Segundo Examen Parcial*

Luis Alfredo, Ixquiac Méndez[1]

[1] *Escuela de Ciencias Físicas y Matemáticas, Universidad de San Carlos, Edificio T1, Ciudad Universitaria, Zona 12, Guatemala.*

Se da solución a dos problemas de la siguiente manera: En el primer problema, se tiene un conjunto de datos que muestran el desempeño deportivo de un atleta. Asumiendo que estos datos se comportan de manera lineal, se utiliza el metodo de minimos cuadrados para encontrar la mejor recta que se ajusta a esos datos. Este proceso se lleva a cabo utilizando el lenguaje de programación C. Con los datos obtenidos, se utiliza la herramienta GNUPlot para analizar gráficamente el modelo matemático obtenido y el conjunto de datos. Finalmente, se realiza un análisis acerca de las predicciones de el modelo para determinado tiempo. En el segundo problema, se tiene una función continua de la cual se sabe que tiene una raíz. Se utiliza el método de la bisección para encontrar una aproximación para dicha raíz. Este método se lleva cabo utilizando el lenguaje de programación C. Luego, se utiliza nuevamente la herramienta GNUPlot para analizar gráficamente la forma de la función y el intervalo donde se sabe que está la raíz.

I. OBJETIVOS

A. Generales

- Desarrollar programas utilizando el lenguaje de programación C y utilizar la herramienta para procesamiento gráfico de datos GNUPlot para resolver problemas que son elementales en el trabajo de un científico.

B. Específicos

- Comparar las predicciones del modelo matematico de mínimos cuadrados con los datos experimentales.
- Encontrar un rango aproximado donde se encuentre una raiz de la funcion arcsin(x), o el valor exacto de la raiz.

II. PROBLEMAS

A. Problema 1

Un atleta necesita pasar la evaluación física para poder ingresar a la selección de su equipo. Como mínimo debe poder de realizar diez sentadillas con 150 Kg de peso. 3 meses antes de la evaluación sufre un desgarre en el cuadricep derecho, lo que le permitió obtener la siguiente evolución

semana	peso (Kg)
1	20
2	26
3	31
4	38
5	45
6	49
7	54

Figura 1: Evolución del atleta

Crear un programa que cumpla con las siguientes condiciones:

- Una gráfica que compara los valores tabulados y la recta que mejor aproxima el crecimiento.
- Estime el peso que logra levantar el atleta después de 3 meses, este logra pasar la prueba para ingresar al equipo.

B. Problema 2

Utilizando el método de bisección, encuentre una raíz de la ecuación f(x)=arcsin(x). Debe de realizar la gráfica de la ecuación y capturar el resultado obtenido con el programa realizado en C.

III. MARCO TEÓRICO

A. Regrecoón lineal; minimos cuadrados

Este consiste en lo siguiente: Se tiene inicialmente un conjunto finito de n datos. Se toma en consideración la ecuación general de una recta. Esta depende siempre de dos parámetros. Si la escribimos en su forma

* Laboratorios de Simulación

pendiente-intercepto con el eje y, estos parámetros se suelen identificar con las letras m y b. Siendo b su intercepto con el eje y, y m el valor de su pendiente. Lo que se hará será lo siguiente: Se expresara los valores de m y b en función de los valores de todos los datos. Esta relación funcional será de acuerdo al criterio de mínimos cuadrados que se explicará a continuación.

$$m = \frac{n\sum_{k=1}^{n}(x_k y_k) - \sum_{k=1}^{n} x_k * \sum_{k=1}^{n} y_k}{n\sum_{k=1}^{n} x_k^2 - (\sum_{k=1}^{n} x_k)^2}$$

$$b = \frac{\sum_{k=1}^{n} y_k - m\sum_{k=1}^{n} x_k}{n}$$

Figura 2: Ecuaciones 1

$$\Delta m = \frac{\sqrt{n}\epsilon}{\sqrt{n\sum_{k=1}^{n} x_k^2 - (\sum_{k=1}^{n} x_k)^2}}$$

