

Avances en el análisis de la recurrencia de fenómenos extremos: Aplicación a grandes vertidos petrolíferos

Alberto Solana Ortega

Geography Department,
San Diego State University, USA

Vicente Solana

Grupo de Inferencia, Probabilidad y Entropía
Instituto de Física Fundamental, CSIC, Madrid

Guión

1. Los fenómenos extremos y el análisis de su recurrencia
2. Dificultades en la modelización de fenómenos extremos
3. Nueva visión basada en un canon de inferencia plausible
4. Inferencia entrópica de modelos de recurrencia
5. Recurrencia de grandes vertidos petrolíferos en Galicia
6. Comentarios finales

Interés creciente de los fenómenos extremos

- **Desastres naturales**
Incremento del número de registros, intensificación de consecuencias
- **Desastres tecnológicos**
Variación constante por unidad de actividad económica en Europa
- **Cambio climático antropogénico**
La atención a cambios en el clima medio se extiende, recientemente, al estudio de la variabilidad de sucesos climáticos extremos
- **Crisis económicas y sociales**
Riesgo de ocurrencia de extremos en los mercados financieros

Ante los fenómenos extremos ...

TRANSICIÓN HACIA UNA SOCIEDAD DEL RIESGO,
caracterizada por:

- Aceptación de la incertidumbre sobre los extremos
- Aceptación de la imposibilidad de controlar el riesgo
- Constatación de que el desarrollo puede generar nuevos riesgos
- Mejora de la transparencia de los procesos de decisión

Accidentalidad catastrófica:

Expresiones como “NUNCA MÁS” (utilizadas en el naufragio del Prestige) no tienen sentido desde el punto de vista lógico.

Variabilidad y recurrencia de extremos

TRES PROBLEMAS

1) Problema de recurrencia simple

La ASIGNACION DE PROBABILIDADES a la cantidad potencial incierta asociada a la magnitud física “*duración de tiempo entre ocurrencias de fenómenos extremos*”, a partir de:

- la información FACTUAL constituida por los datos empíricos de duraciones de tiempo entre registros de sucesivos fenómenos en un lugar dado, y
- la información CONTEXTUAL relativa a las observaciones potenciales posibles,

NOVEDAD: dadas como EVIDENCIA CONTEXTO-FACTUAL conjunta

PROBLEMAS DE RECURRENCIA...

2) Problema de recurrencia condicionada

La modelización se extiende al **CONDICIONAMIENTO DE LAS PROBABILIDADES** previamente asignadas, mediante el condicionamiento de la EVIDENCIA CONTEXTUAL que incorpora la condición de que ha transcurrido un tiempo desde el último suceso extremo registrado

Seguido hasta ahora en el análisis de **FUNCIONES DE RIESGO**

3) Problema general de inferencia dinámica

Se generaliza a **LA ASIGNACION DE PROBABILIDADES** a la cantidad potencial incierta, mediante el condicionamiento previo de la evidencia CONTEXTO - FACTUAL a medida que transcurre el tiempo desde el último suceso extremo.

NOVEDAD: Se propone para el **ANALISIS DINÁMICO DEL RIESGO** y la **VARIABILIDAD DE LA RECURRENCIA DE EXTREMOS.**

La situación de los analistas del riesgo de fenómenos extremos

¿Cómo seleccionar modelos con pocos datos?

RECURSOS:

- Varios marcos de interpretación de la probabilidad paradigmas clásico, lógico, bayesiano, algorítmico, frecuentista, propensivista
- Múltiples formas de representación de los datos empíricos funciones de verosimilitud, ligaduras de momentos (convencionales, ponderados, L-momentos), gráficos de posición de cuantiles, etc.
- Variedad de técnicas de modelización Bondad de ajuste, análisis bayesiano, máxima verosimilitud, métodos entrópicos, etc.

