

Universidad de Sonora

Departamento de Matemáticas

Aplicaciones de la Probabilidad al Análisis
Matemático y al Algebra Lineal



Licenciado en Matemáticas

Presenta:

RENE CESAR LEYVA CONTRERAS

Hermosillo, Sonora

Septiembre 2000

Universidad de Sonora

Repositorio Institucional UNISON



**"El saber de mis hijos
hará mi grandeza"**



Excepto si se señala otra cosa, la licencia del ítem se describe como openAccess

Agradezco al Departamento de Matemáticas de la Universidad de Sonora y al Consejo Nacional de Ciencia y Tecnología (CONACyT), proyecto 3871P-E por el apoyo recibido para la realización de este trabajo.

Contenido

1	Teoría de la Probabilidad y Cadenas de Markov	4
1.1	Introducción	4
1.2	Variables Aleatorias y Vectores Aleatorios	5
1.3	Esperanza y Varianza	12
1.4	Teorema de Feller	18
1.5	Cadenas de Markov	20
2	Aplicaciones de la Probabilidad al Análisis Matemático	25
2.1	Introducción	25
2.2	Teorema de Weierstrass	26
2.3	Una Generalización del Teorema de Taylor	29
2.4	Tasas de Convergencia	32
3	Aplicaciones de la Probabilidad al Álgebra Lineal	37
3.1	Introducción	37
3.2	Matrices Semi-definidas Positivas	38
3.3	Sistemas de Ecuaciones Lineales	40
3.4	Aproximación a la Solución del SEL	42
3.4.1	Algoritmo para la solución del SEL	45
4	Conclusiones	47

Introducción

Desde sus orígenes, la Teoría de la Probabilidad se ha distinguido por ser una herramienta importante en la solución de problemas que se presentan en otras áreas. Basta revisar un poco de la literatura sobre matemáticas aplicadas para darse cuenta que hoy en día la probabilidad juega un papel central en la modelación y /o solución de problemas específicos.

Para llegar a este punto, la probabilidad ha tenido que recurrir a otras ramas de la matemática misma. Es decir, el desarrollo de la probabilidad está sustentado, en gran parte, por las técnicas, teoría y resultados de áreas ajenas a ella. Por ejemplo, la teoría de la medida es parte importante de la fundamentación matemática de la probabilidad, y más concretamente, el cálculo diferencial e integral hace posible evaluar algunas probabilidades. Otro ejemplo lo es el impacto del álgebra lineal en el estudio de las distribuciones multivariadas, y en cadenas de Markov, etc. En general, casi todas las áreas de la matemática (análisis funcional, teoría de ecuaciones diferenciales, álgebra, entre otras) tienen un lugar en el desarrollo de la probabilidad.

¿Qué tanto la teoría de la probabilidad ha contribuido en el desarrollo de otras áreas de la matemática?

El hecho de que la fundamentación matemática de la probabilidad sea relativamente joven (década de los 30's) hace pensar que su contribución ha sido escasa o nula. Sin embargo, en los últimos años, diferentes autores han publicado trabajos donde muestran aplicaciones de la probabilidad al demostrar resultados clásicos de la matemática con procedimientos probabilísticos. Por lo regular, dichas demostraciones muestran ciertas ventajas sobre las tradicionales, en el sentido de que se utilizan argumentos más sencillos y/o proporcionan mayor información respecto al resultado en cuestión.

El objetivo de este trabajo es presentar algunas aplicaciones de la probabilidad al análisis matemático y al álgebra lineal.

En análisis matemático, presentamos una demostración probabilística del famoso Teorema de Aproximación de Weierstrass y una generalización del Teorema de Taylor. Una de las ventajas que se tiene en este caso es que, a diferencia de las demostraciones clásicas, podemos estimar la rapidez de convergencia de la aproximación en cada uno de estos resultados. Este tema se desarrolla en el Capítulo 2.

Respecto a las aplicaciones del álgebra lineal, en el Capítulo 3, presentamos una demostración de un resultado sobre matrices positivamente semidefinidas, conocido en la literatura del álgebra lineal como Teorema de Schur. Además, desarrollamos un algoritmo para la solución de una clase de sistemas de ecuaciones lineales.

Los elementos principales de la Teoría de la Probabilidad que intervienen en las aplicaciones mencionadas son expuestos en el primer capítulo.

Capítulo 1

Teoría de la Probabilidad y Cadenas de Markov

1.1 Introducción

El objetivo de este capítulo es presentar las herramientas necesarias para el desarrollo de la parte principal del trabajo. Estas herramientas son esencialmente resultados y definiciones de la Teoría de la Probabilidad y Procesos Estocásticos que se pueden encontrar en cualquier libro que trate sobre estos temas de nivel licenciatura. Decidimos exponer estos resultados en este primer capítulo para una fácil referencia, y además por el hecho de que algunos de estos son presentados de una forma apropiada, la cual se adapta más al uso que se le dará en los capítulos posteriores.

En términos generales, este capítulo trata los temas de variables aleatorias y vectores aleatorios, así como sus principales características (función de densidad, de probabilidad, de distribución, valor esperado, varianza, etc.); se presentan algunos teoremas límite (Teorema de Feller, Ley de los Grandes Números, etc.) y finalmente se hace un estudio general de algunos elementos de cadenas de Markov.

1.2 Variables Aleatorias y Vectores Aleatorios

1.2.1 Definición.- Un espacio de probabilidad es una terna $(\Omega, \mathfrak{S}, P)$ donde Ω es un conjunto no vacío llamado espacio muestral, \mathfrak{S} es una σ -álgebra de subconjuntos de Ω , y P es una función definida en \mathfrak{S} llamada medida de probabilidad que satisface las siguientes propiedades:

$$(i) 0 \leq P(A) \leq 1, A \in \mathfrak{S};$$

$$(ii) P(\Omega) = 1;$$

(iii) Si $A_1, A_2, \dots \in \mathfrak{S}$ es una colección numerable de eventos tales que $A_i \cap A_j = \phi$ si $i \neq j$, entonces

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

1.2.2 Definición.- Una Variable Aleatoria (v.a.) X sobre el espacio de probabilidad $(\Omega, \mathfrak{S}, P)$ es una función $X : \Omega \rightarrow \mathfrak{R}$ tal que para cada $r \in \mathfrak{R}$

$$\{\omega \in \Omega : X(\omega) \leq r\} \in \mathfrak{S}.$$

En general, el comportamiento de una v.a. X lo podemos estudiar por medio de la función de distribución, definida de la siguiente manera:

1.2.3 Definición.- La Función de Distribución de una v.a. X se define como

$$F_X(x) = P(X \leq x), x \in \mathfrak{R}.$$

Si el contexto en el que se esté trabajando es claro, omitiremos el subíndice de la función de distribución.

Algunas de las propiedades básicas de la función de distribución son las siguientes:

1.2.4 Teorema.- La función de distribución F de una v.a. X satisface las siguientes propiedades:

(i) F es no decreciente, es decir, si $x_1 < x_2$ entonces $F(x_1) \leq F(x_2)$;

(ii) $\lim_{x \rightarrow -\infty} F(x) = 0$ y $\lim_{x \rightarrow \infty} F(x) = 1$;

(iii) F es continua por la derecha, es decir,

$$\lim_{h \rightarrow 0^+} F(x+h) := F(x+) = F(x), \quad \forall x \in \mathfrak{R};$$

(iv) $F(x-) = P[X < x]$, $\forall x \in \mathfrak{R}$, donde

$$F(x-) := \lim_{h \rightarrow 0} F(x-h).$$

Usualmente para estudiar las v.a. se clasifican en discretas, absolutamente continuas y mixtas. En este trabajo nos centraremos en el estudio de las discretas y las absolutamente continuas a las cuales solo las llamaremos continuas.

1.2.5 Definición.- Decimos que una v.a. X es discreta si el conjunto de valores que puede tomar es finito o numerable. Es decir, el rango de X , R_X , toma la forma $R_X = \{x_1, x_2, x_3, \dots\}$.

1.2.6 Definición.- La función de probabilidad de una v.a. discreta X se define como

$$f_X(x) = P[X = x] = P\{\omega \in \Omega : X(\omega) = x\}, \quad x \in \mathfrak{R}.$$

Es fácil ver que la función de probabilidad satisface las siguientes propiedades:

(i) $0 \leq f_X(x) \leq 1$, $\forall x \in R_X$, y

(ii) $\sum_{x \in R_X} f_X(x) = 1$.

1.2.7 Ejemplo: (a) Sea X una v.a. tal que $R_X = \{0, 1, 2, \dots, n\}$, $n \in \mathbb{N}$. Decimos que X tiene una distribución binomial con parámetros n y p , [$X \sim \text{Bin}(n, p)$], si su función

de probabilidad es de la forma:

$$f(x) = \begin{cases} C_x^n p^x (1-p)^{n-x} & \text{si } x = 0, 1, 2, \dots, n \\ 0 & \text{en otro caso} \end{cases}$$

(b) Sea X una v.a. con rango $R_X = \{0, 1, 2, \dots\}$. Decimos que X tiene distribución de Poisson con parámetro λ , [$X \sim P(\lambda)$], si su función de probabilidad es

$$f(x) = \begin{cases} \frac{\lambda^x}{x!} e^{-\lambda} & \text{si } x = 0, 1, 2, \dots \\ 0 & \text{en otro caso} \end{cases}$$

En general, una manera de decir que una v.a. discreta con $R_X = \{x_1, x_2, \dots\}$ tiene determinada distribución o función de probabilidad, es decir que dicha distribución está concentrada en el conjunto $\{x_1, x_2, \dots\}$, en el cual se tiene que $f(x) > 0$.

1.2.8 Proposición.- Sea X una v.a. discreta con función de probabilidad f_X concentrada en el conjunto $\{x_1, x_2, \dots\}$, y sea $Y = aX$, $a \in \mathfrak{R}$. Entonces Y es una v.a. discreta concentrada en el conjunto $\{ax_1, ax_2, \dots\}$, tal que $f_Y(y) = f_X(y/a)$.

Demostración.

Primero tenemos que $f_X(x) > 0$, $\forall x \in \{x_1, x_2, \dots\}$. Ahora,

$$f_Y(y) = P[Y = y] = P[aX = y] = P\left[X = \frac{y}{a}\right] = f_X\left(\frac{y}{a}\right),$$

lo cual es mayor que cero si $\frac{y}{a} \in \{x_1, x_2, \dots\}$, o equivalentemente si $y \in \{ax_1, ax_2, \dots\}$.

Esto demuestra la proposición. ■

1.2.9 Observación.- Del Ejemplo 1.2.7. y la Proposición 1.2.8. tenemos que si $X \sim \text{Bin}(n, p)$ concentrada en el conjunto $\{0, 1, 2, \dots, n\}$, entonces la función de probabilidad

de la v.a. $Y = aX$ está concentrada en el conjunto $\{0, a, 2a, \dots, na\}$, y además

$$f_Y(y) = \begin{cases} C_{(y/a)}^n p^{(y/a)} (1-p)^{n-(y/a)} & \text{si } y = 0, a, 2a, \dots, na. \\ 0 & \text{otro caso} \end{cases}$$

De la misma manera, si $X \sim P(\lambda)$ concentrada en $\{0, 1, 2, \dots\}$, entonces la v.a. $Y = aX$ tiene función de probabilidad concentrada en el conjunto $\{0, a, 2a, \dots\}$ y

$$f_Y(y) = \begin{cases} \frac{\lambda^{(y/a)}}{(y/a)!} e^{-\lambda} & \text{si } y = 0, a, 2a, \dots \\ 0 & \text{otro caso} \end{cases}$$

1.2.10 Definición.- Una v.a. X es continua si existe una función no negativa f definida en \mathfrak{R} tal que para cualquier $A \subset \mathfrak{R}$,

$$P[X \in A] := P\{\omega \in \Omega : X(\omega) \in A\} = \int_A f_X(x) dx.$$

En este caso, a f_X se le llama Función de Densidad de X y satisface

$$\int_{-\infty}^{\infty} f(x) dx = 1.$$

1.2.11 Definición.- Sean X_1, X_2, \dots, X_n v.a.'s definidas en un mismo espacio de probabilidad $(\Omega, \mathfrak{S}, P)$. A $\mathbf{X} = (X_1, X_2, \dots, X_n)$ se le llama Vector Aleatorio de dimensión n . Esto lo podemos representar como una función $\mathbf{X} : \Omega \rightarrow \mathfrak{R}^n$.

