

Lección 5

Introducción a la aproximación

5.1 Conceptos generales sobre aproximación

Los procesos de *interpolación* estudiados anteriormente (véase el problema 2.5.6) enseñan cómo asociar a una función f , perteneciente a un espacio vectorial real X , un interpolante p . Este interpolante se busca en un subespacio S de X , previamente elegido, de modo que los valores de $n + 1$ formas lineales en f y p coincidan ($n + 1$ es la dimensión de S). (¿Quiénes son X , S y las formas lineales en el problema polinómico de Lagrange? ¿Y en el de Taylor? ¿Y en el caso de spline cúbico completo?). Naturalmente, el objeto de construir p radica en la idea de manejarlo en vez de f , digamos a efectos de evaluarlo en un punto, hallar su integral en un cierto intervalo, etc... Es típico de los problemas de interpolación que se disponga de tantas condiciones como parámetros libres hemos de calcular. Pero no siempre es posible plantear la sustitución de una función por otra más sencilla en estos términos.

En la teoría de *aproximación*, que ahora comenzamos, asociaremos a $f \in X$, el elemento $p \in S$ que haga la diferencia $f - p$ “lo menor posible”. Naturalmente para dar sentido a la expresión “lo menor posible” es menester que el espacio X esté dotado de una norma. El elemento p así encontrado, cuando existe, se denomina mejor aproximación a f (por elementos de S). Son pues dos los elementos esenciales de esta teoría: la *FORMA* del aproximante, que decidimos al elegir el conjunto S de los mismos; y la *NORMA* que determina cuál es el mejor de los posibles e incluso su propia existencia y/o unicidad.

Veamos seguidamente algunos ejemplos típicos de situaciones que conducen a problemas de aproximación.

5.1.1 Aproximación funcional. Se desea reemplazar, a efectos de evaluarla, la función $\sin x$, $0 \leq x \leq \pi/2$, por un polinomio p de grado ≤ 10 , de suerte que $\|\sin x - p(x)\|_\infty = \sup_{0 \leq x \leq \pi/2} |\sin x - p(x)|$ sea lo más pequeño posible. Nótese que una formulación alternativa es encontrar $(a_0, a_1, \dots, a_{10})^T \in \mathbb{R}^{11}$ para que la función $F(a_0, a_1, \dots, a_{10})$ sea mínima, con

$$F(a_0, a_1, \dots, a_{10}) = \sup\{|\sin x - a_0 - a_1x - \dots - a_{10}x^{10}| : 0 \leq x \leq \frac{\pi}{2}\}$$

Es un problema en que habitualmente la *NORMA* está preestablecida (la norma del *supremo*) y lo que se decide en cada caso es la *FORMA* del posible aproximante, bien

sea polinomios, funciones trigonométricas, exponenciales, *splines*, etc., pero siempre con la pretensión de que la aproximación sea uniforme.

Pero este problema de mínimos presenta una gran dificultad, y es que no se puede resolver por las técnicas usuales para el cálculo de extremos del cálculo infinitesimal, basadas en igualar a 0 las derivadas. ¿Por qué? De hecho, ni siquiera existen métodos directos de cálculo, que nos proporcionen una solución *exacta*. Entonces, es frecuente renunciar a este tipo de aproximación, en beneficio de otras normas que nos permitan calcular el elemento *óptimo* de forma más eficiente (aunque naturalmente es distinto). Se trata de busca un equilibrio entre la precisión deseada y el costo requerido.

¿Qué función real definida en \mathbb{R}^1 hay que minimizar si trabajamos con la norma L^2 :

$$\|f\|_2 = \left(\int_0^{\pi/2} f(x)^2 dx \right)^{1/2} ?$$

¿Se puede hallar con esta nueva norma la solución anulando derivadas?

