Universidad de Sonora
Departamento de Matemáticas

La Mejor Desigualdad tipo Chebyshev

TESIS

Que para obtener el título de:

Licenciado en Matemáticas

Presenta:

ISMAEL HERNÁNDEZ NORIEGA

Director: Dr. Jesús Adolfo Minjárez Sosa

Hermosillo, Sonora

Agosto 2004

Este trabajo fue financiado por el Consejo Nacional de Ciencia y Tecnología (CONACYT) dentro del proyecto "Control Adaptado de Sistemas Estocásticos", con número de referencia 37239E, bajo la dirección del Dr. Jesús Adolfo Minjárez Sosa.

Índice

Capítulo 1

Introducción

La desigualdad de Chebyshev es uno de los resultados clásicos más importantes de la teoría de probabilidad. Establece que para una variable aleatoria X,

$$P(|X-\mu| \geqslant k) \leqslant \frac{Var(X)}{k^2}, \qquad k>0,$$

donde $\mu = EX$, la cual debe de ser finita.

Generalmente, la demostración de este resultado se basa en la siguiente desigualdad conocida como desigualdad de Markov:

$$P(X \geqslant k) \leqslant \frac{EX^r}{k^r}, \qquad k, r > 0.$$

En la literatura, a este tipo de desigualdades, cuya característica es la comparación de la probabilidad de la cola de la distribución y su valor esperado, se le conoce como *desigualdades tipo Chebyshev.*

Estas desigualdades son la herramienta básica para demostrar resultados no menos importantes como la Ley de los Grandes Números, entre otros. Además tienen aplicaciones en estadística, así como en otras áreas de las matemáticas.

La aplicación directa de las desigualdades tipo Chebyshev es el de aproximar probabilidades por medio de cálculo de cotas.

Es sabido que si conocemos la función de distribución de la variable aleatoria X podemos calcular su valor esperado o esperanza (EX) y su varianza ($Var(X)$), si existen, pero el recíproco no es verdadero, i.e., conociendo EX

y $Var(X)$ no necesariamente se puede reconstruir (o no es tan directo) la función de distribución de X, y por lo tanto, cantidades como $P(|X| > k)$, $k > 0$, son imposibles o difíciles de obtener.

Sin embargo, usar este resultado a veces resulta poco conveniente, ya que en ocasiones la aproximación al valor real de la probabilidad suele dar información insuficiente para nuestro objetivo, i.e., cuando el lado derecho de la desigualdad es mayor o igual a 1. Tal es el caso de una variable aleatoria X tal que $E|X|^r > k^r$. De aquí podemos decir que entre más pequeño sea el valor del lado derecho de la desigualdad tipo Chebyshev, obtenemos información más precisa respecto a las probabilidades.

Este problema ha sido abordado por varios autores, [ver Gauss (1821), Camp y Meidell (1922), Vysochanskii y Petunin (1980), DasGupta (2000)], el cual podemos plantear de la siguiente manera:

Sea X una variable aleatoria distribuida de acuerdo a la función de distribución F. Para $r \geqslant 0$. ¿Cuál es el mejor valor posible de una constante $C^*(r) = C^*(F, r)$ tal que

$$P(|X| > k) \leqslant C^*(r)\frac{E|X|^r}{k^r},$$

para $k > 0$?

El objetivo principal del presente trabajo es obtener una expresión de la mejor constante $C^*(r)$.

El desarrollo de esta teoría se basa principalmente en aplicar técnicas de optimización para el cálculo de estas cotas. Esto nos permite resolver otro tipo de problemas relacionados con la desviación media absoluta de una variable aleatoria, así como problemas de otras áreas de las matemáticas.

La estructura del trabajo es la siguiente. En el Capítulo 2 presentamos la desigualdad de Chebyshev y otras desigualdades relacionadas, así como sus principales aplicaciones. Después, en el Capítulo 3, se desarrollará la teoría que nos lleva al cálculo de la mejor constante en las desigualdades tipo Chebyshev, y en el Capítulo 4 se presentan aplicaciones de esta teoría. Finalmente, en el Capítulo 5 se exponen las conclusiones del trabajo.

Capítulo 2

La Desigualdad de Chebyshev: Aplicaciones y Resultados Relacionados

2.1 La Desigualdad de Chebyshev

La desigualdad de Chebyshev, como ya se ha mencionado, juega un papel importante en la teoría de la probabilidad por sus múltiples aplicaciones. En este capítulo se presentan las más importantes de ellas, así como otros resultados relacionados.

La forma más conocida de la desigualdad de Chebyshev es la siguiente:

Teorema 2.1.1 (Desigualdad de Chebyshev). *Sea X una variable aleatoria con esperanza finita μ. Entonces, para todo $k > 0$,*

$$P(|X - \mu| \geqslant k) \leqslant \frac{Var(X)}{k^2}.$$

En otras palabras, la desigualdad de Chebyshev nos dice que la varianza es una medida de dispersión de los valores de X alrededor de su valor esperado.

La demostración clásica de este teorema que se presenta en la literatura matemática es consecuencia de los siguientes resultados.

Lema 2.1.1. *Sea X una variable aleatoria no negativa. Entonces, para todo $k > 0$,*

$$P(X \geqslant k) \leqslant \frac{EX}{k}.$$

Demostración. Obsérvese que

$$\begin{aligned} EX &= E[XI_{[0,k)}(X) + XI_{[k,\infty)}(X)] \\ &= E[XI_{[0,k)}(X)] + E[XI_{[k,\infty)}(X)] \\ &\geqslant E[XI_{[k,\infty)}(X)], \end{aligned}$$

donde I_A es la función indicadora del conjunto A.

Por otro lado

$$XI_{[k,\infty)}(X) \geqslant kI_{[k,\infty)}(X),$$

lo cual implica

$$\begin{aligned} EX \geqslant E[XI_{[k,\infty)}(X)] &\geqslant E[kI_{[k,\infty)}(X)] \\ &= kE[I_{[k,\infty)}(X)] \\ &= kP(X \geqslant k). \end{aligned}$$

De aquí se obtiene que

$$P(X \geqslant k) \leqslant \frac{EX}{k}.$$

□

Teorema 2.1.2 (Desigualdad de Markov). *Si X es una variable aleatoria no negativa. Entonces, para $k, r > 0$,*

$$P(X \geqslant k) \leqslant \frac{EX^r}{k^r}.$$

Demostración. Se sabe que

$$P(X \geqslant k) = P(X^r \geqslant k^r),$$

y por el lema anterior

$$P(X^r \geqslant k^r) \leqslant \frac{EX^r}{k^r}.$$

Por lo tanto

$$P(X \geqslant k) \leqslant \frac{EX^r}{k^r}.$$

□

Con estos resultados tenemos:

Demostración. (Desigualdad de Chebyshev) Se tiene que $(X-\mu)^2$ es una variable aleatoria no negativa, por lo que podemos aplicar la desigualdad de Markov para $r=2$:

$$P((X-\mu)^2 \geqslant k^2) \leqslant \frac{E(X-\mu)^2}{k^2}.$$

Ahora como $(X-\mu)^2 \geqslant k^2$ si y solo si $|X-\mu| \geqslant k$, entonces

$$P(|X-\mu| \geqslant k) \leqslant \frac{E(X-\mu)^2}{k^2} = \frac{Var(X)}{k^2},$$

lo cual demuestra el Teorema 2.1.1. □

2.2 Aplicaciones de la Desigualdad de Chebyshev

Dentro de las principales aplicaciones de la desigualdad de Chebyshev podemos mencionar las siguientes: *a*) Cálculo de cotas para probabilidades, lo cual es importante cuando es difícil dar un valor exacto de la probabilidad; *b*) Demostración de teoremas límite en probabilidad, y *c*) Cálculo de tamaño de muestra en la aproximación de la media de una población. A continuación haremos una descripción más precisa de cada uno de estos puntos.

