Rev.Div.Amarun, **4** (2025), 63–74

Revista de Divulgación | AMARUN www.amarun.org

Explorando el análisis de sensibilidad a través de desigualdades de Poincaré

David Heredia[1]

Resumen

Este artículo de divulgación ofrece una introducción accesible al análisis de sensibilidad, utilizando las desigualdades de Poincaré como herramienta clave. Se presentan dos índices de sensiblidad: los índices de Sobol y los índices DGSM. A través de un modelo de inundación, se ilustra la utilidad de los índices DGSM como alternativas a índices de Sobol, además de cómo la introducción de pesos determinados por operadores de difusión asociados a desigualdades de Poincaré puede mejorar su precisión.

©2025 Asociación AMARUN

1. Introducción

El tema que presentamos es multidisciplinario y se sitúa en la interfaz de la estadística y el análisis funcional probabilista. Por esta razón, nuestra intención es doble: primero, proveer una introducción amigable al análisis de sensibilidad global, un campo de la estadística que es de interés internacional debido a su aplicación práctica en la ingeniería y la industria; y segundo, destacar la importancia del estudio de las desigualdades de Poincaré con peso para medidas de probabilidad, un concepto que, a pesar de ser fundamentalmente teórico, encuentra una valiosa aplicación práctica en el análisis de sensibilidad.

Con el fin de facilitar la comprensión del lector, en la preparación de este escrito hemos optado por evitar un rigor matemático y obviar ciertos detalles técnicos. Sin embargo, se sugiere contar con conocimientos básicos en probabilidad, los cuales pueden ser consultados en el mini curso de probabilidades de Amarun de 2013 [14], así como ciertos fundamentos en análisis funcional y familiaridad con herramientas numéricas comunes (como técnicas de simulación, resolución de ecuaciones diferenciales, entre otras).

Todo lo que se presenta aquí se basa principalmente en [9] y su artículo predecesor [17]. Además, se recomienda revisar el mini curso de estadística de Amarun de 2024 [8], el cual, a lo largo de cuatro lecciones, profundiza un poco más sobre este tema en el contexto de la desigualdad de Poincaré clásica (es decir, sin peso).

A lo largo de este documento utilizamos indistintamente los términos ley, distribución y medida de probabilidad para referirnos a los mismos objetos. La notación que empleamos para la esperanza de una variable aleatoria real Y es $\mathbb{E}[Y]$, mientras que para su varianza utilizamos $\mathrm{Var}(Y) = \mathbb{E}\left[(Y - \mathbb{E}[Y])^2\right]$.

[1]Institut de mathématiques de Toulouse
dheredia@insa-toulouse.fr

1.1. Un modelo de inundación

Imaginemos que trabajamos en una empresa que planea construir un dique a lo largo del río representado en la Figura 1, con el objetivo de prevenir que sus desbordamientos afecten una zona industrial cercana. Se logran identificar ocho magnitudes que determinan el curso del río, pero lamentablemente, las herramientas utilizadas para medirlas no siempre obtienen resultados precisos. Por ello, resulta conveniente tratar a estas magnitudes como variables aleatorias, cuyas leyes descritas en el Cuadro 1, son congruentes con las distribuciones empíricas observadas durante campañas de recolección de datos. Dado que las magnitudes en el mundo real son siempre finitas, las medidas de probabilidad definidas sobre conjuntos no acotados deben ser truncadas en intervalos finitos.

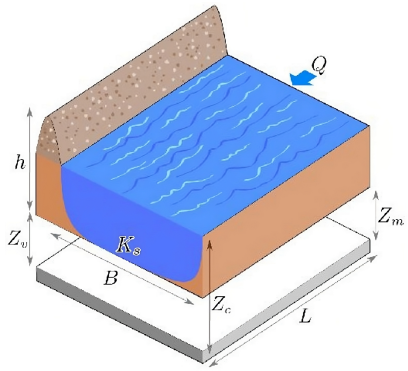

Figura 1: Un río, un dique y una zona industrial escondida en algún lado detrás del dique.

Entrada	Significado	Unidad	Medida de probabilidad
Q	Caudal máximo	m^3/s	Gumbel $\mathcal{G}(1013, 558)\|_{[500,3000]}$
K_s	Coeficiente de Strickler	—	Normal $\mathcal{N}(30, 64)\|_{[15,75]}$
Z_v	Nivel aguas abajo	m	Triangular $\mathcal{T}(49, 50, 51)$
Z_m	Nivel aguas arriba	m	Triangular $\mathcal{T}(54, 55, 56)$
H_d	Altura del dique	m	Uniforme $\mathcal{U}(7, 9)$
C_b	Altura de la orilla	m	Triangular $\mathcal{T}(55, 55{,}5, 56)$
L	Longitud del río	m	Triangular $\mathcal{T}(4990, 5000, 5010)$
B	Ancho del río	m	Triangular $\mathcal{T}(295, 300, 305)$

Cuadro 1: Variables de entrada para el modelo de inundación. La notación $|_I$ indica que la distribución está truncada en el intervalo I.

Arriba, la notación $\mathcal{T}(a, b, c)$ ($a < c < b$) representa a la distribución triangular con valor máximo a, valor mínimo b y moda c. Su nombre se debe a su función de densidad

$$\rho(x) = \begin{cases} \dfrac{2(x-a)}{(b-a)(c-a)} & \text{si } a \leq x \leq c, \\ \dfrac{2(b-x)}{(b-a)(b-c)} & \text{si } c \leq x \leq a, \\ 0 & \text{si no.} \end{cases}$$

la cual describe un triángulo (admire la Figura 2). Esta medida de probabilidad es a menudo utilizada para modelizar variables de una manera sencilla, cuando se disponen de pocas observaciones.

