Nociones de Convergencia en la teoría de Probabilidades

Pablo L. De Nápoli

Departamento de Matemática
Facultad de Ciencias Exactas y Naturales
Universidad de Buenos Aires

Probabilidades y Estadística para Matemática
Segundo cuatrimestre de 2021

Parte I

Introducción: Definiciones y primeros ejemplos

Convergencia en probabilidad

En la teoría de probabilidades se utilizan frecuentemente diferentes nociones de convergencia de una sucesión $(X_n)_{n\in\mathbb{N}}$ de variables aleatorias. Voy a empezar poniendo todas las definiciones para que puedan hacer las prácticas. Después las iremos analizando de a una.

La primera noción que vamos a estudiar es la de convergencia en probabilidad, que aparece en el teorema de Bernoulli (ley débil de los grandes números).

Definición

Sea $(X_n)_{n\in N} : \Omega \to \overline{\mathbb{R}}$ una sucesión de variables aleatorias, definidas sobre un mismo espacio de probabilidad $(\Omega, \mathcal{E}, P)$ y finitas con probabilidad 1. Se dice que (X_n) converge en probabilidad a la variable X si para todo $\delta > 0$, tenemos que

$$P\{|X - X_n| > \delta\} \to 0 \text{ cuando } n \to +\infty$$

Notación:

$$X_n \xrightarrow{P} X$$

Variables con distribución uniforme que se concentran

Ejemplo 1:

Si $X_n \sim \mathcal{U}(-1/n, 1/n)$ y $X = 0$ con probabilidad 1. Entonces $X_n \xrightarrow{P} X$.

$$\begin{aligned} P\{|X_n - X| > \delta\} &= P\{|X_n - X| > \delta / X = 0\} \cdot P\{X = 0\} \\ &+ P\{|X_n - X| > \delta / X \neq 0\} \cdot P\{X \neq 0\} \\ &= P\{|X_n| > \delta\} = 0 \end{aligned}$$

si $\frac{1}{n} < \delta$ o sea $n > \frac{1}{\delta}$, ya que $|X_n| \leq \frac{1}{n}$ con probabilidad 1.

Desigualdad de Tchebyschev clásica

Para el siguiente ejmplo, vamos a necesitar un resultado que vimos en la clase 6.

Proposición (desigualdad de Tchebyschev clásica)

Sea X una variable aleatoria con segundo momento finito entonces

$$P\{|X - E(X)| \geq \lambda\} \leq \frac{Var(X)}{\lambda^2}$$

Lo vimos para variables discretas, pero una vez generalizadas las nociones de esperanza y varianza vale para variables continuas, con idéntica demostración.

Variables con distribución normal que se concentran

Ejemplo 2:

Si $X_n \sim N(0, \sigma_n^2)$ donde $\sigma_n \to 0$ y $X = 0$ con probabilidad 1. Entonces $X_n \xrightarrow{P} X$.

$$\begin{aligned} P\{|X_n - X| > \delta\} &= P\{|X_n - X| > \delta / X = 0\} \cdot P\{X = 0\} \\ &+ P\{|X_n - X| > \delta / X \neq 0\} \cdot P\{X \neq 0\} \\ &= P\{|X_n| > \delta\} \leq \frac{1}{\delta^2}\ \mathrm{Var}(X_n) \\ &= \frac{\sigma_n^2}{\delta^2} \to 0 \end{aligned}$$

cuando $n \to +\infty$, por la desigualdad de de Tchebyschev.

Convergencia casi segura

Otra noción que vamos a estudiar es la de convergencia casi segura.

