Clase 9 Análisis Estadístico de Eventos Extremos 2023

"Precisiones finales y algunas técnicas no desarrolladas en el curso."

Gonzalo Perera
Departamento MEDIA (CURE)

Organización de la presentación.

- 1. OJO! Dependencia vs correlación: ejemplo MUY ilustrativo.
- 2. El impacto sobre los Tiempos de Retorno de falsos ajustes.
- 3. La Teoría de Grandes Desviaciones y algunas aplicaciones.
- 4. El caso multivariado y las Cópulas Extremales.
- 5. El caso de parámetro continuo: el método de Rice-Wschebor.
- 6. Saliendo del horno: como lo sospechaba, HLE en ciertos contextos conduce a un PPNH
- 7. Propuesta de condiciones para trabajo final y defensa.

1. OJO! Dependencia vs correlación: ejemplo MUY ilustrativo.

- En los cursos elementales de Probabilidad y estadística uno siempre se esmera en diferenciar la Correlación Lineal (Corr, de aquí en más) de las medidas de dependencia, en el sentido siguiente:

- **Si X e Y independientes, entonces Corr (X,Y)=0**
- **Si Corr (X,Y)=0, X e Y NO necesariamente son independientes**
-
- **Ejemplo:** Tomar X una N(0,1) e $Y=X^2$, entonces es evidente que X e Y son recontradependientes (hay un vínculo determinístico entre ellas, si se conoce X, se conoce perfectamente Y!!!) y sin embargo es fácil mostrar que Corr(X,Y)=0. Por otro ejemplo detllado ver GP (2011). Probabilidad y Estadística. Editorial Fin de Siglo. ISBN 978-9974-49-509-8, pág. 145, punto 3.
- Por ende, imaginemos que alguien analiza una serie de tiempos estacionaria con colas no muy pesadas yy a partir de una buena cantidad de datos estima las correlaciones para diferentes *lags* k (=1,2,3…), con sus intervalos de confianza y analicemos si razona de éstas dos formas
- **a) En algunos lags tengo correlación significativamente no nula, por ende mis datos no son *iid* , no tengo un "ruido blanco".**
- **b) En todos los lags que probé la correlación no es significativa, así que mis datos son *iid.***

- **En el caso a) razonó correctamente.**
- **En el caso b) confundió no-correlación con independencia, y es un kamikaze de la Estadística, como veremos a continuación.**
- El algoritmo para simular datos que son FUERTEMENTE DEPENDIENTE que vimos en los slides 9 y 10 de la clase 8 tiene una particularidad muy interesante.
- Tal como lo vimos en la clase, es un algoritmo que mezcla distribuciones de Fréchet de distinto orden y en tal caso no corresponde hablar de correlación. Correlación requiere varianza finita y, por ejemplo la Fréchet de orden 1 tiene esperanza finita por lo cual ni siquiera se puede definir la correlación.
- Sin embargo si en el algoritmo se reemplaza la primer Fréchet por una N(0,1) y la segunda por una N(0, σ^2), donde $\sigma>1$ se encuentran los siguientes resultados (podemos calcular exacto pues conocemos distribuciones)
- **La correlación entre los datos da CERO para todo lag de 1 en adelante!!! (el kamikaze estaría chocho con datos que cree independientes!).**
- Nosotros sabemos que hay dependencia fuerte, pero se puede corroborar el error del kamikaze muy gráficamente con una verdadera medida de dependencia:

- Si X e Y son dos variables aleatorias, definimos como sigue el coeficiente de dependencia distribucional d(X,Y) entre ellas por:
- **$d(X,Y)= \sup_{(u,v)\in R^2} |P(X\leq u,Y\leq v)-P(X\leq u)P(Y\leq v)|$**
- Es muy fácil chequear (GP(2011) pág.77) que:
- **Si X e Y independientes, entonces d(X,Y)=0.**
- **Si d(X,Y)=0, entonces X e Y independientes.**
- En el caso de nuestros datos simulados, con un poco de cálculo, puede mostrarse que: P
- **Para todo $k=1,2,3,\ldots..d(X_1,X_k)=d>0$.**
- **Esto significa que nuestros datos tienen, para todo lag, una dependencia entre sí que es constante "hasta la eternidad" (se estrelló el kamikaze!)**
- **MORALEJA: NO HACER EL PROCEDIMIENTO b) DEL KAMIKAZE!!**