2.2.1 Acotamiento de Probabilidades

Esta aplicación es la más directa, y se utiliza para dar una cota superior para $P(|X-\mu| \geqslant k)$, donde $k>0$, usando la esperanza y varianza de la variable aleatoria X, sin necesidad de conocer su distribución.

A continuación se presenta un ejemplo, para ilustrar mejor esta aplicación.

Ejemplo 2.2.1. *Supónga que el número de artículos producidos en una fábrica durante una semana es una variable aleatoria con media 50. Si la varianza de una semana de producción se sabe que es igual a 25, entonces ¿Qué podemos decir acerca de la probabilidad de que en esta semana la producción difiera en más de 10 a la media?*

Solución: Por la desigualdad de Chebyshev

$$P(|X-50| \geqslant 10) \leqslant \frac{Var(X)}{10^2} = \frac{1}{4}$$

entonces la probabilidad de que en la semana de producción el número de artículos exceda en mas de 10 a la media es a lo más 0.25.

2.2.2 Teoremas Límite

La desigualdad de Chebyshev juega un papel muy importante en la demostración de algunos de los más importantes teoremas límite. En esta parte del trabajo se presentarán algunos de ellos.

Sean $X_1, X_2, \ldots, X_n, \ldots,$ variables aleatorias independientes con media y varianza finitas, y denotemos $S_n = \sum_{k=1}^{n} X_k$. Entonces, para todo $n \geqslant 1$

$$E\left(\frac{S_n}{n}\right) = \frac{1}{n}\sum_{k=1}^{n} EX_k,$$

y

$$Var\left(\frac{S_n}{n}\right) = \frac{1}{n^2}\sum_{k=1}^{n} Var(X_k).$$

Ahora, por la desigualdad de Chebyshev,

$$P\left(\left|\frac{S_n}{n} - E\left(\frac{S_n}{n}\right)\right| \geqslant \epsilon\right) \leqslant \frac{Var\left(\frac{S_n}{n}\right)}{\epsilon^2} = \frac{\sum_{k=1}^{n} Var(X_k)}{n^2\epsilon^2}, \qquad (2.1)$$

para todo $n \geqslant 1$.

La relación (2.1) es el punto clave en la demostración de los siguientes resultados.

Teorema 2.2.1 (Ley Débil de los Grandes Números). *Sean $X_1, X_2, \ldots, X_n, \ldots,$ variables aleatorias i.i.d. con media μ y varianza σ^2 finitas. Entonces, para todo $\epsilon > 0$*

$$\lim_{n\to\infty} P\left(\left|\frac{S_n}{n} - \mu\right| \geqslant \epsilon\right) = 0.$$

Demostración. Observemos que $E(\frac{S_n}{n}) = \mu$ y $Var(\frac{S_n}{n}) = \frac{\sigma^2}{n}$. Entonces, por (2.1) obtenemos

$$P\left(\left|\frac{S_n}{n} - \mu\right| \geqslant \epsilon\right) \leqslant \frac{\sigma^2}{n\epsilon^2}.$$

Haciendo $n \longrightarrow \infty$, llegamos a que $P(|\frac{S_n}{n} - \mu| \geqslant \epsilon) \longrightarrow 0$. □

Teorema 2.2.2 (Chebyshev). *Si $X_1, X_2, \ldots, X_n, \ldots$ es una sucesión de variables aleatorias independientes tal que existe alguna $C < \infty$,*

$$Var(X_n) \leqslant C, \quad \forall n \in \mathbb{N},$$

entonces, para cualquier $\epsilon > 0$,

$$\lim_{n\to\infty} P\left(\left|\frac{S_n}{n} - \frac{1}{n}\sum_{k=1}^{n} EX_k\right| < \epsilon\right) = 1.$$

Demostración. Como la sucesión de varianzas es uniformemente acotada, tenemos que

$$Var\left(\frac{S_n}{n}\right) \leqslant \frac{C}{n}, \qquad \forall n \geqslant 1.$$

De acuerdo a la relación (2.1) tenemos que

$$\begin{aligned} P\left(\left|\frac{S_n}{n} - \frac{1}{n}\sum_{k=1}^{n} EX_k\right| < \epsilon\right) &\geqslant 1 - \frac{\frac{C}{n}}{\epsilon^2} \\ &= 1 - \frac{C}{n\epsilon^2}. \end{aligned}$$

Haciendo $n \longrightarrow \infty$, obtenemos

$$\lim_{n\to\infty} P\left(\left|\frac{S_n}{n} - \frac{1}{n}\sum_{k=1}^{n} EX_k\right| < \epsilon\right) \geqslant 1,$$

y como una probabilidad no puede exceder el valor de 1, obtenemos el resultado deseado. □

Como casos particulares del Teorema de Chebyshev tenemos los siguientes.

Teorema 2.2.3. *Suponga que se repite un experimento n veces de forma independiente con dos posibles resultados: éxito y fracaso. Sea X la variable aleatoria que representa el número de éxitos obtenidos.*

1. **(Bernoulli).** *Si la probabilidad de éxito en cada ensayo es p, entonces para cada $\epsilon > 0$,*
$$\lim_{n\to\infty} P\left(\left|\frac{X}{n} - p\right| < \epsilon\right) = 1.$$

2. **(Poisson).** *Si la probabilidad de éxito en el k-ésimo ensayo es p_k, entonces para cada $\epsilon > 0$,*
$$\lim_{n\to\infty} P\left(\left|\frac{X}{n} - \frac{\sum_{m=1}^{n} p_m}{n}\right| < \epsilon\right) = 1.$$

La demostración de este teorema se sigue directamente observando que $X = \sum_{k=1}^{n} X_k$, donde

$$X_k = \begin{cases} 1 & \text{si resulta éxito en el } k\text{-ésimo ensayo} \\ 0 & \text{si resulta fracaso en el } k\text{-ésimo ensayo} \end{cases}$$

y además $EX_k = p_k$, $k = 1, 2, \ldots, n$.

Teorema 2.2.4 (Markov). *Si una sucesión de variables aleatorias X_1, $X_2, \ldots, X_n, \ldots$, es tal que*

$$\frac{1}{n^2} Var\left(\sum_{k=1}^{n} X_k\right) \longrightarrow 0,$$

cuando $n \to \infty$. Entonces para cualquier $\epsilon > 0$

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{k=1}^{n} X_k - \frac{1}{n}\sum_{k=1}^{n} EX_k\right| < \epsilon\right) = 1.$$

Este resultado se demuestra usando argumentos completamente similares a los de los resultados anteriores.

2.2.3 Tamaño de Muestra

La desigualdad de Chebyshev permite encontrar, en términos de la varianza, un tamaño de muestra n que es suficiente para garantizar que la probabilidad de que la desigualdad $|\frac{S_n}{n} - \mu| \geqslant \epsilon$ ocurre es tan pequeña como se desee, lo cual permite obtener una aproximación a la media.

Concretamente, sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de tamaño n, es decir, variables aleatorias i.i.d., y supóngase que tienen media μ y varianza σ^2 finitas. Entonces, por la desigualdad de Chebyshev,

$$P\left(\left|\frac{S_n}{n} - \mu\right| \geqslant \epsilon\right) \leqslant \frac{\sigma^2}{n\epsilon^2}.$$

Sea $\delta > 0$ fijo. De esta manera,

$$P\left(\left|\frac{S_n}{n} - \mu\right| \geqslant \epsilon\right) \leqslant \delta \qquad si \qquad n \geqslant \frac{\sigma^2}{\delta\epsilon^2}.$$

Ejemplo 2.2.2. *Supóngase que $X_1, X_2, \ldots, X_n, \ldots$ son variables aleatorias i.i.d. con distribución de Bernoulli, de tal forma que toman el valor 1 con probabilidad $p = 0.5$. Entonces, ¿De qué tamaño se tiene que tomar la muestra para garantizar que la probabilidad de que la diferencia entre la media aritmética $\frac{S_n}{n}$ y su valor esperado no exceda en más de una décima, sea menor o igual a una centésima?*

Solución: Tenemos que $\mu = p = 0.5$ y $\sigma^2 = p(1-p) = (0.5)^2$. Por la desigualdad de Chebyshev,

$$P\left(\left|\frac{S_n}{n} - \mu\right| \geqslant \epsilon\right) \leqslant \frac{p(1-p)}{n\epsilon^2},$$

para cualquier $\epsilon > 0$. Ahora para $\delta = 0.01$, $\epsilon = 0.1$ se tiene que

$$P\left(\left|\tfrac{S_n}{n} - 0.5\right| \geqslant 0.1\right) \leqslant 0.01 \qquad \text{cuando} \qquad n \geqslant \tfrac{(0.5)^2}{(0.01)^2} = 2500.$$

De esta manera se concluye que se necesitan un tamaño de muestra de al menos 2500 para garantizar que la probabilidad del evento $|\frac{S_n}{n} - 0.5| \geqslant 0.1$ sea menor que 0.01.

