

Propuesta para Proyecto de Grado

Martín Prado Guerra 201922940¹

¹Universidad de los Andes

Asesor de Tesis:

Adolfo Jose Quiroz Salazar¹

Dirección de email: aj.quiroz1079@uniandes.edu.co

RESUMEN

Título Tentativo: Ejemplos y Aplicaciones a las Desigualdades de Concentración.

Las desigualdades de concentración proporcionan cotas a la probabilidad de que una variable aleatoria se encuentre adentro de un rango alrededor de su valor esperado. Algunas de las desigualdades más conocidas son la desigualdad de Markov y la desigualdad de Chebyshev. Estas dos desigualdades dan información valiosa sobre la distribución de las colas de una variable aleatoria, por lo que han sido ampliamente aplicadas en el análisis matemático de diversos problemas. Sin embargo, ante situaciones que precisen de mayor exactitud, se requieren cotas más estrictas, y aquí es donde inicia el estudio de las diferentes clases de desigualdades de concentración. Dentro de este proyecto se recogerán algunas de estas desigualdades y se estudiarán sus aplicaciones en detalle.

Palabras Clave: Desigualdades de Concentración; Ejemplos; Probabilidad; Cotas

MOTIVACIÓN

Ante la imposibilidad de calcular de manera exacta la distribución de una variable aleatoria, las desigualdades de concentración ofrecen información importante sobre la probabilidad de obtener valores atípicos a cualquier distancia de la media. En primer lugar, encontramos las desigualdades de Markov y Chebyshev, las cuales brindan una aproximación inicial a la clase de problemas que abordan estas desigualdades.

Proposición 1 Para cualquier variable aleatoria no negativa X con valor esperado μ , varianza σ^2 y $t > 0$ se tiene que:

$$P\{X \geq t\} \leq \frac{\mu}{t}, \quad (\text{Desigualdad de Markov})$$

$$P\{|X - \mu| \geq t\} \leq \frac{\sigma^2}{t^2}. \quad (\text{Desigualdad de Chebyshev})$$

Las pruebas de estas desigualdades se pueden encontrar en las referencias[5]. El siguiente ejemplo motivará el uso de otra clase de desigualdades.

Sean $X_i \sim \text{Bern}_{1/2}$ variables aleatorias independientes las cuales describen el lanzamiento de una moneda justa. La suma $S_N = \sum_{i=0}^N X_i$ tiene una distribución binomial con media $N/2$ y varianza $N/4$. Si quisiéramos conocer una cota a la probabilidad de obtener más $3/4$ lanzamientos como caras o sellos podríamos utilizar Chebyshev para afirmar

$$P\{S_N \geq 3N/4\} \cup \{S_N \leq N/4\} = P\{|S_N - N/2| \geq N/4\} \leq \frac{N/4}{N^2/16} = \frac{4}{N}.$$

A pesar de que esta cota converge linealmente a 0, en la práctica esta probabilidad disminuye a un ritmo más acelerado. La siguiente desigualdad ofrece una mejor cota a esta probabilidad.

Proposición 2 Sean X_1, \dots, X_N variables aleatorias independientes e idénticamente distribuidas las cuales se encuentran acotadas, es decir que para todo $i \leq N$: $X_i \in [m, M]$, sea μ la media de todas las variables y sea $S_N = \sum_{i=0}^N X_i$. Se sigue que para cualquier $t > 0$,

$$P\{|S_N - N\mu| \geq t\} \leq 2 \exp\left(\frac{-2t^2}{N(M-n)^2}\right). \quad (\text{Desigualdad de Hoeffding})$$

A través de este resultado se puede concluir que la distribución de las colas converge a 0 a una tasa exponencial. En este caso $X_i \in [0, 1]$, y al igual que en el ejemplo pasado, $t = N/4$. Por ende,

$$P\{|S_N - N/2| \geq N/4\} \leq 2\exp(-N/8).$$

17 Similarmente al caso anterior, existen numerosos ejemplos en los que las desigualdades de concentración
18 se encuentran dentro del corazón del problema. Por lo que el objetivo del proyecto será estudiar a gran
19 detalle un gran número de desigualdades y estudiar los casos dónde son utilizadas.