Existe una forma análoga al caso de v.a.'s. para estudiar el comportamiento de los vectores aleatorios como lo muestran las siguientes definiciones:

1.2.12 Definición.- Sean X_1, X_2, \dots, X_n v.a.'s.. Definimos la función de distribución conjunta de X_1, X_2, \dots, X_n o función de distribución del vector $\mathbf{X} = (X_1, X_2, \dots, X_n)$ como

$$F(\mathbf{x}) := F(x_1, x_2, \dots, x_n) = P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n),$$

donde $\mathbf{x} := (x_1, x_2, \dots, x_n) \in \mathfrak{R}^n$.

Los vectores aleatorios tambien los podemos clasificar como discretos y continuos.

1.2.13 Definición.- (i) Un vector aleatorio $\mathbf{X} = (X_1, X_2, \dots, X_n)$ es discreto si los valores que toma en \mathfrak{R}^n forman un conjunto finito o numerable.

(ii) La función de probabilidad $f_{\mathbf{X}}$ del vector aleatorio discreto \mathbf{X} se define como:

$$f_{\mathbf{X}}(\mathbf{x}) := P[\mathbf{X} = \mathbf{x}] = P[X_1 = x_1, X_2 = x_2, \dots, X_n = x_n],$$

donde $\mathbf{x} := (x_1, x_2, \dots, x_n)$.

De esto último podemos decir lo siguiente: Siendo \mathbf{X} un vector aleatorio discreto, entonces la función de probabilidad conjunta $f_{\mathbf{X}}$ determina la distribución de \mathbf{X} , es decir, para $J \subset \mathfrak{R}^n$,

$$P[\mathbf{X} \in J] = \sum_{\mathbf{x} \in J} f_{\mathbf{X}}(\mathbf{x}).$$

1.2.14 Definición.- Decimos que las v.a's. X_1, X_2, \dots, X_n tienen distribución continua conjunta o el vector $\mathbf{X} = (X_1, X_2, \dots, X_n)$ tiene distribución continua, si existe una función $f_{\mathbf{X}}$ no negativa definida en \mathfrak{R}^n tal que para cualquier $A \subset \mathfrak{R}^n$

$$\begin{aligned} P[\mathbf{X} \in A] &= P[(X_1, X_2, \dots, X_n) \in A] \\ &= \int \cdots \int_A f_{\mathbf{X}}(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n. \end{aligned}$$

A $f_{\mathbf{X}}$ se le llama función de densidad conjunta de las v.a's. X_1, X_2, \dots, X_n o función de densidad del vector $\mathbf{X} = (X_1, X_2, \dots, X_n)$.

Sea $\mathbf{X} = (X_1, X_2, \dots, X_n)$ un vector aleatorio discreto de dimensión n con función de probabilidad $f_{\mathbf{X}}$. En este caso, a las funciones de probabilidad de cada una de las v.a's. que componen el vector aleatorio la llamaremos función de probabilidad marginal y se

pueden obtener mediante la formula

$$f_{X_i}(x_i) = \sum_{x_j, j \neq i} f_{\mathbf{X}}(x_1, x_2, \dots, x_n), \quad x_i \in \mathfrak{R}.$$

Es decir, la función de probabilidad de X_i , f_{X_i} , se obtiene sumando los valores de $f_{\mathbf{X}}(x_1, x_2, \dots, x_n)$ sobre todas las x_j con $j \neq i$.

Análogamente, si el vector aleatorio es continuo con función de densidad $f_{\mathbf{X}}$, llamamos a las funciones de densidad $f_{X_i}, i = 1, 2, \dots, n$, de cada v.a. que lo componen, función de densidad marginal y además

$$f_{X_i}(x_i) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{\mathbf{X}}(x_1, x_2, \dots, x_n) dx_1 dx_2 \cdots dx_{i-1} dx_{i+1} \cdots dx_n.$$

Es decir, la densidad de X_i se obtiene integrando la densidad conjunta $f_{\mathbf{X}}(x_1, x_2, \dots, x_n)$ respecto a todas las variables x_j con $j \neq i$.

1.2.15 Definición.- Decimos que las v.a's. X_1, X_2, \dots, X_n son independientes si

$$P[X_1 \in A_1, X_2 \in A_2, \dots, X_n \in A_n] = \prod_{i=1}^n P[X_i \in A_i], \quad \forall A_i \subset \mathfrak{R}, \quad i = 1, 2, \dots, n$$

La independencia de X_1, X_2, \dots, X_n puede caracterizarse de manera sencilla en términos de sus funciones de probabilidad marginales o en términos de sus funciones de densidad marginales, según sea el caso.

1.2.16 Teorema.- Las v.a's. X_1, X_2, \dots, X_n son independientes si y solo si

$$f(x_1, x_2, \dots, x_n) = f_{X_1}(x_1) f_{X_2}(x_2) \cdots f_{X_n}(x_n),$$

donde f es la función de densidad o de probabilidad conjunta de las v.a's. X_1, X_2, \dots, X_n y f_{X_i} es la función de densidad o probabilidad marginal de $X_i, i = 1, 2, \dots, n$, según el caso.

1.2.17 Observación.- (a) Una manera de caracterizar a la distribución binomial es la siguiente. Sean X_1, X_2, \dots, X_n v.a's. independientes e idénticamente distribuidas (i.i.d.) tal que $X_i = 1$ con probabilidad p y $X_i = 0$ con probabilidad $1 - p$, para $i = 1, 2, \dots, n$. Definimos la v.a. $S_n = \sum_{i=1}^n X_i$ la cual representa el número de veces que aparece el 1 en las n v.a's.. De aquí es fácil ver que $S_n \sim \text{Bin}(n, p)$.

(b) Sean X_1, X_2, \dots, X_n v.a's. i.i.d. y $S_n = \sum_{i=1}^n X_i$, donde ahora cada $X_i \sim P(\lambda)$, $i = 1, 2, \dots, n$, entonces se puede deducir fácilmente que $S_n \sim P(n\lambda)$.

Para concluir esta sección daremos la definición de independencia para vectores aleatorios y una de sus caracterizaciones, las cuales generalizan de alguna manera la definición de independencia de v.a's.

1.2.18 Definición.- Sean $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$ vectores aleatorios definidos en un mismo espacio de probabilidad. Entonces $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$ son independientes si y solo si para todos los cubos \mathbf{J}_i de dimensión apropiada se tiene que

$$P[\mathbf{X}_1 \in \mathbf{J}_1, \mathbf{X}_2 \in \mathbf{J}_2, \dots, \mathbf{X}_n \in \mathbf{J}_n] = P[\mathbf{X}_1 \in \mathbf{J}_1] P[\mathbf{X}_2 \in \mathbf{J}_2] \cdots P[\mathbf{X}_n \in \mathbf{J}_n].$$

1.2.19 Teorema.- (i) Sean $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$ vectores aleatorios de dimensión k_1, k_2, \dots, k_n , respectivamente, y sea $d = k_1 + k_2 + \dots + k_n$. Suponga que la distribución de $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n)$ es discreta o continua y sea $f_{\mathbf{X}}$ la función de probabilidad o de densidad de \mathbf{X} , según sea el caso. Entonces $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$ son independientes si y solo si

$$f_{\mathbf{X}}(\mathbf{x}) = \prod_{i=1}^n f_{\mathbf{X}_i}(\mathbf{x}_i), \quad \mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) \in \mathcal{R}^d,$$

donde $f_{\mathbf{X}_i}$ es la función de probabilidad o de densidad de \mathbf{X}_i y $\mathbf{x}_i \in \mathcal{R}^{k_i}$, $i = 1, 2, \dots, n$.

(ii) Si los vectores aleatorios $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$ son independientes y $g_i : \mathcal{R}^{k_i} \rightarrow \mathcal{R}^{s_i}$ son funciones dadas, entonces los vectores $\mathbf{Y}_1 = g_1(\mathbf{X}_1), \mathbf{Y}_2 = g_2(\mathbf{X}_2), \dots, \mathbf{Y}_n = g_n(\mathbf{X}_n)$ son independientes.

1.3 Esperanza y Varianza

En esta sección nos centramos en definir la esperanza y la varianza de v.a's. y sus propiedades básicas.

1.3.1 Definición.- (a) Sea X una v.a. discreta con función de probabilidad f_X , y supongamos que al menos una de las siguientes condiciones (i) o (ii) se satisface:

$$(i) \sum_{x>0} x f_X(x) < \infty;$$

$$(ii) \sum_{x<0} x f_X(x) > -\infty.$$

En este caso, la esperanza (valor esperado o media) de X se define como

$$E[X] := \sum_{x \in R_X} x f_X(x).$$

(b) Si X es una v.a. continua con densidad f_X , la esperanza de X se define como

$$E[X] := \int_{-\infty}^{\infty} x f_X(x) dx,$$

donde se supone que

$$(i) \int_0^{\infty} x f_X(x) dx < \infty, \text{ ó}$$

$$(ii) \int_{-\infty}^0 x f_X(x) dx > -\infty.$$

La esperanza de una v.a. X generalmente se denota por el símbolo μ_X y usamos μ cuando no hay posibilidad de confusión:

$$\mu \equiv \mu_X \equiv E[X]$$

Las propiedades principales de la esperanza se pueden resumir en los siguientes resultados.

1.3.2 Teorema.- Sea $g : \mathfrak{R}^n \rightarrow \mathfrak{R}$ una función real definida en \mathfrak{R}^n y $\mathbf{X} = (X_1, X_2, \dots, X_n)$ un vector aleatorio continuo o discreto, y suponga además que la v.a. $Y = g(\mathbf{X})$ es tal que su esperanza está definida. Entonces,

(i) Si \mathbf{X} es discreto,

$$E[Y] = E[g(\mathbf{X})] = \sum_{\mathbf{x}} g(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}),$$

donde $f_{\mathbf{X}}$ es la función de probabilidad conjunta de las v.a's. X_1, X_2, \dots, X_n .

(ii) Si \mathbf{X} es continuo,

$$E[Y] = E[g(\mathbf{X})] = \int_{\mathfrak{R}^n} g(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x},$$

donde $f_{\mathbf{X}}$ es la función de densidad conjunta de las v.a's. X_1, X_2, \dots, X_n .