2.3 Otras Desigualdades

Existen otras desigualdades relacionadas con la desigualdad de Chebyshev. Como mencionamos al principio del capítulo, una de ellas es la desigualdad de Markov:

$$P(X \geqslant k) \leqslant \frac{EX^r}{k^r}$$

donde X es una variable aleatoria no negativa, $k, r > 0$. Este resultado aparece por primera vez en la edición de 1913 del libro *"The Calculus of Probabilities"* de Markov.

La desigualdad de Markov toma distintas formas. Por ejemplo, sea Y una variable aleatoria no negativa, por lo que $P(Y \geqslant 0) = 1$, y suponga que $EY = \mu < \infty$. Suponga también que $EY^r = \mu_r < \infty$ para algún entero $r > 1$. Entonces,

$$P(Y \geqslant A\mu) \leqslant \frac{1}{A}, \qquad A > 0. \tag{2.2}$$

$$P(Y \geqslant a) \leqslant \frac{\mu}{a}, \qquad a > 0. \tag{2.3}$$

$$P(Y \geqslant A\mu_r) \leqslant \frac{1}{A^r}, \qquad A > 0. \tag{2.4}$$

$$P(Y \geqslant a) \leqslant \frac{\mu_r}{a^r}, \qquad a > 0. \tag{2.5}$$

$$P(Y - \mu \geqslant h) \leqslant \frac{\mu}{\mu + h}, \qquad h > 0. \tag{2.6}$$

En algunos casos la desigualdad de Markov no puede ser mejorada en el sentido de que se puede cumplir la igualdad en (2.6), como es el caso de una variable aleatoria Y con distribución de Bernoulli tal que

$$\begin{aligned} P(Y = 0) &= \frac{h}{\mu + h} \\ P(Y = \mu + h) &= \frac{\mu}{\mu + h}. \end{aligned} \tag{2.7}$$

Desigualdad de Bernstein

La desigualdad de Bernstein establece que

$$P(X \geqslant b) \leqslant \frac{M_X(t)}{e^{bt}},$$

donde $M_X(t)$ es la función generadora de momentos de la variable aleatoria X.

Esta desigualdad es una consecuencia de (2.5). En efecto, sea $Y = e^X$, $a = e^b$ y $r = t$. Como $Y \geqslant 0$ y $a > 0$ se sigue de (2.5) que

$$\begin{aligned} P(e^X \geqslant e^b) &= P(Y \geqslant a) \\ &\leqslant \frac{EY^t}{a^t} \\ &= \frac{Ee^{Xt}}{e^{bt}} \\ &= \frac{M_X(t)}{e^{bt}}, \end{aligned}$$

por lo tanto

$$P(X \geqslant b) \leqslant \frac{M_X(t)}{e^{bt}}.$$

Desigualdad de Gauss

En el año de 1821, Gauss publica este resultado en su libro *"Theoria combinationis observationum erroribus minimis obnoxiae, pars prior"*, el cual dice lo siguiente:

Sea X una variable aleatoria unimodal con moda en cero. Entonces, para $k > 0$,

$$P(|X| > k) \leqslant \frac{4}{9}\frac{EX^2}{k^2}.$$

Desigualdad de Camp-Meidell

En el el año de 1922, B.H. Camp y B. Meidell presentan el siguiente resultado.

Sea X una variable aleatoria unimodal con moda en μ_0. Entonces

$$P(|X - \mu| \geqslant \lambda\sigma) \leqslant \frac{4}{9}\frac{1 + s^2}{(\lambda - s)^2}$$

si $\lambda > s$, donde $\mu = EX$, $\sigma^2 = E(X - \mu)^2$ y $s = \left|\frac{\mu - \mu_0}{\sigma}\right|$.

Con el tiempo este resultado se mejoró hasta obtener que

$$P(|X| \geqslant k) \leqslant \left(\frac{r}{r+1}\right)^r \frac{E|X|^r}{k^r},$$

para toda $r > 0$.

Note que en la desigualdad anterior, si $r = 2$, se obtiene la desigualdad de Gauss.

Capítulo 3

La Mejor Desigualdad tipo Chebyshev

3.1 Planteamiento del Problema

En el capítulo anterior se analizaron varios resultados alrededor de la desigualdad de Chebyshev. Esta desigualdad toma distintas formas dependiendo de la variable aleatoria a analizar (no negativa, valor absoluto, suma parcial de n variables aleatorias i.i.d., etc.), pero todas se deducen de la desigualdad de Markov

$$P(|X| > k) \leqslant \frac{E|X|^r}{k^r}, \qquad k, r > 0.$$

A este tipo de desigualdades se les conoce como *Desigualdades tipo Chebyshev.*

Como se comentó en la Introducción, en algunos casos esta desigualdad no proporciona información relevante sobre la probabilidad que se quiere estimar. Dicha información depende del valor del lado derecho de la desigualdad en el sentido de que entre más pequeño sea éste, es mejor la información obtenida.

Este problema ha sido estudiado por varios autores. Como vimos en el Capítulo 2, Sección 2.3, Gauss (1821) demostró que si la variable aleatoria X es unimodal, con moda en cero, entonces

$$P(|X| \geqslant k) \leqslant \frac{4}{9}\frac{EX^2}{k^2}, \qquad para \quad k > 0. \tag{3.1}$$

Gauss sometió a publicación este resultado en la *Royal Scientific Society* de Göttingen en 1821, precisamente el año en que nació Chebyshev. Esta desigualdad mejora el resultado de Markov en el caso $r = 2$. Después, Camp y Meidell (1922) proponen la siguiente desigualdad:

$$P(|X| > k) \leqslant \left(\frac{r}{r+1}\right)^r \frac{E|X|^r}{k^r}, \qquad \textit{para toda} \quad r > 0,$$

la cual generaliza (3.1). Obsérvese que si $r = 2$ obtenemos (3.1).

Como se observó en el capítulo anterior, en algunos casos las cotas que proporcionan las desigualdades de Markov y Camp-Meidell no pueden ser mejoradas ya que se obtiene la igualdad. Otros ejemplos en esta misma dirección se pueden encontrar en *"Unimodality, Convexity, and Applications"*, publicado en 1988 por Dharmadhikari y Joag-Dev, pero con distribuciones poco comunes y/o muy artificiales.

En el año de 1980, Vysochanskii y Petunin demuestran en su artículo *"Justification of the 3σ rule for unimodal distributions* una desigualdad similar para una variable aleatoria unimodal X, pero con moda no necesariamente en cero. Esta desigualdad toma la forma:

$$P(|X| \geqslant k) \leqslant \max\left\{\frac{4}{9}\frac{EX^2}{k^2}, \frac{4}{3}\frac{EX^2}{k^2} - \frac{1}{3}\right\},$$

ó

$$P(|X| \geqslant k) \leqslant \begin{cases} \frac{4}{9}\frac{EX^2}{k^2}, & si \quad k^2 \geqslant \frac{8}{3}EX^2 \\ \frac{4}{3}\frac{EX^2}{k^2} - \frac{1}{3}, & si \quad k^2 \leqslant \frac{8}{3}EX^2. \end{cases}$$

Por lo tanto, que una desigualdad sea mejor que otra depende de la constante que multiplique al término $\frac{E|X|^r}{k^r}$. La mayoría de los resultados que se han obtenido para mejorar las desigualdades tipo Chebyshev se centran en variables aleatorias unimodales. Sin embargo DasGupta (2000) trata este problema sin esta restricción, lo cual motiva la realización del presente trabajo.