20 PROGRAMA DEL TRABAJO

21 Las 5 referencias de abajo son las principales fuentes que se van a utilizar tanto para la teoría como
22 para los ejemplos. El plan resultará en acabar primero la teoría junto con los ejercicios relevantes en las
23 fuentes [1, 5, 2], para luego entender a detalle los ejemplos de cada una de las fuentes.

24 Cada fuente cubre temas y necesidades diferentes. Por ejemplo, la referencia [2] posee tanto teoría como
25 aplicaciones a optimización y teoría de decisión. El libro [1], por otro lado, se utilizará como referencia
26 teórica para estudiar desigualdades en Martingalas y ejercicios relevantes a este tema. Los artículos [3, 4]
27 serán las referencias de los principales 2 ejemplos para exponer aplicaciones.

28 Finalmente, si el tiempo lo permiten, se incluirá en orden de prioridad, más ejemplos sobre aplicaciones,
29 simulaciones para ilustrar algunas de las aplicaciones y una versión del texto en inglés con la intención de
30 hacer la sustentación también en inglés. Esta última parte no se ha discutido todavía con el asesor, pero, a
31 lo largo del próximo periodo académico se determinará si es viable redactarla en inglés.

32 Aquí adjunto un cronograma para el programa de trabajo. Las fechas se encuentran apretadas con la
33 intención de dejar espacio justo la adición de capítulos adicionales con otras referencias que surjan en el
34 camino.

Fechas	Descripción
05/06/23 - 02/07/23	Lecturas y consultas de referencias [2,5].
03/07/23 - 09/07/23	Boceto de definiciones, teoremas y ejemplos de referencias [2,5].
10/07/23 - 23/07/23	Lectura y ejercicios de referencia [1], capítulos 5,6 y 7.
24/07/23 - 30/07/23	Boceto de definiciones, teoremas y ejemplos de referencias [1].
31/07-23 - 13/08/23	Lecturas y consultas de referencia [3].
14/07-23 - 20/08/23	Boceto de definiciones, teoremas y ejemplos de referencias [3].
21/08/23 - 03/09/23	Lecturas y consultas de referencia [4].
04/08/23 - 10/09/23	Boceto de definiciones, teoremas y ejemplos de referencias [4].
11/09/23 - 17/09/23	Rendición de cuentas y planeación de la estructura del texto en base a los bocetos
18/09/23 - 01/10/23	Escritura y correcciones.
02/10/23 - límite	Adición del contenido opcional adicional + Preparación de la presentación.

36 REFERENCES

- 37 [1] Noga Alon and Joel H Spencer. *The probabilistic method*. John Wiley & Sons, 2016.
- 38 [2] Stéphane Boucheron, Gábor Lugosi, and Olivier Bousquet. Concentration inequalities. *Advanced*
39 *Lectures on Machine Learning: ML Summer Schools 2003, Canberra, Australia, February 2-14, 2003,*
40 *Tübingen, Germany, August 4-16, 2003, Revised Lectures*, pages 208–240, 2004.
- 41 [3] Mateo Díaz, Adolfo J Quiroz, and Mauricio Velasco. Local angles and dimension estimation from data
42 on manifolds. *Journal of Multivariate Analysis*, 173:229–247, 2019.
- 43 [4] H Gzyl, R Jiménez, and AJ Quiroz. The physicist’s approach to the travelling salesman problem—ii.
44 *Mathematical and Computer Modelling*, 13(7):45–48, 1990.
- 45 [5] Roman Vershynin. *High-dimensional probability: An introduction with applications in data science*,
46 volume 47. Cambridge university press, 2018.

47

.....
Martín Prado Guerra
(Aspirante a Proyecto de Grado)

.....
Adolfo José Quiroz Salazar
(Asesor de Tesis)