5.1.2 Sistemas sobredeterminados o incompatibles. Sea A una matriz real con m filas y n columnas. Tratemos de resolver el sistema $A\mathbf{x} = \mathbf{b}$ donde \mathbf{b} es un vector conocido con m componentes y \mathbf{x} , desconocido, tiene n componentes. Sea S un subespacio de $X = \mathbb{R}^m$ generado por los n vectores columna de A . Supongamos que S es un subespacio propio, lo cual ocurre:

a) Siempre que $m > n$, es decir haya más ecuaciones (condiciones impuestas) que incógnitas (parámetros libres).

b) Si $m \leq n$ pero entre las n columnas de A no hay m independientes, es decir haya un número de incógnitas mayor o igual que el de ecuaciones, pero la matriz A no tenga rango máximo.

Si $\mathbf{b} \in \mathbb{R}^m$ no está en S , el sistema carece de solución. Dado que, cuando \mathbf{x} recorre \mathbb{R}^n , $A\mathbf{x}$ nunca coincide con \mathbf{b} , podemos preguntarnos para qué \mathbf{x} es $\|\mathbf{b} - A\mathbf{x}\|$ lo menor posible ($\|\cdot\|$ denota una norma en \mathbb{R}^m previamente elegida). Tales \mathbf{x} juegan el papel de “soluciones” del sistema en un sentido generalizado.

Observemos que su determinación comprende dos etapas:

1. Hallar la mejor aproximación \mathbf{b}^* a \mathbf{b} por elementos de la imagen S . O las mejores, porque en función de la norma utilizada, la solución puede no ser única.
2. Resolver el sistema (o los sistemas) $A\mathbf{x} = \mathbf{b}^*$ en sentido convencional.

En este caso es la *FORMA* del aproximante lo que está totalmente definido (pues es la imagen de la aplicación lineal A), y la *NORMA* lo que hemos de decidir en función de la comodidad y eficiencia de los métodos a emplear, salvo que nos venga impuesta por su significado en el problema.

5.2 Ajuste

Es sin duda el caso más interesante de aproximación desde el punto de vista de los métodos numéricos, y por eso le vamos a dedicar una mayor atención en esta lección

introdutoria. En la práctica, constituye una aproximación funcional discreta, y casi siempre desemboca en la resolución de un sistema sobredeterminado. Nos apoyamos en el siguiente supuesto práctico:

La evolución de la masa m de una muestra radiactiva que se está desintegrando sigue la ley

$$m(t) = \left(\frac{1}{2}\right)^{\frac{t}{r}} m(0) \left(= m(0)e^{-\frac{t}{r} \ln 2}\right), \quad (5.1)$$

donde t es el tiempo y r una constante, que depende del tipo de átomo que se está desintegrando y se llama periodo de semidesintegración o vida media (¿por qué?). Para una muestra de Radio 224 se han efectuado las siguientes medidas

$t(\text{días})$	0	1	3
$m(\text{gramos})$	1.00	0.84	0.57

Se desea conocer r .

5.2.1 Método erróneo de solución. Evidentemente $m(0) = 1$, luego poniendo $t = 1$ en (5.1)

$$\begin{aligned} 0.84 &= e^{(-\ln 2/r)}, \\ \ln 0.84 &= -\ln 2/r, \\ r &= -\ln 2 / \ln 0.84 = 3.97; \end{aligned}$$

y nuestra *respuesta* sería $r = 3.97$ días.

Si ahora pusiéramos $t = 3$

$$0.57 = e^{(-3 \ln 2/r)}, \quad r = 3.70 \text{ días.}$$

Tenemos dos resultados discrepantes ¿Cuál preferir? ¿Cuál es la razón de la discrepancia?