DIFICULTADES:

- Resultados MUY DISCREPANTES desde los mismos datos, especialmente cuando éstos son escasos.
- Problema de SENSIBILIDAD DE LAS COLAS de las distribuciones en la elección de fractiles característicos, los cuales resultan dramáticamente dependientes:
 - del número de datos
 - del tipo de distribución propuesta.
- Elección de modelo NO JUSTIFICADA desde la sola información factual de los datos.
- LIMITACION CONCEPTUAL: No consideran el problema general de inferencia dinámica aplicable a la recurrencia y la variabilidad de fenómenos extremos.

INSUFICIENCIA DE LOS MÉTODOS CORRIENTES

LOS METODOS FRECUENTISTAS

- Son panscópicos, i.e. desarrollados para describir el comportamiento de grandes conjuntos de datos
- No permiten discriminar entre modelos, ni clasificar los modelos según criterios racionales.

Teoría de Valores Extremos: poca ayuda de los argumentos asintóticos porque las suposiciones iniciales **NO SON VERIFICABLES**

LOS METODOS BAYESIANOS

- Subjetividad de las distribuciones de referencia (priors)
- Especificación de la función de verosimilitud.
- No son **METODOS DE ASIGNACION** sino de **ACTUALIZACION** (requieren otros métodos complementarios).

ESTADO DE PERPLEJIDAD DE LOS ANALISTAS DE RIESGOS

- RELATIVISMO en la elección de modelos
- Rechazo explícito del análisis probabilista
- IMPREDECIBILIDAD intrínseca de los desastres
- Limitación a la explicación de las catástrofes una vez sucedidas, sin tener en cuenta que el riesgo se refiere a SUCESOS POTENCIALES.

SIN EMBARGO LA ASIGNACIÓN DE PROBABILIDADES
ES INELUDIBLE:

- toma anticipada de decisiones,
- trazado de mapas dinámicos de riesgo,
- diseño de planes de prevención y contingencia,
- gestión confiable y sostenible de sistemas.

Nueva visión basada en un canon de inferencia

CRISIS EN LA MODELIZACIÓN DEL RIESGO

Obstáculos encontrados por los analistas

Necesidad ubicua en todas las ramas de la ciencia:

- Mejor comprensión de los fenómenos extremos y su variabilidad
- Nuevos enfoques racionales de modelización de la recurrencia

NUEVA PERSPECTIVA

Considerar el razonamiento plausible como un problema fundamental del conocimiento, en el que se debe:

- *examinar cómo seleccionar consistentemente entre modelos y*
- *clarificar su significado y contenido empírico.*

CANON DE INFERENCIA PLAUSIBLE

Conjunto de requisitos esenciales de racionalidad y moralidad que debe satisfacer cualquier teoría de inferencia científica.

(referencias: Keynes, Jeffreys, Carnap, Menger, Jaynes, Tribus)

El canon de inferencia plausible

- R1. Exactitud semántica
- R2. Explicitación de conjeturas y evidencias
- R3. La plausibilidad como función proposicional binaria
- R4. Correspondencia con los esquemas cualitativos elementales del razonamiento
- R5. Autoconsistencia lógica
- R6. Consistencia evidencial
- R7. Universalidad con respecto a los lenguajes y sus interpretaciones

R1. Exactitud semántica

“Todos los conceptos y expresiones utilizados en la formulación de un problema de inferencia plausible deben ser tan claros y exactos como sea posible.”

Deben evitarse la ambigüedad y la vaguedad a fin de que ninguna confusión afecte al planteamiento y a las conclusiones del problema.

- “Navaja de OCKHAM”: No multiplicar las entidades mas allá de lo necesario.
- “Peine de MENGER”: No suprimir las entidades por debajo de la suficiencia.

R2. Explicitación de conjeturas y evidencias

“La evidencia y las conjeturas sobre las que se desea razonar deben ser explicitadas sintácticamente, usando los recursos del lenguaje de la lógica plausible en que se formula un problema”

TRANSPARENCIA. No debe ser escondida o eliminada ninguna parte de la evidencia.

La evidencia contiene todo el conocimiento básico considerado en un problema, expresado en el lenguaje objeto.