Definición

Sea $(X_n)_{n \in N} : \Omega \to \overline{\mathbb{R}}$ una sucesión de variables aleatorias, definidas sobre un mismo espacio de probabilidad $(\Omega, \mathcal{E}, P)$ y finitas con probabilidad 1. Se dice que (X_n) convergen en forma casi segura (o casi con seguridad) a la variable X si

$$X_n \to X \text{ con probabilidad } 1$$

o sea

$$P(\{\omega \in \Omega : X_n(\omega) \to X(\omega)\}) = 1$$

o lo que es equivalente

$$P(\{\omega \in \Omega : X_n(\omega) \not\to X(\omega)\}) = 0$$

Notación:

$$X_n \xrightarrow{c.s.} X$$

Nota: En análisis real, esta noción se denomina convergencia en casi todo punto.

Un ejemplo de convergencia casi segura

Consideremos $U \sim \mathcal{U}(0,1)$, y sea

$$X_n = U^n = \varphi_n(U) \text{ con } \varphi_n(x) = x^n$$

Notemos que la sucesión de funciones $\varphi_n : \mathbb{R} \to \mathbb{R}$ converge puntualmente en $[0,1]$ a la función

$$\varphi(x) = \begin{cases} 1 & \text{si} \quad x = 0 \\ 0 & \text{si} \quad x \neq 0 \end{cases}$$

Entonces $X_n(\omega) \to X(\omega)$ salvo si $U(\omega) = 1$. Pero

$$P\{U = 1\} = P(\{\omega \in \Omega : U(\omega) = 1\}) = 0$$

. Por lo que podemos afirmar que

$$X_n \xrightarrow{c.s.} 0$$

Ejercicio: Prueben que también $X_n \xrightarrow{P} 0$. En general, veremos más adelante que la convergencia casi segura implica la convergencia en probabilidad.

Convergencia en distribución

Una tercera noción que aprece es la de convergencia en distribución.

Definición

Sea $(X_n)_{n \in N} : \Omega \to \overline{\mathbb{R}}$ *una sucesión de variables aleatorias, definidas sobre un mismo espacio de probabilidad* $(\Omega, \mathcal{E}, P)$ *y finitas con probabilidad 1. Consideramos sus funciones de distribución acumulada* F_{X_n}, F_X*. Se dice que* (X_n) *converge en distribución a la variable* X *si*

$$F_{X_n}(x) \to F_X(x)$$

para todo x *que sea un punto de continuidad de* F_X*.* **Notación:**

$$X_n \xrightarrow{D} X$$

Esta noción es más bien una noción de convegencia de distribuciones de probabilidad que de variables aleatorias.

Un ejemplo convergencia de en distribución

Ejemplo 2 (continuación):

Si $X_n \sim N(0, \sigma_n^2)$ donde $\sigma_n \to 0$ y $X = 0$ con probabilidad 1. Entonces $X_n \xrightarrow{D} X$.

Aqui, como vimos en la clase 9,

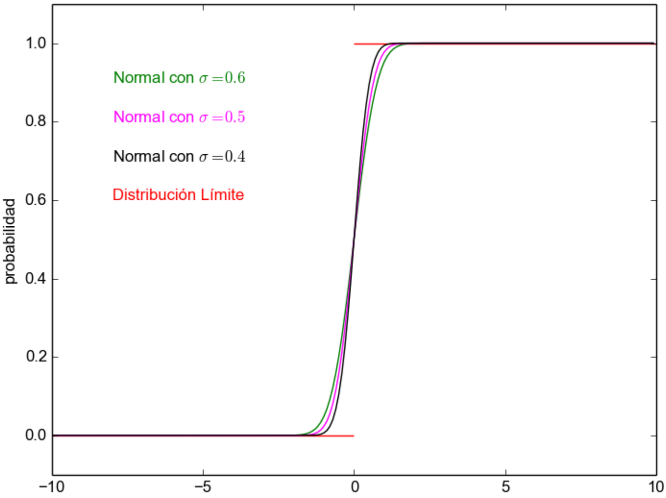
$$F_{X_n}(x) = \Phi\left(\frac{x}{\sigma_n}\right) \quad \text{donde } \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\, dt$$

es la función de distribución de una normal estándar, mientras que F_X es el escalón (función de Heaviside). Vemos que

$$F_{X_n}(x) \to \begin{cases} 0 & \text{si} \quad x < 0 \\ 1 & \text{si} \quad x > 0 \\ 1/2 & \text{si} \quad x = 0 \end{cases}$$

En consecuencia, $F_{X_n}(x) \to F_X(x)$ para todo $x \neq 0$. Luego $X_n \xrightarrow{D} X$. Veremos más adelante, que converger en probabilidad implica converger en distribución.