Quizás (5.1) no sea una ley exacta sino sólo aproximada. Quizá haya habido errores al pesar la muestra o al medir el intervalo entre pesadas. Podríamos preferir la solución 3.97 a la solución 3.70 si supiésemos que los números $t = 0$, $t = 1$, $m = 1.00$, $m = 0.84$ no están afectados de errores de medida y que la ley es exacta de $t = 0$ a $t = 1$. Pero si fuesen las medidas en $t = 0$ y $t = 3$ las exactas habría que preferir 3.70. Naturalmente no hay motivo para preferir unas a otras medidas y en realidad probablemente *todas* las medidas de la tabla serán erróneas (incluyendo el dato $m(0) = 1$ usado en ambas soluciones).

5.2.2 Método correcto: el ajuste. Encontramos m_0 y r tales que

$$\left\| \begin{pmatrix} 1.00 \\ 0.84 \\ 0.57 \end{pmatrix} - \begin{pmatrix} m_0 \\ m_0 e^{(-\ln 2/r)} \\ m_0 e^{(-3 \ln 2/r)} \end{pmatrix} \right\|$$

sea lo menor posible. Este es un problema de mejores aproximaciones ¿Quién es X , S , f ? Se dice que se ha ajustado m_0 y r en el modelo (5.1).

5.2.3 Observación: En la formulación anterior S no es un subespacio. Para *linealizar* el problema, observamos que, de (5.1)

$$\ln m(t) = \ln m(0) - \frac{t}{r} \ln 2 \quad (5.2)$$

Considerando como nuevos parámetros $R = 1/r$, $M = \ln m_0$ llegamos a la formulación: hacer mínimo

$$\left\| \begin{pmatrix} \ln 1.00 \\ \ln 0.84 \\ \ln 0.57 \end{pmatrix} - \begin{pmatrix} M \\ M - R \ln 2 \\ M - 3R \ln 2 \end{pmatrix} \right\|$$

Así estamos resolviendo el sistema lineal incompatible

$$\begin{pmatrix} 0 & 1 \\ -\ln 2 & 1 \\ -3 \ln 2 & 1 \end{pmatrix} \begin{pmatrix} R \\ M \end{pmatrix} = \begin{pmatrix} \ln 1.00 \\ \ln 0.84 \\ \ln 0.57 \end{pmatrix}$$

Resuelto éste, $r = 1/R$.

5.2.4 Relación con la interpolación. Dejemos la formulación (5.2) y volvamos a la (5.1). Disponemos de la familia biparamétrica F de funciones de t de la forma $F(t) = m_0 e^{-\frac{t}{r} \ln 2}$ y hemos estado tratando de encontrar el elemento φ de esta familia tal que $\varphi(0) = 1.00$, $\varphi(1) = 0.84$, $\varphi(2) = 0.57$. Como F es biparamétrica y hay tres condiciones, no cabe esperar exista $\varphi \in F$ que las satisfaga (y de hecho hemos visto que para los valores que la tabla maneja φ no existe). Por tanto *no podemos interpolar*. Conviene entonces, más que satisfacer dos de las tres condiciones, tratar de violar *las tres* lo menos posible: esto es el ajuste. La curva solución $(t, \varphi(t))$ no pasará en general por los puntos (t_j, m_j) que representan los datos, pero se acercará a ellos lo más posible.

En la práctica el número de datos suele ser mucho mayor que el de parámetros a determinar. Los datos forman una *nube de puntos* en el plano (t, m) y la función ajustada se ciñe a dicha nube. Ahora es fácil entender que algunas veces se conozca la interpolación como *aproximación exacta*.

5.2.5 Generalizaciones. En nuestro modelo $m = \varphi(t, m(0), r)$ hay una función real m , una variable independiente real y dos parámetros $m(0)$, r . Más generalmente podremos considerar $\mathbf{y} = \varphi(\mathbf{x}, \boldsymbol{\lambda})$, $\mathbf{y} \in \mathbb{R}^m$, $\mathbf{x} \in \mathbb{R}^n$, $\boldsymbol{\lambda} \in \mathbb{R}^p$ y una tabla de valores $(\mathbf{x}_i, \mathbf{y}_i)$, $i = 1, \dots, N$. Así hay $n \times m$ condiciones a satisfacer con los p parámetros libres.