El problema que abordaremos lo podemos plantear de la siguiente manera: Sea X una variable aleatoria distribuida de acuerdo a alguna función de distribución F. Para $r \geqslant 0$. ¿Cuál es el mejor valor posible de una constante $C^*(r) = C^*(F, r)$ tal que

$$P(|X| > k) \leqslant C^*(r)\frac{E|X|^r}{k^r}$$

para cualquier $k > 0$?

3.2 La Mejor Constante

Teorema 3.2.1. *Sea X una variable aleatoria con función de distribución F. Entonces:*

a) $C^*(r) = \frac{\sup_{x>0}\{x^r P(|X|>x)\}}{E|X|^r}$.

b) Si $|X|$ tiene función de densidad f, entonces

$$C^*(r) \leqslant \frac{\sup_{x>0}\{x^{r+1} f(x)\}}{rE|X|^r} \equiv \overline{C}(r).$$

c) Si F es simétrica y absolutamente continua con función característica $\phi(t)$, entonces $C^*(r)$ es también igual a

$$\frac{2}{\pi E|X|^r} \sup_{x>0} \left\{ x^r \int_0^\infty \frac{sen(tx)}{t}(1-\phi(t))dt \right\}.$$

Observemos que la parte a) del Teorema 3.2.1 nos proporciona una expresión exacta para $C^*(r)$, pero con la desventaja de que esta involucra la cantidad que se quiere aproximar. Sin embargo, una vez calculada se puede usar para demostraciones de resultados ó cálculos posteriores donde la desigualdad de Chebyshev juega un papel importante, como las vistas en el Capítulo 2. Las partes b) y c) del teorema proporcionan caminos alternativos para calcular o acotar $C^*(r)$. Por ejemplo, si la densidad de la variable aleatoria $|X|$ es conocida, es posible usar la parte b). De hecho, como veremos más adelante, en ciertos casos $\overline{C}(r)$ es una buena aproximación al valor de $C^*(r)$ y más fácil de calcular. Por otro lado, cuando no es posible o es difícil obtener la densidad de la variable aleatoria X (e.g. densidades de sumas de variables aleatorias), la parte c) toma mayor importancia debido a que $C^*(r)$ depende principlamente de la función característica de la variable aleatoria X.

Demostración. La parte a) se deduce fácilmente de

$$P(|X| > k) \leqslant C(r)\frac{E|X|^r}{k^r}.$$

Ahora, para la parte $b)$ sea $G(x) = x^r P(|X| > x)$, $x > 0$. Observemos que

$$G(x) = x^r P(|X| > x) = x^r(1 - P(|X| \leqslant x)) = x^r - x^r F_{|X|}(x).$$

Notemos que $G(0) = 0$ y $G(x) \leqslant E|X|^r < \infty$. Supongámos por el momento que $\lim_{x\to\infty} G(x) = 0$. De aquí, $G(x)$ tiene un máximo en $[0, \infty)$.

Ahora, para calcular este máximo, usamos el hecho de que $F'_{|X|}(x) = f(x)$, para calcular la derivada de G:

$$\begin{aligned} G'(x) &= rx^{r-1} - rx^{r-1}F_{|X|}(x) - x^r f(x) \\ &= rx^{r-1}(1 - F_{|X|}(x)) - x^r f(x) \\ &= rx^{r-1}P(|X| > x) - x^r f(x). \end{aligned}$$

De aquí tenemos que $sup_{x>0}G(x)$ se alcanza en un punto $x_* = x_*(F, r)$ tal que

$$P(|X| > x_*) = \frac{x_* f(x_*)}{r},$$

por lo que

$$\sup_{x>0}\{x^r P(|X| > x)\} = \frac{x_*^{r+1} f(x_*)}{r} \leqslant \frac{\sup_{x>0}\{x^{r+1} f(x)\}}{r}.$$

Combinando esta desigualdad con la parte $a)$ del teorema obtenemos la parte $b)$.

Finalmente, $\lim_{x\to\infty} G(x) = 0$ se sigue de lo siguiente:

Sea $Y = |X|$, la cual es no negativa, entonces

$$\begin{aligned} EY^r &= \int_0^\infty y^r f_Y(y)dy \\ &\geqslant \int_x^\infty y^r f_Y(y)dy, \quad x > 0 \\ &\geqslant \int_x^\infty x^r f_Y(y)dy \\ &= x^r P(|X| > x). \end{aligned}$$

De aquí vemos que

$$x^r P(|X| > x) \leqslant \int_x^\infty y^r f_Y(y)dy \leqslant EY^r < \infty$$

y como $\lim_{x\to\infty} \int_x^\infty y^r f_Y(y)dy = 0$, se concluye que $\lim_{x\to\infty} x^r P(|X| > x) = 0$.

Para la parte c) se tiene que por hipótesis $F(x) = 1 - F(-x)$ y que existe la densidad f la cual esta dada como [ver Hoel, Port, Stone, (1971)]

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ixt} \phi(t) dt.$$

Entonces

$$\begin{aligned}
F(x) = \int_{-\infty}^{x} f(s)ds &= \int_{-x}^{\infty} f(s)ds \\
&= \int_{-x}^{\infty} \left[\frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ist}\phi(t)dt \right] ds \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} \phi(t) \left[\int_{-x}^{\infty} e^{-ist} ds \right] dt \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} \phi(t) \frac{e^{-ixt}}{-it} dt \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} \phi(t) \frac{(e^{-ixt} - e^{-ixt} + e^{-ixt})}{-it} dt \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{(e^{-ixt} - e^{-ixt})}{-it} \phi(t)dt + \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{-ixt}}{-it} \phi(t)dt \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{2sen(tx)}{t} \phi(t)dt + F(-x) \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{2sen(tx)}{t} \phi(t)dt + 1 - F(x),
\end{aligned}$$

por lo que

$$F(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{sen(tx)}{t} \phi(t)dt + \frac{1}{2}.$$

Combinando lo anterior con el hecho

$$\int_{-\infty}^{\infty} \frac{sen(u)}{u} du = \pi,$$

se tiene que

$$\begin{aligned}
P(|X| > x) &= 2P(X > x) \\
&= 2(1 - F(x)) \\
&= 2\left(1 - \frac{1}{2} - \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{sen(tx)}{t}\phi(t)dt\right) \\
&= 1 - \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{sen(tx)}{t}\phi(t)dt \\
&= \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{sen(tx)}{t}dt - \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{sen(tx)}{t}\phi(t)dt \\
&= \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{sen(tx)}{t}(1 - \phi(t))dt \\
&= \frac{2}{\pi}\int_{0}^{\infty} \frac{sen(tx)}{t}(1 - \phi(t))dt.
\end{aligned}$$

Por lo tanto, la parte $c)$ se sigue de la parte $a)$. □

En el siguiente ejemplo presentamos una comparación de los valores de $C^*(r)$, $\overline{C}(r)$ y $\left(\frac{r}{r+1}\right)^r$. Esta última es la constante en la desigualdad de Camp-Meidell.