Un ejemplo convergencia de en distribución (2)



Parte II

La Ley Débil de los grandes números

Recordamos de la clase 6: Ley débil de los grandes números

Teorema (Ley débil de los grandes números - caso de variancia finita)

Sean $X_1, X_2, \ldots, X_n, \ldots$ una secuencia infinita de variables aleatorias independientes e idénticamente distribuidas, con

$$E[X_i] = \mu$$

$$Var(X_i) = \sigma^2 < +\infty$$

Entonces si llamamos

$$\overline{X}_n = \frac{X_1 + X_2 + \ldots + X_n}{n}$$

tenemos que

$$\overline{X}_n \xrightarrow{P} \mu$$

Este teorema lo vimos para variables discretas, pero se aplica con la misma demostración para variables continuas. De hecho, el teorema también es cierto sólo con pedir que $E[|X_i|] < +\infty$ pero la demostración en este caso es más complicada.

Demostración (ya la vimos, pero la repazamos)

Por linealidad de la esperanza, $E[\overline{X}_n] = \mu$, y por otro lado

$$\mathrm{Var}(\overline{X}_n) = \frac{\sigma^2}{n}$$

ya que las X_i son independientes. La desigualdad de Tchebyschev, dice entonces que:

$$P\{|\overline{X}_n - \mu| > \delta\} \leq \frac{\sigma^2}{n\delta^2} \to 0 \text{ cuando } n \to +\infty$$

Recordamos: los ensayos de Bernoulli

Considerábamos un experimento aleatorio con dos resultados que convencionalmente se llaman

- éxito (1) con probabilidad p.
- fracaso(0) con probabilidad $q = 1 - p$.

Introdujimos las variables aleatorias de Bernoulli $X_i : \Omega^\infty \to \mathbb{R}$ dadas por

$$X_i = \begin{cases} 1 & \text{si la i-ésima realización del experimento es un éxito} \\ 0 & \text{si la i-ésima realización del experimento es un fracaso} \end{cases}$$

Las X_i son variables aleatorias discretas. También lo es el número de éxitos en n ensayos.

$$S_n = X_1 + X_2 + \ldots + X_n$$

Como ya vimos S_n tiene distribución binomial; $S_n \sim Bi(n, p)$

$$P\{S_n = k\} = b(k, n, p) = \binom{n}{k} p^k q^{n-k}, \quad 0 \leq k \leq n$$

Recordamos: los ensayos de Bernoulli (2)

En este caso

$$\mu = p, \sigma^2 = pq$$

por lo que

$$\frac{S_n}{n} \xrightarrow{P} p \quad \text{teorema de J. Bernoulli}$$

Más precisamente, tenemos la acotación:

Acotación de las colas de la distribución binomial

$$P\left\{ \left| \frac{S_n}{n} - p \right| > \delta \right\} \leq \frac{pq}{n\delta^2}$$

Esta acotación nos proporciona una versión cuantitativa del teorema de Bernoulli, y nos será útil más adelante.

