AZCATL
Revista de divulgación en ciencias, ingeniería e innovación

Coleccionar estampas: un ajuste de curvas para obtener una función que explique por qué es difícil llenar un álbum de estampas

Marco Antonio Rodríguez Andrade
Instituto Politécnico Nacional
algorismo@gmail.com

Resumen

En este artículo se aborda el problema del coleccionista, el cual consiste en coleccionar cierta cantidad de objetos, los cuales son comprados al azar. Se presenta un caso en el cual se compraron sobres, con cinco estampas cada uno, para llenar un álbum de 240 estampas. Se obtiene una función matemática que aproxima la cantidad de estampas coleccionadas en función de la cantidad de sobres que se han comprado. Asimismo, con esta misma función, se determina la cantidad de sobres que se deben comprar para casi llenar el álbum.

Abstract

This article deals with the collector's problem, which consists of collecting a certain number of objects, which are bought randomly. A case is presented in which envelopes were bought, with five stickers each, to fill an album of 240 stickers. A mathematical function is obtained that approximates the number of stickers collected as a function of the number of envelopes purchased. Likewise, with this same function, the number of envelopes that must be purchased to almost fill the album is determined.

Cómo citar este artículo:

Rodríguez, M. (2023). Coleccionar estampas: un ajuste de curvas para obtener una función que explique por qué es difícil llenar un álbum de estampas. *Azcatl. Revista de divulgación en ciencias, ingeniería e innovación, 1* (18-23).

Introducción

El ajuste de curvas es un tema que suele abordarse en los laboratorios no avanzados de Física, donde es común que el docente ya conozca a qué función deben ajustarse los datos que se obtienen del experimento, de tal forma que se ajusten de manera óptima.

Para este experimento se compraron sobres para llenar un álbum de estampas y se llevó un registro de cuántas estampas se han coleccionado contra la cantidad de sobres comprados. Se supone que cada sobre contiene cinco estampas, todas ellas distintas entre sí.

Mediante el uso del *software* LAB Fit Curve Fit, con ajuste a cuatro parámetros, se obtienen varias curvas, las que mejor se ajustan a los datos. El comportamiento cualitativo de las funciones es usado para discriminarlas y sólo con aquellas funciones cuyo comportamiento cualitativo es factible se hace un ajuste de curvas a un sólo parámetro para obtener una mejor función.

Cabe mencionar que la librería de LAB Fit Curve Fit consta de más de 200 funciones y permite decidir cuáles de ellas se ajustan mejor a los datos, ordenándolas de acuerdo con el coeficiente de correlación.

Planteamiento del problema

Una afición común entre los niños y adolescentes es coleccionar estampas de algún tema especial, dichas estampas las adquieren en sobres cerrados. El coleccionista no sabe de antemano las estampas que va a obtener. Si el coleccionista debe reunir n estampas distintas, se pueden plantear dos problemas:

1. En caso de que el coleccionista haya adquirido k sobres, ¿se puede estimar la cantidad de estampas distintas que ha coleccionado?
2. ¿Cuál es el número esperado de sobres que debe comprar un coleccionista para tener al menos un ejemplar de todas las estampas?

El segundo problema se puede considerar resuelto si cada sobre sólo tiene una estampa, en este caso la solución del problema ya existe en el capítulo 28 del libro de Aigner y Ziegler (2010), pero en este mismo texto no aparece una respuesta para el primer problema.

Experimento

La Tabla 1 fue obtenida de la colección de estampas de un álbum en el cual se deben reunir 240 estampas.

Para el ajuste de curvas se ha usado un *software* muy útil llamado LAB Fit Curve Fit2 (LAB Fit ajuste de curvas), aunque existen varios programas libres que se pueden utilizar.

