniciamos de novo o contador do detector e medimos durante 15 min. É importante resaltar que a radiación detectada non só procede do detector, senon que aparece reflexada tamén a radiación natural.



Fig.4. Imaxe da medición

En primeiro lugar, elaboramos un ficheiro de datos en LibreOffice Calc onde almacenamos os datos recollidos polo detector geiger. Así pois creamos unha táboa recollendo o número de deteccións cada 2 , 5 e 10 segundos.

Unha vez que temos recollidos os datos precisamos poder leelos no programa R para poder traballar con eles. Para iso empregamos

o seguinte código:

```
> <span class="enlighter-g1">[</span><span class="enlighter-text">sourcecode language=</span><span class="enlighter-s0">R</span></span><span class="enlighter-g1">]</span>
a<-read.table(file = "clipboard", sep = "\t", header=TRUE)
[/sourcecode]
```

“a” é unha lista na que gardamos os datos da folia de cálculo, os cales se pegan a continuación do código.

Unha vez que xa temos a información en R, procedemos a representala nun histograma; pero antes necesitamos seleccionar a columna e as filas específicas coas que traballar. Isto conséguese da seguinte forma: “a[filas,columnas]”. No noso caso, para as deteccións cada 2 segundos seleccionamos a[1:475,3].

Agora si procedemos a representar os datos coa función *hist()*, obtendo o seguinte resultado:

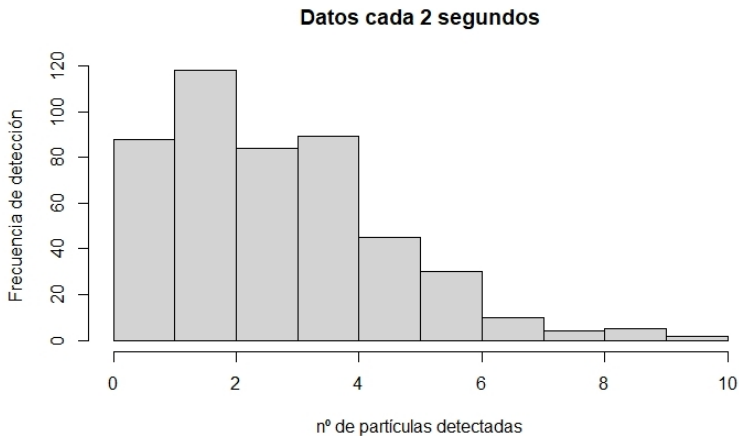


Fig.5. Histograma das deteccións de partículas cada 2 segundos.

Como se pode observar na gráfica anterior, as partículas producidas na desintegración nuclear radiactiva son un exemplo de sucesos que cumpren as condicións dunha distribución de Poisson: só adoptan valores de y naturais, son sucesos raros nos que a desintegración dun átomo non conleva a desintegración dos contiguos, non poden

ocorrer á vez e a taxa na que aparecen é, de media, constante nun intervalo de tempo determinado.

A continuación, facemos os histogramas das deteccións cada 5 segundos e cada 10 segundos. Para obter os datos debemos realizar unha modificación ao comando de lectura, sendo para os datos de 5s a[1:191,6] e para os de 10s a[1:95,9].

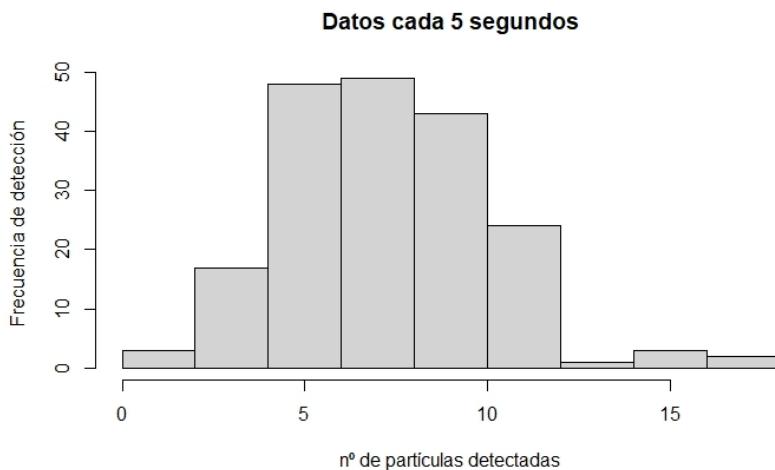


Fig.6.Histograma das partículas detectadas cada 5 segundos.

