

Asignación de la probabilidad del tiempo de recurrencia de un desastre ⁽¹⁾

Por VICENTE SOLANA GOMEZ

Dr. Ing. de Caminos e Investigador Científico del CSIC.

Los modelos probabilísticos de riesgo de ocurrencias de desastres naturales deben ser referidos a la duración del intervalo entre ocurrencias y a los tiempos de recurrencia para el próximo desastre. En estos modelos las probabilidades deben también ser asignadas en base, únicamente, a los datos locales disponibles de intervalos entre ocurrencias.

Se establece un principio de invarianza bajo el condicionamiento del dominio de la variable aleatoria según tiempos de recurrencia. El principio es formalizado usando el paradigma de métodos de inferencia de dos etapas, según requerimientos de invarianza en la información que codifica los datos, y en la selección de la distribución de tiempos de recurrencia.

Un método de inferencia que satisface estos requerimientos es el método de entropía relativa con ligaduras de fractiles descrito en este trabajo.

1. INTRODUCCION

El análisis de riesgo de sistemas en situaciones relacionadas con desastres naturales requiere el uso de modelos de riesgo de ocurrencias. Numerosos modelos han sido propuestos y utilizados en la literatura; sin embargo, muchos de ellos son difícilmente justificables, entre otras razones, por falta de datos [5,6]. Aunque usualmente no referidas, existen, además, algunas dificultades importantes en el empleo de modelos de riesgo de ocurrencias de desastres basados en planteamientos estadísticos convencionales; estas dificultades derivan, simplemente, del hecho de que tales planteamientos corresponden a problemas diferentes de los de inferencia en el análisis de riesgo de sistemas.

Se considera aquí el empleo de modelos probabilistas de riesgos de ocurrencia en los que la variable aleatoria se refiere a la duración del intervalo para la próxima ocurrencia de un desastre catastrófico, en una región o lugar determinado, basados, únicamente, en los datos disponibles de duraciones de intervalo entre ocurrencias de sucesos catastróficos en el mismo lugar y del mismo tipo y características.

Para completar el análisis de riesgo, el modelo probabilista anterior se extiende al tiempo de recurrencia de un desastre, esto es, el intervalo entre un instante cualquiera, no necesariamente coincidente con el último desastre registrado, y la ocurrencia del próximo suceso catastrófico. Ello hace posible la evaluación directa del incremento del riesgo de un suceso catastrófico a medida que transcurre el tiempo desde el último suceso. En otro caso, la no consideración de los tiempos de recurrencia en el análisis de riesgo conduce, generalmente, a la infravaloración de los riesgos de un sistema (Anagnos y Kiremidjian, 1985).

El problema de inferencia consiste en cómo asignar las distribuciones de probabilidad de los diferentes tiempos de recurrencia a partir de los datos de duraciones de intervalos entre ocurrencias de sucesos catastróficos. Este problema tiene dos partes, la primera es la asignación de una distribución a una variable aleatoria partiendo, únicamente, de la información de una muestra de datos de dicha variable, y, la segunda, corresponde al condicionamiento de los valores del dominio de la variable aleatoria que determina los tiempos de recurrencia.

Jeffreys (1939), Cox (1961) y Jaynes (1957) hicieron conjeturas sobre la existencia de un conjunto de principios de consistencia que debiera satisfacer un método de inferencia inductiva.

⁽¹⁾ Presentado en el Encuentro Internacional Catástrofes y Sociedad. Fundación MAPFRE e IPSEMAP, 24-26 octubre 1989. Madrid.

Shore y Johnson (1980) desarrollaron una derivación axiomática del método de entropía relativa en el caso en que la información es dada por los valores de una secuencia de momentos. Sin embargo, tal derivación no es aplicable al caso aquí tratado, en el que la información corresponde a datos de observaciones de una variable aleatoria.

Una vía alternativa de solución del problema de inferencia basada en una muestra de datos de observaciones ha sido desarrollada (Solana y Lind 1989, Solana 1989). En esta vía se han establecido, primero, dos principios fundamentales relacionados con propiedades de invarianza y de monotonicidad [8, 9 y 14] y, a continuación, se ha encontrado un método de inferencia que satisface diferentes formulaciones derivadas de estos principios [10 y 13].