1.3.3 Teorema.- Sean X, X_1, X_2, \dots, X_n v.a's. con esperanza finita. Entonces, las siguientes afirmaciones son válidas:

(i) Si $P[X = a] = 1$ para alguna constante a , $E[X] = a$;

(ii) Para todas las constantes a_1, a_2, \dots, a_n ,

$$E[a_1 X_1 + a_2 X_2 + \dots + a_n X_n] = a_1 E[X_1] + a_2 E[X_2] + \dots + a_n E[X_n].$$

1.3.4 Teorema.- Suponga que X_1, X_2, \dots, X_n son v.a's. independientes con esperanza finita. Entonces

$$E[X_1 X_2 \dots X_n] = E[X_1] E[X_2] \dots E[X_n].$$

1.3.5 Ejemplo.- (a) Consideremos el caso de las v.a's. X_1, X_2, \dots, X_n definidas en la observación 1.2.17(a). Es fácil mostrar que $E[X_i] = p$, $\forall i = 1, 2, \dots, n$. De aquí, por el Teorema 1.3.3(ii), $E[S_n] = np$, es decir, si $X \sim Bin(n, p)$ entonces $E[X] = np$.

(b) Si $X \sim P(\lambda)$, entonces cálculos directos muestran que $E[X] = \lambda$.

1.3.6 Definición.- Sea X una v.a. con esperanza finita μ_X .

(a) La varianza de X se define como

$$\sigma_X^2 \equiv Var(X) := E[(X - \mu_X)^2].$$

(b) La desviación estándar de X es

$$\sigma_X := [Var(X)]^{\frac{1}{2}}.$$

Es fácil mostrar que

$$\sigma_X^2 = E[X^2] - \mu_X^2,$$

la cual es una fórmula que generalmente se usa para calcular la varianza de una v.a.

1.3.7 Definición.- Sean X y Y dos v.a.'s con varianza finita. Entonces, su covarianza se define como

$$Cov(X, Y) := E[(X - \mu_X)(Y - \mu_Y)].$$

Es fácil mostrar que

$$Cov(X, Y) = E[XY] - \mu_X\mu_Y,$$

la cual resulta una fórmula mas conveniente para evaluar $Cov(X, Y)$.

Las propiedades de la varianza y la covarianza se establecen en los siguientes resultados:

1.3.8 Teorema.- Suponga que todas las v.a.'s. a continuación poseen varianza finita.

(i) $Cov(X, X) = Var(X)$,

(ii) $Cov(X, Y) = Cov(Y, X)$,

(iii) $Cov(\cdot, \cdot)$ es lineal en cada uno de sus argumentos, es decir, para todas las con-

stantes a_1, a_2, \dots, a_n ,

$$\text{Cov}(a_1X_1 + a_2X_2 + \dots + a_nX_n, Y) = a_1\text{Cov}(X_1, Y) + a_2\text{Cov}(X_2, Y) + \dots + a_n\text{Cov}(X_n, Y)$$

y

$$\text{Cov}(X, a_1Y_1 + a_2Y_2 + \dots + a_nY_n) = a_1\text{Cov}(X, Y_1) + a_2\text{Cov}(X, Y_2) + \dots + a_n\text{Cov}(X, Y_n).$$

(iv) Además,

$$\text{Cov}(X + a, Y + b) = \text{Cov}(X, Y), \quad \forall a, b \in \mathfrak{R}.$$

1.3.9 Teorema.- Sean X, X_1, X_2, \dots, X_n v.a's. con varianza finita. Entonces las siguientes afirmaciones son válidas:

(i) Si $P[X = a] = 1$ para alguna constante a , entonces $\text{Var}(X) = 0$.

(ii) $\text{Var}(aX) = a^2\text{Var}(X)$ para cualquier constante a .

(iii) Para todas las constantes a_1, a_2, \dots, a_n ,

$$\text{Var}\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i^2 \text{Var}(X_i) + 2 \sum_{i>j} a_i a_j \text{Cov}(X_i, X_j)$$

1.3.10 Teorema.- Si X y Y son v.a's. independientes con esperanza finita, entonces

(i) $\text{Cov}(X, Y) = 0$;

(ii) $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$.

(iii) Si X_1, X_2, \dots, X_n son v.a's. independientes con esperanza finita, entonces

$$\text{Var}(X_1 + X_2 + \dots + X_n) = \sum_{i=1}^n \text{Var}(X_i).$$

1.3.11 Observación.- Es fácil mostrar que si $X \sim \text{Bin}(n, p)$ entonces $\text{Var}(X) = np(1 - p)$, y si $X \sim P(\lambda)$, $\text{Var}(X) = \lambda$.

De los ejemplos más conocidos de distribuciones de vectores aleatorios es la llamada distribución multinormal o normal multivariada. Antes de definirla daremos alguna notación.

Sea $\mathbf{X} = (X_1, X_2, \dots, X_n)$ un vector aleatorio de dimensión n . Denotaremos por $\mu_{\mathbf{X}}$ al vector cuyos componentes son los valores esperados de cada una de las v.a.'s. que componen al vector \mathbf{X} . Es decir,

$$\mu_{\mathbf{X}} = (\mu_{X_1}, \mu_{X_2}, \dots, \mu_{X_n}).$$

Nos referiremos al vector $\mu_{\mathbf{X}}$ como el vector media del vector aleatorio \mathbf{X} . Además, definimos la matriz de covarianza del vector aleatorio \mathbf{X} como:

$$\begin{aligned} \Sigma &= \begin{bmatrix} \text{Cov}(X_1, X_1) & \text{Cov}(X_1, X_2) & \cdots & \text{Cov}(X_1, X_n) \\ \vdots & \vdots & & \vdots \\ \text{Cov}(X_n, X_1) & \text{Cov}(X_n, X_2) & \cdots & \text{Cov}(X_n, X_n) \end{bmatrix} \\ &= \begin{bmatrix} \sigma_{X_1}^2 & \text{Cov}(X_1, X_2) & \cdots & \text{Cov}(X_1, X_n) \\ \vdots & \vdots & & \vdots \\ \text{Cov}(X_n, X_1) & \text{Cov}(X_n, X_2) & \cdots & \sigma_{X_n}^2 \end{bmatrix}, \end{aligned}$$

para la cual estamos suponiendo que el vector media está bien definido.

1.3.12 Ejemplo.- Un vector aleatorio $\mathbf{X} = (X_1, X_2, \dots, X_n)$ tiene una distribución multinormal o normal multivariada si su función de densidad es de la forma

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \exp \left[-\frac{1}{2} (\mathbf{x} - \mu_{\mathbf{X}})^T \Sigma^{-1} (\mathbf{x} - \mu_{\mathbf{X}}) \right]$$

donde $\mu_{\mathbf{X}}$ es el vector media y Σ es la matriz de covarianza, $|\Sigma|$ su determinante y Σ^{-1} su matriz inversa. En este caso usamos la notación $\mathbf{X} \sim N(\mu_{\mathbf{X}}, \Sigma)$.

Observemos que si $n = 1$, obtenemos la distribución normal en \mathfrak{R}

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left[-\frac{(x - \mu)^2}{2\sigma^2} \right], \quad x \in \mathfrak{R},$$

donde $\mu = E[X]$ y $\sigma = \sqrt{Var(X)}$.

A continuación establecemos una propiedad de la matriz de covarianza la cual tiene repercusiones directas en la distribución multinormal. Para esto primero tenemos la siguiente definición.

1.3.13 Definición.- Una matriz simétrica A (es decir, $A = A^t$ donde A^t denota la traspuesta de la matriz A) se dice que es positivamente semidefinida si $x^t Ax \geq 0$ para toda $x \in \mathfrak{R}^n$. Se dice que es positivamente definida si $x^t Ax > 0$, para toda $x \in \mathfrak{R}^n$, $x \neq 0$.

1.3.14 Teorema.- Una matriz Σ es la matriz de covarianza de una distribución multinormal si, y solo si es positivamente semidefinida (ver Definición 1.3.14, abajo).

Este resultado es clásico en la teoría de probabilidad cuando se estudian distribuciones multivariadas. La demostración puede verse en Feller (1985).

Una relación muy importante, desde el punto de vista de las aplicaciones, entre el valor esperado y la varianza de una v.a. es la desigualdad de Chevyshev:

1.3.15 Teorema.- (Desigualdad de Chevyshev) Sea X una v.a. con $E[X]$ y $Var[X]$ finitas. Entonces para cualquier $\varepsilon > 0$

$$P[|X - E[X]| \geq \varepsilon] \leq \frac{Var[X]}{\varepsilon^2}.$$

Para concluir esta sección presentaremos uno de los teoremas límite más importantes de la Probabilidad, el cual juega un papel esencial en las aplicaciones que trataremos en los próximos capítulos.

1.3.16 Teorema.- (Ley Débil de los Grandes Números) Sean X_1, X_2, \dots, X_n v.a's. i.i.d. con media $\mu = E[X_i]$ y varianza finita. Entonces para cada $\varepsilon > 0$

$$P \left\{ \left| \frac{1}{n} \sum_{i=1}^n X_i - \mu \right| \geq \varepsilon \right\} \rightarrow 0.$$

De este teorema podemos ver que la media muestral $\bar{X} := \frac{1}{n} \sum_{i=1}^n X_i$ puede ser usada para aproximar a la media μ . De hecho tenemos que bajo las hipótesis de este teorema $E[\bar{X}] = \mu$, lo cual significa que \bar{X} es un estimador insesgado de μ .

1.4 Teorema de Feller

En esta sección presentaremos el resultado que juega el papel principal en las aplicaciones de la probabilidad al análisis matemático que se presentarán en el próximo capítulo. Este resultado fué propuesto por William Feller [ver, por ejemplo, Feller(1985)] y en términos generales es un resultado sobre convergencia de esperanzas.

Para formular el Teorema de Feller, sea θ un parámetro real que toma valores en un intervalo cerrado y acotado J , y $X_1(\theta), X_2(\theta), \dots$ v.a's. con funciones de densidad $g_1(\cdot; \theta), g_2(\cdot; \theta), \dots$ (o de probabilidad, según sea el caso) respectivamente. Además supongamos que $E[X_n] = \theta$ para toda $n = 1, 2, \dots$ y denotemos $Var[X_n] = \sigma_n^2(\theta)$.

1.4.1 Teorema de Feller.- Sea f una función continua y acotada sobre \mathfrak{R} , y supongamos que $\sigma_n^2(\theta) \rightarrow 0$ uniformemente para $\theta \in J$. Entonces

$$E_{n,\theta}[f(X_n)] \rightarrow f(\theta)$$

uniformemente para $\theta \in J$, donde

$$E_{n,\theta}[f(X_n)] = \int_{\mathfrak{R}} f(x) g_n(x; \theta) dx,$$

para el caso continuo, y

$$E_{n,\theta} [f(X_n)] = \int f(x) g_n(x; \theta) dx$$

para el caso discreto.

Demostración.

La demostración la desarrollaremos suponiendo que las v.a's. X_1, X_2, \dots , son continuas. Para el caso discreto se siguen las mismas ideas.

