UNIVERSIDAD CENTRAL DE VENEZUELA
FACULTAD DE CIENCIAS
ESCUELA DE MATEMÁTICA

Teorema de aproximación de Weierstrass y grandes desviaciones

Trabajo Especial de Grado presentado ante la ilustre Universidad Central de Venezuela por el **Br. Eddre Peña** para optar al título de Licenciado en Matemática.

Tutor: Marisela Domínguez.

Caracas, Venezuela
Marzo, 2006

Nosotros, los abajo firmantes, designados por la Universidad Central de Venezuela como integrantes del Jurado Examinador del Trabajo Especial de Grado titulado "**Teorema de aproximación de Weierstrass y grandes desviaciones**", presentado por el **Br. Eddre Peña**, titular de la Cédula de Identidad **15842054**, certificamos que este trabajo cumple con los requisitos exigidos por nuestra Magna Casa de Estudios para optar al título de **Licenciado en Matemática**.

Marisela Domínguez
Tutor

Margarita Olivares
Jurado

Ramón Bruzual
Jurado

Agradecimiento

Gracias a Dios por darme vida y salud.

Gracias a mis padres y hermanos que de una u otra forma siempre me apoyaron a lo largo de la carrera.

Gracias a la profesora Marisela Domínguez por todo su apoyo y colaboración en la realización de este trabajo.

Gracias a la profesora Mairene Colina por su incondicional apoyo siempre que la necesite.

Gracias a la profesora Margarita Olivares y al profesor Ramón Bruzual por sus sugerencias en la culminación de este trabajo.

Gracias al Consejo de Desarrollo Científico y Humanistíco de la Universidad Central de Venezuela por su colaboración en la edición de este trabajo.

En fin **GRACIAS A TODOS**.......

Índice general

Introducción

El objetivo de este trabajo es comprender el artículo “The Weierstrass approximation theorem and large desviations”de Henryk Gzyl y José Luis Palacios publicado en la revista *American Mathematical Monthly*, 1997.

La prueba de Bernstein (1912) del teorema de aproximación de Weierstrass es una aplicación clásica de teoría de probabilidades al análisis. La herramienta principal utilizada por Bernstein es la desigualdad de Chebyshev. Si el argumento se aplica a funciones que satisfacen la condición de Lipschitz, se puede probar que la velocidad de convergencia de los polinomios de Bernstein a la función es al menos de orden $1/n^{1/3}$.

El resultado del artículo de Gzyl y Palacios establece que si en lugar de usar la desigualdad de Chebyshev se usa otra herramienta probabilística (la teoría de grandes desviaciones) entonces se puede probar que la velocidad de convergencia es al menos de orden $\sqrt{\ln n}/\sqrt{n}$.

Gzyl y Palacios en su publicación presentan un resultado probabilístico, antes de dar la prueba del teorema que muestra dicha convergencia. Para comprender estos resultados es necesario tener conocimientos básicos en teoría de probabilidades, así como también nociones de análisis como lo son la condición de Lipschitz y los polinomios de Bernstein entre otros.

En el primer capítulo se hace un breve repaso sobre algunas definiciones básicas en la teoría de probabilidades como lo son las variables aleatorias, el valor esperado y las desigualdades de Chebyshev y Markov. Además recordamos resultados importantes de probabilidades que son usados para la demostración del teorema de aproximación.

En el segundo capítulo se introducen los polinomios de Bernstein y damos la prueba del teorema de aproximación de Weierstrass usando la desigualdad de Chevyshev.

En el tercer capítulo se desarrolla el artículo de Gzyl y Palacios presentando sus resultados en este trabajo.

CAPíTULO 1

Algunas nociones de probabilidades

1. Espacio de Probabilidad

Cuando realizamos un experimento nos interesa saber si ciertos subconjuntos del espacio muestral Ω, conjunto de resultados, ocurren o no. A estos subconjuntos de interés los llamamos eventos o sucesos.

Un evento $A \subseteq \Omega$ ocurre, al realizar un experimento, si el resultado ω obtenido es tal que $\omega \in A$. Denotaremos a la familia de subconjuntos de Ω formada por los eventos, como $\mathcal{A}$ y con la siguiente definición daremos las propiedades que debe cumplir.

DEFINICIÓN 1.1. Una familia $\mathcal{A}$ de subconjuntos de Ω es una *σ-álgebra* si satisface las siguientes propiedades:

- $\Omega \in \mathcal{A}$, es decir que la realización del experimento produce un resultado. Al conjunto Ω lo llamaremos evento cierto o seguro, ya que este evento siempre ocurre.
- $A \in \mathcal{A}$ implica que $A^c \in \mathcal{A}$, es decir si A es un evento entonces "A no ocurre" también es un evento.
- $A_n \in \mathcal{A}$, para $n = 1, 2...$ entonces $\displaystyle\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$, es decir que la unión de eventos es un evento.

EJEMPLO 1.2. El conjunto de partes de Ω, $\mathcal{P}(\Omega)$ es una σ-álgebra.

DEFINICIÓN 1.3. Sea Ω un espacio muestral y $\mathcal{A}$ una familia de eventos de Ω. Una *medida de probabilidad o probabilidad* sobre Ω, es una función P definida sobre $\mathcal{A}$ que satisface los siguientes axiomas:

- P es no negativa, es decir para todo evento $A \in \mathcal{A}$ se cumple que $P(A) \geq 0$.
- P es $\sigma-$aditiva es decir, si $(A_n)_{n\geq 1} \subseteq \mathcal{A}$ son disjuntos dos a dos (lo que significa $A_i \cap A_j = \emptyset$ para cada $i \neq j$) entonces

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i).$$

- El evento cierto tiene probabilidad 1: $P(\Omega) = 1$.

La terna $(\Omega, \mathcal{A}, P)$ se denomina *espacio de probabilidad*, el valor $P(A)$ se denomina probabilidad de A.

EJEMPLO 1.4. Si lanzamos una moneda al aire entonces podemos describir este evento como un espacio de probabilidad: sea $\Omega = \{cara, sello\}$ el espacio muestral, con $\mathcal{A} = \mathcal{P}(\Omega)$ la σ-álgebra entonces:

$$P(cara) = 1, \quad P(sello) = 0$$

DEFINICIÓN 1.5. Un experimento que tiene dos posibles resultados se denomina *ensayo de Bernoulli*.