Notemos por último que, si bien en el ejemplo conocíamos ‘a priori’ la expresión funcional $m = \varphi(t)$, tal cosa no es necesarias para ajustar. Así, verbigracia, podremos describir una tabla (x_i, y_i) , $i = 1, \dots, 100$, $x_i, y_i \in \mathbb{R}$ ajustándola por una recta $y = Ax + B$ si los puntos (x_i, y_i) en el plano están *grosso modo* alineados, aunque no haya una ley que exprese que y es función lineal de x . Es decir, en el problema genérico de ajuste ni la *FORMA* ni la *NORMA* están predeterminados, siendo muchas veces la intuición o la repetida experimentación lo que nos permite dar con una solución convincente.

5.3 Aproximación óptima: existencia y unicidad

Establezcamos de forma precisa el concepto de aproximación (si no es óptima no es una verdadera aproximación):

5.3.1 Definición. Dados un espacio normado real X , un subconjunto (no necesariamente subespacio) S y un elemento f de X decimos que $p^* \in S$ es una mejor aproximación a f (por elementos de S) si, para cada $p \in S$, $\|f - p^*\| \leq \|f - p\|$, es decir

$$\|f - p^*\| = \inf\{\|f - p\| : p \in S\}.$$

5.3.2 Ejemplo. Si X es el espacio de funciones reales continuas definidas en $[-1,1]$ con la norma del supremo $\|f\|_\infty = \sup\{|f(x)| : -1 \leq x \leq 1\}$, S conjunto de polinomios mónicos de grado n y f es la función idéntica nula, la mejor aproximación es el polinomio de Chebyshev escalado $T_n/2^{n-1}$ y es única. (Corolario 1.2.3)

5.3.3 Ejemplo. Sean $X = \mathbb{R}^2$ con la norma usual $\|(x_1, x_2)^T\| = \sqrt{(x_1^2 + x_2^2)}$, $S = \{(x, 1)^T : x \in \mathbb{R}\}$, $f = (0, 0)^T$. El conjunto $\{\mathbf{x} \in \mathbb{R}^2 \mid \|\mathbf{x} - f\| = r\}$, $r > 0$ es la circunferencia de radio r centrada en el origen. Cuando $r < 1$ no interseca a S . Para $r = 1$ la intersección es el punto $(0, 1)^T$. Por consiguiente $(0, 1)^T$ es la mejor aproximación; no hay elementos de S que disten de f una cantidad $r < 1$.

5.3.4 Ejemplo. Como en el ejemplo 5.3.3 salvo que ahora usaremos la norma $\|(x_1, x_2)^T\| = \max\{|x_1|, |x_2|\}$. El conjunto $\{\mathbf{x} \in \mathbb{R}^2 \mid \|\mathbf{x} - f\| = r\}$ es un cuadrado centrado en el origen y de lado $2r$. Para $r < 1$ no interseca a S . Para $r = 1$ la intersección es el segmento $\{(x, 1)^T : -1 \leq x \leq 1\}$. Cada punto del segmento es una mejor aproximación a f . Note como p^* depende de $\|\cdot\|$.

5.3.5 Ejemplo. $X = \mathbb{R}^2$ con su norma usual $S = \{(x, 0) : x \in \mathbb{R}\}$, $f = (0, -1)^T$. Aquí S y f resultan de los del ejemplo 5.3.3, tras efectuar una translación del vector $(0, -1)^T$. La solución ahora es $(0, 0)^T$, trasladada de la solución del ejemplo 5.3.3 ¿por qué?

La teoría de la aproximación se ocupa de la existencia, unicidad y construcción de las mejores aproximaciones. Se tienen los siguientes resultados básicos:

5.3.6 Proposición

Si X es un espacio normado, S un subconjunto no vacío y compacto de X y f un elemento de X , entonces f tiene, al menos, una mejor aproximación por elementos de S .