En cuanto a la distribución de Gumbel $\mathcal{G}(\eta, \beta)$ (de ubicación $\eta \in \mathbb{R}$ y parámetro de escalamiento $\beta > 0$),

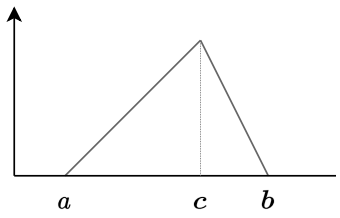

Figura 2: Función de densidad de la distribución triangular $\mathcal{T}(a, b, c)$.

su función de densidad es

$$\rho(x) = \frac{1}{\beta} \exp\left(-\frac{x-\eta}{\beta} - \exp\left(-\frac{x-\eta}{\beta}\right)\right), \quad x \in \mathbb{R}$$

y se utiliza para modelizar magnitudes extremas, principalmente en hidrología. De hecho, de acuerdo al teorema de Fisher–Tippett–Gnedenko (ver *e.g.* [7]), la distribución de Gumbel es una de las tres distribuciones límite que surgen al considerar el máximo de variables aleatorias independientes e idénticamente distribuidas.

Para llevar a cabo la construcción del dique, la empresa se interesa particularmente en dos magnitudes que dependen de las variables de entrada en el Cuadro 1: la altura del desbordamiento máximo anual del río (medido en metros)

$$S = Z_v - H_d - C_b + \left(\frac{Q}{BK_s}\sqrt{\frac{L}{Z_m - Z_v}}\right)^{\frac{3}{5}}. \tag{1}$$

y en el costo anual del mantenimiento del dique, medido en millones de euros

$$C = \mathbb{1}_{S>0} + \left(0{,}2 + 0{,}8\left(1 - e^{-\frac{1000}{S^4}}\right)\right)\mathbb{1}_{S\leq 0} + \frac{1}{20}\,\text{máx}\left\{H_d, 8\right\}. \tag{2}$$

Las variables de entrada junto con las dos cantidades de arriba componen el modelo de inundación, inicialmente por Rocquigny [16] y posteriormente retomado por Ioss y Lêmaitre en [11]. Una descripción detallada de las variables y su papel en las salidas S y C pueden ser consultadas, por ejemplo, en [21].

Para imaginar que nos encontramos en una situación real, debemos fingir que no conocemos estas fórmulas explícitas y que nuestros recursos se reducen a una muestra compuesta por un puñado de observaciones de las variables de entrada, junto con sus respectivos valores de S y C. Con estas limitaciones, nuestra tarea dentro de la empresa es estudiar la influencia de las variables de entrada en ambas magnitudes.

Este es precisamente el objetivo principal del análisis de sensibilidad global (abreviado GSA por su nombre en inglés, *Global Sensitivity Analysis*): cuantificar la influencia de variables aleatorias sobre la salida de una función desconocida $f\colon \mathbb{R}^d \to \mathbb{R}$, cuya evaluación suele ser costosa debido a que la dimensión d es generalmente grande[2]. En particular, el GSA se utiliza para identificar las variables que tienen un mayor impacto en el modelo, así como aquellas cuya relevancia es mínima, con el fin de poder descartarlas.

2. Índices de sensibilidad: índices de Sobol y DGSM

Las herramientas diseñadas para cuantificar la influencia de las variables de entrada en un modelo se conocen como índices de sensibilidad. En este artículo, nos enfocamos en dos familias de ellos: los índices de Sobol e índices DGSM.

[2]En esta ocasión nos enfocamos en funciones de $\mathbb{R}^d$ a $\mathbb{R}$, pero el GSA abarca un contexto mucho más amplio, que incluye espacios de funciones de espacios más generales.

2.1. Índices de Sobol

Para introducir los índices de Sobol es necesario visitar la descomposición de Hoeffding-Sobol, inicialmente propuesta por Wassily Hoeffding [10] y posteriormente reformulada por Ilya Sobol [18]. Para ello, nos situamos en un marco general. Consideremos $X = (X_1, \ldots, X_d) \in \mathbb{R}^d$, un vector de variables aleatorias, y $f\colon \mathbb{R}^d \to \mathbb{R}$, una función que satisface $\mathbb{E}[f(X)^2] < \infty$. Dado un conjunto $I \subseteq \{1, \ldots, d\}$, denotamos por X_I al vector formado por las variables X_j, con $j \in I$. Si bien la hipótesis de independencia se satisface muy rara vez en modelos reales (piénsese, por ejemplo, en las variables del modelo de inundación en el Cuadro 1), esta es necesaria para ciertas propiedades de los índices de Sobol y para la aplicación de las desigualdades de Poincaré unidimensionales, que se introduce más adelante.

La descomposición de Hoeffding-Sobol de $f(X)$ está dada por:

$$f(X) = \sum_{I \subset \{1,\ldots,d\}} f_I(X_I) = \underbrace{\mathbb{E}[f(X)]}_{f_\emptyset} + \sum_{i=1}^{d} f_i(X_i) + \sum_{i<j}^{d} f_{i,j}(X_i, X_j) + \sum_{i<j<k}^{d} f_{i,j,k}(X_i, X_j, X_k) + \ldots, \tag{3}$$

en donde, para cada subconjunto $I \subseteq \{1, \ldots, d\}$, la función f_I depende únicamente del vector X_I (por ejemplo, si $I = \{1, 2, 4\}$, entonces $f_I(X_I) = f_{1,2,4}(X_{1,2,4})$) y satisface la siguiente propiedad con respecto a la esperanza condicional:

$$\mathbb{E}[f_I(X_I) \,|\, X_J] = 0, \quad \text{para cualquier } J \subsetneq I.$$

Por convención denotamos $\mathbb{E}[\,\cdot\,|X_\emptyset] = \mathbb{E}[\,\cdot\,]$.