Ejemplo 3.2.1. *Sea $X \sim N(0,1)$. Calculemos primero el valor de $C^*(r)$. Sea $G(x) = x^r P(|X| > x)$, $x > 0$. Observemos que*

$$\begin{aligned}
G(x) = x^r P(|X| > x) &= 2x^r P(X > x) \\
&= 2x^r(1 - P(X \leqslant x)) \\
&= 2x^r - 2x^r F(x).
\end{aligned}$$

Como $F'(x) = f(x)$ se tiene que

$$\begin{aligned}
G'(x) &= 2rx^{r-1} - 2rx^{r-1}F(x) - 2x^r f(x) \\
&= 2x^{r-1}(r(1 - F(x)) - xf(x)) \\
&= 2x^{r-1}(rP(X > x) - xf(x)).
\end{aligned}$$

De aquí vemos que $sup_{x>0}G(x)$ se alcanza en un punto x_* tal que

$$P(X > x_*) = \frac{x_*\left(\frac{1}{\sqrt{2\pi}}e^{-\frac{x_*^2}{2}}\right)}{r},$$

para cada $r > 0$. Entonces

$$C^*(r) = \frac{sup_{x>0}\{x^r 2P(X > x)\}}{E_F|X|^r} = \frac{x_*^{r+1}\left(\frac{2}{\sqrt{2\pi}}e^{-\frac{x_*^2}{2}}\right)}{rE_F|X|^r}. \qquad (3.2)$$

Por otro lado, para calcular $\overline{C}(r)$, sea $G(x) = x^{r+1}\left(\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}\right)$. Entonces

$$G'(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}x^r((r+1) - x^2).$$

De aquí podemos obtener que $sup_{x>0}G(x)$ se alcanza cuando $x = \sqrt{r+1}$. Por lo tanto

$$\begin{aligned}
\overline{C}(r) &= \frac{(\sqrt{r+1})^{r+1}(f_{|X|}(\sqrt{r+1}))}{rE|X|^r} \\
&= \frac{(r+1)^{\frac{r+1}{2}}\left(\frac{2}{\sqrt{2\pi}}e^{-\frac{1}{2}(r+1)}\right)}{r\int_{-\infty}^{\infty}|t|^r\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}t^2}dt} \\
&= \frac{(r+1)^{\frac{r+1}{2}}e^{-\frac{1}{2}(r+1)}}{r\int_0^{\infty}t^r e^{-\frac{1}{2}t^2}dt} \\
&= \frac{\frac{2}{r}\left(\frac{1}{2}\right)^{\frac{r-1}{2}}\left(\frac{1}{2}\right)(r+1)^{\frac{r+1}{2}}e^{-\frac{1}{2}(r+1)}}{\left(\frac{1}{2}\right)^{\frac{r-1}{2}}\int_0^{\infty}t^r e^{-\frac{1}{2}t^2}dt} \\
&= \frac{\frac{2}{r}\left(\frac{1}{2}\right)^{\frac{r+1}{2}}(r+1)^{\frac{r+1}{2}}e^{-\frac{1}{2}(r+1)}}{\int_0^{\infty}\left(\frac{1}{2}\right)^{\frac{r-1}{2}}t^{r-1}te^{-\frac{1}{2}t^2}dt} \\
&= \frac{\frac{2}{r}\left(\frac{r+1}{2}\right)^{\frac{r+1}{2}}(r+1)e^{-\frac{1}{2}(r+1)}}{\int_0^{\infty}\left(\frac{1}{2}t^2\right)^{\frac{r-1}{2}}te^{-\frac{1}{2}t^2}dt}.
\end{aligned}$$

Haciendo el cambio de variable $x = \frac{1}{2}t^2$, $dx = tdt$ obtenemos

$$\overline{C}(r) = \frac{\frac{2}{r}\left(\frac{r+1}{2}\right)^{\frac{r+1}{2}}(r+1)e^{-\frac{1}{2}(r+1)}}{\Gamma\left(\frac{r+1}{2}\right)},$$

donde Γ denota la función gamma.

La siguiente tabla muestra los valores de las constantes para diferentes valores de r.

r	$C^*(r)$	$\overline{C}(r)$	$\left(\frac{r}{r+1}\right)^r$
1	.426	.736	.5
2	.331	.463	.444
3	.251	.361	.422
4	.251	.305	.410
5	.229	.269	.402
6	.212	.243	.397
7	.201	.224	.393
8	.187	.208	.390
9	.175	.195	.388
10	.164	.184	.386

Tabla 3.1: Comparación de $C^*(r)$, $\overline{C}(r)$ y $\left(\frac{r}{r+1}\right)^r$ para $X \sim N(0,1)$.

Para $r \geqslant 3$, $\overline{C}(r)$ es más pequeña que la constante de Camp-Meidell, y para $r \geqslant 4$ sus valores son muy cercanos al valor de la mejor constante $C^*(r)$. Note que de la Tabla 3.1, si $X \sim N(\mu, \sigma)$, entonces

$$P(|X-\mu| > k\sigma) < \frac{1}{3k^2},$$

para toda $k > 0$.

Ejemplo 3.2.2. *Consideremos de nuevo una variable aleatoria $X \sim N(0,1)$, y tomemos $k = 2$. Entonces $P(|X| > k) = 0.0455$. Además por (3.2)*

$$C^*(r)\frac{E|X|^r}{k^r} = \frac{2x_*^{r+1} f(x_*)}{rk^r}, \tag{3.3}$$

donde x_* es el punto máximo de $x^r P(|X| > x)$ tal que $P(X > x) = \frac{x_* f(x_*)}{r}$. La siguiente tabla muestra los valores de esta cota para diferentes valores de r.

r	1	2	3	4	5	6	7
Cota (3.3)	.1699	.0828	.0564	.0471	.0461	.0492	.0605

Tabla 3.2: Valores de $C^*(r)\frac{E|X|^r}{k^r}$ para $X \sim N(0,1)$ y $k = 2$.

Capítulo 4

Aplicaciones

Concluimos el trabajo presentando aplicaciones de la teoría que se desarrolló en el Capítulo 3. Estas aplicaciones se centran principalmente en calcular cotas inferiores o superiores de expresiones que se presentan no necesariamente en el área de probabilidad.

4.1 Cota para la función zeta de Riemann

La función *zeta de Riemann*, $\zeta : \mathbb{R} \longrightarrow \mathbb{R}$ se define como

$$\zeta(\alpha) = \sum_{k=1}^{\infty} \frac{1}{k^{\alpha}}, \qquad \alpha > 1.$$

Es conocido que $\zeta(\alpha)$ converge uniformemente para $\alpha > 1$. El problema en que estamos interesados es calcular una cota inferior para $\zeta(\alpha)$.

Sea X una variable aleatoria distribuida geométricamente con parámetro p, i.e., $P(X = x) = pq^x$, $x \geqslant 0$, donde $q = 1 - p$. Sea $G(x) = x^r P(X > x)$. Es fácil ver que $G(x) = x^r q^{[x]+1}$. Observemos que la función $G(x)$ es una función discontinua, por el hecho de que la variable aleatoria X es discreta, por lo que el máximo se encuentra en los puntos de discontinuidad de esta función, los cuales son $\{0, 1, 2, \ldots\}$.

Por otro lado, tenemos que

$$\begin{aligned} G'(x) &= rx^{r-1}q^{x+1} + x^r q^{x+1} \ln q \\ &= x^{r-1}q^{x+1}(r + x \ln q), \end{aligned}$$

y tomando en cuenta el comentario anterior se tiene que el máximo de $G(x)$ es alcanzado cuando $x = \left[\frac{-r}{\ln q}\right]$, donde $[\cdot]$ representa la parte entera. Entonces por el Teorema 3.2.1

$$q^{x+1} = P(X > k) \leqslant \frac{\left(\frac{-r}{\ln q}\right)^r q^{1+[-r/\ln q]}}{k^r}.$$

De aquí, para $r > 0$, $0 < q < 1$

$$\frac{1}{k^r} \geqslant \frac{q^{k+1}}{\left(\frac{-r}{\ln q}\right)^r q^{1+[-r/\ln q]}}.$$

En particular, si $q = e^{-\alpha}$ para $r > 0$, entonces

$$\frac{1}{k^r} \geqslant \frac{e^{-\alpha(k+1)}}{\left(\frac{r}{\alpha}\right)^r e^{-\alpha(1+[r/\alpha])}}.$$