Demostración. La función real $\|f - p\|$ definida en S es continua y acotada inferiormente por 0. Por consiguiente alcanza su mínimo. \square

5.3.7 Proposición

Si X es un espacio normado, S un subespacio vectorial de dimensión finita y f un elemento de X , entonces f tiene, al menos, una mejor aproximación por elementos de S .

Demostración. La aproximación óptima, si existe, debe pertenecer al conjunto $S^* = \{p \in S : \|f - 0\| \geq \|f - p\|\}$, ya que si p está en S pero no en S^* no es el elemento buscado pues aproxima a f peor que $0 \in S$. Así, basta probar que existe una mejor aproximación a f por elementos de S^* . Ahora bien, S^* no vacío, cerrado y acotado en un finito-dimensional, y se puede aplicar la proposición precedente. \square

Veremos mas adelante que la hipótesis sobre la dimensión de S no puede suprimirse.

5.4 Convergencia de las mejores aproximaciones. Teorema de Weierstrass

Hasta ahora hemos considerado problemas de aproximación en los que intervenían un espacio normado X , un elemento $f \in X$ y un subconjunto S de X . Un caso importante es aquel en que S es un subespacio vectorial, y muy frecuentemente S es uno de los términos de una sucesión de subespacios S_0, S_1, S_2, \dots expansiva (es decir donde cada subespacio contiene a los anteriores). Así, por ejemplo, S es a menudo el espacio Π_n de polinomios de grado $\leq n$, un miembro de la cadena $\Pi_0, \Pi_1, \Pi_2, \dots$

Supongamos pues dados X , una cadena expansiva de subespacios S_0, S_1, S_2, \dots y un elemento $f \in X$, de forma que existan las aproximaciones óptimas p_i a f por elementos de S_i , $i = 0, 1, 2, \dots$. Obviamente $\|f - p_0\| \geq \|f - p_1\| \geq \dots \geq 0$. Nos vamos a plantear bajo que circunstancias $\lim_i \|f - p_i\| = 0$, es decir, $f = \lim_i p_i$ en X .

Consideremos ante todo el caso en que $X = C[a, b]$, funciones continuas en el intervalo acotado $[a, b]$ con la norma del supremo, y $S_n = \Pi_n$. Si $f \in C[a, b]$ su polinomio $p_n \in \Pi_n$ de mejor aproximación existe, y vamos a demostrar que, para la norma del supremo en $C[a, b]$

$$\lim p_n = f, \quad n \rightarrow \infty, \quad (5.3)$$

es decir, f es límite uniforme de sus polinomios de mejor aproximación.

El resultado (5.3) es consecuencia del siguiente teorema, de importancia notable en todo el Análisis Matemático (véase el problema 5.5.4).

5.4.1 TEOREMA DE WEIERSTRASS (1885).

Si f es una función real continua en un intervalo compacto $[a, b]$, dado $\varepsilon > 0$ existe un polinomio P tal que $|f(x) - P(x)| \leq \varepsilon$ para cada x en $[a, b]$.

Demostración. El teorema posee muchas demostraciones. La que damos aquí se debe a Bernstein (1812) y es constructiva (aunque no práctica, vea problema 5.5.5).

Se supone, sin perder generalidad, que $[a, b] = [0, 1]$, pues cualquier intervalo $[a, b]$ puede transformarse en el $[0, 1]$ mediante un cambio lineal de variable.

Para $n = 1, 2, \dots$ se define el n -ésimo polinomio de Bernstein B_n (relativo a f) mediante la fórmula

$$B_n(x) = \sum_{k=0}^n \binom{n}{k} x^k (1-x)^{n-k} f\left(\frac{k}{n}\right)$$

Demostraremos que los B_n convergen a f , uniformemente en $[0, 1]$.