Esta propiedad permite descomponer a la varianza de $f(X)$ como una suma de varianzas:

$$\mathrm{Var}\,(f(X)) = \sum_{I,J \subseteq \{1,\ldots,d\}} \mathrm{Cov}\,(f_I(X_I), f_J(X_J)) = \sum_{\substack{I \subseteq \{1,\ldots,d\} \\ I \neq \emptyset}} \mathrm{Var}\,(f_I(X_I))\,,$$

indicando que cada $f_I(X_I)$ contiene la información de X_I dentro del modelo $f(X)$, en términos de varianza. En particular para cada $i \in \{1, \ldots, d\}$, el *efecto principal* $f_i(X_i)$ explica la influencia de X_i sin tomar en cuenta su interacción con el resto de variables, mientras que el *efecto total* $f_i^{\mathrm{tot}}(X) = \sum_{I \ni i} f_I(X_I)$ representa su influencia total. Los índices de Sobol surgen naturalmente como proporciones de varianza:

$$S_i = \frac{\mathrm{Var}(f_i(X_i))}{\mathrm{Var}(f(X))} \in [0, 1].$$

Consideramos también los índices de Sobol totales, definidos como

$$S_i^{\mathrm{tot}} = \frac{\mathrm{Var}(f_i^{\mathrm{tot}}(X))}{\mathrm{Var}(f(X))} = \frac{\sum_{I \ni i} \mathrm{Var}(f_I(X_I))}{\mathrm{Var}(f(X))} \in [0, 1]$$

Debido a que estas cantidades pertenecen al intervalo $[0, 1]$, son fácilmente interpretables, ya que podemos pensar en ellas como porcentajes de relevancia de la variable X_i dentro del modelo. Por su simple interpretación, los índices de Sobol suelen ser índices de sensibilidad preferidos en la práctica.

Para cada $i \in \{1, \ldots, d\}$, el índice de Sobol S_i determina la influencia de X_i, sin tomar toma en cuenta las interacciones con el resto de variables. Por otro lado, el índice de Sobol total S_i^{tot} captura tanto la influencia directa de X_i como su contribución en interacciones con otras variables. De estos dos índices, S_i^{tot} es el más relevante, ya que proporciona una visión más completa del impacto de X_i en el modelo. En particular, este puede emplearse con fines de *screening*, identificando variables con una influencia mínima en el modelo si nuestros datos indican que S_i^{tot} está por debajo de cierto umbral, por ejemplo, si $S_i^{\mathrm{tot}} \leq 5\,\%$.

A pesar de la conveniente interpretabilidad de los índices de Sobol, su estimación resulta costosa computacionalmente. Por ello, un aspecto clave de la investigación es la estimación eficiente de los índices o la determinación de cotas, en escenarios con datos limitados. El lector interesado puede encontrar algunos métodos actuales de estimación en [13]. Por nuestra parte, nuestra propuesta es proveer cotas superiores involucrando otro tipo de índice de sensibilidad, como veremos en la siguiente sección.

2.2. Índices DGSM

Cuando se dispone de información sobre las derivadas de la función f, es posible emplear otros índices de sensibildad conocidos como DGSM (por su nombre en inglés *Derivative Global Sensitivity Measure*), cuya fundación teórica se remonta a [19]. El índice DGSM de cada variable X_i está definido como

$$\nu_i = \mathbb{E}\left[\left(\frac{\partial f}{\partial x_i}(X)\right)^2\right]. \tag{4}$$

Gracias a esta simple definición, la estimación de los índices DGSM resulta ser económica en términos de recursos computacionales. Estos pueden ser estimados de manera eficiente, por ejemplo, con métodos de integración de Monte-Carlo.

La idea detrás de los índices DGSM es sencilla: una variable aleatoria X_i se considera influyente según ν_i si en promedio, esta hace que f varíe. Sin embargo, esta noción de variabilidad no es del todo fiable, ya que podría ocurrir que f presente grandes fluctuaciones en dirección a una variable que, en realidad, no sea muy influyente según su índice de Sobol total. Considere, por ejemplo, el modelo descrito por $f(X_1, X_2) = \cos(\pi n X_1) + X_2$, donde X_1 sigue une ley uniforme $\mathcal{U}(0,1)$, X_2 es otra variable independiente tal que $\mathrm{Var}(X_2) > 1/2$ y n es un número entero. En este caso, luego de un cálculo directo obtenemos que

$$S_1^{\mathrm{tot}} = \frac{\frac{1}{2}}{\frac{1}{2} + \mathrm{Var}(X_2)} < \frac{\mathrm{Var}(X_2)}{\frac{1}{2} + \mathrm{Var}(X_2)} = S_2^{\mathrm{tot}}.$$

Sin embargo, esta desigualdad no es compatible con respecto a los índices DGSM clásicos, ya que encontramos que

$$\nu_1 = \frac{1}{2}\pi^2 n^2 > 1 = \nu_2.$$

En lugar de usarlos directamente, la idea es aprovechar de los índices DGSM, que son económicos computacionalmente, para obtener información sobre los índices de Sobol, los cuales tienen un costo más elevado.

2.3. Relación entre los índices de Sobol y DGSM mediante desigualdades de Poincaré

Es en esta parte en donde el análisis funcional interviene con las desigualdades de Poincaré probabilistas, proporcionando cotas superiores de los índices de Sobol que involucran a los índices DGSM. Más precisamente, dado $i \in \{1, \ldots, d\}$, una desigualdad de Poincaré implica la siguiente cota superior de S_i^{tot}:

$$S_i^{\mathrm{tot}} \leq \frac{C_P(\mu_i)}{\mathrm{Var}(f(X))}\nu_i. \tag{5}$$

Arriba, μ_i denota la distribución de X_i y $C_P(\mu_i)$ se la conoce como la constante de Poincaré. La desigualdad será introducida en un marco adecuado en la siguiente sección, para no apartarnos de la discusión sobre los de sensibilidad.

