



UN SOFTWARE CON INTERFAZ WEB PARA EL CÁLCULO DE LOS MOMENTOS INVARIANTES DE WISHART

Antunes Percíncula, Carlos

Facultad de Ingeniería Química FIQ-UNL

Directora: Forzani, Liliana
Codirector: Toledano, Ricardo

Ciencias Exactas

Momentos, Wishart, Software

INTRODUCCIÓN

Cuando modelamos mediciones, nos interesa estudiar si las cantidades obtenidas a partir de ellas tienen propiedades estadísticas que nos permitan responder preguntas de interés. Por ejemplo, supongamos que queremos ver si existe una asociación entre p medidas obtenidas de un mismo individuo. Entonces recolectamos n de dichos individuos, obtenemos su matriz de covarianza que denotaremos W/n de $p \times p$ donde cada posición (i, j) con $i \neq j$ nos mide la asociación empírica entre la i -ésima y j -ésima variable. Para estudiar si dicha asociación es aleatoria o si efectivamente es una asociación real debemos estudiar la distribución muestral de dicha W o, en particular, los momentos, o lo que es equivalente, el promedio poblacional de dichas W . El momento de orden k se define como el valor esperado de W^k y se denota $E[W^k]$. Algunos momentos son distinguidos, como la media de W que es el momento de primer orden, o la varianza de W que es el de orden 2 cuando la media es cero.

En estadística, la distribución Wishart es el modelo paradigmático de la distribución asociada a esa matriz W y ha sido estudiada desde hace mucho tiempo [Haff, 1981].

Título del proyecto: Datos complejos de alta dimensión: regresión eficiente y reducción suficiente de dimensiones para modelos gráficos.

Instrumento: PICT

Año de convocatoria: 2018

Organismo financiador: Agencia Nacional de Promoción de Ciencia y Tecnología

Director/a: Liliana Forzani



El interés por calcular los momentos de W de cualquier orden que son invariantes (en el sentido de que solo dependen de los autovalores de W) no solo está relacionado a los problemas antes mencionados, sino que recobró su importancia con la aparición de *big data* y el estudio de matrices aleatorias. Más precisamente, este tipo de cálculos fueron necesarios, en [Cook y Forzani, 2018] y [Cook y Forzani, 2019] para probar que el algoritmo de cuadrados mínimos parciales (tan utilizado en quimiometría) es consistente cuando la cantidad de predictores p y el tamaño de muestra n crece.

Si bien hace tiempo se cuenta con fórmulas explícitas para calcular estos momentos, estas son sumamente difíciles de usar en la práctica para orden mayor que 3. Más aún, hay escasa o nula bibliografía para consultar si se necesita un momento particular de este tipo de orden mayor que 3 y hasta el momento hay sólo un inicio de catalogación de los momentos de orden bajo [Holgersson y Pielaszkiewicz, 2020]. En el trabajo [Letac y Massam, 2004], proponen una forma de calcular estos momentos. Sin embargo, las dimensiones de los objetos que intervienen y la forma combinatoria que estos momentos presentan, hacen que en la práctica sea muy complicado calcular los momentos manualmente cuando $k \geq 3$. En este trabajo presentamos un paquete de software desarrollado en SAGE y una interfaz web que permiten obtener una expresión de los momentos invariantes de la distribución Wishart (y bajo ciertas condiciones, de sus inversas) en términos de los parámetros de la distribución n y Σ , donde n es un entero positivo y Σ es una matriz simétrica de $p \times p$. Además, se pueden obtener resultado para valores numéricos concretos de estos parámetros, completando el catálogo de momentos de Wishart, tarea comenzada en [Holgersson y Pielaszkiewicz, 2020].

OBJETIVOS

El objetivo es crear una herramienta computacional para calcular los momentos de una variable con distribución Wishart(n, Σ) fácil de usar por investigadores en estadística, cuyo uso se espera sea de *consulta*. Para esto se desarrolló un sitio web donde la investigadora o el investigador pueda calcular los momentos de interés brindando sólo los parámetros n y Σ .

Por otro lado, se pretende hacer disponible los algoritmos programados a través de un paquete de Python/SAGE que permita a los usuarios más expertos incorporar estas funcionalidades en su propio código.

METODOLOGÍA

La programación del cálculo de estos momentos requiere comprensión de ciertos objetos combinatorios como las permutaciones y las funciones simétricas. Se eligió para la implementación de los algoritmos el Sistema de Álgebra Computacional (CAS, por sus siglas en inglés) llamado SAGE, basado en Python, que ya tiene implementados algunos módulos que permiten trabajar con permutaciones, particiones y funciones simétricas. Además consideramos que se contaba con mayor documentación sobre cómo ejecutar el código desarrollado desde una interfaz web, lo que era uno de los objetivos del trabajo.