$$\Delta b = \frac{\epsilon}{\sqrt{n}}$$

Figura 3: Ecuaciones 2

Tomando en cuenta que el área de un cuadrado queda totalmente definida por el tamaño de uno de sus lados, considerar lo siguiente. Considerar el área de un cuadrado definida por las posiciones de los siguientes dos vértices. El primer vértice son las coordenadas de uno de los datos que tenemos en nuestro conjunto. El segundo vértice se obtiene de la siguiente manera: Se traza una recta vertical que pase por el primer vértice. Se encuentra las coordenadas de el punto donde esta recta intersecta a la función de la recta general mencionada arriba. Este punto de intersección será el segundo vértice. Con estos dos vértices, queda definida el área de este cuadrado. Ahora, repetimos el mismo proceso para cada uno de los otros puntos que tenemos en nuestro conjunto de datos. Una vez tenemos calculadas todas las áreas de los n cuadrados, considerar lo siguiente. Considerar a una función de área tal que sea igual a la suma de las áreas de todos los cuadrados. Esta función de área depende entonces del conjunto de todos nuestros datos y los dos parámetros de la recta: m y b. Es entonces una función de 2n+2 variables. Por razones de lo que queremos obtener, los datos de nuestro conjunto serán valores fijos, por lo que todos estos serán constantes. Esta función de área total se vuelve entonces una función de dos variables m y n: Esta función es la función tal que es igual a la suma del área de todos los cuadrados definidos como se explicó arriba. Es una función de dos variables. Utilizando el criterio de derivadas para encontrar valores mínimos de funciones, se procede a encontrar los valores de m y b tales que esta área sea lo menor posible. Esto último es el "Criterio de mínimos cuadrados". Cuando se dice la frase "La recta que mejor se aproxima a los datos" a lo que se refiere es al criterio que se acaba de explicar. En ese sentido es que se aproxima "bien" a los datos. Una vez realizado este proceso, los resultados obtenidos son: (Nota, realizando un análisis más completo se puede tomar en cuenta (como Debe ser!) las incertezas en los datos experimentales. Tomando eso en cuenta, utilizando la teoría de análisis de error también se puede encontrar las incertezas propagadas a los valores de m y b)

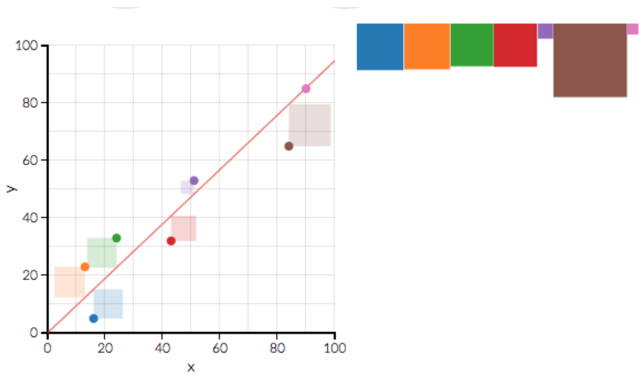


Figura 4: Visualizacion de el area total conformada por la suma de el area de cada cuadrado

B. Método de la bisección

El método de la bisección se basa en un teorema fundamental acerca de funciones continuas. El teorema dice lo siguiente:

Si f es continua en [a,b] y f(a) <0 <f(b), entonces existe un x en [a,b] tal que f(x)=0.

Este teorema nos garantiza que si nuestra función es continua en un intervalo cerrado, entonces existe al menos una raíz. El teorema no da información acerca de cómo encontrarla, solo garantiza su existencia. Por lo que, armados de el conocimiento de que si tenemos una función continua en un intervalo cerrado, entonces existe al menos una raíz de esta función, procedemos a intentar encontrar una raíz para una función de ese tipo con lo siguiente: Nuestras condiciones iniciales son: Tenemos una función de la cual tenemos conocimiento que es continua en el intervalo cerrado [a,b], y ademas la funcion es tal que $f(a)<0$ y $f(b)>0$ (el análisis para el caso $f(a)>0$ y $f(b)<0$ es completamente análogo). Notar lo siguiente. Consideremos al número p tal que es igual al promedio de a y b, i.e $p=(a+b)/2$. Notar que al evaluar la función f en este número p, solo pueden ocurrir 3 casos: $f(p)<0$, $f(p)=0$, $f(p)>0$. Si se da el caso $f(p)=0$, entonces se a tenido suerte y se ha encontrado el valor de la raíz, esta es p. Si por el contrario, este caso no se da, entonces se debe dar alguno de los otros casos. Notar lo siguiente acerca de los siguientes casos:

Si se da el caso $f(p)<0$, entonces ahora tenemos de conocimiento que $f(p)<0$ y $f(b)>0$. Por lo tanto, como hay un teorema que dice que:

Sea [c,d] un intervalo cerrado contenido en un intervalo cerrado [a,b]. Si f es continua en el intervalo cerrado [a,b], entonces también es continua en el intervalo cerrado [c,d].