El método de montecarlo para calcular integrales

Veamos otro ejemplo. Consideramos una integral

$$I = \frac{1}{b-a} \int_a^b f(x)\, dx \text{ donde } f : [a,b] \to \mathbb{R} \text{ es continua}$$

Notamos que esta inegral puede pensarse como la esperanza de la variable $f(U)$ donde $U \sim \mathcal{U}(a,b)$. Un método para calcularla numéricamente consiste en lo siguiente: elegimos puntos al azar independientes

$$X_1, X_2, \ldots X_n$$

con distribución uniforme en $[a,b]$ (lo que podemos simular en la computadora mediante un generador de números pseudo-aleatorios), y consideramos la variable aleatoria

$$I_n = \frac{1}{n} \sum_{k=1}^{n} f(X_k)$$

El método de montecarlo para calcular integrales (2)

Miremos las variables $Y_i = f(X_i)$ serán independientes e idénticamente distribuidas, con esperanza $\mu = I$, y su varianzza es

$$\sigma^2 = E[(Y_i - \mu)^2] = E[(f(X_i) - \mu)^2] = \frac{1}{b-a} \int_a^b [f(x) - \mu]^2 \, dx < \infty$$

Por la ley débil de los grandes números, las I_n convergen en probabilidad a I. ¿Cómo podríamos acotar el error? Si $|f(x)| \leq M$ en $[a, b]$, entonces $|\mu| \leq M$, entonces

$$\sigma^2 \leq \frac{1}{b-a} \int_a^b [2M]^2 \, dx = 4M^2$$

Luego la demostración de la ley débil de los grandes números nos da que:

$$P\{|I_n - I| > \delta\} \leq \frac{\sigma^2}{n\delta^2} = \frac{4M^2}{n\delta^2}$$

Un ejemplo

Supongamos que queremos calcular

$$\pi = \int_0^1 4\sqrt{1-x^2}\,dx$$

Aquí

$$a = 0,\ b = 1,\ f(x) = 4\sqrt{1-x^2},\ M = 4$$

Si queremos 2 decimales exactos con probabilidad mayor a $p = 0{,}99$, plantemaos que queremos

$$\frac{4M^2}{n\delta^2} < q = 1 - p \Leftrightarrow n > \frac{4M^2}{\delta^2 q}$$

con $\delta = 0{,}01$ y $p = 0{,}99$. Nos queda que

$$n \geq 6{,}400{,}000$$

Esta es la cantidad de ensayos que necesitaríamos hacer. ¡Como vemos no es un método muy eficiente para calcular la integral!

Parte IV

Propiedades de la convergencia en probabilidad

Propiedades de la Convergencia en Probabilidad (1)

Recordamos que $X_n \xrightarrow{P} X$ si y sólo si (por definición)

$$P\{|X_n - X| > \delta\} \to 0 \text{ cuando } n \to +\infty$$

Observación

Si (X_n) converge en probabilidad a X, cualquier subsucesión de (X_n) también converge en probabilidad a X.

Proposición (Unicidad del límite)

Si $X_n \xrightarrow{P} X$ y $X_n \xrightarrow{P} Y$, entonces $X = Y$ con probabilidad 1.

Demostración de la unicidad

Por la desigualdad triangular,

$$|X - Y| \leq |X - X_n| + |X_n - Y|$$

Entonces

$$\{|X - Y| > \delta\} \subset \{|X - X_n| > \delta/2\} \cup \{|X_n - Y| > \delta/2\}$$

y

$$P\{|X - Y| > \delta\} \leq P\{|X - X_n| > \delta/2\} + P\{|X_n - Y| > \delta/2\}$$

Deducimos que para todo $\delta > 0$,

$$P\{|X - Y| > \delta\} = 0$$

Como

$$\{X \neq Y\} = \bigcup_{n \in N} \left\{ |X - Y| > \frac{1}{n} \right\}$$

Por la σ-subaditividad de P, deducimos que:

$$P\{X \neq Y\} \leq \sum_{n=1}^{\infty} P\left\{ |X - Y| > \frac{1}{n} \right\} = 0$$

Propiedades de la Convergencia en Probabilidad (2)

Proposición

Si $X_n \xrightarrow{P} X$ *y* $c \in \mathbb{R}$, *entonces* $cX_n \xrightarrow{P} cX$.