Sea $\varepsilon > 0$ arbitrario. Por la hipótesis de que f es acotada en \mathfrak{R} existe $M \in (0, \infty)$ tal que $|f(x)| \leq M, \forall x \in \mathfrak{R}$. Ahora, para $\theta \in J$ y cada $\delta > 0$ definimos

$$A_1 = \{y \in \mathfrak{R} : |y - \theta| < \delta\}$$

y

$$A_2 = A_1^c = \{y \in \mathfrak{R} : |y - \theta| \geq \delta\}.$$

Entonces

$$\begin{aligned} |E_{n,\theta} [f(X_n)] - f(\theta)| &= \left| \int_{\mathfrak{R}} f(x) g_n(x; \theta) dx - f(\theta) \right| \\ &= \left| \int_{\mathfrak{R}} [f(x) - f(\theta)] g_n(x; \theta) dx \right| \leq \int_{\mathfrak{R}} |f(x) - f(\theta)| g_n(x; \theta) dx \\ &= \int_{A_1} |f(x) - f(\theta)| g_n(x; \theta) dx + \int_{A_2} |f(x) - f(\theta)| g_n(x; \theta) dx =: I_1 + I_2. \end{aligned} \quad (1.1)$$

Por otro lado, como f es continua en el intervalo cerrado y acotado J , entonces es uniformemente continua en J [ver, por ejemplo, Rudin (1987)]. De aquí, elegimos $\delta > 0$ tal que $|f(x) - f(\theta)| < \frac{\varepsilon}{2}$ para $\theta, x \in J$ y $|x - \theta| < \delta$. Fijando este δ , ahora elegimos $N = N_\varepsilon$ tal que

$$n \geq N \quad \text{implica} \quad \sigma_n^2(\theta) \leq \frac{\delta^2 \varepsilon}{4M}. \quad (1.2)$$

Esto es posible por la convergencia uniforme a cero de $\sigma_n^2(\theta)$ en J . De esta manera

$$I_1 \leq \frac{\varepsilon}{2} \int_{A_1} g_n(x; \theta) dx \leq \frac{\varepsilon}{2}. \quad (1.3)$$

Además, por la desigualdad de Chebyshev,

$$I_2 \leq 2M \int_{A_2} g_n(x; \theta) dx = 2MP [|X_n - \theta| \geq \delta]$$

$$\leq \frac{2M\sigma_n^2(\theta)}{\delta^2} < \frac{\varepsilon}{2} \text{ si } n \geq N. \quad (1.4)$$

La última desigualdad se sigue de (1.2). Finalmente, combinando (1.1), (1.3) y (1.4) obtenemos el resultado deseado. ■

1.5 Cadenas de Markov

Para concluir el capítulo de prerequisites de probabilidad, en esta sección presentaremos los elementos necesarios sobre cadenas de Markov sin pretender abundar mucho sobre el tema. Esta parte será aplicada en la solución de sistemas de ecuaciones lineales que presentaremos en el último capítulo.

1.5.1 Definición.- Un proceso estocástico es una familia $X = \{X(t), t \in T\}$ de v.a's., definidas sobre un mismo espacio de probabilidad, donde T es el conjunto de parámetros o tiempo.

En términos generales, la definición anterior se refiere a que, para cada $t \in T$, $X(t)$ es una v.a. definida en $(\Omega, \mathfrak{S}, P)$, es decir, $X(t) : \Omega \rightarrow \mathfrak{R}$. Esto significa que en realidad el proceso depende de las variables $t \in T$, $\omega \in \Omega$, de tal forma que lo podemos ver como una colección de funciones reales definidas sobre $T \times \Omega$:

$$X(\cdot, \cdot) : T \times \Omega \rightarrow \mathfrak{R},$$

que satisfacen las siguientes propiedades:

(i) para cada $t \in T$ fijo, $X(t, \cdot)$ es una v.a. en Ω ;

(ii) para cada $\omega \in \Omega$ fijo, $X(\omega, t) : T \rightarrow \mathfrak{R}$ es una función de $t \in T$, llamada trayectoria del proceso X .

Los procesos estocásticos se pueden clasificar dependiendo de la naturaleza del conjunto donde toman valores las v.a's. $X(t)$, $t \in T$, al cual llamamos *espacio de estados*, y dependiendo del tipo de conjunto T . Denotaremos por S al espacio de estados, y a cada punto de S se le llama estado del proceso. Si S es finito o numerable decimos que el proceso es discreto; si S es un intervalo, decimos que el proceso es continuo. Además, si T es numerable, por ejemplo, $T = \{0, 1, 2, \dots\}$, $T = \{0, \pm 1, \pm 2, \dots\}$, etc., decimos que es un proceso estocástico en tiempo discreto. Si T es un intervalo, digamos, $T = [a, b]$, $T = [0, \infty)$, $T = (-\infty, \infty)$, etc., decimos que es un proceso en tiempo continuo.

Si $X = \{X(t), t \in T\}$ es un proceso en tiempo discreto, digamos, $T = \{0, 1, 2, \dots\}$, escribimos $X(t) = X_t$, de tal forma que el proceso estocástico es una sucesión de v.a's.: $X = \{X_0, X_1, X_2, \dots\}$.

Una clase importante de procesos estocásticos la constituyen las cadenas de Markov. Estos son procesos con espacios de estados discreto, aunque el conjunto T puede ser arbitrario (discreto o continuo), y su característica principal es que las observaciones futuras solo dependen del presente y no del pasado.

Para los fines que se persiguen en este trabajo, es suficiente considerar cadenas de Markov con espacio de estados finitos y tiempo discreto. En este sentido, para facilitar la exposición, suponemos que $S = \{0, 1, 2, \dots, n\}$ y $T = \mathbb{N}_0 := \{0, 1, 2, \dots\}$.

1.5.2 Definición.- Un proceso estocástico $X = \{X_t; t \in \mathbb{N}_0\}$ es una cadena de Markov si satisface la propiedad

$$P[X_{t+1} = j \mid X_0 = i_0, X_1 = i_1, \dots, X_{t-1} = i_{t-1}, X_t = i] = P[X_{t+1} = j \mid X_t = i], \quad (1.5)$$

para todo $i_0, i_1, \dots, i_{t-1}, i, j \in S$ y $t \in \mathbb{N}_0$.

A la relación (1.5) se le conoce como propiedad de Markov y su interpretación es la siguiente: para cada tiempo $t \in \mathbb{N}_0$, la distribución del “futuro” X_{t+1} es condicionalmente independiente del “pasado” X_0, X_1, \dots, X_{t-1} , dado el “presente” X_t .

A las probabilidades $P[X_{t+1} = j / X_t = i]$ se les conoce como probabilidades de transición en un paso de la cadena X . Si estas probabilidades no dependen de t , decimos que X es una cadena homogénea en el tiempo o estacionaria, y escribimos

$$p_{ij} = P[X_1 = j / X_0 = i] = P[X_{t+1} = j / X_t = i],$$

para todo $i, j \in S$, $t \in \mathbb{N}_0$. Aquí solo consideramos cadenas estacionarias, de tal forma que en lo que sigue nos referiremos a $X = \{X_t; t \in \mathbb{N}_0\}$ como una cadena de Markov estacionaria, con espacio de estados finitos.

A la matriz $P = [p_{ij}]_{n \times n}$ formada por las probabilidades de transición en un paso se le llama matriz de transición de la cadena X . Las probabilidades de transición satisfacen las siguientes propiedades:

(a) $p_{ij} \geq 0$ para todo $i, j \in S$;

(b) $\sum_{j=1}^n p_{ij} = 1$, $i = 1, 2, \dots, n$;

(c) $p_{ij}^{(l+m)} = \sum_{k=1}^n p_{ik}^{(l)} p_{kj}^{(m)}$, para todo $l, m \in \mathbb{N}_0$, $i, j \in S$,

en donde $p_{ij}^{(m)}$ es la probabilidad de transición en m pasos, $m > 0$, definida como

$$P_{ij}^{(m)} := P[X_m = j / X_0 = i] = P[X_{k+m} = i / X_k = i],$$

para todo $k \geq 0$.

Si $m = 1$, $p_{ij} = p_{ij}^{(1)}$; y si $m = 0$ definimos

$$p_{ij}^{(0)} = \delta_{ij} = \begin{cases} 1 & \text{si } i = j \\ 0 & \text{si } i \neq j \end{cases}$$

A la propiedad (c) se le conoce como Ecuación de Chapman-Kolmogorov y su demostración se puede ver, por ejemplo, en Hoel, Port, Stone(1972).

De la demostración de este resultado tambien podemos deducir que las probabilidades $p_{ij}^{(m)}$ son los elementos correspondientes de la matriz P^m (la matriz P elevada a la m -ésima potencia), en el sentido de que $p_{ij}^{(m)}$ es el elemento en la i -ésima fila y la j -ésima columna de P^m :

$$P^m = \left[p_{ij}^{(m)} \right]_{n \times n}$$

La interpretación general de una cadena de Markov es la siguiente. El proceso empieza en un estado inicial i_0 elegido de acuerdo a una distribución de probabilidades inicial $\pi_{i_0} = P[X_0 = i_0]$, $i_0 = 1, 2, \dots, n$. Después el proceso transita a un estado i_1 de acuerdo a la probabilidad de transición $p_{i_0 i_1}$, y este procedimiento se repite. De esta manera, transiciones subsecuentes producen estados i_1, i_2, \dots , de acuerdo con las probabilidades de transición p_{ij} . Por la propiedad de Markov, es fácil ver que la distribución conjunta de las v.a's. X_1, X_2, \dots, X_n toma la forma

$$P[X_0 = i_0, X_1 = i_1, \dots, X_n = i_n] = \pi_{i_0} p_{i_0 i_1} p_{i_1 i_2} \cdots p_{i_{n-1} i_n}, \quad i_0, i_1, \dots, i_n \in S. \quad (1.6)$$

Observemos que la distribución inicial de la cadena X , $\pi_{i_0} = P[X_0 = i_0]$, $i_0 = 1, 2, \dots, n$, satisface:

$$\sum_{i_0=1}^n \pi_{i_0} = 1 \quad \text{y} \quad \pi_{i_0} \geq 0, \quad i_0 = 1, 2, \dots, n.$$

Dada la cadena de Markov $X = \{X_t : t \in IN_0\}$, decimos que dos estados i y j se comunican si existen dos enteros k_1 y k_2 tales que $p_{ij}^{(k_1)} > 0$ y $p_{ji}^{(k_2)} > 0$. Esto es, los estados i y j se comunican si uno puede ser alcanzado desde el otro con probabilidad positiva.

Sea $\tilde{S} \subset S$ un subconjunto de estados tales que

- (i) todos los estados en \tilde{S} se comunican;
- (ii) si $i \in \tilde{S}$ y $j \notin \tilde{S}$, entonces $p_{ij}^k = 0$ para todo k .

Entonces decimos que \tilde{S} forma una clase ergódica de estados.

1.5.3 Definición.- Si S forma, por sí misma, una clase ergódica (es decir, todos los estados se comunican con cada uno de los otros) entonces decimos que la cadena de Markov X es irreducible.

Capítulo 2

Aplicaciones de la Probabilidad al Análisis Matemático

2.1 Introducción

En este capítulo presentaremos algunas de las aplicaciones de la probabilidad al análisis matemático, las cuales se centran principalmente en la teoría de aproximación de funciones continuas a través de dos resultados bien conocidos: El Teorema de Aproximación de Weierstrass y el Teorema de Taylor para aproximación de funciones.

Primero presentaremos una demostración probabilística del Teorema de Weierstrass. Esta resulta más atractiva que la tradicional pues los argumentos usados son más sencillos y permiten obtener más información de la aproximación. Enseguida presentaremos una generalización del Teorema de Taylor, que se obtiene por medio de propiedades sencillas de la distribución de Poisson.

Una de las ventajas de usar la teoría de la probabilidad para analizar ambos resultados es la posibilidad de calcular la tasa de convergencia de las aproximaciones correspondientes. Este punto será analizado al final del capítulo.

Los argumentos probabilísticos aplicados para mostrar los resultados anteriores están

basados principalmente en los siguientes hechos, los cuales son una consecuencia del Teorema de Feller que se expuso en la Sección 1.4 del capítulo anterior.

Sean $X_1(\theta), X_2(\theta), \dots, X_n(\theta)$, v.a's. i.i.d. con media $E[X_i] = \theta$ para $i = 1, 2, 3, \dots$. Suponemos que θ es un parámetro real que toma valores en un intervalo cerrado y acotado J . Suponemos además que $K := \sup_{\theta \in J} \text{Var}[X_i] = \sup_{\theta \in J} E[(X_i - \theta)^2] < \infty$, para cada $i = 1, 2, \dots$.