Ahora sumando sobre todos los valores de k se tiene que

$$\begin{aligned} \sum_{k=1}^{\infty} \frac{1}{k^r} &\geqslant \frac{\sum_{k=1}^{\infty} e^{-\alpha(k+1)}}{\left(\frac{r}{\alpha}\right)^r e^{-\alpha(1+[r/\alpha])}} \\ &= \frac{\sum_{k=1}^{\infty} e^{-\alpha k}}{\left(\frac{r}{\alpha}\right)^r e^{-\alpha[r/\alpha]}} \\ &= \frac{\alpha^r e^{\alpha[r/\alpha]}}{r^r(e^\alpha - 1)}. \end{aligned} \tag{4.1}$$

De esta manera, si en (4.1) tomamos $\alpha = r$ tal que $r > 1$, entonces obtenemos una cota inferior para la función zeta de Riemann

$$\zeta(\alpha) = \sum_{k=1}^{\infty} \frac{1}{k^\alpha} \geqslant \frac{\alpha^\alpha e^{\alpha[\alpha/\alpha]}}{\alpha^\alpha(e^\alpha - 1)} = \frac{e^\alpha}{e^\alpha - 1}. \tag{4.2}$$

La siguiente tabla proporciona los valores de la cota inferior en (4.2) para distintos valores de α.

Observación: Otra propiedad de la función zeta de Riemann, la cual será usada en la próxima sección es la siguiente [ver Zaldívar (2002)]:

$$\zeta(r) = \prod_{p\ primo} \frac{1}{1 - \frac{1}{p^r}}, \qquad r > 1. \tag{4.3}$$

α	2	4	6	8	10	12
$\zeta(\alpha)$	1.645	1.082	1.017	1.004	1.001	1.000...
Cota (4.2)	1.157	1.019	1.003	1.001	1.000	1.000...

Tabla 4.1: Valores de la función zeta de Riemann y la cota en (4.2) para ciertos valores de α.

4.2 Exponencial de primos

Es fácil demostrar que la serie $\sum_p e^{-\lambda p}$, donde p es un número primo, converge. Sin embargo, es difícil contar con una expresión de su suma total, por lo cual toma importancia el cálculo de cotas superiores que nos sirvan para análisis posteriores.

Sea $f(x)$ una función con distribución exponencial, i.e., $f(x) = \lambda e^{-\lambda x}$ para $\lambda > 0$ y $x \geqslant 0$. Utilizando la parte *b)* del Teorema 3.2.1, sea $G(x) = x^{r+1}\lambda e^{-\lambda x}$. Entonces derivando con respecto a x se obtiene

$$\begin{aligned} G'(x) &= (r+1)x^r\lambda e^{-\lambda x} - \lambda^2 e^{-\lambda x}x^{r+1} \\ &= \lambda e^{-\lambda x}x^r((r+1) - \lambda x). \end{aligned}$$

De esta manera, $\sup_{x>0} G(x)$ es alcanzado cuando $x = \frac{(r+1)}{\lambda}$. Combinando lo anterior con el hecho de que $P(X > x) = e^{-\lambda x}$, se tiene que

$$e^{-\lambda x} \leqslant \frac{\left(\frac{r+1}{\lambda}\right)^{r+1}\lambda e^{-\lambda\left(\frac{r+1}{\lambda}\right)}}{rx^r},$$

o bien,

$$1 - \frac{1}{x^r} \leqslant 1 - \frac{r\lambda^r e^{r+1}}{(r+1)^{r+1}}e^{-\lambda x},$$

para todo $x, r > 0$. Por otra parte obsérvese que para p primo

$$1 \geqslant \frac{1}{p^r} \geqslant \frac{r\lambda^r e^{r+1}}{(r+1)^{r+1}}e^{-\lambda p} \geqslant 0,$$

entonces

$$0 \leqslant 1 - \frac{1}{p^r} \leqslant 1 - \frac{r\lambda^r e^{r+1}}{(r+1)^{r+1}}e^{-\lambda p} \leqslant 1,$$

por lo que

$$\frac{1}{1 - \frac{1}{p^r}} \geqslant \frac{1}{1 - \frac{r\lambda^r e^{r+1}}{(r+1)^{r+1}}e^{-\lambda p}} \geqslant 1. \tag{4.4}$$

Utilizando (4.3) y (4.4) obtenemos

$$\begin{aligned}\zeta(r) &\geqslant \prod_p \left(\frac{1}{1 - \frac{r\lambda^r e^{r+1}}{(r+1)^{r+1}} e^{-\lambda p}} \right) \\ &\geqslant \prod_p \left(1 + \frac{r\lambda^r e^{r+1}}{(r+1)^{r+1}} e^{-\lambda p} \right) \\ &\geqslant 1 + \frac{r\lambda^r e^{r+1}}{(r+1)^{r+1}} \sum_p e^{-\lambda p}.\end{aligned}$$

De aquí llegamos a que

$$\sum_p e^{-\lambda p} \leqslant \frac{(r+1)^{r+1}}{r\lambda^r e^{r+1}} (\zeta(r) - 1)$$

para todo $r > 1$.

4.3 Cotas para probabilidades de colas

La distribución normal $N(\mu, \sigma^2)$ tiene la propiedad [ver Dorogovtsev (1997)]

$$P(|X - \mu| > k\sigma) \leqslant \frac{2\phi(k)}{k}, \qquad k > 0,$$

donde $\phi(k)$ es la densidad de una variable aleatoria normal estándar.

Como podemos observar, esta expresión nos permite obtener una comparación entre la función de distribución y la densidad de la variable aleatoria normal. Aún más, nos permite calcular cotas para la probabilidad de la cola de la distribución en términos de la densidad, lo cual para fines prácticos sería muy útil por su fácil evaluación.

La pregunta que surge es si expresiones de este tipo se pueden obtener para otras distribuciones, es decir, si existe una constante $D(k) := D(k, f)$, $k > 0$, tal que

$$P(|X| > k) \leqslant D(k) f(k)$$

para una variable aleatoria X con densidad f.

El siguiente resultado nos ofrece una respuesta.

Proposición 4.3.1. *Sea X una variable aleatoria con función de distribución $F(x)$ y $|X|$ con función de densidad $f(x)$. Suponga que f es diferenciable y que $\psi(x) := -\frac{xf'(x)}{f(x)}$ es creciente para $x > 0$. Entonces para cualquier k tal que $\psi(k) > 1$*

$$P(|X| > k) \leqslant \frac{k}{\psi(k) - 1} f(k). \tag{4.5}$$

Demostración. De la parte $b)$ del Teorema 3.2.1 se tiene que para cada $r > 0$

$$P(|X| > k) \leqslant \frac{\sup_{x>0}\{x^{r+1} f(x)\}}{rk^r}. \tag{4.6}$$

En particular tomemos $r = \psi(k) - 1$. Por otro lado, sea $G(x) = x^{r+1} f(x)$ y derivando con respecto a x se llega a que

$$\begin{aligned} G'(x) &= (r+1)x^r f(x) + x^{r+1} f'(x) \\ &= x^r((r+1)f(x) + xf'(x)), \end{aligned}$$

por lo que $G(x)$ se maximiza cuando

$$r + 1 = -\frac{xf'(x)}{f(x)} = \psi(x).$$

De aquí, $\psi(k) - 1 + 1 = \psi(x)$, i.e., $\psi(k) = \psi(x)$, y como ψ es creciente para $x > 0$, se concluye que $k = x$. Por lo tanto

$$P(|X| > k) \leqslant \frac{k^{r+1} f(k)}{rk^r} = \frac{kf(k)}{\psi(k) - 1}$$

□

A continuación se muestran algunos ejemplos para ilustrar la Proposición 4.3.1.