Demostración.

Si $c \neq 0$, tenemos que

$$P\{|cX_n - cX| > \delta\} = P\left\{|X_n - X| > \frac{\delta}{|c|}\right\} \to 0 \text{ cuando } n \to +\infty$$

Si $c = 0$ es trivial. $\square$

Propiedades de la Convergencia en Probabilidad (3)

Proposición

Si $X_n \xrightarrow{P} X$ *e* $Y_n \xrightarrow{P} Y$, *entonces* $X_n + Y_n \xrightarrow{P} X + Y$.

Demostración.

Tenemos la desigualdad:

$$|(X+Y)-(X_n+Y_n)| \leq |X-X_n| + |Y-Y_n|$$

Eso se traduce en

$$\{|(X+Y)-(X_n+Y_n)| > \delta\} \subset \{|X-X_n| > \delta/2\} \cup P\{|Y-Y_n| > \delta/2\}$$

y por lo tanto (por la subaditividad de la probabilidad),

$$P\{|(X+Y)-(X_n+Y_n)| > \delta\} \leq P\{|X-X_n| > \delta/2\} + P\{|Y-Y_n| > \delta/2\}$$

$\square$

Propiedades de la Convergencia en Probabilidad (4)

Lema

Si $X : \Omega \to \overline{\mathbb{R}}$ es una variable aleatoria finita con probabilidad 1, entonces dado $\varepsilon > 0$, existirá un k_0 tal que

$$P\{|X| \geq k_0\} < \frac{\varepsilon}{2}$$

Notamos que esto dice que las colas de una variable aleatoria finita en casi todo punto tienden a cero en probabilidad.

Demostración

Notamos que como X es finita con probabilidad 1 por hipótesis,

$$\sum_{k=1}^{\infty} P\{k-1 \leq |X| < k\} = 1$$

es una serie convergente, por consiguiente dado $\varepsilon > 0$, existirá un k_0 tal que:

$$\sum_{k=k_0+1}^{\infty} P\{k-1 \leq |X| \leq k\} < \frac{\varepsilon}{2}$$

Es decir que:

$$P\{|X| \geq k_0\} < \frac{\varepsilon}{2}$$

Propiedades de la Convergencia en Probabilidad (5)

Lema

Si $X_n \xrightarrow{P} X$, *entonces* (X_n) *está acotada en probabilidad, en el siguiente sentido, dado* $\varepsilon > 0$ *existen* $M = M_\varepsilon$ *y* $n_0 = n_0(\varepsilon)$ *tales que*

$$\forall\, n \geq n_0(\varepsilon) : \; P\{|X_n| > M\} < \varepsilon$$

Demostración

Elijamos k_0 como en el lema anterior De la desigualdad triangular,

$$|X_n| \leq |X_n - X| + |X|$$

Deducimos que:

$$P\{|X_n| > k_0 + \delta\} \leq P\{|X_n - X| > \delta\} + P\{|X| > k_0\}$$

y en consecuencia que

$$P\{|X_n| > k_0 + \delta\} \leq \varepsilon$$

si $n \geq n_0(\varepsilon)$. Esto prueba la afirmación del lema, con $M = k_0 + \delta$.

Propiedades de la Convergencia en Probabilidad (6)

Lema

Si $X_n \xrightarrow{P} 0$ *e* Y_n *"está acotada en probabilidad", entonces* $X_n Y_n \xrightarrow{P} 0$.