Probemos ante todo las relaciones

$$\sum_{k=0}^n \binom{n}{k} x^k (1-x)^{n-k} = 1 \quad (5.4)$$

$$\sum_{k=0}^n \binom{n}{k} x^k (1-x)^{n-k} \frac{k}{n} = x \quad (5.5)$$

$$\sum_{k=0}^n \binom{n}{k} x^k (1-x)^{n-k} \left(\frac{k}{n}\right)^2 = \left(1 - \frac{1}{n}\right) x^2 + \frac{1}{n} x \quad (5.6)$$

que calculan los polinomios de Bernstein de las funciones 1, x , x^2 . Escribiendo

$$(p+q)^n = \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} \quad (5.7)$$

y tomando $p = x$, $q = 1 - x$ resulta (5.4). Derivando (5.7) respecto a p y poniendo en el resultado $p = x$, $q = 1 - x$, obtenemos (5.5). Dos derivaciones respecto a p en (5.7) conducen a (5.6).

La identidad

$$\sum_{k=0}^n \left(\frac{k}{n} - x\right)^2 \binom{n}{k} x^k (1-x)^{n-k} = \frac{x(1-x)}{n} \quad (5.8)$$

es consecuencia de (5.4) - (5.6), como se ve al desarrollar el cuadrado $(k/n - x)^2$.

Tras estos preliminares, multiplicamos (5.6) por $f(x)$ y restamos $B_n(x)$ al resultado para tener

$$f(x) - B_n(x) = \sum_{k=0}^n [f(x) - f(k/n)] \binom{n}{k} x^k (1-x)^{n-k} \quad (5.9)$$

Al ser f continua en un compacto es uniformemente continua y acotada, y existen $\delta > 0$, $M > 0$ tales que $|f(x) - f(y)| < \varepsilon/2$ si $|x - y| < \delta$, $x, y \in [0, 1]$ y además $|f(x)| < M$ para cada $x \in [0, 1]$. Fijados x en $[0, 1]$ y n , dividamos los índices k en (5.9) en dos subconjuntos A y B como sigue: k está en A si $|k/n - x| < \delta$ y k está en B en caso contrario. (Nótese que A, B dependen de x, n y δ ; a su vez δ depende de ε y f .)

Ahora

$$\left| \sum_A [f(x) - f(k/n)] \binom{n}{k} x^k (1-x)^{n-k} \right| \leq \frac{\varepsilon}{2} \sum_A \binom{n}{k} x^k (1-x)^{n-k} \leq \frac{\varepsilon}{2} \quad (5.10)$$

y por otro lado, por las elecciones de M, B y por (5.8):

$$\begin{aligned} \left| \sum_B [f(x) - f(k/n)] \binom{n}{k} x^k (1-x)^{n-k} \right| &\leq 2M \sum_B \binom{n}{k} x^k (1-x)^{n-k} \\ &= 2M \sum_B \frac{(k/n - x)^2}{(k/n - x)^2} \binom{n}{k} x^k (1-x)^{n-k} \leq \frac{2M}{\delta^2} \frac{x(1-x)}{n} \leq \frac{M}{2n\delta^2} \end{aligned} \quad (5.11)$$

(hemos usado que $x(1-x) \leq \frac{1}{4}$ para $0 \leq x \leq 1$). Tomemos n suficientemente avanzado para que $\frac{M}{(2n\delta^2)} \leq \frac{\varepsilon}{2}$. Entonces, llevando (5.10) y (5.11) a (5.9) es $|f(x) - B_n(x)| \leq \varepsilon$ y el teorema está probado. \square