Sea $Y_n = \frac{1}{n} \sum_{i=1}^n X_i$, $n \in \mathbb{N}$, y denotemos por $g_n(\cdot, \theta)$ su función de densidad. Entonces de las propiedades de la esperanza y de la varianza (Teoremas 1.3.3 y 1.3.10) tenemos

$$E[Y_n] = \frac{1}{n} \sum_{i=1}^n E[X_i] = \theta; \quad \forall \theta \in J, \quad n \in \mathbb{N}; \quad (2.1)$$

$$\text{Var}[Y_n] = \frac{1}{n^2} \sum_{i=1}^n \text{Var}[X_i] \leq \frac{K}{n}, \quad \forall n \in \mathbb{N}. \quad (2.2)$$

Observemos que $\text{Var}[Y_n] \rightarrow 0$ uniformemente en J , cuando $n \rightarrow \infty$. Ahora por el Teorema de Feller, para cada función continua y acotada f sobre \mathfrak{R}

$$E_{n,\theta} \left[f \left(\frac{1}{n} \sum_{i=1}^n X_i \right) \right] \equiv E \left[f \left(\frac{1}{n} \sum_{i=1}^n X_i \right) \right] \rightarrow f(\theta), \quad (2.3)$$

uniformemente en J , cuando $n \rightarrow \infty$.

Las aplicaciones de la probabilidad en las que estamos interesados, las obtendremos de la relación (2.3) al elegir familias particulares de densidades.

2.2 Teorema de Weierstrass

La primera aplicación proporciona una demostración alternativa, por medio de la teoría de la probabilidad, del conocido Teorema de Aproximación de Weierstrass.

2.2.1 Teorema.- Si f es una función real continua en un intervalo $J = [a, b]$, existe una

sucesión de polinomios P_n tal que

$$\lim_{n \rightarrow \infty} P_n(\theta) = f(\theta); \quad (2.4)$$

uniformemente en $\theta \in J$.

La demostración de este teorema es constructiva. En algunos libros [e.g. Apostol(1972), Kolmogorov y Fomin (1975)] se presenta una sucesión de polinomios trigonométricos que convergen a la función f . Rudín (1980) proporciona una sucesión de polinomios distinta a la de las últimas referencias.

Sin pérdida de generalidad tomaremos $J = [0, 1]$. Esto lo podemos hacer ya que si (2.4) es válido para $\theta \in [0, 1]$, la aplicación de este resultados a la función $g(\theta) = f[(b-a)\theta + a]$, $\theta \in [0, 1]$, nos lleva al caso general, ya que además, $P_n[(b-a)\theta + a]$ también es un polinomio en θ .

Demostración del Teorema 2.2.1

Sea $\{X_i, i = 1, 2, 3, \dots\}$ una familia de v.a's. i.i.d. tal que $X_i = 1$ con probabilidad θ y $X_i = 0$ con probabilidad $1 - \theta$, $\theta \in [0, 1]$. Del Ejemplo 1.3.5, la distribución de la variable aleatoria $\sum_{i=1}^n X_i$ tiene distribución binomial con parámetros n y θ , para todo $n \in \mathbb{N}$.

Por otro lado tenemos que $E[X_i] = \theta$ y $Var[X_i] = \theta(1 - \theta)$, $i = 1, 2, 3, \dots$. En consecuencia,

$$E[Y_n] = E\left[\frac{1}{n} \sum_{i=1}^n X_i\right] = \theta,$$

y

$$Var[Y_n] = Var\left[\frac{1}{n} \sum_{i=1}^n X_i\right] = \frac{\theta(1 - \theta)}{n} \leq \frac{1}{n} \rightarrow 0,$$

cuando $n \rightarrow \infty$, uniformemente en $[0, 1]$.

Entonces, por el Teorema de Feller, [ver (2.1),(2.2),(2.3)] para cualquier función con-

tinua f en $[0, 1]$

$$E[f(Y_n)] \rightarrow f(\theta), \quad (2.5)$$

uniformemente en $[0, 1]$ cuando $n \rightarrow \infty$.

Por otro lado, de la Observación 1.2.9, la media aritmética $Y_n = \frac{1}{n} \sum_{i=1}^n X_i$ tiene función de probabilidad concentrada en los puntos $\frac{k}{n}$, $k = 0, 1, 2, 3, \dots$. Entonces de la Observación 1.2.9 y el Teorema 1.3.2(i), tenemos

$$E[f(Y_n)] = \sum_{k=0}^n f\left(\frac{k}{n}\right) C_k^n \theta^k (1-\theta)^{n-k}, \quad n \in \mathbb{N}. \quad (2.6)$$

Combinando (2.5) y (2.6) concluimos que

$$\sum_{k=0}^n f\left(\frac{k}{n}\right) C_k^n \theta^k (1-\theta)^{n-k} \rightarrow f(\theta),$$

cuando $n \rightarrow \infty$, y la convergencia es uniforme en $[0, 1]$. Esto demuestra el teorema. ■

2.2.2. Observación.- El Teorema 2.2.1 establece que cualquier función continua f en un intervalo cerrado puede ser aproximada uniformemente por polinomios. De hecho, la expresión (2.6) presenta de forma explícita los polinomios que aproximan a f . A estos polinomios se les llama Polinomios de Bernstein, y los denotaremos por $B_n(\theta; f)$. Es decir,

$$B_n(\theta; f) := \sum_{k=0}^n f\left(\frac{k}{n}\right) C_k^n \theta^k (1-\theta)^{n-k}.$$

Con esta notación, el Teorema 2.2.1 puede reformularse como

$$\sup_{\theta \in [0,1]} |f(\theta) - B_n(\theta; f)| \rightarrow 0, \quad (2.7)$$

cuando $n \rightarrow \infty$.

2.3 Una Generalización del Teorema de Taylor

Sea f una función de valor real definida en un intervalo J de \mathfrak{R} . Si f tiene derivadas continuas de todos los ordenes, $f^{(n)}$, $n = 1, 2, 3, \dots$, en cada punto de J , escribiremos $f \in C^\infty$ en J .

Ahora, supongamos que $f \in C^\infty$ en alguna vecindad de un punto $c \in J$. A la serie de potencias

$$\sum_{n=0}^{\infty} \frac{f^{(n)}(c)}{n!} (x-c)^n,$$

se le llama serie de Taylor alrededor de c generada por f .

El Teorema de Taylor se puede formular de la siguiente manera [ver Apostol (1972)]:

2.3.1. Teorema.- Sea $f \in C^\infty$ en $[a, b]$ y sea $c \in [a, b]$. Supóngase que existe una vecindad $B(c)$ de c , y una constante M tal que $|f^{(n)}(x)| \leq M^n$ para cada $x \in B(c) \cap [a, b]$ y cada $n \in \mathbb{N}$. Entonces, para cada $x \in B(c) \cap [a, b]$, tenemos

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(c)}{n!} (x-c)^n. \quad (2.8)$$

La siguiente aplicación del Teorema de Feller nos permitirá dar una generalización del Teorema de Taylor.

Sea $\{X_i, i = 1, 2, 3, \dots\}$ una familia de v.a.'s. i.i.d. que tienen distribución de Poisson con parámetro $\theta \in [0, M_1]$, donde $M_1 > 0$ es fijo pero arbitrario. Entonces, por la Observación 1.2.17 (b), la variable aleatoria $\sum_{i=1}^n X_i$ tiene una distribución de Poisson con parámetro $n\theta$.

Por otro lado, puesto que $E[X_i] = \theta$ y $Var[X_i] = \theta$, $i = 1, 2, 3, \dots, n$ (Observación 1.3.11) tenemos que

$$E[Y_n] = E\left[\frac{1}{n} \sum_{i=1}^n X_i\right] = \theta$$

y

$$\text{Var} [Y_n] = \text{Var} \left[\frac{1}{n} \sum_{i=1}^n X_i \right] = \frac{\theta}{n} \leq \frac{M_1}{n} \rightarrow 0$$

uniformemente en $[0, M_1]$, cuando $n \rightarrow \infty$.

Entonces, por el Teorema de Feller [ver (2.3)], para cada función continua y acotada f en \mathfrak{R} , tenemos que

$$E [f (Y_n)] \rightarrow f (\theta), \quad (2.9)$$

uniformemente en $[0, M_1]$, cuando $n \rightarrow \infty$.

Ahora de la Observación 1.2.9, la variable aleatoria $Y_n = \frac{1}{n} \sum_{i=1}^n X_i$ tiene una función de probabilidad concentrada en los puntos $\frac{k}{n}$, $k = 0, 1, 2, 3, \dots$. Entonces, por el Teorema 1.3.2(i),

$$E [f (Y_n)] = \sum_{k=0}^{\infty} f \left(\frac{k}{n} \right) \frac{(n\theta)^k}{k!} \exp (-n\theta), \quad \forall n \in \mathbb{N}, \theta \in [0, M_1]. \quad (2.10)$$

Por otro lado, como

$$\exp (-n\theta) = \sum_{j=0}^{\infty} \frac{(-1)^j (n\theta)^j}{j!},$$

y tomando $h = \frac{1}{n}$, la expresión (2.10) se transforma

$$E \left[f \left(Y_{\frac{1}{h}} \right) \right] = \sum_{k=0}^{\infty} \sum_{j=0}^{\infty} \frac{(-1)^j}{j!} f (kh) \left(\frac{\theta}{h} \right)^{j+k} \frac{1}{k!}.$$

Además, si $m = j + k$ y recordando que $C_k^m = \frac{m!}{k!(m-k)!}$, obtenemos

$$E \left[f \left(Y_{\frac{1}{h}} \right) \right] = \sum_{m=0}^{\infty} \left(\frac{\theta}{h} \right)^m \frac{1}{m!} \sum_{k=0}^m (-1)^{m-k} C_k^m f (kh). \quad (2.11)$$

Por lo tanto, de (2.9) y (2.11) concluimos que

$$\sum_{m=0}^{\infty} \left(\frac{\theta}{h} \right)^m \frac{1}{m!} \sum_{k=0}^m (-1)^{m-k} C_k^m f (kh) \rightarrow f (\theta), \quad (2.12)$$

cuando $h \rightarrow 0$, uniformemente en $\theta \in [0, M_1]$.

Por otro lado, sea f una función real arbitraria. Definimos el operador cociente en diferencia Δh como

$$(\Delta_h f)(x) = \frac{f(x+h) - f(x)}{h}, \quad h \in \mathfrak{R}.$$

Observe que

$$\begin{aligned} (\Delta_h^2 f)(x) &:= \Delta_h \{(\Delta_h f)(x)\} = \frac{(\Delta_h f)(x+h) - (\Delta_h f)(x)}{h} \\ &= \frac{\frac{f(x+2h) - f(x+h)}{h} - \frac{f(x+h) - f(x)}{h}}{h} \\ &= \frac{f(x+2h) - 2f(x+h) + f(x)}{h^2} \end{aligned}$$

En general, se puede mostrar por inducción que

$$(\Delta_h^m f)(x) = \Delta_h \{(\Delta_h^{m-1} f)(x)\} = \frac{1}{h^m} \sum_{k=0}^m (-1)^{m-k} C_k^m f(x+kh), \quad m \in \mathbb{N}. \quad (2.13)$$

Aplicando la relación (2.13) en (2.11) obtenemos

$$\sum_{m=0}^{\infty} \frac{\theta^m}{m!} \Delta_h^m f(0) \rightarrow f(\theta), \quad (2.14)$$

uniformemente en $[0, M_1]$, cuando $h \rightarrow 0$.