Observemos entonces que, si la muestra disponible indica que el lado derecho de (5) es suficientemente pequeño, podemos inferir que X_i no tiene una influencia significativa en el modelo, sin necesidad de calcular S_i^{tot}. Esto hace de los índices DGSM una alternativa para realizar *screening* con bajo costo computacional. En efecto, la varianza de $f(X)$ y el índice ν_i pueden estimarse de manera efectiva utilizando integración de Monte-Carlo. Además, como se explica en [17], la constante de Poincaré $C_p(\mu_i)$ puede ser aproximada con gran precisión mediante una discretización en elementos finitos que involucra a la función de densidad de μ_i. Como tal, al ser un método determinista, no depende de las observaciones.

2.4. Aplicación al modelo de inundación

Veamos en acción la cota superior (5) aplicada a nuestro modelo de inundación presentado en la introducción. Recordemos que las variables aleatorias involucradas se detallan en el Cuadro 1 y que nuestra tarea es estudiar la influencia de dos cantidades (en principio desconocidas, en nuestra calidad de practicantes): desbordamiento máximo del río S y del costo de mantenimiento del dique C, descritos en (1) y (2).

Contamos con una muestra de 150 observaciones, donde cada una de ellas corresponde a un vector formado por las 150 variables entrada, lo que implica que cada variable se observa 150 veces. Además, contamos con las evaluaciones de esas 150 observaciones tanto en S como en C. Esto es suficiente para calcular la cota superior de los índices de Sobol.

Presentamos los resultados para las dos magnitudes simultáneamente. En la Figura 3, se muestra la cota superior (5) para cada variable. Los diagramas de caja representan 100 replicaciones bootstrap para ilustrar intervalos de confianza. Con fines ilustrativos, hemos incluido los "verdaderos" índices de Sobol, los cuales son en realidad estimaciones de alta precisión obtenidas mediante el método *pick and freeze* (véase [13]), utilizando una muestra de tamaño 10 000.

Observe que, en este ejemplo práctico, las cotas superiores logran identificar las variables aleatorias que no influyen de manera significativa en ambas magnitudes (Z_m, C_b, L, y B). Sin embargo, las cotas superiores no son satisfactorias para la variable Q en la altura máxima S, ni para las variables H_d y K_s en el costo C.

En nuestro trabajo [9], proponemos mejorar estas cotas superiores mediante la introducción de pesos en los índices DGSM, dando como origen a los denominados índices DGSM con peso. Al igual que en el caso clásico, una desigualdad funcional permite establecer la cota superior que los vincula con los índices de Sobol: las desigualdades de Poincaré con peso (como era de esperar). Este enfoque será explicado en la siguiente sección.

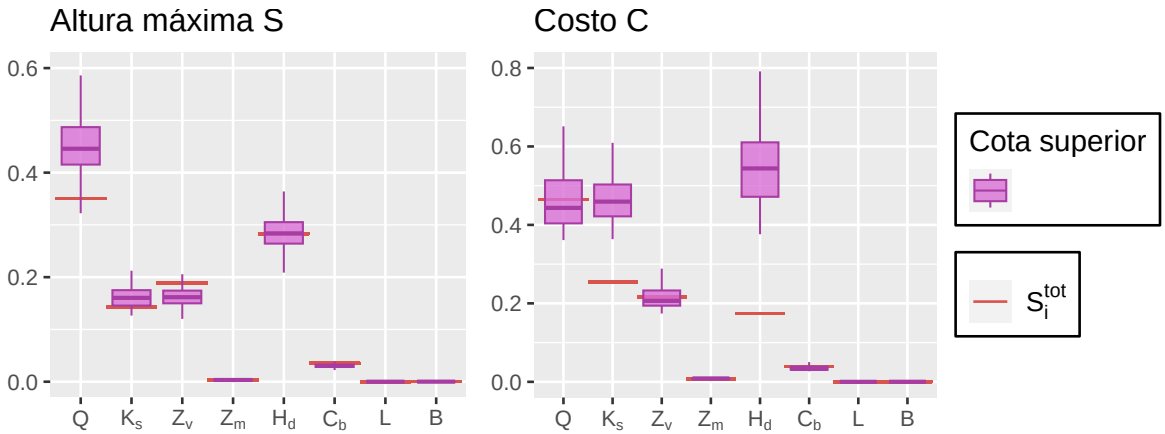

Figura 3: Cotas superiores de los índices de Sobol para el modelo de inundación. Las líneas rojas horizontales indican los verdaderos valores.

3. Desigualdades de Poincaré con peso para el GSA

3.1. Desigualdades de Poincaré con peso

En esta parte adoptamos un tono enfocado en el análisis funcional. Los principales protagonistas son una medida de probabilidad μ definida sobre un intervalo (a, b) (con $-\infty \leq a < b \leq \infty$), con función de densidad $\rho\colon (a, b) \to \mathbb{R}$, y una función no negativa $w\colon (a, b) \to \mathbb{R}^+$ que denominaremos peso. En lo que sigue, supondremos que ρ y w son suficientemente regulares, de manera que todas las derivadas que aparezcan estén bien definidas. Denotamos a la integral de una función h bajo la medida μ por $\int_a^b h(x)\,\mu(dx)$ o

simplemente por $\int_a^b h\, d\mu$.

Las desigualdades de Poincaré en específicamente teoría de probabilidad deben su nombre a la desigualdad homónima en espacios funcionales de diferenciabilidad débil, específicamente en los espacios de Sobolev (véase por ejemplo [3]). Esta última establece que si $\Omega \subseteq \mathbb{R}^n$ es un subconjunto acotado en al menos una dirección, entonces existe una constante $C > 0$ tal que

$$\int_\Omega (h(x))^2\, dx \leq C \int_\Omega \sum_{i=1}^n \left(\frac{\partial h}{\partial x_i}(x)\right)^2 dx, \tag{6}$$

para toda función f para la cual todo lo escrito arriba tenga sentido y que, además, "se anule" en la frontera de Ω.