Con este teorema ahora sabemos que f es tal que es continua en el intervalo cerrado [p,b] y como $f(p)<0$ y $f(b)>0$, entonces la funcion cumple nuevamente las condiciones necesarias para aplicar el teorema fundamental mencionado al inicio. Ahora estamos seguros que hay un número x en [p,b] tal que $f(x)=0$. Es decir, seguimos estando seguros que en ese intervalo hay al menos una raíz de la función f. Lo interesante detrás de este hecho es que ahora nuestro rango de incertidumbre acerca de donde se encuentra la raíz a disminuido. Justo este hecho es la segunda idea clave detrás de el método de la bisección. Para el caso $f(p)>0$, hubiese ocurrido exactamente lo mismo. Por lo tanto, sabiendo que a medida que apliquemos este procedimiento, vamos a ir obteniendo cada vez intervalos más pequeños en los cuales podremos estar siempre seguros que habrá una raíz allí, nos motivamos a intentar encontrar la raíz de una función utilizando este proceso. Solo es de repetir este proceso varias veces, y a mayor cantidad de veces que se aplique este proceso, nuestro rango de incertidumbre será menor. Por lo tanto, puede que nunca encontremos el valor exacto de la raíz, pero si puede ser posible que podamos encontrar un valor cercano a este con la precisión que se desee.

IV. METODOLOGIA

A. Problema 1

El primer problema consiste en realizar un análisis de datos acerca de la evolución en el rendimiento deportivo de un atleta. Se cuenta con un conjunto finito de 7 parejas de datos. Se desea aplicar el método de mínimos cuadrados a este conjunto de datos para obtener un modelo matemático que trate de modelar cómo continuará siendo su rendimiento en un futuro. Con los resultados obtenidos, se realizará también un análisis gráfico en el cual se compara el modelo y los datos experimentales. También se dará un pronóstico de cuál será el rendimiento deportivo del atleta luego de 3 meses, para decidir si este podrá o no aprobar la evaluación para ingresar a su equipo.

El proceso de solución consta de dos bloques. En el primero se realiza el tratamiento de datos para obtener el modelo matemático dado por el método de mínimos cuadrados. En este tratamiento también se realiza el pronóstico del rendimiento deportivo del atleta. Este análisis se llevará a cabo utilizando el lenguaje de programación C. En el segundo bloque, se realizará un análisis gráfico de los datos experimentales y el modelo matemático obtenido. Este análisis se realizará utilizando la herramienta de GNUPlot. Ambos bloques se automatizan para que puedan ser ejecutados en un solo proceso por medio de un script. A continuación se da una explicación más detallada de cómo se realizará cada bloque.

Para el primer bloque se contará inicialmente con lo siguiente: Se tendrá un archivo de texto con todos los datos experimentales. Se tendrá el ejecutable del programa en C que funcionara de la siguiente manera. Al iniciar el programa, este tendrá ingresados los valores de los datos experimentales. Como se explicó en el marco teórico, todos los parámetros del modelo matemático de mínimos cuadrados se pueden determinar en función los datos experimentales. Por lo que, una simple aplicación de las funciones que determinan estos nos permiten obtener dichos parámetros. Una vez obtenidos dichos parámetros, se procede a realizar 2 cosas. Lo primero, es pronosticar cuál es el rendimiento deportivo de el atleta luego de 3 meses de haber iniciado su recuperación. Este dato se obtiene con un cálculo simple utilizando el modelo matemático ya obtenido. Este dato se compara con el valor que se esperaría que el atleta alcanza para poder ingresar a la selección de su equipo. En función de que su avance sea el apropiado o no, se imprimirá en pantalla un mensaje que indique si el atleta podrá cumplir su objetivo o no según el modelo matemático. Lo segundo que se realizará será crear un archivo de texto. En este archivo de texto se utilizara para ser ejecutado con la herramienta de gnuplot. En este archivo iran todos los comandos que delimitaran los parámetros adecuados para un correcto análisis de los datos. En este archivo estará contenida toda la información para poder graficar el modelo matemático. En este archivo también están las instrucciones para graficar en el mismo marco los datos experimentales para así poder compararlos. Una vez terminado de crear dicho archivo de texto, el programa de C finaliza.