2. Variable aleatoria.

Definición 1.6. Sea $(\Omega, \mathcal{P}(\Omega), P)$ un espacio de probabilidad discreto. Una *variable aleatoria* real es una función

$$X : \Omega \longrightarrow \mathbb{R}$$

que a cada $\omega \in \Omega$ le asocia un único valor real $X(\omega)$.

Definición 1.7. Una variable aleatoria se denomina *discreta*, si su rango es un conjunto discreto (finito o numerable). En este caso existe un conjunto $(x_n)_{n\geq 1}$ (conjunto de valores de X) tal que

$$\sum_{n\geq 1} P(X = x_n) = 1.$$

En particular, las variables definidas sobre espacios de probabilidad discretos son discretas. Sin embargo una variable aleatoria definida sobre un espacio de probabilidad no discreto puede ser discreta.

Ejemplo 1.8.

- Sea $c \in \mathbb{R}$ un número fijo. Se define $X(\omega) = c$ para todo $\omega \in \Omega$, es decir X es constante. Entonces para todo intervalo $I \subseteq \mathbb{R}$ se tiene que

$$X^{-1}(I) = [\omega \in \Omega : X(\omega) \in I] = \begin{cases} \Omega & si \quad c \in I \\ \emptyset & si \quad c \in I^c. \end{cases}$$

  Por lo tanto

$$P\left(X^{-1}(I)\right) = P\left([\omega \in \Omega : X(\omega) \in I]\right) = \begin{cases} 1 & si \quad c \in I \\ 0 & si \quad c \in I^c. \end{cases}$$

- Si en el ejemplo anterior consideramos al espacio muestral $\Omega = [a, b] \subseteq \mathbb{R}$, donde $a < b$, se obtienen variables aleatorias discretas definidas sobre espacios no discretos.

Variable aleatoria de Bernoulli

Identifiquemos los resultados de un ensayo de Bernoulli como éxito y fracaso; y sea $p \in (0, 1)$ la probabilidad de obtener éxito en una realización del experimento. Definimos la variable aleatoria $X = 1$ cuando el resultado es éxito y $X = 0$ cuando el resultado es fracaso. Es decir, tenemos

$$\Omega = \{\text{ éxito, fracaso }\}$$

$$X : \Omega \to \mathbb{R}$$

$$X(\text{ éxito }) = 1$$

$$X(\text{ fracaso }) = 0$$

Una variable aleatoria de este tipo se denomina *variable aleatoria de Bernoulli.*

Variable aleatoria binomial

Supongamos que se realizan n ensayos de Bernoulli independientes y con probabilidad de éxito p. La variable aleatoria Y definida como el número de éxitos en los n ensayos, se denomina *variable aleatoria binomial* con parámetros n y p. Esto también se expresa mediante la notación $Y \sim B(n,p)$. Es decir sea

$$\Omega = \{\text{ cero éxito, un éxito,...,n éxitos}\}$$

$$Y : \Omega \to \mathbb{R}$$

$$\begin{aligned} Y(\text{ cero éxito }) &= 0 \\ Y(\text{ un éxito }) &= 1 \\ &. \\ &. \\ &. \\ Y(\text{ n éxitos }) &= n. \end{aligned}$$

Probabilidad de una variable aleatoria.

Definición 1.9. Dada una variable aleatoria discreta X cuyo conjunto de valores es un conjunto $\{x_n\}_{n\geq 1}$, la función

$$\mathbb{P} : (x_n)_{n\geq 1} \longrightarrow [0,1]$$

$$x_n \longrightarrow \mathbb{P}(n) = p_n = P(X = x_n)$$

se denomina *función de probabilidad de la variable aleatoria X.*

OBSERVACIÓN 1.10. Esta función satisface la condición

$$\sum_{n\geq 1} p_n = \sum_{n\geq 1} P(X = x_n) = 1.$$

Probabilidad de una variable aleatoria de Bernoulli.

Si X es una variable aleatoria Bernoulli con probabilidad de éxito p tenemos que su función de probabilidad es

$$\begin{aligned} \mathbb{P} : \{0,1\} &\mapsto [0,1] \\ 1 &\mapsto p \\ 0 &\mapsto 1-p \end{aligned}$$

Es decir,

$$\begin{aligned} \mathbb{P}(1) = p_1 = P\{X = 1\} = p \\ \mathbb{P}(0) = p_0 = P\{X = 0\} = 1-p \end{aligned}$$

Probabilidad de una variable aleatoria binomial.

Si Y es una variable aleatoria binomial con parámetros n y p, es decir $Y \sim B(n,p)$ tenemos que su función de *distribución* es

$$b(i,n,p) = \mathbb{P}(i) = p_i = P(Y = i) = \binom{n}{i} p^i (1-p)^{n-i}$$

para $i = 0, ..., n$.

Observemos que la función dada por $(p_i)_{i=1}^{n}$ es una función de probabilidad ya que

$$\sum_{i=0}^{n} p_i = \sum_{i=1}^{n} p^i (1-p)^{n-i} = (p+1-p)^n = 1$$

3. Esperanza y varianza

Esperanza de una variable aleatoria.

DEFINICIÓN 1.11. Sea X una variable aleatoria discreta tal que su conjunto de valores es $(x_n)_{n\geq 1}$ y su función de probabilidad $(p_n)_{n\geq 1}$, donde $p_n = P(X = x_n)$. La *esperanza matemática, valor esperado o media* de X es el número denotado por $E[X]$ y definido como

$$E[X] = \int_\Omega X(\omega)dP(\omega) = \sum_{n=0}^{\infty} x_n P(X = x_n) = \sum_{n=0}^{\infty} x_n p_n$$

siempre y cuando la serie anterior sea convergente. En este caso se dice que existe la esperanza matemática de X.