Con el desarrollo anterior hemos demostrado el siguiente teorema:

2.3.2 Teorema.- Para cualquier función continua y acotada g en \mathfrak{R} , y $c \in \mathfrak{R}$,

$$\sum_{m=0}^{\infty} \frac{\theta^m}{m!} \Delta_h^m g(c) \rightarrow g(c+\theta), \quad (2.15)$$

cuando $h \rightarrow 0$, uniformemente en cualquier intervalo cerrado y acotado.

En efecto, tomando $f(\theta) = g(c+\theta)$ para $x \in \mathfrak{R}$, en la relación (2.14) obtenemos (2.15).

uaues. Además supongamos que la esperanza y la varianza de una variable aleatoria con

La relación (2.15) representa un desarrollo de la serie de Taylor generalizada de g [ver (2.8)] aplicando el operador cociente de diferencia Δ_h en lugar de las derivadas. Obsérvese que en este caso g no necesita ser diferenciable, como es el caso del Teorema 2.3.1.

2.4 Tasas de Convergencia

En las secciones anteriores analizamos el problema de aproximación uniforme de funciones continuas f por medio de una sucesión de funciones $\{f_n\}$. En este caso sólo mostramos la convergencia, pero un problema interesante que surge es el referente a la rapidez de la misma. Concretamente nuestro problema se puede plantear de la siguiente manera:

Suponga que una sucesión de funciones $\{f_n\}$ converge uniformemente a una función f en un intervalo cerrado y acotado J . Es decir, para cada $\varepsilon > 0$ existe N_ε tal que

$$\sup_{\theta \in J} |f(\theta) - f_n(\theta)| \leq \varepsilon \quad \forall n \geq N_\varepsilon$$

Considerando lo visto en las secciones anteriores sabemos que N_ε existe, por ejemplo para el Teorema de Weierstrass y el Teorema de Taylor, y nuestro problema entonces es encontrarlo explícitamente como una función de ε .

El análisis de la tasa de convergencia toma mayor importancia cuando se quiere aproximar funciones por métodos numéricos. El siguiente teorema proporciona información sobre la tasa de convergencia cuando la función que se desea aproximar pertenece a cierta clase de funciones. Ejemplificaremos este resultado con los ejemplos de las secciones anteriores.

Sea $\theta \in J$ un parámetro real, y sea $\{g_n(\cdot; \theta); n = 1, 2, 3, \dots\}$ una familia de densidades. Además supongamos que la esperanza y la varianza de una variable aleatoria con

densidad g_n son θ y $\sigma_n^2(\theta)$, respectivamente. Es decir:

$$\int x g_n(x; \theta) dx = \theta, \quad n \in \mathbb{N}, \theta \in J;$$

$$\int (x - \theta)^2 g_n(x; \theta) dx = \sigma_n^2(\theta), \quad n \in \mathbb{N}, \theta \in J.$$

2.4.1 Teorema.- Sea f una función acotada en \mathfrak{R} que satisface la condición de Lipschitz:

$$|f(x) - f(y)| \leq L|x - y|, \quad \forall x, y \in \mathfrak{R}. \quad (2.16)$$

donde M es cota de f y L es una constante positiva. Además supongamos que $\sigma_n^2(\theta) \rightarrow 0$ cuando $n \rightarrow \infty$ uniformemente en $\theta \in J$. Sea $\varepsilon > 0$ y N_ε tal que

$$\sigma_n^2(\theta) \leq \frac{\varepsilon^3}{16L^2M} \quad \forall n \geq N_\varepsilon,$$

Entonces

$$\left| f(\theta) - \int f(x) g_n(x; \theta) dx \right| \leq \varepsilon \quad \forall n \geq N_\varepsilon$$

Demostración.

Sea $\delta > 0$ y defina

$$I_1 := \int_{A_1} |f(x) - f(\theta)| g_n(x; \theta) dx,$$

$$I_2 := \int_{A_2} |f(x) - f(\theta)| g_n(x; \theta) dx$$

donde

$$A_1 := \{y : |y - \theta| < \delta\} \quad \text{y} \quad A_2 = A_1^c = \{y : |y - \theta| \geq \delta\}.$$

Puesto que $\int g_n(x; \theta) dx = 1$, para cada $\delta > 0$ tenemos

$$\left| f(\theta) - \int_{\mathfrak{R}} f(x) g_n(x; \theta) dx \right| = \left| \int_{\mathfrak{R}} [f(\theta) - f(x)] g_n(x; \theta) dx \right|$$

$$\begin{aligned} &\leq \int_{\mathbb{R}} |f(x) - f(\theta)| g_n(x; \theta) dx \\ &= I_1 + I_2. \end{aligned}$$

Ahora de (2.16),

$$I_1 = \int_{A_1} |f(x) - f(\theta)| g_n(x; \theta) dx \leq L\delta \int_{A_1} g_n(x; \theta) dx \leq L\delta.$$

Por otro lado, $|f(x) - f(\theta)| \leq 2M$ y por la Desigualdad de Chebyshev (Teorema 1.3.15), tenemos

$$\int_{A_2} g_n(x; \theta) dx = P[|X_n - \theta| \geq \delta] \leq \frac{\sigma_n^2(\theta)}{\delta^2}.$$

Entonces

$$I_2 = \int_{A_2} |f(x) - f(\theta)| g_n(x; \theta) dx \leq \frac{2M\sigma_n^2(\theta)}{\delta^2}.$$

Si $\delta = \frac{\varepsilon}{2L}$, obtenemos

$$I_1 \leq \frac{\varepsilon}{2}. \quad (2.17)$$

Por otro lado, recuerde que N_ε es tal que $\sigma_n^2(\theta) \leq \frac{\varepsilon^3}{16L^2M} \forall n \geq N_\varepsilon$. Luego, como $\delta = \frac{\varepsilon}{2L}$,

$$\frac{2M\sigma_n^2(\theta)}{\delta^2} \leq \frac{2M\varepsilon^3}{16L^2M} = \frac{\varepsilon}{2}.$$

Por lo tanto

$$I_2 \leq \frac{\varepsilon}{2}. \quad (2.18)$$

Combinando las relaciones (2.17) y (2.18) obtenemos el resultado deseado. ■

2.4.2 Ejemplo.- (a) Recuerde que el Teorema 2.2.1 garantiza que una función f continua y acotada puede aproximarse por medio de los polinomios de Bernstein. Para calcular la tasa de convergencia de dicha aproximación, supondremos que la función f es de Lipschitz con constante L . Observemos que las hipótesis del Teorema 2.4.1. se satisfacen, y además

[ver (2.6.)]

$$\int f(x) g_n(x; \theta) dx = E[f(Y_n)] = B_n(\theta; f), \quad n \in \mathbb{N}. \quad (2.19)$$

Sea $\varepsilon > 0$ fijo pero arbitrario. Del hecho de que $\sigma_n^2(\theta) \leq \frac{1}{n}$ tenemos que

$$\frac{1}{n} \leq \frac{\varepsilon^3}{16L^2M} \text{ implica que } \sigma_n^2(\theta) \leq \frac{\varepsilon^3}{16L^2M},$$

donde M es una cota de f . Es decir, si $N_\varepsilon := \left\lceil \frac{16L^2M}{\varepsilon^3} \right\rceil$, donde $[x]$ denota el primer entero que es más grande que x , tenemos que

$$n \geq N_\varepsilon \text{ implica que } \sigma_n^2(\theta) \leq \frac{\varepsilon^3}{16L^2M}.$$

Por lo tanto, por el Teorema 2.4.1, la relación (2.19) y la definición de N_ε concluimos que

$$\sup_{\theta \in [0,1]} |f(\theta) - B_n(\theta; f)| < \left(\frac{16L^2M}{n} \right)^{1/3} = O(n^{-1/3}),$$

para cada función de Lipschitz en $[0, 1]$.

(b) Consideremos el problema de aproximación de una función f continua y acotada, por medio de series generalizadas de Taylor. De nuevo, supondremos que f es de Lipschitz con constante L y recuerde que [ver (2.11), (2.12), (2.13), (2.14)]

$$\int f(x) g_n(x; \theta) = E[f(Y_n)] = \sum_{m=0}^{\infty} \frac{\theta^m}{m!} \Delta_{1/n}^m f(0) \rightarrow f(\theta),$$

cuando $n \rightarrow \infty$ uniformemente en $\theta \in J$. De nuevo nuestro problema es calcular la tasa de convergencia. Para esto procederemos de manera similar al ejemplo anterior.

Sea $\varepsilon > 0$ fijo y arbitrario. Como $\sigma_n^2(\theta) \leq \frac{M_1}{n}$, $\theta \in [0, M_1]$, tenemos que

$$\frac{M_1}{n} \leq \frac{\varepsilon^3}{16L^2M} \text{ implica que } \sigma_n^2(\theta) \leq \frac{\varepsilon^3}{16L^2M},$$

donde M es una cota de f . Es decir,

$$\text{si } n \geq N_\varepsilon := \left\lceil \frac{16L^2MM_1}{\varepsilon^3} \right\rceil \text{ entonces } \sigma_n^2(\theta) \leq \frac{\varepsilon^3}{16L^2M}.$$

Por el Teorema 2.4.1 y la definición de N_ε llegamos a que

$$\sup_{\theta \in [0,1]} \left| f(\theta) - \sum_{m=0}^{\infty} \frac{\theta^m}{m!} \Delta_{1/n}^m f(0) \right| < \left(\frac{16L^2MM_1}{n} \right)^{1/3} = O(n^{-1/3}),$$

Es decir

$$\sup_{\theta \in [0,1]} \left| f(\theta) - \sum_{m=0}^{\infty} \frac{\theta^m}{m!} \Delta_{1/n}^m f(0) \right| = O(n^{-1/3}).$$

Capítulo 3

Aplicaciones de la Probabilidad al Álgebra Lineal

3.1 Introducción

El objetivo de este capítulo es mostrar dos aplicaciones de la teoría de la probabilidad al álgebra lineal. La primera es una demostración probabilística de un teorema que establece ciertas propiedades de las matrices positivamente semidefinidas.

La otra aplicación está relacionada con la solución de una clase de Sistemas de Ecuaciones Lineales (SEL). En este caso podemos decir que dicha aplicación también se extiende al área del análisis numérico, debido a que el método presentado también proporciona un algoritmo que puede ser implementado computacionalmente por medio de técnicas de simulación conocidas generalmente como Método de Monte Carlo. Esta aplicación se presenta en las Secciones 3.3 y 3.4, y nos reintingiremos a la presentación del método de solución del SEL y su algoritmo, sin la implementación en computadora.

3.2 Matrices Semi-definidas Positivas

Generalmente, para verificar que una matriz A simétrica es semi-definida positiva se recurre directamente a la definición (Definición 1.3.13) o bien, se trata de verificar que sus valores propios, digamos $\lambda_1(A), \lambda_2(A), \dots, \lambda_n(A)$, son no-negativos (positivos) [ver, por ejemplo, Lang (1978), Strang (1978)]. Ambos procedimientos pueden resultar tediosos o difíciles, como es el caso de la demostración del resultado siguiente:

3.2.1 Teorema. Si $A = (a_{ij})_{n \times n}$ y $B = (b_{ij})_{n \times n}$ son matrices positivamente semi-definidas, entonces la matriz $C = (a_{ij}b_{ij})$ es positivamente semidefinida .

La demostración que presentaremos de este teorema esta basada en argumentos puramente probabilísticos, especialmente en el hecho (Teorema 1.3.14) de que si Σ es una matriz positivamente semidefinida, entonces existen v.a's. con distribución normal que tienen a Σ como su matriz de covarianza. En este caso, sin pérdida de generalidad, podemos suponer que dichas v.a's. tienen media cero, ya que la covarianza es invariante bajo traslaciones (Teorema 1.3.8 (iv)), y en particular tendríamos que:

$$Cov(X, Y) = Cov(X - \mu_X, Y - \mu_Y)$$

para cualesquiera v.a's. X y Y .