El diagrama de flujo de como funciona este programa es el siguiente:

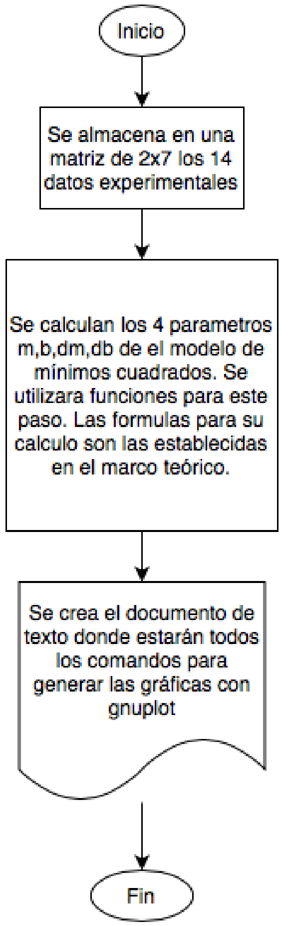


Figura 5: Diagrama de flujo programa 1

Ahora inicia el segundo bloque. En este bloque se ejecuta el archivo de texto que recien se creo con el programa en C. Este archivo contiene los comandos que dan el formato con la información y parámetros adecuados para poder visualizar el modelo matemático y los datos experimentales. Este archivo está programado también para que genere una gráfica con los datos experimentales y el modelo teórico. Esta gráfica se guardará como una imagen en formato jpeg en la carpeta donde se ejecute en script.

La entradas y salidas de este programa serán:

Entradas:
2n=2*7=14 datos experimentales

Salidas:
Un dato numérico que dará el desempeño del atleta luego de 3 meses de que haya iniciado su recuperación. Un archivo de imagen en formato jpeg que tendrá graficados los datos experimentales y el modelo matemático para poder compararlos.

B. Problema 2

La estructura general detrás de la solución del problema 2 es la misma que se utilizó para el problema 1. Esta consistirá principalmente en dos bloques cuya ejecución se automatizará por medio de un script.

El segundo problema consiste en encontrar la raíz de la función f(x)=arcsin(x) utilizando el método de la bisección.

El primer bloque inicia ejecutando el programa en lenguaje C. Se sabe que esta función es continua. El programa tiene ingresado los valores de los extremos de un intervalo, extremos tales que la funcion cumple con los requerimentos de el teorema mencionado en el marco teórico. Por esta razón se puede utilizar el método de la bisección. En este programa se tiene definida una función que ejecuta el algoritmo descrito en el marco teórico. Esta es una función tal que: tiene como entradas los extremos de un intervalo cerrado inicial, y como "salida" modifica los extremos de ese intervalo inicial de tal manera que devuelve un nuevo intervalo más pequeño. Utilizando un ciclo for se automatiza para que se pueda ejecutar dicha función un número n de veces. Esta cantidad de iteraciones es un dato que el usuario puede escoger a la hora de iniciar el programa. Como se mencionó en el marco teórico, puede haber dos tipos de resultados a cada vez que uno ejecuta el algoritmo. Puede que se obtenga el valor exacto de la raíz en alguna iteración o puede que al finalizar todas las iteraciones tengamos los extremos de un intervalo cerrado en el cual sabemos que hay una raíz. Lo que sigue a continuación es generar un archivo de texto que se utilizara en el bloque dos para mostrar en la gráfica de de la función el intervalo(o el valor exacto, si fuera el caso) donde el método de la bisección asegura que se encuentra una raíz. Este archivo de texto contendrá las coordenadas de los puntos sobre el eje x que son los extremos de dicho intervalo(o la coordenada de un solo punto si es que se encontró la raíz exacta). Al terminar de generar dicho archivo de texto, el programa en C finaliza. y se procede al segundo bloque.

El diagrama de flujo de como funciona este programa es el siguiente:

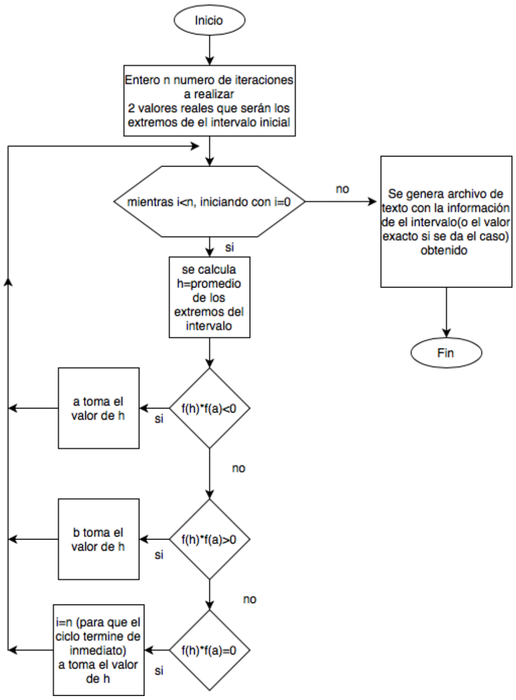


Figura 6: Diagrama de flujo programa 2

El segundo bloque consiste en ejecutar un archivo de texto que ya a sido creado con anterioridad. Este archivo de texto contiene las instrucciones para generar una imagen en formato jpg donde aparecerá la gráfica de la función en cuestión y también aparecerá indicado el intervalo(o punto) encontrado en el programa de C.

Una vez generada la imagen, se procede a borrar el archivo de texto que se utilizó para dar información para crear dicha imagen.

Las entradas y salidas de este programa son:

Entradas:

Número entero de iteraciones que se desea efectuar.

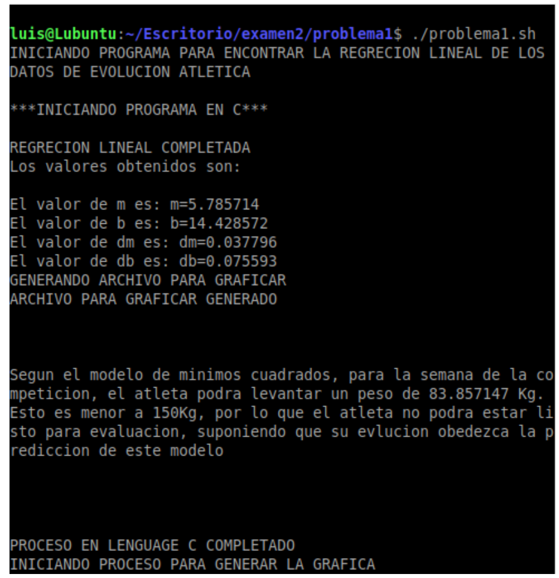
Salidas:

Una imagen en formato jpg que muestra una gráfica de la función en cuestión y el intervalo donde el método de la bisección indica que se encuentra una raíz(o el punto exacto si se da ese caso).

V. RESULTADOS

A. Problema 1

Al ejecutar el programa, los datos obtenidos son:



```
luis@Lubuntu:~/Escritorio/examen2/problema1$ ./problema1.sh
INICIANDO PROGRAMA PARA ENCONTRAR LA REGRECION LINEAL DE LOS
DATOS DE EVOLUCION ATLETICA

***INICIANDO PROGRAMA EN C***

REGRECION LINEAL COMPLETADA
Los valores obtenidos son:

El valor de m es: m=5.785714
El valor de b es: b=14.428572
El valor de dm es: dm=0.037796
El valor de db es: db=0.075593
GENERANDO ARCHIVO PARA GRAFICAR
ARCHIVO PARA GRAFICAR GENERADO


Segun el modelo de minimos cuadrados, para la semana de la co
mpeticion, el atleta podra levantar un peso de 83.857147 Kg.
Esto es menor a 150Kg, por lo que el atleta no podra estar li
sto para evaluacion, suponiendo que su evlucion obedezca la p
rediccion de este modelo



PROCESO EN LENGUAGE C COMPLETADO
INICIANDO PROCESO PARA GENERAR LA GRAFICA
```

Figura 7: Consola

La gráfica correspondiente es:

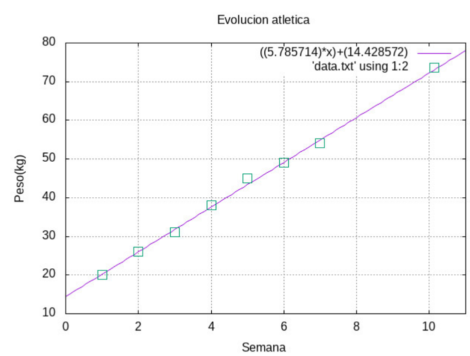
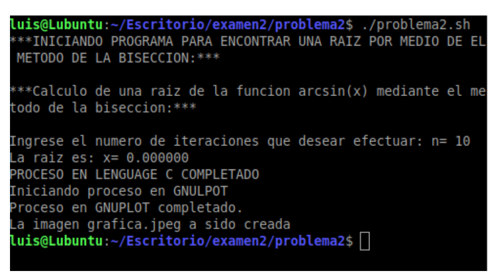