DEFINICIÓN 1.12. Sea X una variable aleatoria discreta (como en la definición anterior). Sea $h : \mathbb{R} \longrightarrow \mathbb{R}$ una función entonces la composición $h \circ X$ también es una variable aleatoria discreta y su conjunto de valores es $\{h(x_n)\}_{n\geq 1}$. Además la esperanza de $h \circ X$ se denota por $E(h(X))$ y

$$\begin{aligned} E(h(X)) &= \int_\Omega h(X(\omega))dP(\omega) \\ &= \sum_{n=0}^{\infty} h(x_n)P\{X = x_n\} \\ &= \sum_{n=0}^{\infty} h(x_n)p_n. \end{aligned}$$

TEOREMA 1.13. *Si $X_1, X_2, ..., X_n$ son variables aleatorias con esperanzas finitas, entonces la esperanza de su suma existe y es igual a la suma de sus esperanzas:*

$$E(X_1 + ... + X_n) = E(X_1) + ... + E(X_n)$$

DEMOSTRACIÓN. Es suficiente demostrarlo para dos variables X y Y. Utilizando la definición de esperanza y la linealidad de la integral podemos escribir:

$$E(X) + E(Y) = \int_\Omega X(\omega)dP(\omega) + \int_\Omega Y(\omega)dP(\omega) = \int_\Omega (X(\omega) + Y(\omega))dP(\omega) = E(X + Y).$$

lo cual completa la demostración. $\square$

TEOREMA 1.14. *Si X, Y son variables aleatorias independientes, con esperanza finita, entonces su producto es una variable aleatoria con esperanza finita, y*

$$E(XY) = E(X)E(Y).$$

DEMOSTRACIÓN. Para calcular $E(XY)$ debemos multiplicar cada valor posible $x_j y_k$ por su probabilidad $f(x_j)g(y_k)$ correspondiente. Por consiguiente

$$E(XY) = \sum_{j,k} x_j y_k f(x_j) g(y_k) = \left(\sum_j x_j f(x_j)\right)\left(\sum_k y_k g(y_k)\right)$$

donde el rearreglo se justifica, dado que la serie es absolutamente convergente. □

OBSERVACIÓN 1.15. La misma regla de multiplicación es válida, por inducción, para cualquier número de variables aleatorias independientes.

EJEMPLO 1.16. A continuación estudiaremos la esperanza de la variable aleatoria binomial. Sea S_n el número de éxitos en n ensayos de Bernoulli que tienen probabilidad p de éxito. Sabemos que S_n tiene la distribución binomial, $S_n \sim B(n,p)$ de donde tenemos que

$$E(S_n) = \sum_{i=0}^{n} i b(i,n,p) = np \sum_{i=0}^{n} b(i-1, n-1, p).$$

La última suma incluye todos los términos de la distribución binomial para $n-1$ ensayos y, por lo tanto, es igual a 1. En consecuencia la esperanza de la distribución binomial es:

$$E(S_n) = np$$

Varianza de una variable aleatoria

DEFINICIÓN 1.17. Si X es una variable aleatoria con media $E[X] = \mu$, entonces la *varianza* de X, denotada $Var[X]$, se define como

$$Var[X] = E[X^2] - (E[X])^2 = E[X^2] - \mu^2.$$

La varianza de X es una medida del dispersion de los valores de X alrededor de la media.

Ejemplo 1.18. Calcularemos la varianza para una variable aleatoria binomial.
Sea $X \sim B(n,p)$, sabemos que $E[X] = np$ entonces

$$\begin{aligned}
E[X^2] &= \sum_{k=0}^{n} k^2 P(X=k) = \sum_{k=0}^{n} k^2 \binom{n}{k} p^k (1-p)^{n-k} \\
&= \sum_{k=0}^{n} \frac{k^2 n!}{k!(n-k)!} p^k (1-p)^{n-k} = \sum_{k=0}^{n} (k^2 - k + k) \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \\
&= \sum_{k=0}^{n} k(k-1) \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} + E[X] \\
&= \sum_{k=2}^{n} \frac{n!}{(k-2)!(n-k)!} p^k (1-p)^{n-k} + E[X] \\
&= p^2 n(n-1) \sum_{k=2}^{n} \frac{(n-2)!}{(k-2)!(n-k)!} p^{k-2} (1-p)^{n-k} + E[X] \\
&= p^2 n(n-1) \sum_{k=2}^{n} \binom{n-2}{k-2} p^{k-2} (1-p)^{n-k} + E[X] \\
&= p^2 n(n-1)(p + (1-p))^{n-2} + E[X] \\
&= p^2 n(n-1) + np = np(1 + (n-1)p).
\end{aligned}$$

Por lo tanto

$$Var[X] = E[X^2] - (E[X])^2 = np(1 + (n-1)p) - (np)^2 = np(1-p)$$

4. Desigualdades de Markov y de Chebyshev

Desigualdad de Markov

PROPOSICIÓN 1.19. *Si X es una variable aleatoria con valores no negativos, entonces para $a > 0$*

$$P(X > a) \leq \frac{E(X)}{a}$$

DEMOSTRACIÓN. Sea $A = \{j : x_j \leq a\}$ y $P(X = x_j) = p_j$ entonces:

$$P(X > a) = \sum_{j \in A} p_j$$

$$E(X) = \sum_{j \in A} x_j p_j + \sum_{j \in A^c} x_j p_j \geq \sum_{j \in A} x_j p_j \geq \sum_{j \in A} a p_j = aP(X > a)$$

$$P(X > a) \leq \frac{E(X)}{a}$$

□

Desigualdad de Chebyshev:

PROPOSICIÓN 1.20. *Si X es una variable aleatoria con media finita μ y varianza σ^2, entonces para cualquier $k > 0$*

$$P[|X - \mu| \geq k] \leq \frac{\sigma^2}{k^2} = \frac{E[(X - \mu)^2]}{k^2}$$

(con $E[X] < \infty$).

DEMOSTRACIÓN. Como $Y^2 = (X - \mu)^2 > 0$ es una variable aleatoria y definiendo $a = k^2 > 0$ se puede usar la desigualdad de Markov. Obteniendo:

$$P(Y \geq a) = P(|X - \mu| \geq k) = P((X - \mu)^2 \geq k^2) \leq \frac{E[(X - \mu)^2]}{k^2} = \frac{\sigma^2}{k^2}.$$

□

OBSERVACIÓN 1.21. Estas desigualdades dan cotas de la probabilidad sólo cuando la esperanza o la varianza son finitas.

5. Convergencia en probabilidad y ley débil de los grandes números

Definición 1.22. Sea $\{X_n\}_{n\geq 1}$ una sucesión de variables aleatorias en $(\Omega, \mathcal{A}, P)$ y X una variable aleatoria en $(\Omega, \mathcal{A}, P)$. Decimos que $\{X_n\}_{n\geq 1}$ *converge a X en probabilidad* si dado $\alpha > 0$

$$\lim_{n\to\infty} P\{|X_n - X| > \alpha\} = 0.$$

Es decir, dados $\alpha, \varepsilon > 0$ existe N tal que

$$P\{|X_n - X| > \alpha\} < \varepsilon \text{ si } n \geq N.$$

Teorema 1.23 (Ley débil de los grandes números).