Demostración del Teorema 3.2.1.

Como A y B son matrices positivamente semidefinidas, existen vectores aleatorios independientes $X = (X_1, X_2, \dots, X_n)$ y $Y = (Y_1, Y_2, \dots, Y_n)$ con media cero y matrices de covarianza A y B , respectivamente. Es decir,

$$A = (a_{ij})_{n \times n} := (Cov(X_i, X_j))_{n \times n} \text{ y } B = (b_{ij})_{n \times n} := (Cov(Y_i, Y_j))_{n \times n}.$$

Ahora

$$a_{ij}b_{ij} = Cov(X_i, X_j) Cov(Y_i, Y_j)$$

$$\begin{aligned}
&= \left(E[X_i X_j] - \mu_{X_i} \mu_{X_j} \right) \left(E[Y_i Y_j] - \mu_{Y_i} \mu_{Y_j} \right) \\
&= E[X_i X_j] E[Y_i Y_j], \quad i, j = 1, 2, \dots, n.
\end{aligned} \tag{3.1}$$

Definamos las v.a's. $Z_i := X_i Y_i$, $i = 1, 2, \dots, n$; por la independencia de los vectores X y Y , tenemos que

$$E[X_i X_j] E[Y_i Y_j] = E([X_i Y_i][X_j Y_j]) = \text{Cov}(Z_i, Z_j), \tag{3.2}$$

donde la última igualdad se debe a que $E[Z_i] = 0$, $i = 1, 2, \dots, n$.

Finalmente, combinando (3.1) y (3.2) llegamos a que

$$C = (c_{ij})_{n \times n} = (\text{Cov}(Z_i, Z_j))_{n \times n},$$

lo cual implica, por el Teorema 1.3.13, que C es una matriz positivamente semidefinida.

■

De acuerdo a la demostración probabilística del Teorema de Schur, la idea que se propone para mostrar que una matriz es positivamente semidefinida, es mostrar que es una matriz de covarianza. El siguiente ejemplo muestra el uso de estas ideas.

3.2.2 Ejemplo.- Sean a_1, a_2, \dots, a_n números reales y B una matriz definida como

$$B = \begin{bmatrix}
a_1^2 & a_1^2 & a_1^2 & \cdots & a_1^2 \\
a_1^2 & a_1^2 + a_2^2 & a_1^2 + a_2^2 & \cdots & a_1^2 + a_2^2 \\
& a_1^2 + a_2^2 & a_1^2 + a_2^2 + a_3^2 & \cdots & a_1^2 + a_2^2 + a_3^2 \\
& & a_1^2 + a_2^2 + a_3^2 & & a_1^2 + a_2^2 + a_3^2 + a_4^2 \\
\vdots & \vdots & \vdots & & \vdots \\
a_1^2 & a_1^2 + a_2^2 & a_1^2 + a_2^2 + a_3^2 & \cdots & a_1^2 + a_2^2 + \dots + a_n^2
\end{bmatrix}$$

Para mostrar que B es positivamente semidefinida, es suficiente probar que es la matriz

de covarianza de ciertas v.a's.. En efecto, es fácil ver que B es la matriz de covarianza de las v.a's.

$$Z_1 = X_1, Z_2 = X_1 + X_2, \dots, Z_n = X_1 + X_2 + \dots + X_n$$

donde X_1, X_2, \dots, X_n son v.a's. independientes con media cero y varianzas $a_1^2, a_2^2, \dots, a_n^2$, respectivamente.

3.3 Sistemas de Ecuaciones Lineales

Consideremos un Sistema de Ecuaciones Lineales (SEL) de la forma

$$Bx = c, \tag{3.3}$$

donde $x^t = (x_1, x_2, \dots, x_n)$ es el vector de variables, la matriz $B = (b_{ij})_{n \times n}$ y el vector $c^t = (c_1, c_2, \dots, c_n)$ son conocidos.

Es fácil ver que el SEL

$$x = Ax + c, \tag{3.4}$$

es equivalente al SEL (3.3), donde $I - A = B$ e I denota la matriz identidad.

Existen muchas técnicas para resolver un SEL de la forma (3.3) o (3.4). La dificultad para hacerlo se presenta en la práctica cuando el número de variables es demasiado grande.

Desafortunadamente, desde el punto de vista de las aplicaciones, los SEL de mayor interés son precisamente los de dimensiones grandes, lo cual exige el desarro de algoritmos numéricos fáciles de programar que proporcionen buenas aproximaciones a la solución del SEL.

Un método numérico bien conocido en la literatura del álgebra lineal para resolver el SEL (3.3) o (3.4) es el siguiente.

Supongamos que B es no singular y que

$$\max_i \sum_{j=1}^n |a_{ij}| < 1. \quad (3.5)$$

Bajo esta hipótesis, una solución al SEL (3.3) o (3.4) se obtiene, aplicando el siguiente ecuación recursiva,

$$\begin{aligned} x^{(0)} &\equiv 0; \\ x^{(k+1)} &\equiv Ax^{(k)} + c; \quad k = 0, 1, 2, \dots, \end{aligned} \quad (3.6)$$

donde $x^{(k)}$ son vectores de dimensión n . En efecto suponiendo que $A^{(0)} = I$ e iterando la ecuación (3.6) tenemos

$$\begin{aligned} x^{(k+1)} &= (I + A + A^{(2)} + \dots + A^{(k-1)} + A^{(k)}) c \\ &= \sum_{m=0}^k A^m c. \end{aligned} \quad (3.7)$$

Ahora tomando límite cuando $k \rightarrow \infty$ en (3.7) y usando el hecho de que B es no singular y la condición (3.5) llegamos a que (Rubinstein (1981)),

$$\lim_{k \rightarrow \infty} x^{(k)} = \lim_{k \rightarrow \infty} \sum_{m=0}^k A^m c = (I - A)^{-1} c = B^{-1} c = x. \quad (3.8)$$

Desarrollando (3.7) podemos obtener, en particular, la j -ésima coordenada del vector $x^{(k+1)}$:

$$x_j^{(k+1)} = c_j + \sum_{i_1} a_{ji_1} c_{i_1} + \sum_{i_1, i_2} a_{ji_1} a_{i_1 i_2} c_{i_2} + \dots + \sum_{i_1, i_2, \dots, i_k} a_{ji_1} a_{i_1 i_2} \dots a_{i_{k-1} i_k} c_{i_k}.$$

Una desventaja de este algoritmo proviene de la cantidad de operaciones que requiere su ejecución, y que naturalmente se acentúa cuando la cantidad de variables del SEL es grande.

Un método alternativo, que de cierta manera soluciona el problema antes mencionado,

consiste en representar a la solución del SEL como un parámetro desconocido de una población estadística, y aplicar argumentos probabilísticos para estimarlo. Es decir, la idea del método es expresar al vector de variables x como un parámetro de una población estadística que estimaremos vía observaciones de v.a's.. En este caso, estas v.a's constituirán una cadena de Markov con ciertas propiedades, y las observaciones correspondientes se obtienen simulando la evolución de la cadena.

En este trabajo sólo demostraremos que el método es convergente, sin discutir los aspectos prácticos de la simulación.

3.4 Aproximación a la Solución del SEL

Sea $h^t = (h_1, h_2, \dots, h_n)$ un vector arbitrario y x una solución al SEL (3.3) o (3.4). Consideremos el producto interior

$$\langle h, x \rangle = h_1x_1 + h_2x_2 + \dots + h_nx_n.$$

Tomando $h^t = e_j = \left(\underbrace{0, 0, \dots, 0}_j, 1, 0, \dots, 0 \right)$, tenemos que $\langle h, x \rangle = x_j$, es decir, obtenemos la j -ésima coordenada del vector x . De acuerdo a este hecho, nuestro problema se reduce por lo tanto a encontrar un estimador del producto interior $\langle h, x^{(k+1)} \rangle$, con $x^{(k+1)}$ definido en (3.6), lo cual en vista de (3.8) es una aproximación de $\langle h, x \rangle$.

Sea $X = \{X_t; t = 0, 1, 2, \dots\}$ una cadena de Markov irreducible (Definición 1.5.3) con espacios de estados $S = \{1, 2, \dots, n\}$ y matriz de transición $P = (p_{ij})_{n \times n}$. Recordemos que p_{ij} es la probabilidad de pasar del estado i al estado j en un paso, y denotemos por π_i la distribución inicial de la cadena de Markov X .

Supondremos que la cadena de Markov satisface las siguientes condiciones:

1. $-\pi_i > 0$, si $h_i \neq 0$; $i = 1, 2, 3, \dots, n$;

2. $-p_{ij} > 0$, si $a_{ij} \neq 0$; $i, j = 1, 2, \dots, n$.

Sea k un entero arbitrario. Para construir un estimador del producto interior $\langle h, x^{(k+1)} \rangle$ denotaremos por (i_0, i_1, \dots, i_k) una trayectoria de tamaño k generada por las v.a's. (X_0, X_1, \dots, X_k) . Definimos la variable aleatoria

$$W_m := \frac{a_{i_0 i_1} a_{i_1 i_2} \cdots a_{i_{m-1} i_m}}{p_{i_0 i_1} p_{i_1 i_2} \cdots p_{i_{m-1} i_m}}, \quad m = 1, 2, \dots, k,$$

la cual, tomando $W_0 \equiv 1$, puede expresarse de manera recursiva como:

$$W_m = W_{m-1} \frac{a_{i_{m-1} i_m}}{p_{i_{m-1} i_m}}, \quad m = 1, 2, \dots, k. \quad (3.9)$$

Además, definimos la variable aleatoria

$$\tilde{\eta}_k(h) := \frac{h_{i_0}}{\pi_{i_0}} \sum_{m=0}^k W_m c_{i_m}; \quad (3.10)$$

la cual está asociada a la trayectoria $i_0 \rightarrow i_1 \rightarrow \cdots \rightarrow i_k$ que tiene probabilidad $\pi_{i_0} p_{i_0 i_1} p_{i_1 i_2} \cdots p_{i_{k-1} i_k}$ [ver (1.6)].

3.4.1 Teorema.- Para cada $k \in \mathbb{N}$

$$\mu_{\eta_k} := E[\tilde{\eta}_k(h)] = \langle h, \sum_{m=0}^k A^m c \rangle = \langle h, x^{(k+1)} \rangle.$$

Demostración.

Por el Teorema 1.3.2 (i) tenemos que

$$\begin{aligned} E[\tilde{\eta}_k(h)] &= \sum_{i_0=1}^n \cdots \sum_{i_k=1}^n \eta_k(h) P[X_0 = i_0, X_1 = i_1, \dots, X_k = i_k] \\ &= \sum_{i_0=1}^n \cdots \sum_{i_k=1}^n \eta_k(h) \pi_{i_0} p_{i_0 i_1} p_{i_1 i_2} \cdots p_{i_{k-1} i_k}. \end{aligned} \quad (3.11)$$

En (3.11), $\eta_k(h)$ denota los valores de la variable aleatoria $\tilde{\eta}_k(h)$ asociada con la trayec-

toría $i_0 \rightarrow i_1 \rightarrow \dots \rightarrow i_k$, la cual tiene probabilidad $\pi_{i_0} p_{i_0 i_1} p_{i_1 i_2} \dots p_{i_{k-1} i_k}$.