R3. La plausibilidad como función proposicional binaria

“La plausibilidad debe formularse como una función binaria de proposiciones, siendo sus argumentos asimétricos una conjetura y una evidencia, expresadas en el lenguaje en el que se plantea el problema”.

IDEAS que representa la plausibilidad:

- grado de soporte, confirmación o certeza
- grado de implicación parcial
- contenido relativo de evidencia

R4. Correspondencia con los esquemas cualitativos del razonamiento

“La plausibilidad y su representación numérica deben satisfacer esquemas elementales comparativos del razonamiento plausible que relacionan las plausibilidades individuales expresadas en el mismo lenguaje”.

Especialmente deben satisfacer los esquemas asociados con diferentes problemas de inferencia cuyas evidencias se relacionan:

- Los esquemas elementales de PÓLYA
- Los esquemas de variación monótona de las plausibilidades con los datos.

R5. Autoconsistencia lógica

“La inferencia plausible debe estar basada en su lógica interna; debe ser lógicamente consistente de modo que no surjan contradicciones cuando el conocimiento básico del problema representa la misma evidencia u otras equivalentes”.

INVARIANZA. Los métodos y reglas que relacionan plausibilidades individuales aplicados en forma alternativa desde la misma evidencia u otras equivalentes, deben conducir a los mismos resultados.

R6. Consistencia evidencial

“Los procedimientos de inferencia plausible deben ser consistentes con la evidencia adoptada como conocimiento básico en un problema, de modo que:

- la evidencia original NO SEA MODIFICADA y*
- pueda ser TOTALMENTE RECUPERADA.*

Es totalmente nuevo: Impone la INVERTIBILIDAD y LA RECUPERABILIDAD DE LA EVIDENCIA.

Distinto del requerimiento de JAYNES de “verificación de la información testable”.

R7. Universalidad con respecto a los lenguajes y sus interpretaciones

“Una teoría de inferencia plausible debe ser universal al menos en los dos sentidos de ser aplicable CUALQUIERA QUE SEAN:

- el lenguaje lógico formal en que se formula el problema,*
- la interpretación semántica del lenguaje formal”.*

Particularización:

La teoría y los procedimientos de inferencia deben ser aplicables CUALQUIERA QUE SEA EL NÚMERO DE DATOS considerados como parte de la evidencia.

Inferencia entrópica

Asignación directa de probabilidades, seleccionando la distribución inferida que extremiza una medida logarítmica de la información, sujeta a ligaduras sobre probabilidades, comúnmente expresadas como momentos de distribuciones.

Clases:

- Método de MÁXIMA ENTROPÍA.
Extremiza el funcional de entropía absoluta de Jaynes
- Método de MÍNIMA ENTROPÍA RELATIVA.
Extremiza el funcional de entropía relativa de Kullback-Leibler.

Dificultades:

- No unicidad de funcionales
- Selección de distribuciones de referencia
- ¿Cómo fijar las ligaduras desde los datos empíricos observacionales?

El método REF de entropía relativa con ligaduras fractiles

Doble extremización del funcional de entropía relativa sujeta a ligaduras de equiprobabilidad de cuantiles anclados a los datos de observaciones, en vez de ligaduras de momentos

Dos tipos de evidencias:

- Evidencia contextual e_z
- Evidencia contexto-factual (D, e_z) , donde D es el conjunto de datos ordenados de los tiempos entre registros $\{z_i^{obs}\}$, $i=1, \dots, n$.

Selección directa del par de distribuciones $\{P(z|e_z), Q(z|D, e_z)\}$

- $P(z|e_z)$ pertenece al espacio de distribuciones de referencia
 - $Q(z|D, e_z)$ es una distribución inferida
- Ambas son condicionales respecto a las evidencias.