Antes de introducir la desigualdad de Poincaré en el contexto probabilístico señalamos dos diferencias importantes con respecto a la la desigualdad 6. La primera de ellas es estructural: en probabilidades, una desigualdad de Poincaré es una propiedad que satisface una medida de probabilidad, en lugar de una propiedad relacionada a un espacio funcional. La segunda diferencia reside en que, en una desigualdad de Poincaré probabilista, no es necesario que las funciones satisfagan la condición de Dirichlet (es decir, no es necesario que se anulen en la frontera del dominio).

Además de estas dos diferencias, en la versión probabilista de la desigualdad nosotros consideramos la introducción de una función continua no negativa w, la cual denominamos peso. Así pues, decimos que la medida μ satisface una desigualdad con peso w si se satisface:

$$\int_a^b (h)^2\, d\mu \leq C_P(\mu, w) \int_a^b w\, \left(h'\right)^2 d\mu, \tag{7}$$

para toda función tal que $\int_a^b f\, d\mu = 0$ (decimos que f es centrada) para la cual todo lo escrito arriba tenga sentido. Arriba, $C_P(\mu, w)$ denota la constante admisible más pequeña y es llamada constante (óptima) de Poincaré. Si existe una función f con la cual (9) se cumple con igualdad, decimos que esta satura la desigualdad de Poincaré.

La elección del peso constante $w \equiv 1$ conduce a la desigualdad de Poincaré (probabilista) clásica, la cual ha sido ampliamente estudiada en la literatura. Véase por ejemplo [1], la referencia fundamental sobre operadores de difusión de Markov y desigualdades funcionales para medidas de probabilidad, o [17], en donde se aborda su aplicación en GSA.

Por otro lado, la noción de desigualdad de Poincaré con peso se extiende naturalmente al caso multidimensional, en donde el peso corresponde a una matriz en lugar de una función escalar. Sin embargo, el caso multi-dimensional resulta significativamente más complejo de abordar, y los resultados disponibles en la literatura son aún limitados.

La desigualdad de Poincaré (7) está estrechamente vinculada a un operador de difusión (de la forma $Lh(x) = \alpha(x)h''(x) + \beta(x)'h'(x)$), definido para funciones h suficientemente regulares como:

$$Lh = \frac{1}{\rho}(wh'\rho)' = wh'' + \left(w' + w\log(\rho)'\right) h'.$$

La conexión radica en la propiedad de simetría:

$$\int_a^b (-Lh)g\, d\mu = \int_a^b h(-Lg)\, d\mu = \int_a^b w\, h\, g\, d\mu,$$

la cual se cumple para cualquier función h y g, siempre que h (o g) satisfaga $wh'\rho(a) = wh'\rho(b)$. A este requerimiento lo conocemos como condiciones de frontera de tipo Neumann, lo cual resalta aún más la

diferencia entre esta desigualdad de Poincaré y su homónima (6), cuyo operador asociado impone condiciones de Dirichlet.

Una propiedad fundamental del operador L es su relación con la constante de Poincaré $C_P(\mu, w)$. El hecho es que ésta se caracteriza como $C_P(\mu, w) = 1/\lambda$, donde λ es el ínfimo de los valor propio no nulos de $-L$. Dicho de otra manera, $C_P(\mu, w)$ es el ínfimo de los $\lambda > 0$ para los cuales el siguiente problema tiene solución: encontrar h tal que $\int_a^b (h)^2 \, d\mu < \infty$ que satisfaga

$$\begin{cases} Lh(x) = \dfrac{1}{\rho(x)}(wh'\rho)(x) = -\lambda h(x), & \text{para todo } x \in (a, b) \\ wh'\rho(a) = wh'\rho(b) = 0. \end{cases} \tag{8}$$

En el caso de que el ínfimo sea alcanzado (tratándose entonces del mínimo), la función propia g asociada a $1/C_P(\mu, w)$ es aquella que satura la desigualdad de Poincaré (7). Además, de acuerdo con [5], se sabe que g se caracteriza como la única función propia que satisface $|g'| > 0$ en (a, b). Esta idea es clave para determinar pesos que permitan establecer una desigualdad de Poincaré.

Planteemos ahora el problema desde una perspectiva inversa: en lugar de buscar la función propia g a partir de un peso w, consideremos el caso opuesto. Suponemos que g es una función conocida, centrada y que satisface $|g'| > 0$. Nos interesa determinar el peso w_g para el cual g es una función propia. De este modo, al ser g estrictamente monótona, su valor propio necesariamente es $1/C_P(\mu, w_g)$ y ésta satura la desigualdad de Poincaré. Inspeccionando (8), notamos que tal peso w_g es único salvo un factor de escalamiento que depende de λ. Este se vuelve único si fijamos, por ejemplo, $\lambda = C_P(\mu, w_g) = 1$. De esta manera, al resolver (8) en función de g, obtenemos el peso

$$w_g(x) = -\frac{1}{g'(x)\rho(x)} \int_a^x g(y)\rho(y) \, dy, \qquad \text{para todo } x \in (a, b),$$

el cual ya aparece en [6].

Esta estrategia nos otorga poder de construir el peso asociado de acuerdo a la función g que queramos. Por ejemplo, si tomamos g definido por $g(x) = x - \int_a^b y \, \mu(dy)$, recuperamos algunos operadores de difusión L clásicos, dependiendo la elección de μ: el operador de Laguerre cuando μ sigue una distribucón gamma, el operador de Jacobi asociado a la ley beta, y el operador de Hermite en el caso de la distribución normal (refiérase a [1]). De hecho, para este último caso, obtenemos el peso constante $w_g \equiv 1$, por lo que la desigualdad de Poincaré asociada coincide con la desigualdad clásica.