Al capturar los datos en el *software*, éste nos proporciona una gama de posibilidades para ajustar los datos anteriores; en este caso, el ajuste se hizo con cuatro parámetros. Los coeficientes de correlación son superiores a 0.98 y el programa organiza las funciones de orden decreciente de acuerdo con su coeficiente de correlación. Usando este criterio todas las funciones son factibles.

Al considerar el comportamiento cualitativo de las funciones, si $f(k)$ es la función que estima la cantidad de estampas distintas, entonces esta función debe cumplir que

- es no decreciente;
- está acotada, siendo n una cota superior, y
- $\lim_{k\to\infty} f(k) = n.$

Las funciones que se recomiendan se presentan en la Tabla 2.

En Aigner y Ziegler (2005), en el problema del coleccionista donde se desean coleccionar n objetos, se encuentra la relación

$$n \ln n$$

Tabla 1.

k	5	11	19	23	29	46	54	60	65	71	77	83	89	97	105
D.O.E*	24	49	76	99	113	145	161	173	180	186	194	200	205	210	215

* Datos obtenidos experimentalmente.

compras, con gran probabilidad de que se hayan coleccionado los n objetos. Cabe aclarar que cada compra es de sólo un objeto, el cual es elegido al azar.

En el caso de que estemos trabajando $n = 240$:

$$n \ln n \approx 1315.4 \qquad \text{Ecuación 1}$$

La Tabla 3 muestra las funciones que mejor se ajustan a los datos que tienden a 240 cuando k tiende a infinito y se tiene un valor de t para el cual $f(s) \geq 239$, para $s \geq t$ y el coeficiente de correlación.

Lo anterior se hizo para que las tres funciones tendieran a 240, cuando x tiende a $y = Ak^{\frac{B}{x}}$ porque en este caso la función es decreciente.

En cada sobre vienen 5 estampas, por lo cual, para la función

$$f(k) = 240(1 - e^{-0.0212389k})$$

se tiene que $f(s) \geq 239$, para $s \geq t = 258$, es decir, hay que adquirir $5(258) = 1290$ estampas para coleccionar 240 estampas. En caso de que se compraran una por una al azar, de acuerdo con la ecuación 1, se requiere adquirir 1315 para reunir aproximadamente 239 estampas.

En otro contexto, si hay sobres con 5 estampas a 5 pesos el sobre, prácticamente tiene el mismo costo que comprar una por una al azar con un costo de 1 peso cada una.

Tabla 2.

$f(k)$	$\lim_{k\to\infty} f(k)$	**Factible**
$y = 361.68\, k^{-7.1783k^{-0.8978}}$	361.68	Sí, pero llevándola a un ajuste de dos parámetros
$y = \frac{10.786}{k} + 9.4497 \ln k$	∞	No
$y = \frac{321.44}{1 + 10.823\, k^{-0.76366}}$	321.44	Sí, pero llevándola a un ajuste de dos parámetros
$y = \frac{k}{0.21848 + 28974\, k - 28974\, k^2}$	0	No
$y = 241.503\,(1 - e^{-0.0292k})$	241.503	Sí, pero llevándola a un ajuste de un parámetro
$y = \frac{1}{238.38 + 16.257\, k - \frac{19035}{k}}$	0	No
$y = 36.94\, k^{0.57609k - \frac{2.9853}{k}} - 2.4453k$	Tiende a infinito	No
$y = 211.25\,(e^{-0.009639k} - e^{-0.02351k})$	0	No
$y = 4.032\,(\ln(k + 2.4495)^{2.9896}) - 1.8113\, k$	Tiende a infinito	No
$y = 4.3764\,(\ln(k + 1.5252)^{2.6507}) - 0.00493\, k^2$	Tiende a infinito	No

Tabla 3.

$f(k)$	t	**Factible**
$y = 240\, k^{-19.201294k^{-1.3495}}$	2370	0.98067
$y = \frac{240}{1 + 160.007\, k^{-1.4844}}$	1220	0.9917
$y = 240\,(1 - e^{-0.0212389k})$	258	0.998
$y = 204\, k^{\frac{5.55501}{k}}$	12500	0.98047

En otras palabras, adquirir sobres de 5 estampas, en las cuales todas son distintas, realmente no ofrece una ventaja significativa sobre comprar una por una al azar.