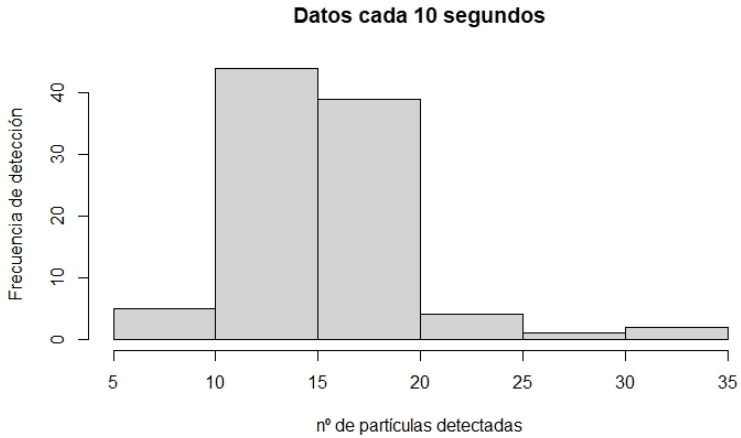


Fig.7. Histograma das deteccións cada 10 segundos.

Unha vez feita a representación dos datos obtidos proseguimos co análise estatístico facendo as medias e as desviacións típicas. Para isto seguimos utilizando R , esta vez os comandos *mean* para calcular a media e *sqrt(var())* para a desviación típica (xa que a desviación típica é a raíz cadrada da varianza), e obtivemos os seguintes resultados:

```
sqrt(var(a[1:475,3]))      mean(a[1:475,3])
1.886403                  3.105263
```

Fig.8 e 9. Media e desviación típica das deteccións cada 2 segundos.

```
sqrt(var(a[1:190,6]))      mean(a[1:190,6])
2.832485                  7.763158
```

Fig.10 e 11. Media e desviación típica das deteccións cada 5 segundos.

```
sqrt(var(a[1:95,9]))  
4.209518
```

```
mean(a[1:95,9])  
15.52632
```

Fig.12 e 13. Media e desviación típica das deteccións cada 10 segundos.

Unha vez que coñecemos os valores da media e da desviación típica para cada caso, xa somos quen de representar a curva normal e a Poisson de cada experimento para poder estudar o axuste do noso experimento a estas predicións teóricas.

Para poder representar as novas curvas conxuntamente co histograma anterior debemos usar a función *lines* metendo no eixo x un vector ou secuencia que conteña os valores de partículas recollidos polo detector geiger e no eixo y a curva normal ou a Poisson.

Mais antes de nada precisamos reescalar o histograma, xa que tanto a normal coma a Poisson van vir dadas en termos de probabilidade. Esta reescala facémola co parámetro *prob=TRUE*.

Para realizar a curva normal empregaremos a función *dnorm* do seguinte xeito:

lines(x,dnorm(x,mean=m,sd=d),col=2,). Dentro da curva normal temos que volver a introducir o vector ou secuencia cos valores de partículas detectadas e a continuación a media e a desviación típica que calculamos anteriormente e que gardamos nas funcións “m” e “d” respectivamente. Por último cambiamos a cor da gráfica a cor vermello con *col=2*.

A representación da distribución de Poisson ten un procedemento semellante ao da curva normal. Neste caso empregamos a función *dpois* en conxunto con *lines* do seguinte xeito: *lines(x,dpois(x,m))*.

O que estamos introducindo dentro de *dpois* simplemente é de novo os valores de partículas detectados coa variable “x” e a media dos valores que xa temos gardada na variable “m”.