Encuentro de Jóvenes Investigadores

Si bien podríamos considerar que el principal objetivo es calcular $E[W^k]$ para cualquier k , los resultados de Letac y Massam permiten calcular algunas otras esperanzas, que son funciones de potencias de W y de potencias de trazas de potencias de W . Por ejemplo, para $k = 2$ los momentos que pueden calcularse, con sus respectivas fórmulas, se ven en el lado izquierdo de la ecuación (1).

$$E \begin{pmatrix} W \operatorname{tr} W \\ W^2 \end{pmatrix} = \begin{bmatrix} B_k^{-1} \end{bmatrix} \begin{bmatrix} D_k(n/2) \end{bmatrix} \begin{bmatrix} B_k \end{bmatrix} \begin{pmatrix} \Sigma \operatorname{tr} \Sigma \\ \Sigma^2 \end{pmatrix} \quad (1)$$

El cálculo de las matrices B_k y B_k^{-1} en el lado derecho de la ecuación (1) requiere el cómputo de ciertas funciones simétricas, llamadas *polinomios de Jack de orden k* , en una base particular, la base *potencia* o *power-sum*. Hasta este trabajo SAGE no contaba con métodos programados que calculen estos polinomios en su generalidad total. Este es uno de los aportes del autor, así como el cómputo de algunas otras cantidades de interés combinatorio que aparecen en [Lapointe et al, 2000] y que necesitaron revisión.

Los algoritmos desarrollados se organizaron en una lista de funciones que se ofrecen en forma de un paquete de Python. Este paquete se puede instalar en SAGE y permite calcular estos polinomios en varias bases y también pasar de una a otra y obtener las esperanzas que aparecen en [Letac y Massam, 2004], entre otros aportes. En la literatura hay recopilaciones de polinomios de Jack para los primeros órdenes, así como de los primeros momentos de W , lo cual fue utilizado para validar nuestra implementación.

Simultáneamente se diseñó el sitio web donde se ofrece estas funcionalidades, el cual está todavía en construcción. Éste se encuentra alojado provisoriamente en el repositorio de Github del autor: <https://antunescarles.github.io/wishart-moments-calculator/>.

CONCLUSIONES

El desarrollo hecho en este trabajo ha sido fructífero en dos direcciones. Por un lado, llena el vacío de información sobre los momentos de Wishart, haciéndolos accesibles para su consulta y para su uso computacional. En la Figura 1 se muestra la interfaz desarrollada, donde se puede ver la respuesta del sistema al calcular $E[W^3]$. Por otro, algunos subproductos pueden ser de importancia para otros investigadores y desarrolladores, como por ejemplo las implementaciones del cálculo de los polinomios de Jack.

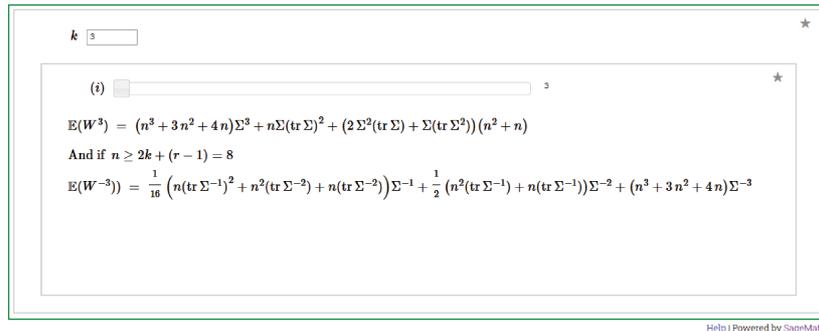


Figura 1: Interfaz web interactiva para el cálculo de los momentos

BIBLIOGRAFÍA BÁSICA

- **Cook RD, Forzani L.** 2018. "Big data and partial least-squares prediction". Canadian Journal of Statistics, 46(1), 62-78.
- **Cook RD, Forzani L.** 2019. "Partial least-squares prediction in high-dimensional regression". The Annals of Statistics, 47(2), 884-908.
- **Haff LR.** 1981. "Further identities for the Wishart distribution with applications in regression". Canadian Journal of Statistics, 9(4), 531-544.
- **Holgersson T, Pielaszkievicz J.** 2020. "A Collection of Moments of the Wishart Distribution". Recent Developments in Multivariate and Random Matrix Analysis : Festschrift in Honour of Dietrich von Rosen, 147-162. Springer International Publishing.
- **Lapointe L, Lascoux A, Morse J.** 2000a. "Determinantal expression and recursion for Jack polynomials". Electronic Journal of Combinatorics, 7.
- **Letac G, Massam H.** 2004. "All invariant Moments of the Wishart Distribution". Scandinavian Journal of Statistics, 31(2), 295-318.