Sea $\{X_n\}_{n\geq 1}$ una sucesión de variables aleatorias independientes, idénticamente distribuidas cada uno con esperanza $E(X_i) = \mu$ y varianza σ^2 (ambas finitas). Entonces para cada $\varepsilon > 0$

$$P\left\{\left|\frac{X_1 + ... + X_n}{n} - \mu\right| \geq \varepsilon\right\} \to 0 \quad \textit{cuando} \quad n \to \infty.$$

Teorema 1.24 (Teorema de la convergencia dominada de Lebesgue).

Sea $X_n : \Omega \to \mathbb{R}, n \in \mathbb{N}$, una sucesión de variables aleatorias que converge puntualmente a una variable aleatoria X. Supongamos que existe una variable aleatoria Y tal que $E(Y)$ es finita y

$$\forall n, \forall \omega, \qquad |X_n(\omega)| \leq Y(\omega).$$

(Y domina a las X_n). Entonces, cada X_n es integrable, X es integrable y se cumple que

$$\lim_{n\to\infty} \int_\Omega X_n dP(\omega) = \int_\Omega \lim_{n\to\infty} X_n dP(\omega) = \int_\Omega X dP(\omega).$$

Mas aún, se tiene que

$$\lim_{n\to\infty} \int_\Omega |X - X_n| dP(\omega) = 0.$$

Teorema 1.25 (Teorema de la convergencia dominada en probabilidad).

Sea $X_n : \Omega \to \mathbb{R}, n \in \mathbb{N}$, *una sucesión de variables aleatorias que converge en probabilidad a una variable aleatoria* X. *Supongamos que existe una variable aleatoria* Y *tal que* $E(Y)$ *es finita y*

$$\forall n, \forall \omega, \qquad |X_n(\omega)| \leq Y(\omega).$$

(Y *domina a las* X_n*). Entonces, cada* X_n *es integrable,* X *es integrable y se cumple que*

$$\lim_{n\to\infty} \int_\Omega X_n dP(\omega) = \int_\Omega \lim_{n\to\infty} X_n dP(\omega) = \int_\Omega X dP(\omega).$$

Mas aún, se tiene que

$$\lim_{n\to\infty} \int_\Omega |X - X_n| dP(\omega) = 0.$$

CAPíTULO 2

Polinomios de Bernstein y teorema de aproximación de Weierstrass

Polinomio de Bernstein

DEFINICIÓN 2.1. Sea $f : [0,1] \to \mathbb{R}$ una función continua. El *polinomio n-ésimo de Bernstein* P_n de la función f se define como:

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

donde $\displaystyle\binom{n}{k} = \frac{n!}{k!(n-k)!}$

A continuación veremos una demostración probabilística del teorema de Weierstrass referente a la aproximación de funciones continuas por polinomios.

TEOREMA 2.2. *Sea f una función continua en $[0,1]$ y sea P_n el n-ésimo polinomio de Bernstein de f entonces P_n converge a f uniformemente en $[0,1]$.*

DEMOSTRACIÓN. Sea x fijo veamos que $P_n(x)$ converge a $f(x)$ puntualmente. Sea (X_n) una sucesión de variables aleatorias independientes Bernoulli

$$X_n = \begin{cases} 1 & \text{con probabilidad} \quad x \\ 0 & \text{con probabilidad} \quad 1-x. \end{cases}$$

Sea $S_n = \displaystyle\sum_{i=1}^{n} X_i$ sabemos que

$$P(S_n = k) = \binom{n}{k}x^k(1-x)^{n-k}.$$

Como los X_i son independientes, equidistribuidos y de cuadrado integrable, por la ley débil de los grandes números $\dfrac{S_n}{n}$ converge a x en probabilidad lo que implica que $f\left(\frac{S_n}{n}\right)$ converge a $f(x)$ en probabilidad, de donde $E\left(f\left(\frac{S_n}{n}\right)\right)$ converge a $E(f(x)) = f(x)$ es decir:

$$\sum_{k=1}^{n} f\left(\frac{k}{n}\right)\binom{n}{k}x^k(1-x)^{n-k} \text{ converge a } f(x) \text{ puntualmente.}$$

Veamos ahora la convergencia uniforme:

Para $\delta > 0$, usando el binomio de Newton aplicado a $(x + (1-x))^n = 1$ obtenemos

$$\begin{aligned}
|P_n(x) - f(x)| &= \left|\sum_{k=0}^{n} f\left(\frac{k}{n}\right)\binom{n}{k}x^k(1-x)^{n-k} - f(x)\sum_{k=0}^{n}\binom{n}{k}x^k(1-x)^{n-k}\right| \\
&= \left|\sum_{k=0}^{n}\left(f\left(\frac{k}{n}\right) - f(x)\right)\binom{n}{k}x^k(1-x)^{n-k}\right| \\
&\leq \sum_{k=0}^{n}\left|f\left(\frac{k}{n}\right) - f(x)\right|\binom{n}{k}x^k(1-x)^{n-k} \\
&= E\left(\left|f\left(\frac{S_n}{n}\right) - f(x)\right|\right) \\
&= \int_{\Omega}\left(\left|f\left(\frac{S_n(\omega)}{n}\right) - f(x)\right|\right)dP(\omega) \\
&= \int_{\{|\frac{S_n}{n}-x|>\delta\}}\left(\left|f\left(\frac{S_n}{n}\right) - f(x)\right|\right)dP \\
&\quad + \int_{\{|\frac{S_n}{n}-x|\leq\delta\}}\left(\left|f\left(\frac{S_n}{n}\right) - f(x)\right|\right)dP.
\end{aligned}$$

Como f es continua en el compacto $[0,1]$ entonces f es uniformemente continua por lo tanto dado $\varepsilon > 0$ podemos tomar δ tal que :

$|x-y| \leq \delta$ implica $|f(x) - f(y)| < \dfrac{\varepsilon}{2}$ para todo $x, y \in [0,1]$.