3.2. Pesos para GSA

En el contexto de GSA, el peso w_g se adapta a modelos $f : \mathbb{R}^d \to \mathbb{R}$ cuyo comportamiento (con respecto a una variable específica X_i) se asemeja al de la función g. Más concretamente, como se muestra en [9], la desigualdad de Poincaré establece la cota superior para S_i^{tot}:

$$S_i^{\text{tot}} \leq \frac{1}{\text{Var}(f(X))} \mathbb{E}\left[w_g(x_i) \left(\frac{\partial f}{\partial x_i}(X) \right)^2 \right], \tag{9}$$

la cual se espera que sea precisa si g es similar al efecto principal f_i del modelo objetivo (definido en la descomposición de Hoeffding-Sobol, en (3)). El término que involucra la esperanza se denomina índice DGSM con peso, el cual generaliza el índice clásico ν_i definido en (4). En particular, si elegimos g como una función lineal, esta cota superior conducirá a buenos resultados en los casos cuando f_i sea aproximadamente lineal. Este enfoque de usar un peso adaptado a modelos con tendencias lineales fue inicialmente propuesto en [20].

En un contexto más general (no lineal), idealmente tomaríamos g igual al efecto principal f_i. Sin embargo, éste es desconocido, ya que en la práctica no tenemos acceso a la descripción explícita de f, y f_i no es necesariamente monótono (recordamos que esta estrategia solo es aplicable cuando g es una función estrictamente monótona). Ante la incapacidad de construir el peso w_{f_i} asociado a f_i, la idea entonces consiste en utilizar un estimador estrictamente monótono de f_i, al cual denotamos por $\widehat{f_i}$. Como tal, el peso resultante $\widehat{w_i} = w_{\widehat{f_i}}$ no es determinista, ya que depende de la muestra utilizada para estimar f_i. En el caso cuando f_i es una función estrictamente monótona, $\widehat{w_i}$ es en sí mismo un estimador del peso w_{f_i}.

El costo computacional para calcular el peso $\widehat{w_i}$ no es alto, y su calidad como estimador depende directamente de la precisión del estimador del efecto principal f_i. De hecho, aunque la estimación de los efectos principales es una etapa estándar económica dentro de la metodología de GSA, la literatura sobre métodos para obtener un estimador estrictamente monótono no es amplia y se requiere mayor investigación en este campo. Una vez $\widehat{f_i}$ haya sido estimado, su peso asociado $\widehat{w_i}$ se aproxima resolviendo numéricamente la ecuación diferencial (8). Esto puede ser llevado a cabo, por ejemplo, con el método de Runge-Kutta de orden 4 (más detalles se encuentran en [9] y [4] es un buen libro de métodos numéricos).

3.3. Aplicación en un modelo juguete

Resaltamos la relevancia de cota superior (9), aplicándola en el siguiente modelo juguete:

$$f(X) = \prod_{i=1}^{5} \left(\frac{X_i^4 - \frac{1}{5}}{1 + a_i} + 1 \right), \tag{10}$$

donde el vector $X = (X_1, \ldots, X_5)$ está compuesto de variables aleatorias independientes de distribución uniforme estándar $\mathcal{U}(0,1)$ y el vector $a = (1, 2, 4{,}5, 90, 90)$ juega un rol en la importancia de las variables. Cuanto mayor es a_i, menor es la influencia de X_i en el modelo. Este modelo está inspirado en la función g-Sobol, clásica en GSA, dada por la expresión

$$f(X) = \prod_{i=1}^{d} \left(\frac{4\,|X_i - 1/2| - 1}{1 + a_i} + 1 \right),$$

con $a \in \mathbb{R}^d$ un vector. Una de las ventajas de utilizar este tipo de modelos multiplicativos, es que se conocen de antemano los valores exactos de los índices de Sobol S_i^{tot} y los efectos principales f_i. En el caso del modelo (10), cada efecto principal está dado por la función monótona $x_i \mapsto f_i(x_i) = \left(x_i^4 - \frac{1}{5}\right)/(1 + a_i)$.

Para cada variable X_i, consideramos el peso que depende de los datos $\widehat{w_i}$ y el peso asociado a g lineal, al cual denotamos por $w_{\text{i,lin}}$. Los estimadores de los efectos principales que definen cada peso $\widehat{w_i}$ se obtienen mediante el ajuste de un SCAM (*Shape Constrained Additive Model*, segùn su nombre en inglés; véase [15]) a los datos. Comparamos cada cota superior la versión sin peso, escrita en (5). Las simulaciones numéricas fueron realizadas con una muestra de tamaño 150.

La Figura 4 indica las cotas superiores de los índices de Sobol. Observe que, para cada variable X_i, los mejores resultados se obtienen con el peso $\widehat{w_i}$. Esto era de esperarse, ya que $f(X)$ es un modelo con efectos principales monótonos, de manera que los estimadores $\hat{f}_i$ son precisos (mire la Figura 5). Por otro lado, las cotas superiores en ausencia de peso generan los peores resultados.

3.4. De vuelta al modelo de inundación

Para dar conclusión a este escrito, investigamos el desempeño de la cota superior (9) en el modelo de inundación. Mantenemos la misma configuración numérica que en la sección 2.4 y nos enfocamos únicamente en las cuatro variables más influyentes sobre las dos salidas S y C (descritas en (1) y (2)).

Los resultados se exhiben en la Figura 6. Se observa que las cotas superiores utilizando pesos mejoran las cotas sin peso para todas las variable, con excepción de H_d en el costo C. De hecho, para esta variable,

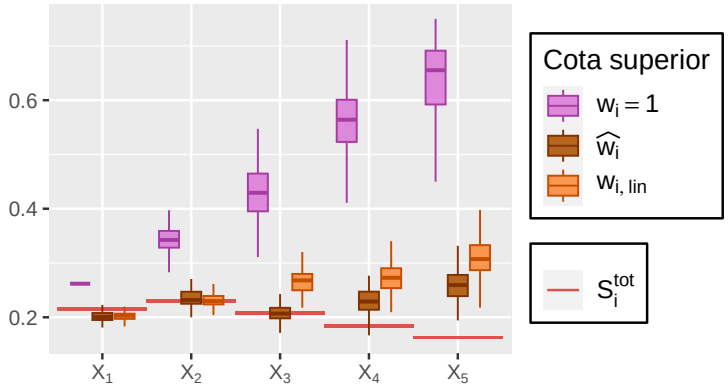

Figura 4: Cotas superiores de los índices de Sobol para el modelo $f(X)$. Las líneas horizontales indican los verdaderos valores.