Modelización entrópica de la recurrencia simple

SELECCIÓN del par de distribuciones de referencia e inferida $\{P(z|e_z), Q(z|D, e_z)\}$ que codifican el conocimiento de las duraciones de tiempo potenciales y actuales proporcionadas por el par de evidencias contextual y contexto-factual $\{e_z(D, e_z)\}$,

mediante la DOBLE EXTREMIZACION:

- del funcional de entropía relativa de KULLBACK - LEIBLER

$$S[q(z|D, e_z), p(z|e_z)] = \int q(z|D, e_z) \log \frac{q(z|D, e_z)}{p(z|e_z)} dz$$

- sujeto a las ligaduras de equiprobabilidad en los intervalos entre cada par de sucesivas observaciones factuales.

Inferencia dinámica de la variabilidad de la recurrencia de extremos

SELECCION del par de distribuciones de referencia e inferidas condicionadas $\{P(z|e_z, Z > z_c), Q(z|(D, e_z), Z > z_c)\}$ que codifican, para cada tiempo z_c desde la última observación, el conocimiento de las duraciones de tiempo potenciales y actuales dadas por el par de evidencias contextual y contexto-factual condicionadas $\{(e_z, Z > z_c), ((D, e_z), Z > z_c)\}$,

Mediante la TRIPLE EXTREMIZACION del funcional paramétrico de entropía relativa condicionada de KULLBACK-LEIBLER

$$S[q, p, \alpha, z_c] = \int q(z|(D, e_z), Z > z_c) \log \frac{q(z|(D, e_z), Z > z_c)}{p(z|e_z, Z > z_c)} dz$$

Sujeto a las ligaduras de equiprobabilidad condicionadas por el parámetro

$[z_c] = Q(z_c|D, e_z)$, correspondiente a la probabilidad en el primer dato de D mayor o igual que la duración de tiempo transcurrida z_c .

Aplicación: Recurrencia de grandes vertidos petrolíferos en Galicia

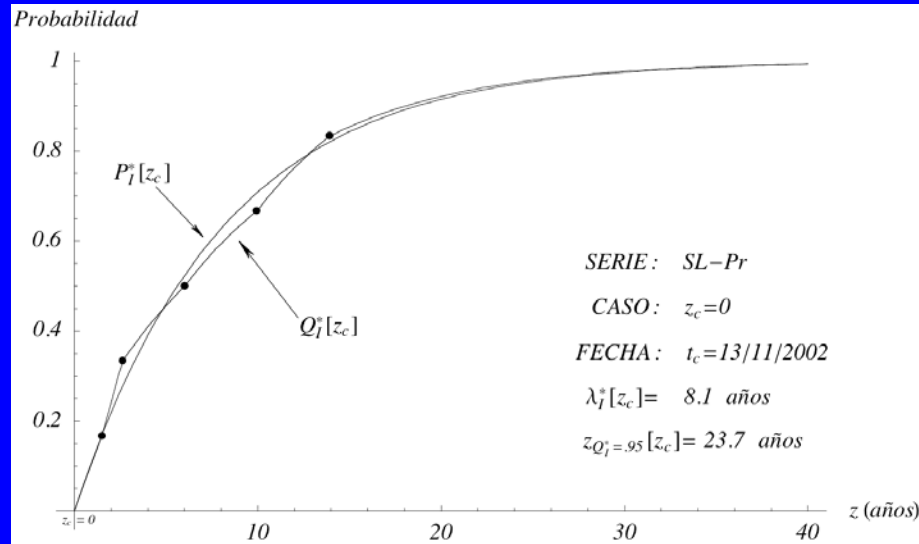
Principales accidentes contaminantes (ITOPF, CEPRECO, CEDRE)

Fecha	Buque	Lugar	C. transportada (Tm)	Carga vertida (Tm)	Fallecidos
18/01/1957	Janina	¿Sur de Vigo?	10300	10300	—
1965	Yanxilas	Vigo	16000	—	—
08/11/1968	Spyros Lemos	Finisterre	20000	20000	¿5?
05/05/1970	Polycommander	Vigo	50000	50000	¿23?
12/05/1976	Urquiola	La Coruña	100000	100000	1
31/12/1978	Andros Patria	Cabo Villano	208000	208000	34
03/12/1992	Aegean Sea	La Coruña	79000	79000	—
13/11/2002	Prestige	Finisterre	77000	63000	—