Unha vez realizado este proceso cos tres experimentos, obtivemos os seguintes resultados:

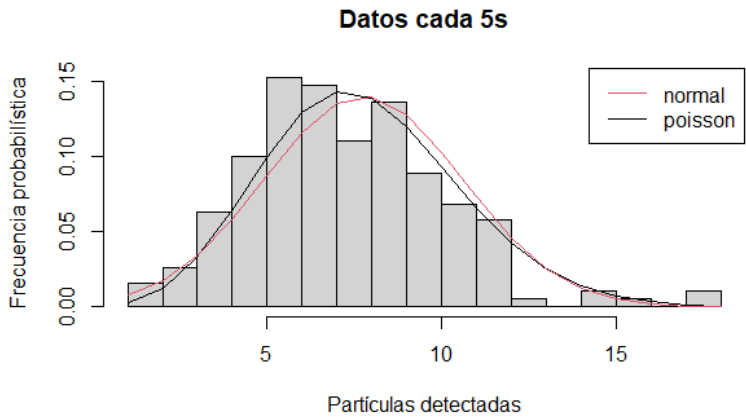
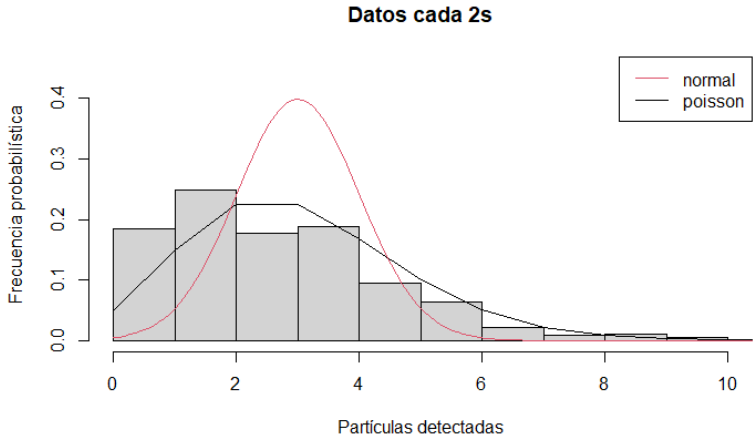


Fig.14. Representación dos datos cada 2s coa súa distribución normal e Poisson.
Fig.15. Representación dos datos cada 5s coa súa distribución normal e Poisson.

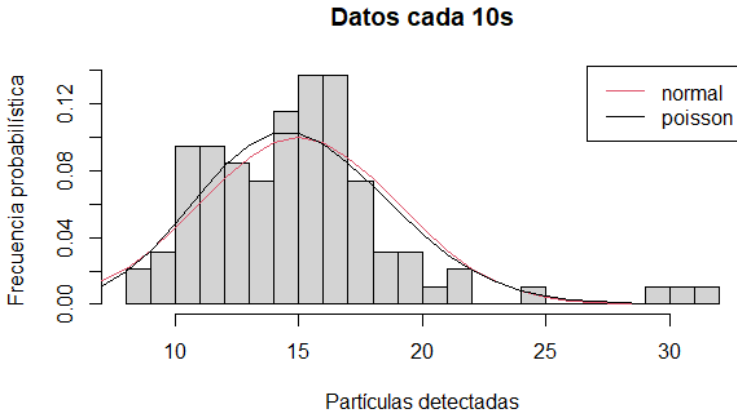


Fig.16. Representación dos datos cada 10s coa súa distribución normal e Poisson.

Podemos levar o teorema central do límite ao caso da desintegración nuclear radiactiva. Se realizamos unha gráfica mostral das partículas detectadas nun intervalo de dous segundos, esta segue unha distribución de tipo Poisson como xa observamos na gráfica correspondente. Con esta investigación pretendemos detectar se ampliando ese intervalo de tempo a cinco ou dez segundos, cunha mostra suficiente, esa distribución de Poisson se achega a unha distribución normal.

Por último, para ver a importancia de realizar máis ou menos medicións á hora de ter uns resultados máis fiables e exactos imos calcular o erro en función do número de medicións.