Ahora, se tiene que $\frac{5}{240} - 0.0122389 = -3.7906\times 10^{-4}$, de lo cual se puede inferir que una buena opción es la función

$$f(k) = 240\,(1 - e^{-\frac{5}{240}k}).$$

Una generalización del problema

Suponga que desea coleccionar M objetos, los cuales se van a adquirir, al azar, en paquetes de 5 (todos distintos); proponemos la función

$$f(k) = M(1 - e^{-\frac{5}{M}k})$$

la cual estima la cantidad de objetos que se han coleccionado al realizar k compras, al establecer la ecuación

$$f(k) = M(1 - e^{-\frac{5}{M}k}) = M - 1,$$

cuya solución es

$$k = \frac{M}{5}\ln M.$$

Como en cada compra se obtienen 5 sobres, entonces se deben adquirir $M \ln M$ objetos para que aproximadamente $M-1$ sean distintos entre ellos. El resultado anterior es coherente con el encontrado en Aigner y Ziegler (2010).

Simulación

En el lenguaje Python se hicieron dos programas para simular a 20 coleccionistas que desean llenar un álbum con M estampas para $M = 100, 300, 500$.

En la Tabla 4 se muestran las gráficas generadas y en la Tabla 5, el código correspondiente.

Tabla 4.

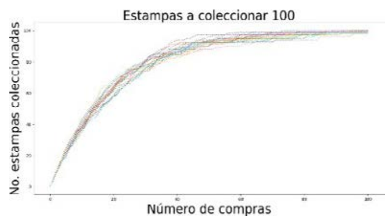
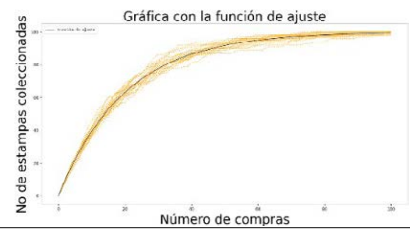
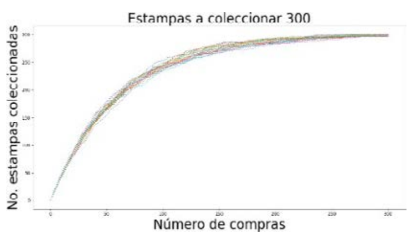
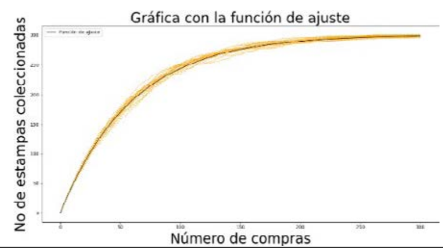
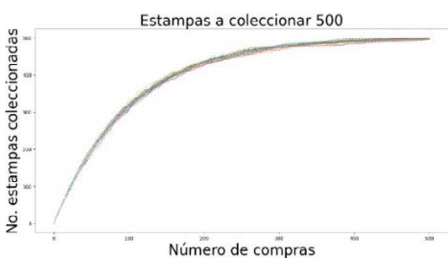
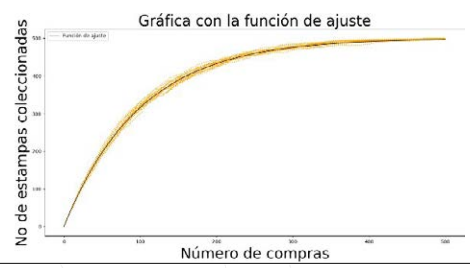
Gráfica sin la función de ajuste	Gráfica con la función $f(k) = M\,(1 - e^{-\frac{5}{M}k})$
	
	
	

Tabla 5.