Figura 8: Grafica Peso(Kg) en funcion del tiempo(Semanas)

B. Problema 2

Se mostrarán los resultados con dos intervalos iniciales diferentes:

Resultados con intervalo inicial de [-0.5,0.5]:



```
luis@Lubuntu:~/Escritorio/examen2/problema2$ ./problema2.sh
***INICIANDO PROGRAMA PARA ENCONTRAR UNA RAIZ POR MEDIO DE EL
 METODO DE LA BISECCION:***

***Calculo de una raiz de la funcion arcsin(x) mediante el me
todo de la biseccion:***

Ingrese el numero de iteraciones que desear efectuar: n= 10
La raiz es: x= 0.000000
PROCESO EN LENGUAGE C COMPLETADO
Iniciando proceso en GNULPOT
Proceso en GNUPLOT completado.
La imagen grafica.jpeg a sido creada
luis@Lubuntu:~/Escritorio/examen2/problema2$
```

Figura 9: Consola

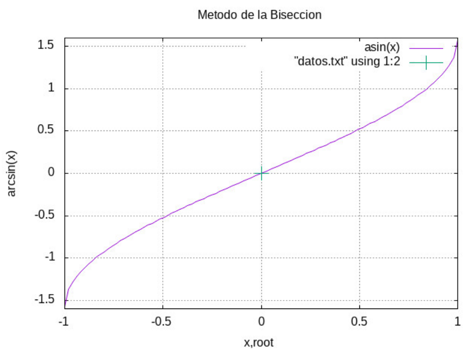
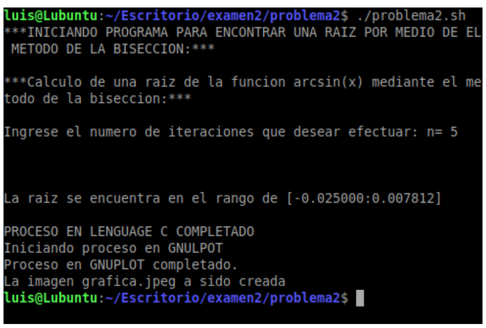


Figura 10: Grafica problema dos. Magnitud de extremos de intervalo iguales.

Resultados con Intervalo Inicial de [-0.55,0.5]:



```
luis@Lubuntu:~/Escritorio/examen2/problema2$ ./problema2.sh
***INICIANDO PROGRAMA PARA ENCONTRAR UNA RAIZ POR MEDIO DE EL
 METODO DE LA BISECCION:***

***Calculo de una raiz de la funcion arcsin(x) mediante el me
todo de la biseccion:***

Ingrese el numero de iteraciones que desear efectuar: n= 5



La raiz se encuentra en el rango de [-0.025000:0.007812]

PROCESO EN LENGUAGE C COMPLETADO
Iniciando proceso en GNULPOT
Proceso en GNUPLOT completado.
La imagen grafica.jpeg a sido creada
luis@Lubuntu:~/Escritorio/examen2/problema2$
```

Figura 11: Consola

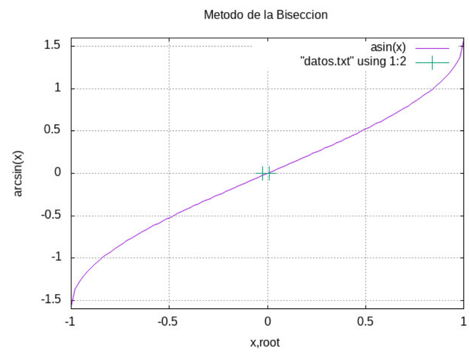


Figura 12: Grafica problema dos. Magnitud de extremos de intervalo inicial distintos. n=5 iteraciones.

VI. CONCLUSIONES

1. El lenguaje de programación C y la herramienta de análisis gráfico GNUPlot son muy útiles para trabajar con problemas como los que se resolvieron aquí.
2. El modelo de minimos cuadrados para el rendimiento de el atleta se acoplaba vastante bien a los datos experimentales.
3. Según el modelo de mínimos cuadrados, el atleta no podrá entrar a la selección de su equipo.
4. Una raíz de la función arcsen(x) es x=0

Powell, V., Lehe, L. Ordinary Least Squares Regression explained visually. Retrieved 27 April 2020, from https://setosa.io/ev/ordinary-least-squares-regression/