El primero de estos principios fundamentales, denominado **principio de invarianza fuerte**, ha sido formulado como sigue [9]:

En toda asignación de probabilidad, modos alternativos de utilización de los mismos datos deben conducir al mismo resultado.

En este trabajo se hace aplicación del principio de invarianza fuerte al problema de inferencia de tiempos de recurrencia para un suceso catastrófico próximo. En la sección 2 se establece un principio de invarianza en la asignación de probabilidades de los tiempos de recurrencia aplicable en la modelación de ocurrencias de desastres. Este principio estipula requerimientos de consistencia que son exigibles a un método de inferencia basada en datos de duraciones de intervalos entre ocurrencias de sucesos catstróficos. En la sección 3 se describe un nuevo método de mínima entropía que satisface los requerimientos, aquí indicados, que constituyen el principio de invarianza en la inferencia de tiempos de recurrencia; dicho método ha sido recientemente aplicado al caso de modelos de intervalos entre ocurrencias de grandes terremotos y de inundaciones [10 y 11].

2. EL PRINCIPIO DE INVARIANZA EN LA INFERENCIA DE LOS TIEMPOS DE RECURRENCIA

Este principio se refiere a la invarianza en la aplicación de un método de inferencia a la asig-

nación de probabilidades de los tiempos de recurrencia para un nuevo suceso catastrófico.

Para su establecimiento se considera el paradigma de un método genérico de inferencia basada en datos de observaciones, el cual consiste en dos etapas, una primera de codificación de datos y otra, posterior, de selección de una distribución de probabilidad (Solana y Lind 1989).

En esta aplicación los datos son un conjunto de observaciones de duraciones de intervalos entre ocurrencias sucesivas. En ella, primeramente, el método de inferencia asigna, a partir de los datos, una distribución de probabilidad de la duración del intervalo entre ocurrencias correspondiente a un suceso futuro próximo. El paradigma de método de inferencia será aquí descrito, igual que en [9] y [12], utilizando operadores en cada una de sus etapas.

Posteriormente se formaliza el principio de invarianza considerando los operadores del método de inferencia juntamente con el condicionamiento del dominio de la variable aleatoria. Para ello, se examinan todas las posibilidades de modos alternativos de condicionamiento tanto en la codificación de datos como en la selección de la distribución de probabilidad relativas a los tiempos de recurrencia.

Sea la variable aleatoria W que representa el tiempo transcurrido para una nueva ocurrencia desde un suceso catastrófico anterior T_i , cuyos valores son $w = t - T_i, \forall t \geq T_i$; y sean los datos de la variable W obtenidos de una serie temporal de observaciones de sucesos catastróficos, tales que $w_i = T_{i+1} - T_i, i = 1, 2, \dots, n$.

Sean los datos del problema de inferencia constituidos por el conjunto $S = \{w_j\}, j = 1, 2, \dots, n$, de datos de duración de los intervalos entre ocurrencias reordenados según valores crecientes de w_j , y por el dominio E de valores de W , tales que $w \geq 0$.

En la etapa primera de **codificación de datos**, el método de inferencia crea un conjunto de ligaduras a partir de los datos dados. Sea H un operador que codifica los datos en ligaduras. Esta etapa puede ser representada simbólicamente por

$$I_W(S) = H(S, w \geq 0) \quad (1)$$

siendo $I_W(S)$ la información sobre W , codificada por H a partir de la muestra S y el dominio E dados.

En la etapa segunda de **selección de probabi-**

lidades, el método asigna una distribución de probabilidad de la variable W entre todas las posibles distribuciones que, de acuerdo con la única información codificada disponible, satisfacen el conjunto de ligaduras (1). Sea F el operador que selecciona tal distribución de probabilidad. Esta etapa puede ser representada simbólicamente por

$$Q_W(w/S) = F[I_W(S), P_W(w)], \quad (2)$$

donde $P_W(w)$ es una distribución de referencia y w representa la duración del intervalo para una próxima ocurrencia, tal que $w = t - T_{n+1}, \forall t \geq T_{n+1}$.

Sean ahora los valores de W condicionados por el subdominio $E_c, E_c \subset E$, tales que $w \geq w_c$. Los tiempos de recurrencia para un nuevo suceso, dado que ha transcurrido un tiempo w_c desde el suceso anterior, representados por $w - w_c, \forall w \geq w_c$, corresponden a una traslación de módulo w_c de los valores condicionados de W .