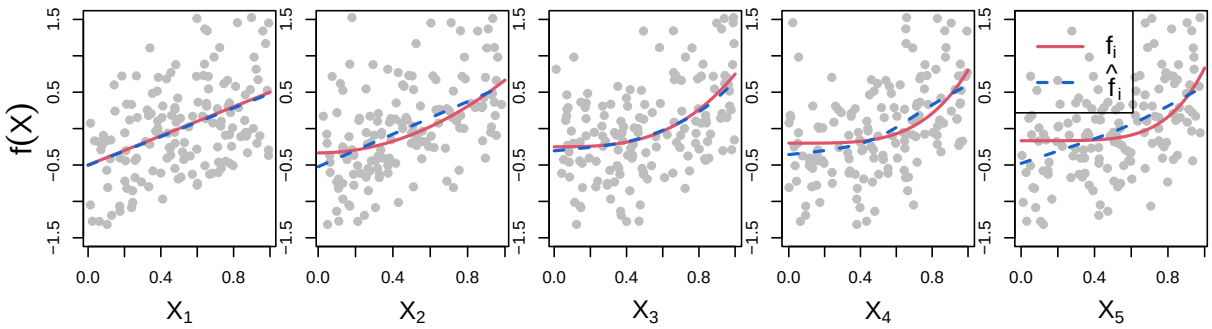

Figura 5: Estimadores monótonos $\widehat{f}_i$ de los efectos principales f_i en modelo $f(X)$ (10). Hemos recentrado el modelo en esta figura.

ninguna cota superior es satisfactoria. Esta situación se explica por los efectos principales. En efecto, en la Figura 7, se puede ver que los efectos principales y sus estimadores son casi lineales, salvo el de la variable H_d en C, el cual sigue una curva cuadrática. Como tal, su estimador carece de precisión.

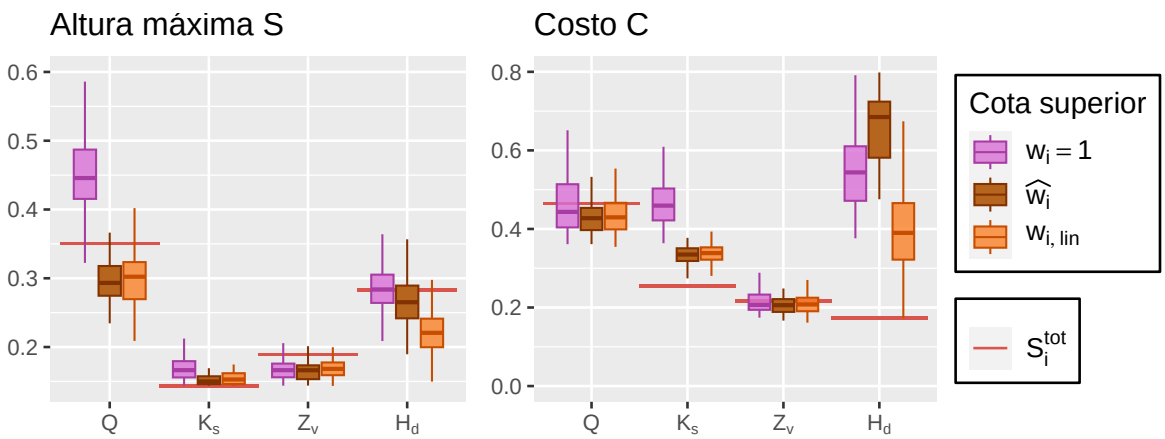

Figura 6: Cotas superiores de los índices para el modelo de inundación. Las líneas rojas horizontales indican los verdaderos valores.

4. Conclusiones

Como hemos evidenciado, los índices DGSM son herramientas idóneas para detectar variables que tienen poca influencia en un modelo probabilista. En particular, mostramos que su precisión puede mejo-

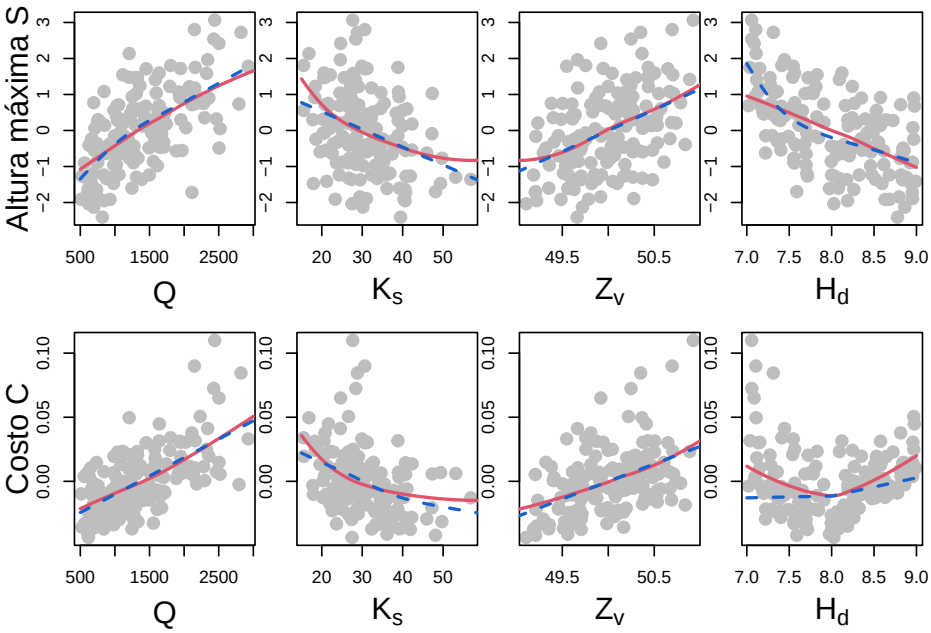

Figura 7: Efectos principales f_i (líneas rojas) y sus estimadores $\widehat{f_i}$ (líneas azules) de las magnitudes S y C. Hemos recentrado ambas magnitudes en esta figura.

rar gracias a la introducción de pesos, cuya construcción está inspirada en las desigualdades de Poincaré. Esto justifica el estudio teórico de las desigualdades de Poincaré con peso, motivado por necesidades del análisis de sensibilidad global, cuya misión principal es cuantificar la influencia de las variables de entrada.