Datos de duración del tiempo entre ocurrencias registradas (años)

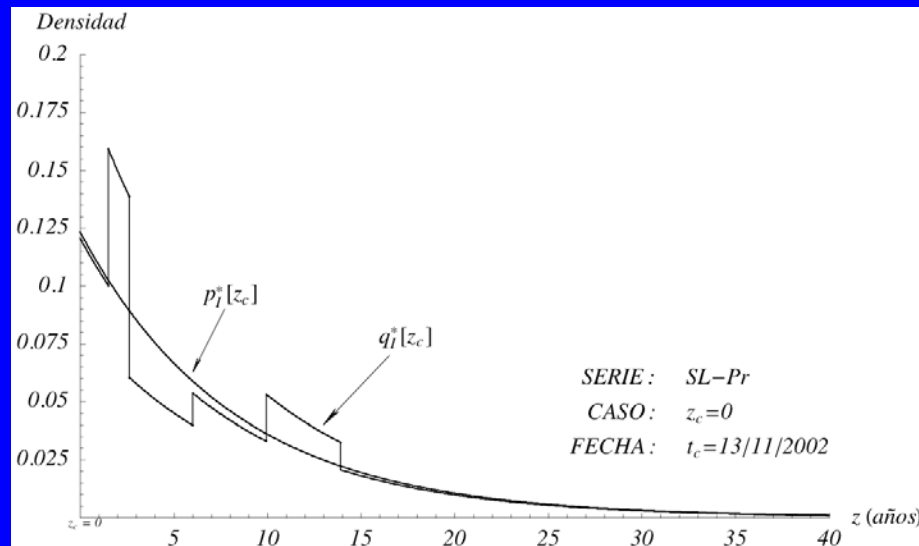
$$D_{Galicia}^{SL-Pr} = \{1.515, 2.637, 6.019, 9.946, 13.924\}$$

Distribuciones de probabilidad tras accidente del Prestige



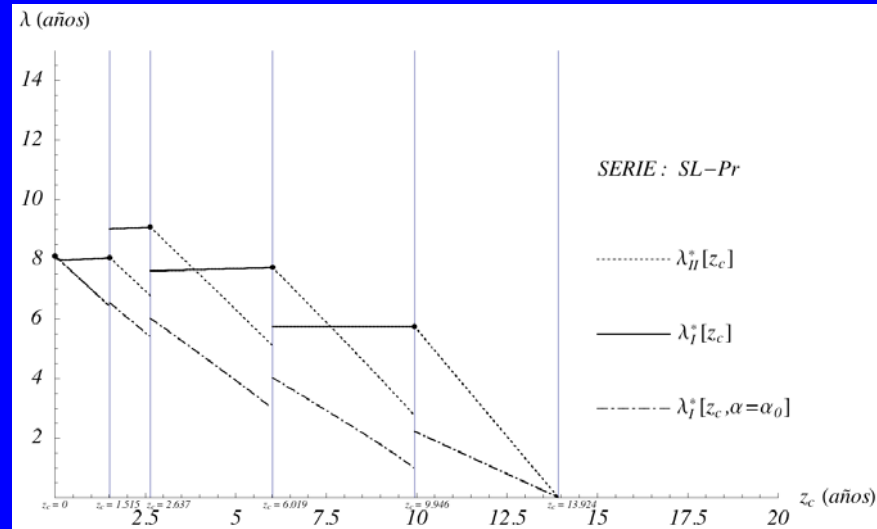
Pares de ligaduras fractiles que codifican las evidencias y distribuciones de referencia e inferida