Sea el conjunto de datos $S_c = \{w_j\}, j = s + 1, s + 2, \dots, n, 0 < s < n$, relativo al condicionamiento de los intervalos entre ocurrencias tal que $w_j \geq w_c$, donde w_{s+1} es el primer dato de S superior al valor $w = w_c$.

Existen modos alternativos de obtención de la información codificada y de la distribución de probabilidad referidas a la muestra de datos condicionados S_c y al subdominio E_c . El principio de invarianza fuerte referido en la introducción prescribe, en tales casos, que se obtenga el mismo resultado usando los mismos datos según modos alternativos. La aplicación de este principio a la inferencia de tiempos de recurrencia debe ser examinada en cada etapa del método de inferencia y en todas las posibilidades que determinan modos alternativos de condicionamiento del dominio de W .

Codificación de datos

Hay dos modos posibles de condicionar la información proporcionada por los datos:

I) Primero, la información codificada $I_W(S)$ es condicionada por E_c tal que $w \geq w_c$. De este modo la información resultante será un conjunto de ligaduras condicionadas que pueden ser representadas por $I_W(S/w \geq w_c)$.

II) Segundo, el nuevo conjunto de ligaduras se obtiene aplicando el operador codificación H a

la muestra condicionada S_c y al subdominio E_c . La información codificada de este modo puede ser representada por $I_W(S_c) = H[S_c, w \geq w_c]$.

Selección de probabilidades

En esta etapa surgen dos posibilidades de modos alternativos de condicionamiento de la distribución de probabilidad.

La primera posibilidad corresponde al desdoblamiento del modo (I) seguido en la codificación de datos, en el cual la información inicial $I_W(S)$ es condicionada por E_c . En esta posibilidad los dos modos de selección de una distribución de probabilidad condicionada son los siguientes:

I.1) Primero, la distribución de probabilidad se obtiene a partir de la distribución, dada por (2), $Q_W(w/S)$, relativa a la información inicial, la cual es condicionada por $w \geq w_c$ usando la regla clásica de condicionamiento de probabilidades. La distribución así condicionada puede ser representada por $Q_W(w/S, w \geq w_c)$.

I.2) Segundo, la distribución de probabilidad es seleccionada usando el operador F entre todas las posibles distribuciones que satisfacen las ligaduras condicionadas $I_W(S/w \geq w_c)$ y la distribución de referencia condicionada. La distribución resultante es

$$Q_W[W|S(w \geq w_c)] = F[I_W(S|w \geq w_c), P_W(w|w \geq w_c)]. \quad (3)$$

La segunda posibilidad corresponde a modos alternativos de condicionamiento de distribuciones, siguiendo, en cada uno de ellos, el mismo orden en la aplicación de operadores que en los modos de codificación de datos (I) y (II). En esta posibilidad los modos alternativos de selección de una distribución de probabilidad condicionada son:

I.1) Primero, la misma distribución condicionada $Q_W[w|S(w \geq w_c)]$ que ha sido ya considerada en la primera posibilidad.

I.2) Segundo, la distribución de probabilidad seleccionada por el operador F entre todas las distribuciones que satisfacen el nuevo conjunto de ligaduras $I_W(S_c)$, deducido de la muestra condicionada S_c , y la distribución de referencia condicionada. Esta distribución es:

$$Q_W[w|S_c] = F[I_W(S_c), P_W(w|w \geq w_c)] \quad (4)$$

Al hacer aplicación del principio de invarianza fuerte es muy importante asegurar que los modos alternativos impliquen exactamente el uso de los mismos datos. De otra manera no debe hacerse aplicación del principio. Ello hace necesaria en la aplicación del principio de invarianza fuerte al caso de condicionamiento del dominio de la variable aleatoria, la consideración de dos situaciones diferentes, según que el condicionamiento corresponda a cualquier subdominio o, únicamente, a la clase de subdominios limitados por valores iguales a alguno de los datos (Solana 1989). En la aplicación a la inferencia de tiempos de recurrencia, la clase de subdominios limitados por algún dato corresponde a los subdominios E_s , $w \geq w_s$, tales que $w_s = w$, $0 < s < n$.

