

El tema general de este trabajo se enmarca en la teoría de procesos empíricos y sus aplicaciones a la estadística. Sin embargo, conviene señalar que las herramientas utilizadas e incluso algunas de las contribuciones obtenidas podrían perfectamente incluirse en áreas como el análisis funcional, teoría de la medida y probabilidad.

El origen de la teoría de procesos empíricos se remonta a los años treinta aunque su mayor desarrollo y popularidad llegó a principios de los años ochenta, impulsados en parte por el libro de David Pollard [Pollard, 2012], cuya primera edición se publicó en 1984. En estado actual de desarrollo, la teoría de procesos empíricos es un tema altamente técnico que requiere profundizar en otras ramas de las matemáticas como el análisis funcional o la teoría de la medida. La idea básica subyacente es una de las más elementales en la teoría de la probabilidad: la aproximación asintótica de las frecuencias relativas de un cierto suceso en pruebas repetidas e independientes a la probabilidad del mismo. Este hecho básico se ha formulado a lo largo de la historia en diferentes formas, desde la clásica “ley del azar” o “postulado de Bernoulli” sobre estabilidad de las frecuencias, hasta su deducción como teorema en el marco de la construcción axiomática de Kolmogorov. En resumen, puede decirse que la teoría de procesos empíricos se ocupa de las generalizaciones de esta idea básica, especialmente las relacionadas con la validez “uniforme” de la aproximación, que surgen en los ámbitos más variados, incluyendo los desarrollos más recientes de estadística en variedades y estadística con datos funcionales.

El método plug-in para la norma del supremo

En muchas aplicaciones prácticas es preciso estimar la norma del supremo o alguna cantidad relacionada. Esto se hace, por ejemplo, en el test de Kolmogorov-Smirnov: se estima la distancia de Kolmogorov entre dos variables aleatorias, distancia entre sus funciones de distribución en el espacio de las funciones acotadas, mediante la distancia en términos de la norma del supremo entre las contrapartes empíricas. Esta idea es conocida en estadística como *método plug-in*.

El objetivo de este trabajo es mostrar la potencia del método delta funcional y la teoría moderna de procesos empíricos, tratados en [van der Vaart & Wellner, 1996], para obtener expresiones asintóticas de estimadores inspirados en el método plug-in en los que interviene la norma del supremo. Este enfoque separa el estudio de los estadísticos en dos partes: la convergencia débil de los elementos aleatorios involucrados, normalmente en dimensión infinita, y la suavidad de los funcionales que dan lugar a los estimadores. De esta forma, podemos dar resultados muy generales evitando hacer un estudio particular en cada ejemplo considerado.

Estado del arte

En el artículo [Raghavachari, 1973] se obtiene el comportamiento asintótico del estimador plug-in de la distancia de Kolmogorov bajo la alternativa, es decir, cuando se trata de estimar la distancia entre dos distribuciones distintas. Para alcanzarlo, se asume que las funciones de distribución poblacionales, normalmente desconocidas, son continuas. Esta condición de continuidad de las funciones involucradas ha sido desde entonces impuesta en diversos artículos (véase por ejemplo [Álvarez-Esteban *et al.*, 2016, Equation (11)], [Dette *et al.*, 2018, Assumption 7.4] y [Freitag *et al.*, 2006, Section 2]). Se debe principalmente a que la forma de proceder en todos ellos es reiterar la metodología de [Raghavachari, 1973]: realizar un profundo y cuidadoso estudio del estimador, centrándose en él únicamente como variable aleatoria real. En esta candidatura proponemos un cambio: mirar a los estimadores como la evaluación de un funcional en un proceso estocástico apropiado. Por ejemplo, en este caso la norma del supremo y el proceso empírico. Con este nuevo paradigma, la hipótesis de regularidad sobre las funciones de distribución se torna innecesaria.

Desde un punto de vista más general, hemos explotado esta idea para obtener resultados en otros contextos. Algunos ejemplos son la convergencia de estadísticos de tipo Berk-Jones, la estimación de cópulas o los contrastes de hipótesis basados en métricas de probabilidad. Todos ellos tienen en común que se usa el supremo o la norma uniforme.