De (3.9) y (3.10) tenemos

$$\begin{aligned} E[\tilde{\eta}_k(h)] &= E\left[\frac{h_{i_0}}{\pi_{i_0}} \sum_{m=0}^k W_m c_{i_m}\right] \\ &= \sum_{i_0=1}^n \dots \sum_{i_k=1}^n h_{i_0} \sum_{m=0}^k a_{i_0 i_1} a_{i_1 i_2} \dots a_{i_{m-1} i_m} c_{i_m} p_{i_m i_{m+1}} \dots p_{i_{k-1} i_k}. \end{aligned} \quad (3.12)$$

Usando el hecho de que $\sum_{j=1}^n p_{ij} = 1$, la expresión (3.12) la podemos escribir como:

$$E[\tilde{\eta}_k(h)] = \sum_{m=0}^k \sum_{i_0=1}^n \dots \sum_{i_m=1}^n h_{i_0} a_{i_0 i_1} a_{i_1 i_2} \dots a_{i_{m-1} i_m}. \quad (3.13)$$

Por otro lado tenemos que

$$A^2 = (a_{i_0 i_1})_{n \times n} (a_{i_1 i_2})_{n \times n} = (d_{i_0 i_2})_{n \times n}$$

donde

$$d_{i_0 i_2} = \sum_{i_1=1}^n a_{i_0 i_1} a_{i_1 i_2}, \quad i_0, i_2 = 1, 2, \dots, n.$$

En general

$$A^m = (d_{i_0 i_m})_{n \times n}$$

donde

$$d_{i_0 i_m} = \sum_{i_1=1}^n \dots \sum_{i_{m-1}=1}^n a_{i_0 i_1} a_{i_1 i_2} \dots a_{i_{m-1} i_m}, \quad i_0, i_m = 1, 2, \dots, n. \quad (3.14)$$

Sustituyendo (3.14) en (3.13), un cálculo directo nos lleva a que

$$E[\tilde{\eta}_k(h)] = \langle h, \sum_{m=0}^k A^m c \rangle.$$

Tomando en cuenta lo anterior y la relación (3.7) concluimos:

$$E[\tilde{\eta}_k(h)] = \langle h, \sum_{m=0}^k A^m c \rangle = \langle h, x^{(k+1)} \rangle . \blacksquare$$

3.4.1 Algoritmo para la solución del SEL

Tomando en cuenta que la media aritmética es un estimador insesgado del valor esperado (Teorema 1.3.16), podemos estimar $\langle h, x^{(k+1)} \rangle$ simulando N trayectorias de la forma $i_0^{(s)} \rightarrow i_1^{(s)} \rightarrow \dots \rightarrow i_k^{(s)}$, $s = 1, 2, \dots, N$, de longitud k , de la cadena de Markov, y luego calcular su media muestral:

$$\theta_k = \frac{1}{N} \sum_{s=1}^N \eta_k^{(s)}(h) \approx \langle h, x^{(k+1)} \rangle,$$

donde $\eta_k^{(s)}(h)$ es una observación de la variable aleatoria $\eta_k(h)$ correspondiente a la trayectoria generada.

Algoritmo para generar $\langle h, x^{(k+1)} \rangle$

1.- Elegimos cualquier entero $k > 0$.

2.- Simulamos N trayectorias independientes $i_0^{(s)} \rightarrow i_1^{(s)} \rightarrow \dots \rightarrow i_k^{(s)}$, $s = 1, 2, \dots, N$, de la cadena de Markov.

3.- Calculamos

$$\eta_k^{(s)}(h) = \frac{h_{i_0^{(s)}}}{\pi_{i_0}} \sum_{m=0}^k W_m^{(s)} c_{i_m^{(s)}}, \quad s = 1, 2, \dots, N,$$

donde

$$W_m^{(s)} = \frac{a_{i_0^{(s)} i_1^{(s)}} a_{i_1^{(s)} i_2^{(s)}} \cdots a_{i_{m-1}^{(s)} i_m^{(s)}}}{p_{i_0^{(s)} i_1^{(s)}} p_{i_1^{(s)} i_2^{(s)}} \cdots p_{i_{m-1}^{(s)} i_m^{(s)}}}, \quad W_0^{(s)} \equiv 1. \quad (3.15)$$

4.- Calculamos

$$\theta_k = \frac{1}{N} \sum_{s=1}^N \eta_k^{(s)}(h)$$

el cual es un estimador insesgado del producto interior $\langle h, x^{(k+1)} \rangle$.

3.4.2 Observación.- De las expresiones (3.7), (3.8) y el Teorema 3.4.1 tenemos que

$$\lim_{k \rightarrow \infty} E[\eta_k(h)] = E[\eta_\infty(h)] = \langle h, x \rangle.$$

De aquí que el estimador de $\langle h, x \rangle$ toma la forma

$$\theta_\infty = \frac{1}{N} \sum_{s=1}^N \eta_\infty^{(s)}(h),$$

donde

$$\eta_\infty^{(s)}(h) = \frac{h_{i_0^{(s)}}}{\pi_{i_0^{(s)}}} \sum_{m=0}^{\infty} W_m^{(s)} c_{i_m^{(s)}}, \quad s = 1, 2, \dots, N,$$

y $W_m^{(s)}$ como en (3.15).

Como se comentó al principio de esta sección, para encontrar una solución al SEL (3.3) o (3.4) tomamos al vector $h^t = e_i = \left(\underbrace{0, 0, \dots, 0}_j, 1, 0, \dots, 0 \right)$ para tener que $\langle h, x \rangle = x_j$ (j -ésima coordenada del vector x). Ahora consideremos que el estado inicial de la cadena de Markov es j , es decir, $P[X_0 = j] = \pi_j = 1$. De aquí, las trayectorias en las que estamos interesados son de la forma $j \rightarrow i_1 \rightarrow i_2 \rightarrow \dots \rightarrow i_n$. Con esto llegamos al siguiente corolario el cual resuelve el problema planteado originalmente.

3.4.3 Corolario.-

$$E[\eta_k(e_i)] = x_j^{(k+1)}$$

y

$$\theta_k(e_j) = \frac{1}{N} \sum \eta_k^{(s)}(e_j) \approx x_j^{(k+1)}$$

donde

$$\eta_k^{(s)}(e_j) = \sum W_m^{(s)} c_{i_m^{(s)}}$$

y

$$W_m^{(s)} = \frac{a_{ji_1} a_{i_1 i_2} \dots a_{i_{m-1} i_m}}{p_{ji_1} p_{i_1 i_2} \dots p_{i_{m-1} i_m}}, \quad W_0^{(s)} \equiv 1.$$

Capítulo 4

Conclusiones

En este trabajo hemos estudiado algunas aplicaciones de la probabilidad al análisis matemático y al álgebra lineal. Concretamente, se presentaron demostraciones probabilísticas de los Teoremas de Weierstrass, de Taylor, así como para Matrices Semi-definidas Positivas, y se desarrolló un algoritmo para calcular la solución de una clase de sistemas de ecuaciones lineales.

Las demostraciones probabilísticas de estos teoremas presentan ciertas ventajas que se pueden apreciar en el hecho de que los argumentos utilizados son más sencillos, y que proporcionan mayor información respecto a los resultados en cuestión. Por ejemplo, en el Teorema de Weierstrass podemos conocer explícitamente los polinomios que aproximan a la función continua, lo cual a su vez permite calcular la tasa de convergencia. En este mismo sentido, el uso de técnicas de probabilidad nos lleva a proponer una generalización del Teorema de Taylor, sin requerir que la función a aproximar sea diferenciable; más aún, al igual que con el Teorema de Weierstrass, podemos calcular la tasa de convergencia de esta aproximación.

En lo que respecta al algoritmo presentado en el Capítulo 3 para la solución de sistemas de ecuaciones, éste tiene la ventaja de que el número de operaciones que se realizan al aplicarlo es mucho menor respecto a los métodos tradicionales, lo cual es

un aspecto que no se debe descuidar cuando se trabaja con sistemas con un número grande de variables. El precio que se paga en este caso, es que la solución que propone este algoritmo, está representada por el valor esperado de una variable aleatoria. Esto implica que para poder proponer una solución concreta, debemos de tener observaciones de la variable aleatoria, para después calcular su media aritmética, la cual sería una estimación de la solución. Por lo tanto, por cada conjunto de observaciones de la variable aleatoria se tendrá una propuesta de solución. En este sentido, un trabajo interesante podría ser el calcular el nivel de confianza con el que se obtienen dichas soluciones, con lo cual obtendríamos información sobre la aproximación.

Existen otros trabajos que tratan aplicaciones de la probabilidad a otras áreas de la matemática misma. Por ejemplo, el problema de la aguja de Buffon, el cual es tratado en varios trabajos [ver, e.g., Cruse y Lehman (1982), Tuckwell (1988)], se puede adaptar para obtener una aproximación del número π ; Luque-Vásquez (1979) presenta una demostración del teorema de Radon-Nikodym con argumentos probabilísticos; en Rubinstein (1981), Sosa-León (1997), Yakowitz (1970), entre otros, se estudia el problema de integración numérica por el método de Monte Carlo; en Yakowitz (1977) se estudian algoritmos probabilísticos para resolver ecuaciones diferenciales, etc.

Para concluir, haremos un comentario sobre la bibliografía utilizada. La parte sobre teoría de probabilidad se tomó de los libros Cavazos-Cadena (1991), Feller (1985) y Hoel, Port, Stone (1972). Las aplicaciones al análisis matemático se estudiaron en Goldstein (1975) y Feller (1985); al álgebra lineal en Olkin (1985) y Rubinstein (1981).

Bibliografía

- [1] Apostol, T. (1972). *Análisis Matemático*. Reverté S. A., Barcelona, España.
- [2] Cavazos-Cadena, R. (1991) *Fundamentos de Estadística - Parte I*, VII Coloquio del Departamento de Matemáticas, CINVESTAV.
- [3] Cruse, A., Lehman, M. (1982) *Lecciones de Cálculo 2*, Fondo Educativo Interamericano, México.
- [4] Feller, W. (1985) *Introducción a la Teoría de Probabilidades y sus Aplicaciones*, Vol. 2, LIMUSA, 2da. Edición.
- [5] Goldstein, Jerome A.(1975). *Some Applications of the Large Numbers*. Bol. Soc. Bras. Mat., Vol. 6. pp 25-38.
- [6] Hoel, P.G, Port, S.C., Stone, J.Ch. (1972) *Introduction to Stochastic Processes*, Houghton Mifflin, USA.
- [7] Lang, S. (1976). *Álgebra Lineal*. Fondo Educativo Interamericano, México.
- [8] Luque-Vásquez, F. (1979) *Demostración Probabilística del Teorema de Radom-Nikodym*, Tesis de Licenciatura, Departamento de Matemáticas, Universidad de Sonora.
- [9] Olkin, I. (1985) *A probabilistic proof of a theorem of Schur*, The American Mathematical Monthly, 92, pp. 50-51.

- [10] Rubinstein, R. (1981) *Simulation and the Monte Carlo Method*, John Wiley & Sons, USA.
- [11] Rudin, W. (1980). *Principios de Análisis Matemático*. McGraw-Hill, 3ª ed.. México.
- [12] Sosa-León, S. (1997) *Integración Numérica por el Método de Monte Carlo*, Tesis de Licenciatura, Departamento de Matemáticas, Universidad de Sonora.
- [13] Strang, G. (1986) *Algebra Lineal y sus Aplicaciones*, SITESA, México.
- [14] Tuckwell, H.C. (1988) *Elementary Applications of Probability Theory*, Chapman and Hall, Great Britain.
- [15] Yakowitz, S.J. (1970) *Computational Probability and Simulation*, Addeson-Wesley, USA.

RIS T1730