Demostración.

$$\begin{aligned}
P\{|X_n Y_n| > \delta\} &= P\left\{|X_n| > \frac{\delta}{|Y_n|}\right\} \\
&= P\left\{|X_n| > \frac{\delta}{|Y_n|} \ \wedge \ |Y_n| \leq M\right\} + P\left\{|X_n| > \frac{\delta}{|Y_n|} \ \wedge \ |Y_n| > M\right\} \\
&\leq P\left\{|X_n| > \frac{\delta}{M}\right\} + P\{|Y_n| > M\} \to 0
\end{aligned}$$

cuando $n \to +\infty$. □

Propiedades de la Convergencia en Probabilidad (6)

Corolario

Si $X_n \xrightarrow{P} X$ *e* $Y_n \xrightarrow{P} Y$, *entonces* $X_n Y_n \xrightarrow{P} XY$.

Demostración.

Utilizamos el truco habitual de "sumar y restar":

$$XY - X_n Y_n = XY - X_n Y + X_n Y - X_n Y_n = (X - X_n)Y + X_n(Y_n - Y)$$

Entonces como $X - X_n \xrightarrow{P} 0$ e Y está acotada en probabilidad, deducimos que $(X - X_n)Y \xrightarrow{P} 0$. Similarmente, como $Y_n - Y \xrightarrow{P} 0$ y X_n está acotada en probabilidad (por la proposición 2, deducimos que $(X - X_n)Y \xrightarrow{P} 0$. Tenemos entonces que $X_n Y_n - XY \xrightarrow{P} 0$, y en consecuencia $X_n Y_n \xrightarrow{P} XY$ □

Parte V

Una aplicación de la ley débil de los grandes números al análisis: los polinomios de Bernstein

El teorema de Weierstrass

En esta sección expondremos una prueba del teorema de Weierstrass sobre aproximación a funciones continuas por polinomios, debida a S.N. Bernstein:

Teorema (Weierstrass)

Sea $f \in C[0,1]$ una función continua $f : [0,1] \to \mathbb{R}$, entonces existe una sucesión de polinomios $P_n(t)$ tal que $P_n(t) \to f(t)$ uniformemente para $t \in [0,1]$.

En un lenguaje más moderno, el teorema de Weierstrass dice que los polinomios son densos en el espacio $C[0,1]$ de las funciones continuas (con la norma del supremo).

La prueba de S.N. Berstein (1912) de este teorema, consiste en utilizar la distribución binomial, para construir explícitamente una sucesión de polinomios que converge uniformemente a f, utilizando ideas de probabilidades realacionadas con la ley débil de los grandes números.

Polinomios de Bernstein: Una prueba del teorema de Weierstrass (1)

Veamos primero la idea intuitiva de la demostración: sea $p \in [0,1]$ y sea como antes S_n el número de éxitos en n ensayos de Bernoulli con probabilidad p. La ley de los grandes números afirma que:

$$\frac{S_n}{n} \to p \text{ (en probabilidad)}$$

y como f es continua es razonable esperar que:

$$f\left(\frac{S_n}{n}\right) \to f(p)$$

(De vuelta, esto no es estrictamente cierto para toda sucesión de ensayos de Bernoulli, pero sí vale en probabilidad.) Por lo que esperamos que:

$$E\left[f\left(\frac{S_n}{n}\right)\right] \to E[f(p)] = f(p)$$

Notemos que:

$$B_n(p) = E\left[f\left(\frac{S_n}{n}\right)\right] = \sum_{k=0}^{n} f\left(\frac{k}{n}\right) b(k,n,p)$$

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

es un polinomio en la variable p. Se lo denomina el n-ésimo polinomio de Bernstein.

La demostración de S.N. Bernstein, consiste en probar que $B_n(p) \to f(p)$ uniformemente para $p \in [0,1]$ (Los argumentos anteriores no constituyen una prueba rigurosa, pero explican intuitivamente por qué esta afirmación es cierta).

De hecho, la demostración de esta afirmación se basa en argumentos muy similares a los que nos llevaron a la prueba del teorema de Bernoulli.