Programa 1

```
import random

import matplotlib.pyplot as plt


def generar_vector():

   numeros = random.sample(range(1,
   241), 5)return numeros


for a in
range(1,21):
x=
list(range(1
,241))

 vector= generar_vector()

 conjunto= set(vector)  # Convierte el vector en un conjunto

 y=[len(conjunto)]


 for i in range(1,240):

   vector2=generar_vector()

   conjunto=conjunto.union(set(vector2))

   y.append(len(conjunto))

 plt.plot(x, y,',')
```

Programa 2

```
 #Importamos librerias


 import matplotlib.pyplot as plt


 import
 random
 import
 numpy
 import
 math
 #K es el numero de sobres
 compradosk=500
 #L es la cantidad de coleccionistas en el experimentol=20
 #M es la cantidad de estampillas diferentes que hay

M=500

#H es la cantidad de estampillas por sobreh=5

#Iniciamos la apertura de sobres por coleccionistafor p in

range (1):

   #Creamos arrays, uno que contenga al punto inicial y otrovacia

que contenga los puntos

   puntos =

set([])a=[0]

#Para cada sobre abierto

   for i in range(k):

      #Creamos una lista de 5 enteros aleatorios entre 1 y M

      numeros = set([random.randint(1, M) for i in range(h)])#Los

      unimos al conjunto que tenemos puntos=puntos|numeros

      #Checamos cardinalidad y lo agregamos al historial a

      a.append(len(puntos))

      #Graficamos el array a que es el historial de la

cardinalidad por apertura de sobre

      plt.plot(a,linestyle='dotted',color="orange")#Para la

función

x=numpy.array(range(10*k))*0.1

y=numpy.zeros(len(x))

for i in range(len(x)):

   #Gráfica de la función ajuste

   y[i]=M*(1-math.exp(-(h/M)*(x[i])))

#Graficamos la función

plt.plot(x,y,linewidth=1,color="black",label='Función deajuste')

#Mostramos y embellecemos la tabla

plt.title("Gráfica con la función de ajuste",fontsize=30)

plt.xlabel("Número de compras",fontsize=30) plt.ylabel("No de

estampas coleccionadas",fontsize=30) plt.legend()

 plt.show()
```

Conclusiones

Una idea a priori, entre la población que colecciona estampas, es que hay una o varias estampas difíciles de conseguir, es decir, que algunas de éstas las imprimen en menor cantidad para dificultar el llenado del álbum. Una de las conclusiones de este ensayo es que no se requiere tal artimaña, el azar es suficiente para garantizar que cada coleccionista compre muchas estampas para casi llenar o llenar por completo el álbum.

Otra conclusión es que el hecho de que vendan sobres con 5 estampas, todas ellas distintas entre sí, en la práctica no proporciona mayor ventaja que ir comprando una por una al azar.

Desde el punto de vista didáctico y de divulgación, el problema presentado en este artículo proporciona una experiencia en la cual el alumno puede percatarse de lo siguiente:

1. El coeficiente de correlación no es un criterio suficiente para decidir cuál es la función más indicada que explique el problema tratado, pues pueden existir muchas funciones que se ajustan, con un alto índice de correlación, a los datos obtenidos, pero cualitativamente son funciones muy distintas.
2. Tener propiedades cualitativas del fenómeno abordado puede ayudar a discriminar entre diferentes opciones que se ajustan bien a los datos.
3. Considerando que los parámetros de ajuste son aproximaciones numéricas, el contexto del problema puede ayudar a ir mejorando el modelo matemático.
4. Las herramientas tecnológicas son útiles para resolver problemas, pero requieren de una correcta interpretación por parte del usuario para dar resultados satisfactorios.

Referencias

Aigner, M. y Ziegler, G. (2005). *El libro de las demostraciones*. Nivola Libros Ediciones.

Aigner, M. y Ziegler, G. (2010). *Proofs from* THE BOOK. Springer.