Según fue establecido en [12], la aplicación del principio de invarianza fuerte en los casos de codificación de datos según los modos alternativos (I) y (II), y de selección de probabilidades según los modos alternativos (I.1) y (II), es exigible únicamente para la clase de subdominios limitados por datos. La aplicación del mismo principio en el caso de selección de probabilidades según los modos alternativos (I.1) y (I.2) es, sin embargo, exigible para todo subdominio E_s .

Resumiendo, la aplicación del principio de invarianza fuerte a la inferencia de tiempos de recurrencia establece los siguientes requerimientos:

En la etapa de codificación de datos:

— "Idéntico conjunto de ligaduras debería ser obtenido si las ligaduras codificadas a partir de la muestra de datos iniciales de duración de intervalos entre ocurrencias, son condicionadas por un subdominio limitado por alguno de los datos, o el conjunto de ligaduras es directamente codificado a partir de la muestra de datos condicionados por el mismo subdominio";

En la etapa de selección de probabilidades:

— "Idéntica distribución de probabilidad de los tiempos de recurrencia debería ser obtenida si la distribución de probabilidad de duración de intervalos entre ocurrencias, asignada inicialmente es condicionada por cualquier subdominio, o la distribución es asignada directamente a partir de las ligaduras condicionadas para el mismo subdominio".

— "Idéntica distribución de probabilidad de los tiempos de recurrencia debería ser obtenida si

la distribución de probabilidad de duración de intervalos, asignada inicialmente, es condicionada por cualquier subdominio limitado por alguno de los datos, o la distribución es directamente asignada a partir de las ligaduras codificadas relativas a la muestra de datos condicionados por el mismo subdominio".

Esta aplicación del principio de invarianza fuerte constituye el principio de invarianza en la inferencia de tiempos de recurrencia. Este nuevo principio estipula, por una parte, la invarianza de la información codificada, cuando ésta es condicionada por un subdominio limitado por alguno de los datos de duración de intervalos entre ocurrencias y, por otra parte, la invarianza de la distribución de probabilidad de tiempos de recurrencia.

Los requerimientos anteriores correspondientes al principio de consistencia en la inferencia de tiempos de recurrencia, pueden expresarse formalmente: en la etapa de codificación de datos por

$$I_W(S_s) = I_W(S|w \geq w_s), s = 1, 2, \dots, n, \quad (5)$$

y en la etapa de selección de probabilidades, por

$$Q_W(w|S, w \geq w_s) = Q_W(w|(S|w \geq w_s)) \quad (6)$$

$$Q_W(w|S, w \geq w_s) = Q_W(w|S_s), s = 1, 2, \dots, n \quad (7)$$

3. EL METODO DE ENTROPIA RELATIVA CON LIGADURAS DE FRACTILES

Un método de inferencia probabilística basada en una muestra de datos de observaciones, denominado el método de entropía relativa con ligaduras de fractiles, ha sido desarrollado por Lind y Solana (1988). El método, que utiliza conceptos de entropía relativa (Shore y Johnson 1980), corresponde a una nueva versión de los métodos de entropía establecidos por Jaynes [3].

El método consta de las dos etapas de codificación de datos y de selección de probabilidades, y es aquí presentado en el caso de una variable aleatoria escalar, aplicado a la inferencia de la duración del intervalo para una nueva ocurrencia. Para ello, se consideran nuevamente los datos del problema de inferencia, que están dados: por el conjunto $S = \{w_j\}$, $j = 1, 2, \dots, n$, cuyos elementos, ordenados según valores crecientes de w_j , se refieren a la duración de intervalos entre ocurrencias de sucesos; y por el dominio E , $w > 0$, de la variable aleatoria W .

Codificación de datos

El procedimiento seguido en la codificación de datos se basa en la conocida propiedad de simetría de una muestra aleatoria de la variable W (Feller, 1971). Sea w la duración entre ocurrencias correspondiente a un nuevo suceso catastrófico; esta duración w tiene la misma probabilidad de estar comprendida en cada una de las $n + 1$ regiones en que los elementos de la muestra dada S dividen el dominio E . Esto puede ser formulado, en términos de fractiles de probabilidad, según la regla siguiente [7]:

"Los elementos w_j de la muestra ordenada S , de tamaño n , son los fractiles $j/(n + 1)$, $j = 1, 2, \dots, n$, de la probabilidad asignada a la duración del intervalo para una nueva ocurrencia, en base, únicamente, a las proposiciones de los mismos datos incluidos en S y el dominio E ."