Para iso calculamos a raíz cadrada do número de deteccións e dividímolos entre as deteccións. Para facelo máis visual, usando o comando *plot* elaboramos unha representación do erro en función do número de medicións, obtendo o seguinte resultado:

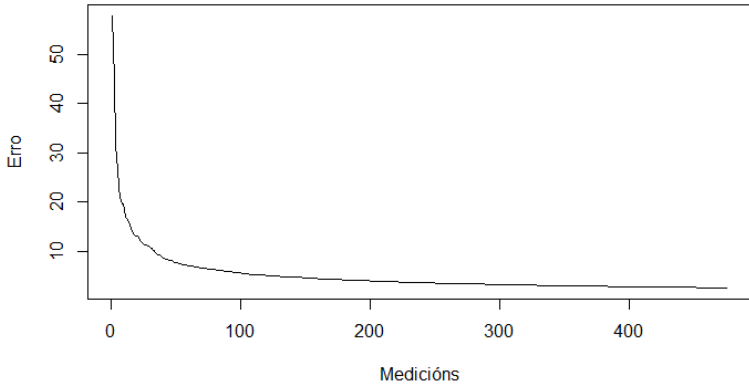


Fig.17. Representación do erro (en tanto por cen) en función do número de medicións.

3. Conclusións

Unha vez realizadas as gráficas onde representamos os datos e a predición da distribución normal e de Poisson, observase como si que se cumpren de xeito aproximado os dous obxectivos deste traballo:

Por unha parte vese como o número de partículas detectadas se axusta á predición da distribución de Poisson.

Ademais tamén se comproba o teorema central do límite, xa que ao aumentar os intervalos de tempo aumenta tamén a frecuencia media coa que ocorren os sucesos; polo que a Poisson e o noso histograma aseméllanse máis a unha distribución normal.

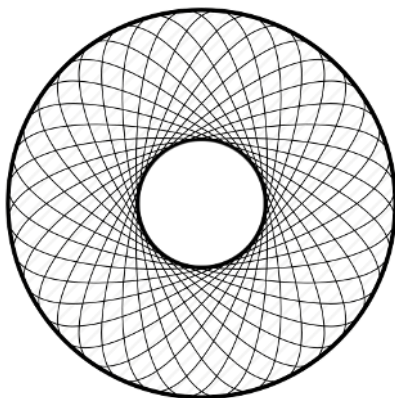
Por outra parte, grazas á gráfica da variación do erro en función das medidas tomadas pódese apreciar moi claramente a importancia da realización de moitas medidas á hora de obter resultados fiables e que se aproximen máis á realidade.