Ejemplo 4.3.1. *Sea X una variable aleatoria con densidad t de Student $f_\alpha(x)$ con α grados de libertad. Mostraremos que para todo $k > 1$,*

$$P(X > k) \leqslant \frac{k(\alpha + k^2)}{\alpha(k^2 - 1)} f_\alpha(k). \tag{4.7}$$

Primero, por definición tenemos que, para $x \geqslant 0$,

$$f_\alpha(x) = \frac{\Gamma\left(\frac{\alpha+1}{2}\right)\left(1 + \frac{x^2}{\alpha}\right)^{-\frac{(\alpha+1)}{2}}}{\sqrt{\alpha\pi}\,\Gamma\left(\frac{\alpha}{2}\right)}$$

y

$$f'_\alpha(x) = -\frac{\Gamma\left(\frac{\alpha+1}{2}\right)\frac{(\alpha+1)}{2}\left(1+\frac{x^2}{\alpha}\right)^{-\frac{(\alpha+1)}{2}-1}\frac{2x}{\alpha}}{\sqrt{\alpha\pi}\,\Gamma\left(\frac{\alpha}{2}\right)}.$$

De aquí

$$\begin{aligned}\psi(x) &= \frac{x\Gamma\left(\frac{\alpha+1}{2}\right)\frac{(\alpha+1)}{2}\left(1+\frac{x^2}{\alpha}\right)^{-\frac{(\alpha+1)}{2}-1}\frac{2x}{\alpha}}{\Gamma\left(\frac{\alpha+1}{2}\right)\left(1+\frac{x^2}{\alpha}\right)^{-\frac{(\alpha+1)}{2}}}\\ &= \frac{x\left(\frac{\alpha+1}{2}\right)\left(\frac{2x}{\alpha}\right)}{1+\frac{x^2}{\alpha}}\\ &= \frac{x^2(\alpha+1)}{\alpha+x^2},\end{aligned}$$

lo cual implica

$$\psi(x) - 1 = \frac{\alpha(x^2-1)}{\alpha+x^2}.$$

Por lo tanto de (4.5) *se concluye que*

$$P(X > k) \leqslant \frac{k(\alpha+k^2)}{\alpha(k^2-1)} f_\alpha(k)$$

para todo $k > 1$.

Ejemplo 4.3.2. *Sea* X *una variable aleatoria con densidad gamma* $f_{\alpha,\lambda}(x) = e^{-\lambda x}x^{\alpha-1}\frac{\lambda^\alpha}{\Gamma(\alpha)}$. *En este caso tenemos que para todo* $k > E(X) = \frac{\alpha}{\lambda}$

$$P(X > k) \leqslant \frac{k}{k\lambda-\alpha} f_{\alpha,\lambda}(k). \tag{4.8}$$

En efecto, derivando la función $f_{\alpha,\lambda}(x)$ *obtenemos*

$$f'_{\alpha,\lambda}(x) = e^{-\lambda x}x^{\alpha-2}\frac{\lambda^\alpha}{\Gamma(\alpha)}(-\lambda x+\alpha-1).$$

De aquí,

$$\begin{aligned}\psi(x) &= \frac{-xe^{-\lambda x}x^{\alpha-2}\frac{\lambda^\alpha}{\Gamma(\alpha)}(-\lambda x+\alpha-1)}{e^{-\lambda x}x^{\alpha-1}\frac{\lambda^\alpha}{\Gamma(\alpha)}}\\ &= \lambda x - \alpha + 1,\end{aligned}$$

y por lo tanto, de (4.5) *se obtiene que*

$$P(X > k) \leqslant \frac{k}{k\lambda - \alpha} f_{\alpha,\lambda}(k),$$

para todo $k > E(X) = \frac{\alpha}{\lambda}$.

Aplicando argumentos similares a los Ejemplos 4.3.1 y 4.3.2 podemos obtener las siguientes cotas para las distribuciones F y *Beta*.

Ejemplo 4.3.3. *Sea* X *una variable aleatoria con* F, $f_{\alpha,\beta}(x)$ *con* α *y* β *grados de libertad. Entonces para* $k > 1$

$$P(X > k) \leqslant \frac{2k(\beta + \alpha k)}{\alpha\beta(k-1)} f_{\alpha,\beta}(k). \tag{4.9}$$

Ejemplo 4.3.4. *Sea* X *una variable aleatoria con densidad* $Beta(\alpha,\beta), f_{\alpha,\beta}$. *Si* $\beta > 1$, *entonces para todo* $k > \frac{\alpha}{\alpha+\beta-1}$

$$P(X > k) \leqslant \frac{k(1-k)}{k(\alpha+\beta-1) - \alpha} f_{\alpha,\beta}(k). \tag{4.10}$$

4.4 Acotamiento de la desviación media absoluta

El resultado del Teorema 3.2.1 se puede extender a una infinidad de variables aleatorias, como es el caso de $Y = |X - \mu|$, donde $\mu = EX$, por lo que

$$P(|X - \mu| > k) \leqslant C_\mu^*(r) \frac{E|X-\mu|^r}{k^r}, \tag{4.11}$$

donde $C_\mu^*(r) = \frac{\sup_{x>0}\{x^r P(|X-\mu|>x)\}}{E|X-\mu|^r}$. Esta cota es muy útil para encontrar una familia de cotas superiores para la desviación media absoluta $E|X - \mu|$.

Para lo siguiente, usaremos la siguiente notación:

$p = P(X = \mu)(= 0$ si X es continua$)$.

$C_\mu^*(r) = \frac{\sup_{x>0}\{x^r P(|X-\mu|>x)\}}{E|X-\mu|^r}$.

$\alpha(r) = \frac{r}{r-1}(1-p)^{1-(1/r)}(C_\mu^*(r))^{1/r}$, para $r > 1$.

Teorema 4.4.1. *Para toda* $r > 1$

$$E|X - \mu| \leqslant \alpha(r)(E|X - \mu|^r)^{1/r}. \tag{4.12}$$

Antes de empezar con la demostración, nótese que si $\alpha(r) = 1$ en (4.12) obtenemos la desigualdad de Hölder, por lo que para nuestros fines, (4.12) tendrá sentido si $\alpha(r) < 1$.

Demostración. Sean $r_1, r_2 \in \mathbb{R}$ tal que $0 \leqslant r_1 < 1 < r_2 < \infty$, entonces tenemos que

$$P(|X - \mu| > k) \leqslant C_\mu^*(r_1)\frac{E|X - \mu|^{r_1}}{k^{r_1}}$$

y

$$P(|X - \mu| > k) \leqslant C_\mu^*(r_2)\frac{E|X - \mu|^{r_2}}{k^{r_2}}.$$

Ahora fijemos $\delta > 0$, entonces

$$\begin{aligned}
E|X - \mu| &= \int_0^\infty P(|X - \mu| > x)dx \\
&= \int_0^\delta P(|X - \mu| > x)dx + \int_\delta^\infty P(|X - \mu| > x)dx \\
&\leqslant C_\mu^*(r_1)E|X - \mu|^{r_1}\int_0^\delta \frac{dx}{x^{r_1}} + C_\mu^*(r_2)E|X - \mu|^{r_2}\int_\delta^\infty \frac{dx}{x^{r_2}} \\
&= C_\mu^*(r_1)E|X - \mu|^{r_1}\frac{\delta^{1-r_1}}{1 - r_1} + C_\mu^*(r_2)E|X - \mu|^{r_2}\frac{\delta^{1-r_2}}{r_2 - 1}.
\end{aligned}$$

En particular, para todo $\delta > 0$ y $r_2 > 1$ y tomando $r_1 = 0$

$$\begin{aligned}
E|X - \mu| &\leqslant C_\mu^*(0)\delta + C_\mu^*(r_2)E|X - \mu|^{r_2}\frac{\delta^{1-r_2}}{r_2 - 1} \\
&= \sup_{x>0}\{P(|X - \mu| > x)\}\delta + C_\mu^*(r_2)E|X - \mu|^{r_2}\frac{\delta^{1-r_2}}{r_2 - 1} \\
&= P(|X - \mu| > 0)\delta + C_\mu^*(r_2)E|X - \mu|^{r_2}\frac{\delta^{1-r_2}}{r_2 - 1} \\
&= (1 - P(X = \mu))\delta + C_\mu^*(r_2)E|X - \mu|^{r_2}\frac{\delta^{1-r_2}}{r_2 - 1} \\
&= (1 - p)\delta + C_\mu^*(r_2)E|X - \mu|^{r_2}\frac{\delta^{1-r_2}}{r_2 - 1}.
\end{aligned} \tag{4.13}$$

Por otra parte, para $a, b > 0$ sea $G(\delta) = a\delta + b\delta^{1-r_2}$, entonces derivando esta función con respecto a δ tenemos

$$G'(\delta) = a + (1 - r_2)b\delta^{-r_2},$$

la cual se minimiza cuando $\delta = \left(\frac{b(r_2-1)}{a}\right)^{1/r_2}$.