Ilustación de la aproximación mediante los polinomios de Bernstein

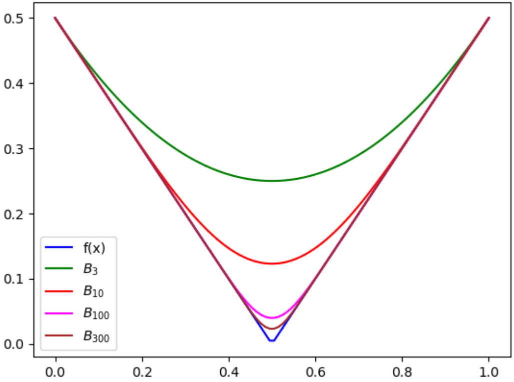


Figura: Aproximación a $f(x) = |\frac{1}{2} - x|$ mediante los polinomios de Bernstein.

Propiedades de las funciones continuas que vamos a utilizar

Para la prueba del teorema de Weierstrass utilizaremos, dos propiedades claves de las funciones continuas en un intervalo cerrado de la recta, a saber:

1. Una función continua en un intervalo cerrado de la recta, es acotada: existe una constante $M > 0$ tal que:

$$|f(p)| \leq M \; \forall \; p \in [0,1]$$

2. Una función continua en un intervalo cerrado de la recta, es uniformemente continua: dado $\varepsilon > 0$ existe $\delta > 0$ tal que si $x, y \in [0,1]$ y si $|x - y| \leq \delta$, entonces $|f(x) - f(y)| < \varepsilon$.

Acotación de las colas de la distribución binomial

Recordamos la acotación de las colas de la distribución binomial: que obtuvimos antes:

$$P\left\{\left|\frac{S_n}{n} - p\right| > \delta\right\} \leq \frac{pq}{n\delta^2} \leq \frac{1}{4n\delta^2}$$

ya que:

$$pq = p(1-p) \leq \frac{1}{4} \; \forall \; p \in [0,1]$$

Más explícitamente podemos escribir esto como:

$$\sum_{|k/n-p|>\delta} b(k,n,p) = \sum_{|k/n-p|>\delta} P\{S_n = k\} \leq \frac{1}{4n\delta^2}$$

Demostración del teorema (1)

Queremos acotar la diferencia:

$$B_n(p) - f(p) = \sum_{k=0}^{n} \left[f\left(\frac{k}{n}\right) b(k,n,p) \right] - f(p) = \sum_{k=0}^{n} \left[f\left(\frac{k}{n}\right) - f(p) \right] b(k,n,p)$$

pues

$$\sum_{k=0}^{n} b(k,n,p) = 1$$

(¡Es una distribución de probabilidades!). En consecuencia,

$$|B_n(p) - f(p)| \leq \sum_{k=0}^{n} \left| f\left(\frac{k}{n}\right) - f(p) \right| b(k,n,p)$$

En esta suma separamos dos partes, la suma sobre los k donde $|k/n - p| \leq \delta$ (con el δ dado por la continuidad uniforme), y la parte donde $|k/n - p| > \delta$.

Demostración del teorema (2)

La primer parte la acotamoss fácilmente:

$$\sum_{k:|k/n-p|\leq\delta} \left| f\left(\frac{k}{n}\right) - f(p) \right| b(k,n,p) \leq \sum_{k:|k/n-p|\leq\delta} \varepsilon\, b(k,n,p) \leq \varepsilon$$

pues los $b(k,n,p)$ suman 1.

La otra parte de la suma la acotamos usando nuestra estimación de las colas de la distribución binomial:

$$\sum_{k:|k/n-p|>\delta} \left| f\left(\frac{k}{n}\right) - f(p) \right| b(k,n,p) \leq 2M \sum_{|k/n-p|>\delta} b(k,n,p) < \frac{2M}{4n\delta^2} < \varepsilon$$

si $n \geq n_0(\varepsilon)$. En consecuencia, $|B_n(p) - f(p)| < 2\varepsilon$ si $n \geq n_0(\varepsilon)$, para todo $p \in [0,1]$. Esto concluye la prueba del teorema de Weierstrass.