Finalmente, cabe recalcar que para facer estas medicións máis

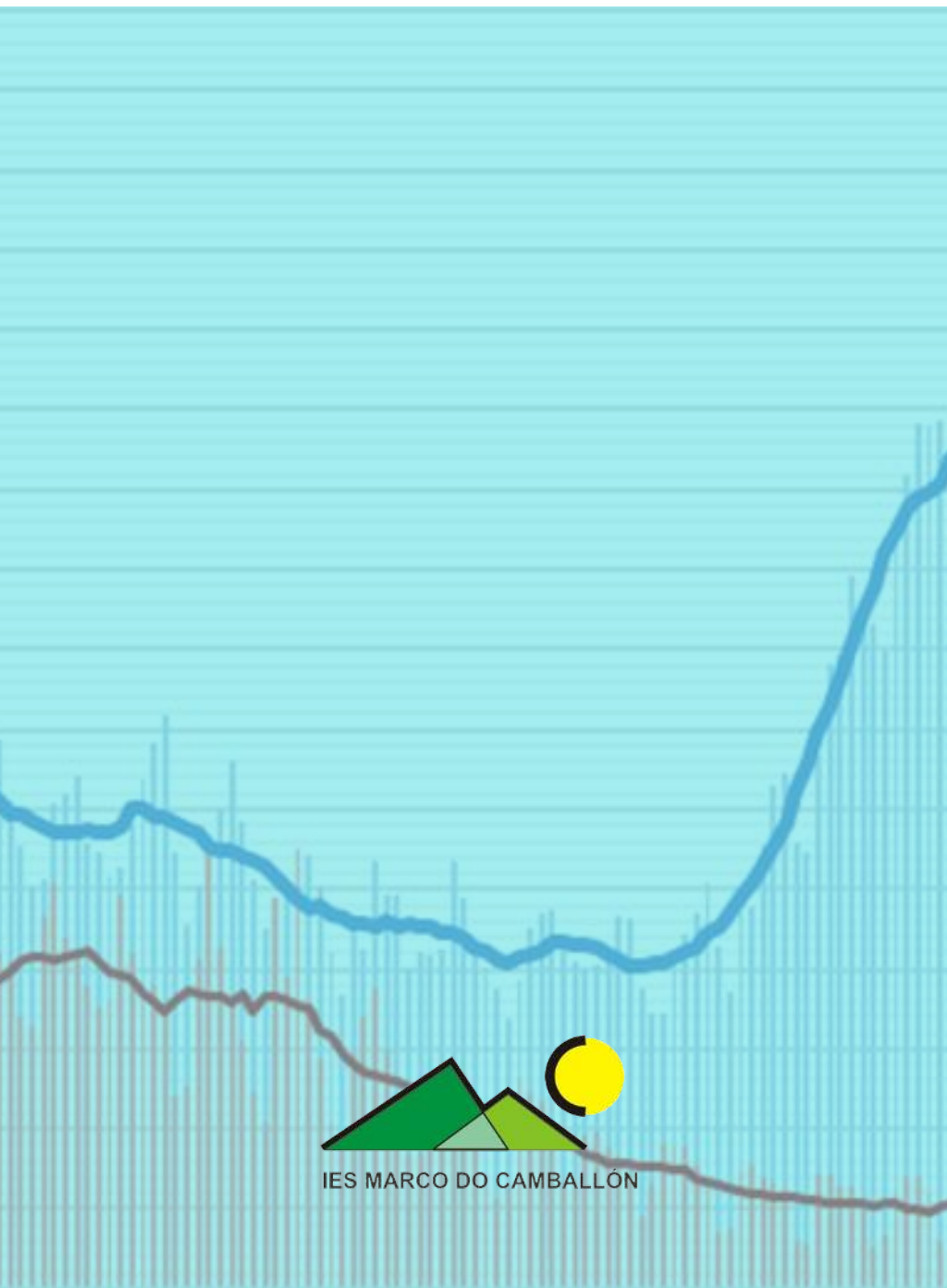
reais deberíase ter recollido tamén a medición de radiación de fondo que existía á hora de realizar a nosa detección e que non procedía da nosa fonte radiactiva.

4. Webgrafía

- Equipo Superprof (s.f.). Qué significa distribución normal en Matemáticas. Recuperado o 03 de marzo do 2022, do sitio Web: <https://www.superprof.es/diccionario/matematicas/probabilidades/distribucion-normal.html>
- Dpto. Física Atómica, Molecular y Nuclear. Univ. Complutense de Madrid (s.f.). Radiactividad. Recuperado o 03 de marzo do 2022, do sitio Web: <http://nuclear.fis.ucm.es/FERIA/FERIA2.html>
- Canal Díaz, Neus (2006). Distribuciones de probabilidad. El teorema central del límite. Recuperado o 03 de marzo do 2022, do sitio Web: <https://www.revistaseden.org/files/8-CAP%208.pdf>
- Quintela del Rio, Alejandro (2006). Estadística Básica Edulcorada. Recuperado o 03 de marzo do 2022, do sitio Web: <https://bookdown.org/aquintela/EBE/el-teorema-central-del-limite.html>
- Proyecto CEACES (s.f.). Distribución de Poisson. Recuperado o 03 de marzo do 2022, do sitio Web: <https://www.uv.es/ceaces/base/modelos%20de%20probabilidad/poisson.htm>
- Wikipedia (2021). Distribución de Poisson. Recuperado o 03 de marzo do 2022, do sitio Web: https://es.wikipedia.org/wiki/Distribuci%C3%B3n_de_Poisson



O quinto número
de *Marco Numérico*
saíu do prelo
em maio de 2021.



IES MARCO DO CABBALLÓN