Sea $Q_w(w|S)$ la distribución de probabilidad inferida a partir de S y E , teniendo la función de densidad $q_w(w|S)$. La regla anterior fija los fractiles de probabilidad en los puntos $w = w_j$ ($j = 1, 2, \dots, n$) y los valores convencionales 0 y 1 en los extremos de E .

El procedimiento determina el conjunto de ligaduras de fractiles, dadas según los pares de valores fractiles

$$\{w_j, Q_w(w|S)|_{w=w_j}\} = \{w_j, Q_j\}, j = 0, 1, \dots, n + 1; \quad (8)$$

dónde los fractiles de probabilidad Q_j son dados por

$$Prob[w < w_j|S] = j/(n + 1), j = 0, 1, \dots, n + 1. \quad (9)$$

Selección de probabilidades

Sean $P_w(w)$ la distribución de probabilidad de referencia, teniendo función de densidad $p_w(w)$, y P_j el valor de $P_w(w)$ en el fractil de abscisa $w = w_j$. Sean $E_j = [w_j, w_{j+1})$ las regiones en que los elementos dividen el dominio E .

El procedimiento empleado en la selección de la distribución de probabilidad de la duración del intervalo para una nueva ocurrencia, es encontrar la distribución $Q_w(w|S)$ que minimiza el funcional de entropía relativa de Kullback-Leibler

$$D(q, p) = \int_E q_w(w|S) [\log q_w(w|S) - \log p_w(w)] dw, \quad (10)$$

y satisface el conjunto de ligaduras de fractiles (8,9).

La solución general ha sido encontrada utilizando el método de multiplicadores de Lagrange [7]. Esta puede ser expresada en cada región E_j , en términos de los siguientes multiplicadores

$$\mu_j = (Q_{j+1} - Q_j)/(P_{j+1} - P_j), j = 1, 2, \dots, n, \quad (11)$$

La solución es

$$Q_w(w|S) =$$

$$Q_j + \mu_j [P_w(w) - P_j], w \in E_j, j = 1, 2, \dots, n, \quad (12)$$

con función de densidad

$$q_w(w|S) = \mu_j p_w(w), w \in E_j, j = 1, 2, \dots, n, \quad (13)$$

La función de densidad seleccionada $q_w(w|S)$, dada por (13), es discontinua en los fractiles de abscisa $w = w_j$ y corresponde, en cada región E_j , a los valores de la función de densidad de referencia multiplicados por el factor de escala constante μ_j .

Este método de inferencia satisface los requerimientos de consistencia que estipula el principio de invarianza en el condicionamiento del dominio de la variable aleatoria, establecido en [12]. En esta aplicación, el método satisface asimismo los requerimientos establecidos en la sección 2, relativos a la invarianza en la inferencia de tiempos de recurrencia, formalizados según las expresiones (5, 6 y 7).

4. CONCLUSIONES

Se considera en este trabajo el empleo de modelos probabilísticos de riesgo de ocurrencias de desastres naturales, los cuales se refieren a la duración del intervalo entre ocurrencias y al tiempo de recurrencia para un nuevo suceso catastrófico. La asignación de probabilidades en este modelo se realiza, exclusivamente, en base a datos locales disponibles de duraciones de intervalos de ocurrencias.

Se establece un principio de invarianza en la asignación de probabilidades de tiempos de recurrencia, aplicable a la modelación de desastres. Este principio ha sido derivado como una consecuencia del principio de invarianza fuerte en la inferencia de probabilidades.

El nuevo principio, que ha sido formalizado según el paradigma de inferencia que considera

dos etapas de codificación de datos y selección de probabilidades, establece los requerimientos de invarianza de la información codificada y de invarianza de la distribución de probabilidad de tiempos de recurrencia.

Existe un método que satisface los requerimientos anteriores de invarianza. Este es el método de entropía relativa con ligaduras de fractiles, que ha sido descrito en este trabajo. Aparentemente, no hay otro método de inferencia disponible que satisfaga estos requerimientos de invarianza.