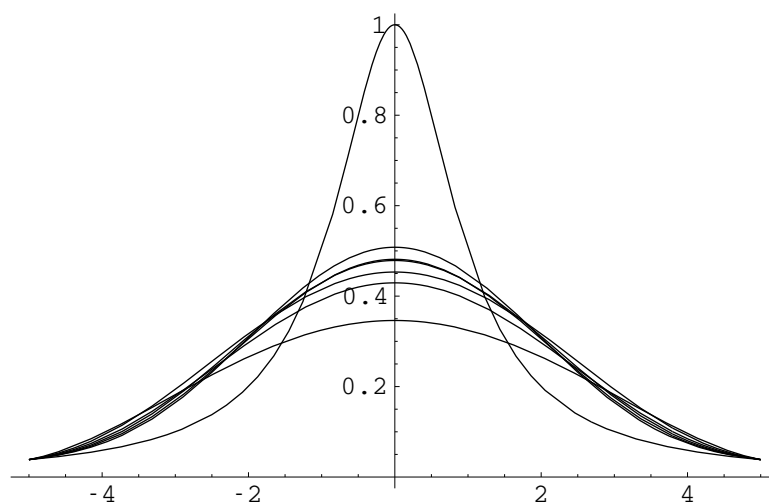


Figura 5.1: Aproximantes de grados 5 a 10 para la función de Runge

5.5 Cuestiones y problemas

5.5.1 Sea X el espacio de funciones f reales definidas en $-\infty < a \leq x \leq b < +\infty$ tales que

$$\int_a^b |f|^2 dx < \infty$$

En X se usa la norma

$$\|f\| = \left(\int_a^b |f|^2 dx \right)^{1/2}$$

Demuestre que, si $f \in X$, su mejor aproximación por una constante es su valor medio

$$\int_a^b f(x) dx / (b - a).$$

5.5.2 Sea ahora X el espacio de funciones reales acotadas definidas en $-\infty \leq a \leq x \leq b \leq +\infty$ con la norma $\|f\| = \sup\{|f(x)| : a \leq x \leq b\}$ ¿Cuál es la mejor aproximación a f por una constante?

5.5.3 Si $y = f(x)$ es una función real definida, estrictamente creciente y continua en $[0,1]$ ¿Cuál es la constante que mejor la aproxima en la norma L_1 (integral del módulo)? (Indicación: la distancia de f a una constante dada corresponde al área de cierta región del plano (x, y) ; exprese tal área como una integral definida en que la variable de integración sea y .)

5.5.4 Use el teorema de Weierstrass para demostrar la fórmula (5.3) de la lección.

5.5.5 ¿Qué debe valer n para que, cuando $f(x) = x^2$, $\|B_n f - f\|_\infty < 10^{-8}$? ¿Cree que los polinomios de Bernstein proporcionan un medio *práctico* de obtener polinomios de aproximación? Aquí y más abajo $B_n f$ denota el polinomio de Bernstein de grado $\leq n$ asociado a f .

5.5.6 Probar que $\|B_n f\|_\infty \leq \|f\|_\infty$.

5.5.7 Si f es un polinomio de grado $\leq k$, pruebe que $B_n f$ lo es para cada $n = 0, 1, 2, \dots$

5.5.8 Pruebe que B_n es un operador monótono, es decir que $B_n f \leq B_n g$ si $f \leq g$.

5.5.9 Sea f convexa en $[0,1]$. Pruebe que para $n = 2, 3, \dots$, $0 < x < 1$ se tiene $B_{n-1} f(x) \geq B_n f(x)$. Si f es continua la desigualdad es estricta, a menos que f sea lineal en cada subintervalo

$$\left[\frac{k-1}{n-1}, \frac{k}{n-1} \right], \quad k = 1, \dots, n-1$$

en cuyo caso $B_{n-1} f = B_n f$.

5.5.10 Obtenga explícitamente $B_n f$, si $f(x) = x^3$. Pruebe que $\lim_n n(B_n f - f) = 3x^2(1-x)$.

5.5.11 Si f y su derivada son continuas en $[a, b]$, $a, b \in \mathbb{R}$, entonces para cada $\varepsilon > 0$, existe un polinomio tal que $\|f - p\|_\infty \leq \varepsilon$, $\|f' - p'\| \leq \varepsilon$.