Así

$$\begin{aligned}
|P_n(x) - f(x)| &\leq \int_{(|\frac{S_n}{n}-x|>\delta)}\left(\left|f\left(\frac{S_n}{n}\right) - f(x)\right|\right)dP + \frac{\varepsilon}{2} \\
&\leq 2||f||_\infty P\left(\left|\frac{S_n}{n} - x\right| > \delta\right) + \frac{\varepsilon}{2}
\end{aligned}$$

Como $E(S_n) = nx$ y

$$Var(S_n) = \sigma^2(S_n) = n(1-x)x$$

por la desigualdad de Chebyshev:

$$\begin{aligned} P\left(\left|\frac{S_n}{n} - x\right| > \delta\right) &\leq \frac{E\left(\frac{S_n}{n} - x\right)^2}{\delta^2} \\ &= \frac{\sigma^2\left(\frac{S_n}{n}\right)}{\delta^2} \\ &= \frac{\frac{1}{n^2}\sigma^2(S_n)}{\delta^2} \\ &= \frac{n(1-x)x}{n^2\delta^2} \\ &= \frac{x(1-x)}{n\delta^2} \leq \frac{1}{4}\frac{1}{n\delta^2}. \end{aligned}$$

Pero $\frac{1}{4}\frac{1}{n\delta^2}$ converge a 0 cuando $n \to \infty$. Por lo tanto $P_n(x)$ converge a $f(x)$ uniformemente.

$\square$

CAPíTULO 3

Teorema de aproximación de Weierstrass y grandes desviaciones

1. Métodos probabilísticos

Comenzamos recordando un resultado básico de análisis matemático.

TEOREMA 3.1. *Si $D \subset \mathbb{R}$ y D es un compacto, si $f : D \longrightarrow \mathbb{R}$ es una función continua. Sean*

$$M = \sup\{f(x) : x \in D\}, \quad m = \inf\{f(x) : x \in D\}.$$

Entonces existen $p, q \in D$ tales que $f(p) = M, \quad f(q) = m$.

OBSERVACIÓN 3.2. Un resultado análogo es cierto para un compacto $D \subseteq \mathbb{R}^n$.

OBSERVACIÓN 3.3. Claramente un intervalo de la forma $[a, b]$, con $a, b \in \mathbb{R}$ es un subconjunto compacto de $\mathbb{R}$. Utilizando este hecho y el teorema anterior se puede afirmar que toda función continua en un intervalo cerrado y acotado alcanza su valor máximo y su valor mínimo en ese intervalo.

Todo el material usado referente a grandes desviaciones puede hallarse en [1].

TEOREMA 3.4. *Sea $B(n, x)$ una variable aleatoria binomial con n eventos independientes y probabilidad de éxito x con $x \in [0, 1]$ para cada uno de ellos y sea $a > 0$ arbitrario entonces*

$$P(|B(n, x) - nx| > a) \leq 2 \exp\left(\frac{-2a^2}{n}\right) \tag{3.1}$$

Idea de la prueba: Sea $X_1, \cdots, X_n$ variables aleatorias independientes idénticamente distribuidas con:

$$P(X_i = 1 - x) = x,$$

$$P(X_i = -x) = 1 - x,$$

y sea $X = X_1 + ... + X_n$. Veamos que X tiene distribución $B(n,x) - nx$.

Sea

$$Y_i(\omega) = \begin{cases} 1 & \text{si} \quad X_i(\omega) = 1 - x \\ 0 & \text{si} \quad X_i(\omega) = -x. \end{cases}$$

Es claro que los Y_i tienen distribución de Bernoulli donde:

$$P(Y_i = 1) = P(X_i = 1 - x) = x$$

$$P(Y_i = 0) = P(X_i = -x) = 1 - x.$$

Sea $Y = Y_1 + ... + Y_n$ luego

$$P(Y = k) = P\left(\sum_{i=1}^{n} Y_i = k\right) = \binom{n}{k}[P(Y_i = 1)]^k[P(Y_i = 0)]^{n-k} = \binom{n}{k}x^k(1-x)^{n-k}.$$

Sea ω tal que $Y(\omega) = n$ con $n = h + k$ donde k es el número de veces que $Y_i(\omega) = 1$ y h es el número de veces que $Y_i = 0$, por lo tanto hay k veces en que $X_i(\omega) = 1 - x$ y h veces en que $X_i(\omega) = -x$ luego:

$$\begin{aligned} X(\omega) = \sum_{i=1}^{n} X_i(\omega) &= \sum_{i=1}^{h+k} X_i(\omega) \\ &= k(1-x) + h(-x) \\ &= k - x(h+k) \\ &= k - nx. \end{aligned}$$

Como estamos considerando el caso en que $Y(\omega) = k$ entonces

$$X(\omega) = Y(\omega) - nx$$

Por lo tanto $X \sim B(n,x) - nx$.

Quiero probar que

$$\begin{aligned} P(|B(n,x) - nx| > a) &\leq 2\exp\left(\frac{-2a^2}{n}\right) \\ P(|X| > a) &\leq 2\exp\left(\frac{-2a^2}{n}\right). \end{aligned}$$

Sabemos que:

$$|X(\omega)| > a \text{ implica } X(\omega) > a \text{ ó } -X(\omega) > a$$

$$\{\omega \in \Omega : |X(\omega)| > a\} \subseteq \{\omega \in \Omega : X(\omega) > a\} \cup \{\omega \in \Omega : -X(\omega) > a\}$$

$$P\{\omega \in \Omega : |X(\omega)| > a\} \leq P\{\omega \in \Omega : X(\omega) > a\} + P\{\omega \in \Omega : -X(\omega) > a\}$$

como X es simétrico

$$P\{\omega \in \Omega : X(\omega) > a\} = P\{\omega \in \Omega : -X(\omega) > a\}$$

$$P\{\omega \in \Omega : |X(\omega)| > a\} \leq 2P\{\omega \in \Omega : X(\omega) > a\}$$

Es decir nuestro problema se reduce a demostrar que

$$P(X > a) \leq \exp\left(\frac{-2a^2}{n}\right)$$

Lo cual demostraremos utilizando los siguientes tres lemas.

Lema 3.5. *Sean $\alpha, \beta \in \mathbb{R}$ con $|\alpha| \leq 1$, tenemos*

$$\cosh(\beta) + \alpha \sinh(\beta) \leq \exp\left(\alpha\beta + \frac{\beta^2}{2}\right) \tag{3.2}$$

Demostración. Recordemos que

$$\cosh(\beta) = \frac{\exp(\beta) + \exp(-\beta)}{2}$$

$$\sinh(\beta) = \frac{\exp(\beta) - \exp(-\beta)}{2}$$

Cuando $\alpha = 1$ tenemos,

$$\begin{aligned} \cosh(\beta) + \alpha \sinh(\beta) &= \frac{\exp(\beta) + \exp(-\beta)}{2} + \frac{\exp(\beta) - \exp(-\beta)}{2} \\ &= \exp(\beta) < \exp\left(\beta + \frac{\beta^2}{2}\right). \end{aligned}$$

Si $\alpha = -1$ entonces:

$$\begin{aligned} \cosh(\beta) + \alpha \sinh(\beta) &= \frac{\exp(\beta) + \exp(-\beta)}{2} - \frac{\exp(\beta) - \exp(-\beta)}{2} \\ &= \exp(-\beta) < \exp\left(-\beta + \frac{\beta^2}{2}\right). \end{aligned}$$

En lo que sigue consideramos dos situaciones: primero lo que ocurre fuera del compacto $[-100, 100]$ y luego lo que ocurre dentro de este conjunto.