Agradecimiento

El trabajo aquí presentado es parte del proyecto de investigación PB 87-0458 parcialmente financiado por la DGICYT.

REFERENCIAS

- [1] JEFFREYS, H.: *Theory of probability*, Oxford University Press, London, 1939 (2nd and 3th ed., 1948, 1961).
- [2] COX, R. T.: *The Algebra of Probable Inference*, Oxford University Press, London, 1961.
- [3] JAYNES, E. T.: *Papers on Probability, Statistical Physics and Statistics*, Rosencrantz R. D. (ed.), Reidel Publishing Co. Dordrecht, Netherlands, 1983.
- [4] SHORE, J. E. and JOHNSON, R. W.: "Axiomatic Derivation of Principle of Maximum Entropy and the Principle of Minimum Cross-Entropy", *IEEE Transactions on Information Theory*, IT-26, 1, 26-37, 1980.
- [5] ANAGNOS, T. and KIREMIDJIAN, A.S.: "A Review of Earthquake Occurrence Models for Seismic Hazard Analysis", *Probabilistic Engineering Mechanics*, Vol 3, 1, 3-11, 1988.
- [6] JARA, J. M. and ROSEMBLUETH, E.: "Probability Distribution of Times between Characteristic Subduction Earthquakes", *Earthquake Spectra*, Vol 4, 3, 499-529, 1988.
- [7] LIND, N. C. and SOLANA, V.: "Estimation of Random Variables with Fractile Constraints", IRR No. 11, Institute for Risk Research, University of Waterloo, Waterloo, Ontario, Canada, 1988.
- [8] SOLANA, V. and LIND, N. C.: "A Monotonic Property of Distributions Based on Entropy with Fractile Constraints", 8th. International Maximum Entropy Conference, St. John's College, Cambridge, England 1988. *Maximum Entropy and Bayesian Methods*, John Skilling (ed.),

Kluwer Academic Publications, Dordrecht, Netherland, 1989.

[9] SOLANA, V. and LIND, N. C.: "Two Principles for Data Based Probabilistic System Analysis", Proceedings 5th. International Conference on Structural Safety and Reliability, San Francisco, August 1989. *Structural Safety and Reliability*, Shinozuka, Ang and Schueller (eds.), ASCE American Society of Civil Engineers, Vol III, 1667, 1990.

[10] LIND, N. C., HONG, H. P. and SOLANA, V.: "Probabilistic assignment of earthquake interarrival times using cross-entropy", 4th. International Conference on Soils Dynamics and Earthquake Engineering. UNAM, México D. F. 23-26 Octubre 1989.

[11] LIND, N. C., HONG, H. P. and SOLANA, V.: "A cross-entropy method for flood probability analysis". *Stochastic Hydraulic and Hydrology*. Vol 3, No 3, 1989.

[12] SOLANA, V.: "Consistency principle for data based probabilistic inference". 9th. International Maxent Workshop. Dartmouth College, New Hampshire. Agosto 1989. *Maximum Entropy and Bayesian Methods in Science and Engineering*. Paul Fougere (ed.), Kluwer Academic Publications, Dordrecht, Netherlands, 1990. (en prensa).

[13] LIND, N. C. and SOLANA, V.: "Fractile Constrained Entropy Estimation of Distributions based on Scarced Data". *Journal of Civil Engineering Systems*. 1990 (en prensa).

[14] SOLANA, V.: "Data-Monotonicity of Logical Probabilities using Bayesian and Entropy Methods". 10 th International Workshop on entropy and Bayesian Methods. University of Wyoming. Agosto 1990.

Vicente Solana Gómez



Doctor Ingeniero de Caminos de la Universidad Politécnica de Madrid (1968), realizó su doctorado en el campo de la Fiabilidad de Estructuras.

Investigador científico de CSIC. Trabajó en el Instituto Torroja. Ha sido Profesor Asociado en la Universidad de Waterloo.

Actualmente adscrito a la Unidad de Matemáticas del Consejo Superior de Investigaciones Científicas, en Madrid, desarrolla su actividad científica en el área de Sistemas. Es investigador principal del proyecto de investigación "Inferencia Probable y Modelación de Incertezas usando Entropía" (1990-93), que comprende también aplicaciones al Análisis de Riesgo y Fiabilidad de Sistemas.