Aportaciones del trabajo

La principal aportación de este trabajo es mostrar que la norma del supremo se puede ver como un funcional diferenciable Hadamard direccional (consultar [Shapiro, 1991]). Esta perspectiva permite aplicar el método delta en su versión funcional a estadísticos que se puedan escribir como la norma del supremo de un elemento aleatorio, como es el caso de los estimadores plug-in. De esta manera el estudio del comportamiento asintótico del estadístico se reduce al análisis de los elementos aleatorios subyacentes. Este problema usualmente se centra en la convergencia débil, típicamente en un espacio de dimensión infinita, de un proceso estocástico adecuado.

Cuando se trata del proceso empírico, el teorema de Donsker da respuesta a esta parte. Este es el caso de la extensión que damos de los resultados de [Raghavachari, 1973] y uno nuevo que aportamos sobre la estimación de métricas integrales de probabilidad. Sin embargo, no en todas las aplicaciones se utiliza el proceso empírico. Esta es la situación de la pregunta abierta de Jager y Wellner sobre estadísticos de tipo Berk-Jones ([Jager & Wellner, 2004, Question 2, p.329]). En este caso usamos la convergencia débil del proceso empírico ponderado. Final-

mente, en el caso de las cópulas la convergencia se basa en un resultado análogo al teorema de Donsker para el proceso cópula empírico ([Segers, 2012]).

Como consecuencia de este nuevo punto de vista se obtienen teoremas clásicos como los expuestos en [Raghavachari, 1973] con una mayor sencillez y limpieza, a la vez que se debilitan las hipótesis impuestas. Al mismo tiempo se derivan otros nuevos, como los mencionados anteriormente: la respuesta a la pregunta abierta de Jager y Wellner ([Jager & Wellner, 2004, Question 2, p.329]), un resultado asintótico sobre la estimación de cópulas y un resultado asintótico sobre estimación de métricas de probabilidad. Aún a día de hoy seguimos trabajando en otros problemas aplicando estas mismas técnicas.

References

- [Álvarez-Esteban *et al.*, 2016] Álvarez-Esteban, P. C., Del Barrio, E., Cuesta-Albertos, J. A., & Matrán, C. (2016). A contamination model for the stochastic order. *Test*, 25(4), 751-774.
- [Dette *et al.*, 2018] Dette, H., Möllenhoff, K., Volgushev, S., & Bretz, F. (2018). Equivalence of regression curves. *Journal of the American Statistical Association*, 113(522), 711-729.
- [Fermanian, 2013] Fermanian, J. D. (2013). An overview of the goodness-of-fit test problem for copulas. In *Copulae in Mathematical and Quantitative Finance* (pp. 61-89). Springer, Berlin, Heidelberg.
- [Freitag *et al.*, 2006] Freitag, G., Lange, S., & Munk, A. (2006). Non-parametric assessment of non-inferiority with censored data. *Statistics in medicine*, 25(7), 1201-1217.
- [Jager & Wellner, 2004] Jager, L., & Wellner, J. A. (2004). On the "Poisson boundaries" of the family of weighted Kolmogorov statistics. In *A festschrift for Herman Rubin* (pp. 319-331). Institute of Mathematical Statistics.
- [Pollard, 2012] Pollard, D. (2012). *Convergence of stochastic processes*. Springer Science & Business Media.
- [Raghavachari, 1973] Raghavachari, M. (1973). Limiting distributions of Kolmogorov-Smirnov type statistics under the alternative. *The Annals of Statistics*, 1(1), 67-73.
- [Segers, 2012] Segers, J. (2012). Asymptotics of empirical copula processes under non-restrictive smoothness assumptions. *Bernoulli*, 18(3), 764-782.

[Shapiro, 1991] Shapiro, A. (1991). Asymptotic analysis of stochastic programs.
Annals of Operations Research, 30(1), 169-186.

[van der Vaart & Wellner, 1996] van der Vaart, A.W., & Wellner, J.A. (1996).
Weak Convergence and Empirical Processes: With Applications to Statistics.
Springer.