Si $|\beta| \geq 100$ entonces

$$\beta \geq 100 \text{ ó } \beta \leq -100.$$

Si $\beta \geq 100$ entonces

$$\alpha\beta + \frac{\beta^2}{2} < \beta + \frac{\beta^2}{2}$$

para $|\alpha| < 1$ luego:

$$\begin{aligned}\cosh(\beta) + \alpha\sinh(\beta) &= \frac{\exp(\beta)+\exp(-\beta)}{2} + \frac{\alpha\exp(\beta)-\alpha\exp(-\beta)}{2}\\ &= \frac{\exp(\beta)(1+\alpha)+\exp(-\beta)(1-\alpha)}{2}.\end{aligned}$$

Por otro lado tenemos:

$$-1 < \alpha < 1 \text{ implica } 0 < 1+\alpha < 2.$$

Además,

$$-1 < \alpha < 1 \text{ implica } 2 > 1-\alpha > 0.$$

De donde

$$\begin{aligned}\frac{\exp(\beta)(1+\alpha)+\exp(-\beta)(1-\alpha)}{2} &< 2\frac{\exp(\beta)+\exp(-\beta)}{2}\\ &= \exp(\beta)+\exp(-\beta).\end{aligned}$$

Como $\exp(-\beta) < \exp(\beta) \quad$ con $\quad \beta \geq 100$ entonces

$$\exp(\beta)+\exp(-\beta) < 2\exp(\beta)$$

además

$$\begin{aligned}\beta < \frac{\beta^2}{2} &\quad \text{implica} \quad \exp(\beta) < \exp\left(\frac{\beta^2}{2}\right)\\ 2 < \beta &\quad \text{implica} \quad 2 < \exp(\beta).\end{aligned}$$

De donde

$$2\exp(\beta) < \exp\left(\frac{\beta^2}{2}\right)\exp(\beta).$$

Finalmente

$$\cosh(\beta) + \alpha\sinh(\beta) \leq \exp\left(\alpha\beta + \frac{\beta^2}{2}\right).$$

Si $\beta \leq -100$ entonces

$$\alpha\beta + \frac{\beta^2}{2} < \frac{\beta^2}{2} - \beta$$

para $|\alpha| < 1$ luego:

$$\begin{aligned}\frac{\exp(\beta)(1+\alpha)+\exp(-\beta)(1-\alpha)}{2} &< 2\frac{\exp(\beta)+\exp(-\beta)}{2}\\ &= \exp(\beta)+\exp(-\beta).\end{aligned}$$

Como

$$\exp(-\beta) > \exp(\beta) \text{ con } \beta \leq -100$$

entonces

$$\exp(\beta) + \exp(-\beta) < 2\exp(-\beta)$$

además

$$\begin{aligned} -\beta < \frac{\beta^2}{2} \quad &\text{implica} \quad \exp(-\beta) < \exp\left(\frac{\beta^2}{2}\right) \\ 2 < -\beta \quad &\text{implica} \quad 2 < \exp(-\beta). \end{aligned}$$

De donde

$$2\exp(-\beta) < \exp\left(\frac{\beta^2}{2}\right)\exp(-\beta).$$

Finalmente

$$\cosh(\beta) + \alpha\sinh(\beta) \leq \exp\left(\alpha\beta + \frac{\beta^2}{2}\right).$$

A continuación trabajaremos con $\beta \in [-100, 100]$. Supongamos que (3.2) no se satisface entonces

$$\cosh(\beta) + \alpha\sinh(\beta) > \exp\left(\alpha\beta + \frac{\beta^2}{2}\right)$$

la función

$$f(\alpha, \beta) = \cosh(\beta) + \alpha\sinh(\beta) - \exp\left(\alpha\beta + \frac{\beta^2}{2}\right)$$

donde $f(\alpha, \beta) > 0$. Como f es continua en el compacto $[-1, 1] \times [-100, 100]$ entonces f alcanza un mínimo y un máximo en ese conjunto. Analicemos los puntos críticos de f tomando derivadas parciales iguales a cero:

$$\frac{\partial f(\alpha, \beta)}{\partial \alpha} = \sinh(\beta) - \beta\exp\left(\alpha\beta + \frac{\beta^2}{2}\right)$$

de donde

$$\sinh(\beta) = \beta\exp\left(\alpha\beta + \frac{\beta^2}{2}\right).$$

$$\frac{\partial f(\alpha, \beta)}{\partial \beta} = \sinh(\beta) + \alpha\cosh\beta - (\alpha + \beta)\exp\left(\alpha\beta + \frac{\beta^2}{2}\right)$$

sustituyendo el valor de $\sinh(\beta)$ encontrado tenemos:

$$\begin{aligned} \beta\exp\left(\alpha\beta + \frac{\beta^2}{2}\right) + \alpha\cosh\beta &= (\alpha + \beta)\exp\left(\alpha\beta + \frac{\beta^2}{2}\right) \\ \alpha\cosh\beta &= (\alpha + \beta)\exp\left(\alpha\beta + \frac{\beta^2}{2}\right) - \beta\exp\left(\alpha\beta + \frac{\beta^2}{2}\right) \\ \alpha\cosh\beta &= \alpha\exp\left(\alpha\beta + \frac{\beta^2}{2}\right) \\ \cosh\beta &= \exp\left(\alpha\beta + \frac{\beta^2}{2}\right). \end{aligned}$$

Luego

$$\tanh\beta = \frac{\sinh\beta}{\cosh\beta} = \frac{\beta\exp\left(\alpha\beta + \frac{\beta^2}{2}\right)}{\exp\left(\alpha\beta + \frac{\beta^2}{2}\right)} = \beta.$$