Con el resultado anterior y por (4.13) se sigue

$$\begin{aligned}
E|X-\mu| &\leqslant (C_\mu^*(r_2))^{1/r_2}(E|X-\mu|^{r_2})^{1/r_2}(1-p)^{\frac{r_2-1}{r_2}} \\
&\quad + \frac{(C_\mu^*(r_2))^{1/r_2}(E|X-\mu|^{r_2})^{1/r_2}}{r_2-1}(1-p)^{\frac{r_2-1}{r_2}} \\
&= (C_\mu^*(r_2))^{1/r_2}(E|X-\mu|^{r_2})^{1/r_2}(1-p)^{\frac{r_2-1}{r_2}}\left(1+\frac{1}{r_2-1}\right) \\
&= (C_\mu^*(r_2))^{1/r_2}(E|X-\mu|^{r_2})^{1/r_2}(1-p)^{\frac{r_2-1}{r_2}}\left(\frac{r_2}{r_2-1}\right) \\
&= \alpha(r_2)(E|X-\mu|^{r_2})^{1/r_2}.
\end{aligned}$$

Como $r_2 > 1$, entonces para cualquier $r > 1$

$$E|X-\mu| \leqslant \alpha(r)(E|X-\mu|^r)^{1/r}.$$

□

La cota (4.12) es muy usada cuando la expresión exacta de $E|X-\mu|$ es muy difícil de obtener o mas aún cuando no se puede calcular, y en cambio se cuenta con una expresión mas sencilla para $E|X-\mu|^r$, como es el caso cuando r es par, y de esta manera obtener una cota para $E|X-\mu|$.

Capítulo 5

Conclusiones

En este trabajo se estudiaron las desigualdades tipo Chebyshev, así como sus principales aplicaciones y algunos resultados relacionados.

La expresión general de las desigualdades tipo Chebyshev es

$$P(|X| > k) \leqslant C^*(r)\frac{E|X|^r}{k^r}, \qquad k > 0,$$

la cual, dependiendo de las condiciones sobre la variable aleatoria X, toma distintas formas. Por ejemplo

$$P(|X - \mu| \geqslant k) \leqslant \frac{Var(X)}{k^2}, \tag{5.1}$$

$$P(|X| > k) \leqslant \frac{E|X|^r}{k^r}, \tag{5.2}$$

$$P\left(\left|\frac{S_n}{n} - E\left(\frac{S_n}{n}\right)\right| \geqslant \epsilon\right) \leqslant \frac{\sum_{k=1}^{n} Var(X_k)}{n^2\epsilon^2}, \tag{5.3}$$

$$P(|X| > k) \leqslant \frac{4}{9}\frac{EX^2}{k^2}, \tag{5.4}$$

$$P(|X| \geqslant k) \leqslant \left(\frac{r}{r+1}\right)^r \frac{E|X|^r}{k^r}, \tag{5.5}$$

$$P(|X| \geqslant k) \leqslant \max\left\{\frac{4}{9}\frac{EX^2}{k^2}, \frac{4}{3}\frac{EX^2}{k^2} - \frac{1}{3}\right\}, \tag{5.6}$$

$$P(|X| \geqslant k) \leqslant \begin{cases} \frac{4}{9}\frac{EX^2}{k^2}, & si \quad k^2 \geqslant \frac{8}{3}EX^2 \\ \frac{4}{3}\frac{EX^2}{k^2} - \frac{1}{3}, & si \quad k^2 \leqslant \frac{8}{3}EX^2 \end{cases} \tag{5.7}$$

siendo (5.1) y (5.2) las desigualdades mas conocidas.

Como es sabido, en muchos casos las desigualdades (5.1) y (5.2) no proporcionan información respecto a la probabilidad $P(|X| > k)$, ya que el lado derecho de estas desigualdades resulta ser mayor o igual a 1. En este sentido, los resultados (5.4) – (5.7) tienen mayor importancia. Es decir, como podemos observar, entre más pequeña sea la constante $C^*(r)$ es más probable que se obtenga información útil respecto a la probabilidad $P(|X| > k)$, es decir, obtener una mejor desigualdad tipo Chebyshev.

En este trabajo fue analizado este problema tomando como punto de partida la pregunta ¿Cuál es el mejor valor posible de una constante $C^*(r) = C^*(F, r)$ tal que

$$P(|X| > k) \leqslant C^*(r)\frac{E|X|^r}{k^r},$$

para $k > 0$? Este problema fue resuelto por DasGupta (2000) bajo hipótesis específicas sobre la distribución de la variable aleatoria X, [ver Teorema 3.2.1] y fue la base para el desarrollo del presente trabajo.

La solución a estos problemas se centró principalmente en analizar las propiedades de la función $G(x) = x^r P(|X| > x)$. Este mismo análisis sirvió para establecer otras aplicaciones, como por ejemplo, obtener cotas para $P(|X| > x)$ en términos de la densidad de la variable aleatoria, y para la desviación media absoluta de una variable aleatoria. Además se obtuvieron resultados en otras áreas de las matemáticas, como cotas para la función zeta de Riemann y la serie exponencial de primos.

Por último, los resultados de nuestro trabajo se establecieron sólo para variables aleatorias univariadas, lo cual sigue siendo válido para variables aleatorias definidas como funciones de vectores aleatorios. Es decir, sea $\mathbb{X} = (X_1, X_2, \ldots, X_n)$ un vector aleatorio de dimensión n y $F : \mathbb{R}^n \longrightarrow \mathbb{R}$ una función arbitraria. Entonces podemos aplicar los resultados a la variable aleatoria $Y = F(X_1, X_2, \ldots, X_n)$.

Se pueden extender los resultados expuestos al caso de vectores aleatorios, pero considerando clases muy particulares de distribuciones multivariadas, como por ejemplo, para las distribuciones unimodales esféricamente simétricas. Un ejemplo de este tipo de distribuciones es la normal multivariada. La forma que toman las desigualdades tipo Chebyshev, por ejemplo, para el caso $n = 2$, es [ver DasGupta (2000)]:

$$P(||\mathbb{X}|| > k) \leqslant C^*_{\mathbb{X}}(r)\frac{E||\mathbb{X}||^r}{k^r}, \quad k > 0.$$

Bibliografía

[1] DASGUPTA, A. (2000), *Best constants in Chebyshev inequalities with various applications*, Metrika, 51, 186-201.

[2] DOROGOVTSEV, A. Ya. (1997), *Probability Theory: Collection of Problems*, Transalation of Mathematical Monographs, AMS, 163, 87.

[3] GNEDENKO. B. V. (1963), *The Theory of Probability*, 2da. ed., Nueva York: Chelsea Publishing.

[4] HOEL, P., PORT, S. y STONE, C. (1971), *Introduction to Probability Theory*, Boston: Hougthon Mifflin.

[5] KOTZ, S. y JHONSON, N. (1982), *Encyclopedia of Statistical Sciences*, Volume 1, Nueva York: Jhon Wiley & Sons.

[6] SELLKE, T.M. y SELLKE, S.H. (1997) *Chebyshev inequalities for unimodal distributions"*, American Statistical Association, 51, 34-40.

[7] ZALDÍVAR, F. (2002), *La función zeta de Riemann*, Miscelánea Matemática, 36, 66-67.