Referencia:
distribución exponencial uniparamétrica

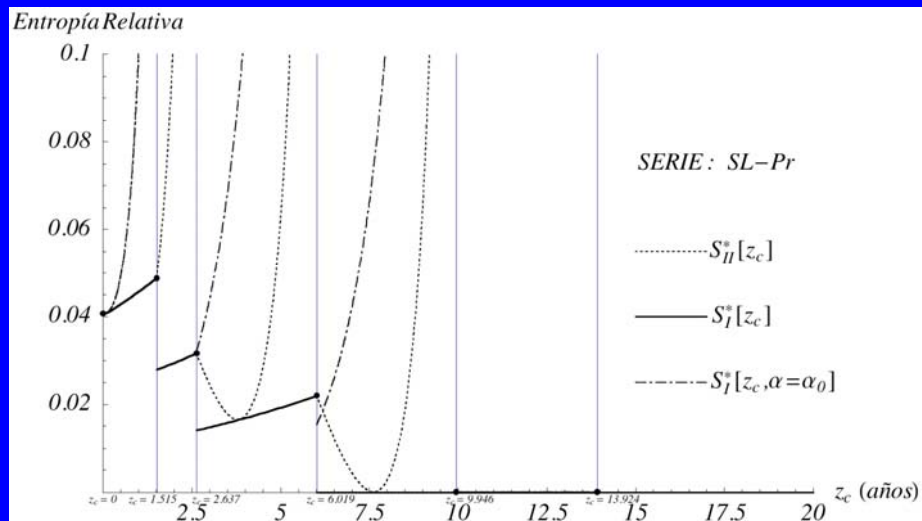


Funciones de densidad de probabilidad de referencia e inferida

Inferencia dinámica: Variabilidad de los modelos de recurrencia con el tiempo transcurrido



Variabilidad de los tiempos medios de recurrencia de referencia con z_c



Variabilidad de las entropías relativas mínimas con z_c

Tiempos de recurrencia de referencia esperados y fractiles inferidos característicos para diversos períodos en Galicia

Período	Tiempo esperado de referencia (años)	Fractil .95 inferido (años)
14/11/2002 – 20/05/2004	7.9	23.5
21/05/2004 – 04/07/2005	9.0	25.7
05/07/2005 – 20/11/2008	7.7	22.1
21/11/2008 – 24/10/2012	5.7	17.2
25/10/2012 –	indefinido	indefinido

Comentarios finales

Consecuencias del análisis dinámico de la recurrencia:

1. *Aparición de una estructura en la variabilidad que muestra la influencia de los datos en la inferencia, a través de una MEMORIA OBSERVACIONAL.*

Dicha memoria se da incluso en el caso analizado de distribuciones de referencia exponenciales, consideradas usualmente como distribuciones sin memoria, con tiempos de recurrencia constantes.

2. *Aparición de una limitación predictiva esencial cuando el tiempo transcurrido es mayor que el penúltimo dato de duración de tiempo entre ocurrencias. NO ES POSIBLE SELECCIONAR distribuciones, ya que la entropía mínima se anula.*

No es una desventaja sino un signo de HONESTIDAD Y CONSISTENCIA con respecto a la información de las evidencias consideradas en el problema de inferencia.

Principales referencias

A. Solana-Ortega y V. Solana (2010), “Towards an infometric solution to the risk modelling crisis: The canon of plausible inference and the representation of observed data”,

Info-Metrics: Theory and Applications in the Social Sciences.

Próxima publicación en *Econometric Reviews*.

A. Solana-Ortega y V. Solana (2008): “Análisis de la recurrencia de grandes vertidos petrolíferos y de la accidentalidad en los sistemas de transporte basados en entropía”,

Estudios de Construcción y Transporte 108: 147-182.

A. Solana-Ortega y V. Solana (2007), “What comes after the Prestige disaster? An entropic approach to modelling the recurrence of major oil tanker spills in Galicia”,

Risk Analysis 27: 901-920.

A. Solana-Ortega y V. Solana (2006): “Entropic inference for assigning probabilities, some difficulties in axiomatics and applications”,

Bayesian Inference and Maximum Entropy Methods in Science and Engineering, AIP, NY, pp. 449-458.