Lo que implica que $\beta = 0$ de donde f alcanza el máximo y el mínimo cuando $\beta = 0$. Pero $f(\alpha, 0) = 0$ y esto contradice $f(\alpha, \beta) > 0$. La contradicción viene de suponer que (3.2) no se satisface. □

LEMA 3.6. *Para todo $\theta \in [0,1]$ y para todo λ*

$$\theta\exp(\lambda(1-\theta)) + (1-\theta)\exp(-\lambda\theta) \leq \exp\left(\frac{\lambda^2}{8}\right)$$

DEMOSTRACIÓN. Sea $\theta \in [0,1]$ entonces $|2\theta - 1| \leq 1$.
Aplicando el Lema 1 a $\alpha = 2\theta - 1$ y $\beta = \frac{\lambda}{2}$ obtenemos

$$\cosh\left(\frac{\lambda}{2}\right) + (2\theta-1)\sinh\left(\frac{\lambda}{2}\right) \leq \exp\left(\frac{(\frac{\lambda}{2})^2}{2} + (2\theta-1)\frac{\lambda}{2}\right)$$

De esto se siguen las siguientes desigualdades:

$$\begin{aligned}
\frac{\exp\left(\frac{\lambda}{2}\right) + \exp\left(\frac{-\lambda}{2}\right)}{2} + (2\theta-1)\frac{\exp\left(\frac{\lambda}{2}\right) - \exp\left(\frac{-\lambda}{2}\right)}{2} &\leq \exp\left(\frac{\lambda^2}{8} + \theta\lambda - \frac{\lambda}{2}\right) \\
\frac{\exp\left(\frac{\lambda}{2}\right)(1+2\theta-1) + \exp\left(\frac{-\lambda}{2}\right)(1-2\theta+1)}{2} &\leq \exp\left(\frac{\lambda^2}{8}\right)\exp\left(\theta\lambda - \frac{\lambda}{2}\right) \\
\theta\exp\left(\frac{\lambda}{2}\right) + (1-\theta)\exp\left(\frac{-\lambda}{2}\right) &\leq \exp\left(\frac{\lambda^2}{8}\right)\exp\left(\theta\lambda - \frac{\lambda}{2}\right) \\
\frac{\theta\exp\left(\frac{\lambda}{2}\right) + (1-\theta)\exp\left(\frac{-\lambda}{2}\right)}{\exp\left(\theta\lambda - \frac{\lambda}{2}\right)} &\leq \exp\left(\frac{\lambda^2}{8}\right) \\
\theta\exp\left(\frac{\lambda}{2} - \theta\lambda + \frac{\lambda}{2}\right) + (1-\theta)\exp\left(\frac{-\lambda}{2} - \theta\lambda + \frac{\lambda}{2}\right) &\leq \exp\left(\frac{\lambda^2}{8}\right) \\
\theta\exp\left(\lambda(1-\theta)\right) + (1-\theta)\exp(-\lambda\theta) &\leq \exp\left(\frac{\lambda^2}{8}\right).
\end{aligned}$$

□

LEMA 3.7. *Sea X_i tal que:*

$$P(X_i = 1 - x) = x,$$

$$P(X_i = -x) = 1 - x,$$

para $\lambda > 0$ *se tiene que*

$$E[\exp(\lambda X_i)] = x\exp(\lambda(1-x))) + (1-x)\exp(-\lambda x) \leq \exp\left(\frac{\lambda^2}{8}\right)$$

donde $E[.]$ *es el valor esperado.*

DEMOSTRACIÓN. Por definición tenemos:

$$E[g(X)] = \sum_{x\in\Re(X)} g(x)P(X=x) = \sum_{n=1}^{\infty} g(x_n)P(X=x_n)$$

$$E[\exp(\lambda X)] = \sum_{n=1}^{\infty} \exp(\lambda x_n)P(X=x_n)$$

Como

$$P(X_i = 1-x) = x \ ,$$

$$P(X_i = -x) = 1-x$$

entonces

$$\begin{aligned} E[\exp(\lambda X_i)] &= \exp(-\lambda x)P(X_i=-x) + \exp(\lambda(1-x))P(X_i = 1-x) \\ &= (1-x)\exp(-\lambda x) + x\exp(\lambda(1-x)). \end{aligned}$$

Por el Lema 2 y dado que $x \in [0,1]$ y $\lambda > 0$ tenemos que

$$E[\exp(\lambda X_i)] = (1-x)\exp(-\lambda x) + x\exp(\lambda(1-x)) \leq \exp\left(\frac{\lambda^2}{8}\right)$$

$\square$

Ya podemos probar el Teorema 3.4.

DEMOSTRACIÓN. Sabemos que

$$E[\exp(\lambda X)] = E[\exp(\lambda(X_1 + X_2 ... + X_n))] = E\left(\prod_{i=1}^{n}\exp(\lambda X_i)\right) = \prod_{i=1}^{n} E(\exp(\lambda X_i))$$

Hemos podido intercambiar el producto con la esperanza debido a que los X_i son independientes. Como además están idénticamente distribuidos obtenemos:

$$E[\exp(\lambda X)] = E(\exp(\lambda X_1))^n \leq \left(\exp\left(\frac{\lambda^2}{8}\right)\right)^n = \exp\left(\frac{\lambda^2 n}{8}\right)$$

Ahora aplicando la desigualdad de Markov,

$$P(X > a) = P(\exp(\lambda X) > \exp(\lambda a)) \leq \frac{E[\lambda X]}{\exp(\lambda a)} \leq \exp\left(\frac{\lambda^2 n}{8} - \lambda a\right)$$

Finalmente tomando $\lambda = \dfrac{4a}{n}$ obtenemos:

$$P(X > a) \leq \exp\left(\frac{\frac{16a^2n}{n^2}}{8} - \frac{4a^2}{n}\right)$$

luego

$$P(X > a) \leq \exp\left(\frac{-2a^2}{n}\right)$$

□

2. Velocidad de convergencia para funciones de Lipschitz

Condición de Lipschitz

DEFINICIÓN 3.8. Una función real f en $[0,1]$ satisface una *condición de Lipschitz* si existe una constante positiva C tal que

$$|f(x) - f(y)| < C|x - y| \qquad \forall x, y \in [0,1]$$

TEOREMA 3.9. *Para una función f que satisface la condición de Lipschitz, tenemos una sucesión de polinomios P_n donde el grado de P_n es n y una constante k la cual depende de f, entonces*

$$||P_n - f||_\infty \leq k\sqrt{\frac{\ln(n)}{n}}$$

donde $||.||_\infty$ es la norma del supremo.