En nuestro enfoque, estos pesos se construyen a partir de estimadores monótonos de los efectos principales del modelo. Cuando nos enfrentamos a variables que presentan efectos principales no monótonos, es necesario recurrir a estrategias más sofisticadas. Por ejemplo, entre los enfoques disponibles, la *descomposición en caos de Poincaré* (véase [9, 12]), que también se basa en desigualdades de Poincaré, permite estimar con gran precisión los índices de Sobol total, bajo la hipótesis de independencia entre variables de entrada.

Finalmente, vale la pena destacar que diversas metodologías, incluida la nuestra, requieren la hipótesis de independencia. Si esta independencia no se cumple, podemos aplicar desigualdades de Poincaré multidimensionales para estimar los costosos índices de Sobol. Sin embargo, en el contexto de la dimensión superior, el estudio teórico de estas desigualdades es desafiante, y los resultados actuales todavía no se ajustan del todo a las necesidades del análisis de sensibilidad. El lector puede encontrar un compendio de varios resultados en [2].

Referencias

[1] D. Bakry, I. Gentil, and M. Ledoux. *Analysis and Geometry of Markov Diffusion Operators*. Grundlehren der mathematischen Wissenschaften. Springer International Publishing, 2013.

[2] M. Bonnefont. Poincaré inequality with explicit constant in dimension $d \geq 1$. Global sensitivity analysis and poincaré inequalities. July 2022.

[3] H. Brezis. *Functional Analysis, Sobolev Spaces and Partial Differential Equations*. Universitext. New York, United States: Springer, first edition, 2011.

[4] R.L. Burden, J.D. Faires, and A.M Burden. *Numerical Analysis*. Cengage Learning, tenth edition edition, 2015.

[5] M.F. Chen and F.Y. Wang. Estimation of spectral gap for elliptic operators. *Trans. Am. Math. Soc.*, 349:1239–1267, 1997.

[6] M. Ernst, G. Reinert, and Y. Swan. First-order covariance inequalities via stein's method. *Bernoulli*, 26:2051–2081, 2020.

[7] L. Haan and A. Ferreira. *Extreme Value Theory: An Introduction*. Springer Series in Operations Research and Financial Engineering. 2006.

[8] D. Heredia. Análisis de sensibilidad global con desigualdades de poincaré, 2024. Minicurso de estadística de Amarun. Disponible en `https://www.amarun.org/index.php/materiales/lecciones-y-ejercicios`.

[9] D. Heredia, A. Joulin, and O. Roustant. On one dimensional weighted poincare inequalities for global sensitivity analysis, 2024. Disponible en arXiv:2412.04918.

[10] W. Hoeffding. A Class of Statistics with Asymptotically Normal Distribution. *Ann. Math. Stat.*, 19(3):293 – 325, 1948.

[11] B. Iooss and P. Lemaître. *A Review on Global Sensitivity Analysis Methods*, pages 101–122. Springer US, Boston, MA, 2015.

[12] N. Lüthen, O. Roustant, F. Gamboa, B. Iooss, S. Marelli, and B. Sudret. Global sensitivity analysis using derivative-based sparse Poincaré chaos expansions. *Int. J. Uncertain. Quantif.*, 13:57–82, 2023.

[13] C. Prieur and S. Tarantola. *Variance-Based Sensitivity Analysis: Theory and Estimation Algorithms*, pages 1217–1239. Springer International Publishing, Cham, 2017.

[14] C. Profeta. Minicurso de probabilidades de Amarun, 2013. Disponible en `https://www.amarun.org/index.php/materiales/lecciones-y-ejercicios`.

[15] N. Pya and S.N. Wood. Shape constrained additive models. *Comput. Stat.*, 25:543–559, 2015.

[16] É. D. Rocquigny. La maîtrise des incertitudes dans un contexte industriel. 2nde partie : revue des méthodes de modélisation statistique physique et numérique. *Journal de la société française de statistique*, 3:33 – 71, 2006.

[17] O. Roustant, F. Barthe, and B. Iooss. Poincaré inequalities on intervals – application to sensitivity analysis. *Electron. J. Stat.*, 11(2):3081 – 3119, 2017.

[18] I. M. Sobol and A. Gershman. On an alternative global sensitivity analysis estimator. *Proceedings of SAMO*, 1(1):33 – 34, 1995.

[19] I. M. Sobol and Kucherenko. Derivative based global sensitivity measures and their link with global sensitivity indices. *Math. Comput. Simul.*, 79(10):3009–3017, 2009.

[20] S. Song, Z. Tong, L. Wang, S. Kucherenko, and Z. Lu. Derivative-based new upper bound of sobol' sensitivity measure. *Reliab. Eng. Syst. Saf.*, 187:142–148, 05 2018.

[21] C. Steiner. *Sur l'utilisation des relations d'entrelacement dans l'étude des générateurs de Markov auto-adjoints. : Application aux inégalités spectrales et fonctionnelles et à l'analyse de sensibilité.* Phd thesis, Université Paul Sabatier - Toulouse III, 2022. Analyse fonctionnelle [math.FA].