Nota: *Como la función f es Lipchitz, entonces es uniformemente continua y acotada por una constante M tal que $||f||_\infty \leq M$.*

DEMOSTRACIÓN. Definamos el polinomio de Bernstein como:

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

Luego separando la sumatoria

$$\begin{aligned} |P_n(x) - f(x)| &= \left| \sum_{i=0}^{n} \binom{n}{i} x^i (1-x)^{n-i} f\left(\frac{i}{n}\right) - f(x) \right| \\ &= \left| \sum_{i:|i-nx|\leq a} \binom{n}{i} x^i (1-x)^{n-i} f\left(\frac{i}{n}\right) - f(x) \right| \\ &+ \left| \sum_{i:|i-nx|> a} \binom{n}{i} x^i (1-x)^{n-i} f\left(\frac{i}{n}\right) - f(x) \right|. \end{aligned}$$

Aplicando la desigualdad triangular obtenemos:

$$\left| \sum_{i:|i-nx|\leq a} \binom{n}{i} x^i (1-x)^{n-i} f\left(\frac{i}{n}\right) - f(x) \right| \leq \sum_{i:|i-nx|\leq a} \binom{n}{i} x^i (1-x)^{n-i} \left| f\left(\frac{i}{n}\right) - f(x) \right|.$$

Luego

$$\left|\sum_{i:|i-nx|>a} \binom{n}{i} x^i(1-x)^{n-i} f\left(\frac{i}{n}\right) - f(x)\right| \leq \sum_{i:|i-nx|>a} \binom{n}{i} x^i(1-x)^{n-i} \left(\left|f\left(\frac{i}{n}\right)\right| + |f(x)|\right).$$

Sabemos que f es Lipschitz es decir

$$|f(x) - f(y)| \leq C|x-y|$$

y como

$$\sum_{i=0}^{n} \binom{n}{i} x^i(1-x)^{n-i} = 1$$

tenemos:

$$\sum_{i:|i-nx|\leq a} \binom{n}{i} x^i(1-x)^{n-i} \left|f\left(\frac{i}{n}\right) - f(x)\right| \leq \frac{aC}{n}$$

Además como $||f||_\infty \leq M$ y

$$P(|B(n,x) - nx| > a) = \sum_{i:|i-nx|>a} \binom{n}{i} x^i(1-x)^{n-i}$$

entonces:

$$\sum_{i:|i-nx|>a} \binom{n}{i} x^i(1-x)^{n-i} \left(\left|f\left(\frac{i}{n}\right)\right| + |f(x)|\right) \leq 2MP(|B(n,x) - nx| > a)$$

Luego:

$$|P_n(x) - f(x)| \leq \frac{aC}{n} + 2MP(|B(n,x) - nx| > a)$$

Finalmente por el teorema 3.4:

$$|P_n(x) - f(x)| \leq \frac{aC}{n} + 2M\exp\left(\frac{-2a^2}{n}\right).$$

Ahora optimicemos el parámetro a en términos de n.

Consideremos la función:

$$F(x) = \frac{-xC}{n} + 2M\exp\left(\frac{-2x^2}{n}\right),$$

entonces:

$$F'(x) = \frac{-C}{n} + \frac{8Mx}{n}\exp\left(\frac{-2x^2}{n}\right).$$

Sea x tal que $F'(x) = 0$ despejando la ecuación exponencial obtenemos:

$$x\exp\left(\frac{-2x^2}{n}\right) = \frac{C}{8M}$$

No podemos dar una solución exacta para esta ecuación exponencial. Sin embargo ella conduce a la solución asintótica

$$x = \frac{\sqrt{n \ln n}}{2}.$$

Tomando $a = \dfrac{\sqrt{n \ln n}}{2}$ se obtiene

$$\begin{aligned}
|P_n(x) - f(x)| &\leq \frac{aC}{n} + 2M \exp\left(\frac{-2a^2}{n}\right) \\
&\leq \frac{\left(\frac{\sqrt{n \ln n}}{2}\right) C}{n} + 2M \exp\left(\frac{-2\left(\frac{\sqrt{n \ln n}}{2}\right)^2}{n}\right) \\
&\leq \frac{C\sqrt{n \ln n}}{2n} + 2M \frac{1}{\sqrt{n}} \\
&\leq \sqrt{\frac{\ln n}{n}} \frac{C}{2} + \sqrt{\frac{\ln n}{n}} \frac{2M}{\sqrt{\ln n}} \\
&\leq \left(\frac{C}{2} + \frac{2M}{\sqrt{\ln n}}\right) \sqrt{\frac{\ln n}{n}}.
\end{aligned}$$

Con $K = \dfrac{C}{2} + \dfrac{2M}{\sqrt{\ln n}}$ y para $n \geq 3$ $\square$

Bibliografía

[1] ALON, N. AND SPENCER, J. The Probabilistic method, Wiley, New York, 1992.

[2] ARRIOJAS, M. “Notas de teoría de probabilidades”. Caracas. Edic. de la Escuela de Matemáticas U.C.V.

[3] BACHMAN, G. AND NARICI, L. Functional analysis. Academic Press, 1966.

[4] BETZ, C. Introducción a la teoría de la medida e integración. U.C.V.1992

[5] CHUNG K.L. A Course in probability theory 2nd ed., Academic Press, New York, 1974.

[6] CHUNG, K.L. Elementary probability theory with stochastic process. Spinger Verlag, 1978.

[7] FELLER, W. Introducción a la teoría de probabilidad y sus aplicaciones. Editorial Limusa. Vol.1.

[8] GZYL, H. AND PALACIOS, J.L.“The Weierstrass Aproximation theorem and large desviations”. American Mathematical Montly, 104 pag 650-653, 1997.

[9] HALMOS, P. Teoría intuitiva de los conjuntos. CECSA, 1971.

[10] KOLMOGOROV, A. AND FOMIN, S. Elementos de la teoría de funciones y de análisis funcional. MIR, 1975.

[11] MEYER, P. Probabilidad y aplicaciones estadísticas. Bogota; Fondo Educativo Interamericano. 1973.

[12] OLIVARES, M. “Curso de probabilidades”. Caracas. Edic. de la Escuela de Matemáticas U.C.V. Octubre 2002.

[13] ROYDEN, H.L. Real analysis. Collier Macmillan, 1968.

[14] RUDIN, W. Real and complex analysis. Mc Graw-Hill, 1966.