2. El impacto sobre los Tiempos de Retorno de falsos ajustes.

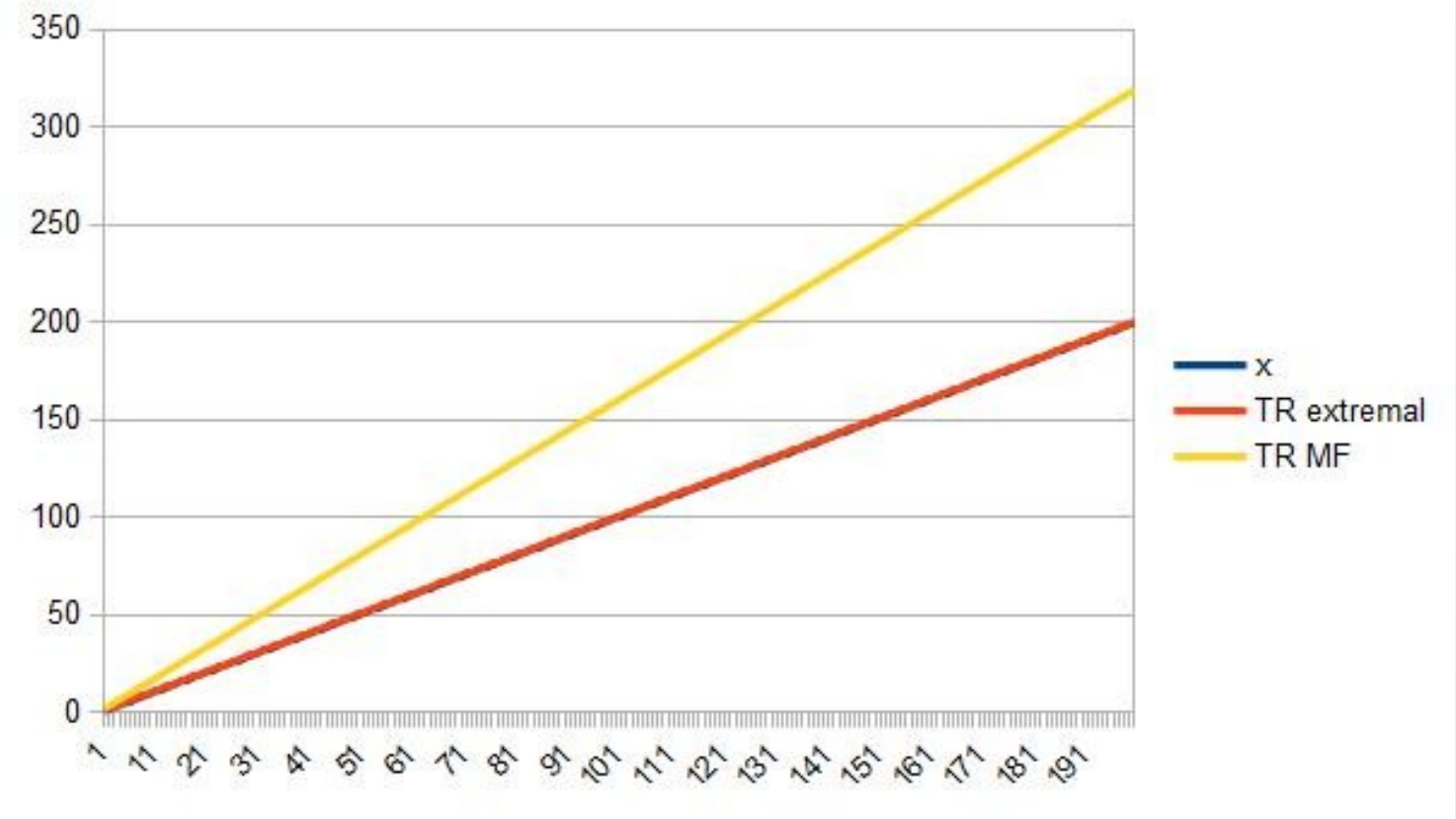
- Esta sección podría llamarse “La pregunta planteada por Angel la clase pasada” (© Angel Segura, S.R.L.).
- Se trata de lo que vimos en el slide 20 de la clase 8, donde una extremal no ajusta a los datos y una mezcla de dos sí lo hace, y lo que vimos en el slide 23, donde una mezcla de dos extremales no ajusta a los datos y una mezcla de tres sí lo hace. Llamaremos al primer escenario “Una vs. Mezcla de 2” y “Mezcla de 2 vs. Mezcla de 3”.
- **Angel preguntaba si la diferencia entre los modelos que no ajustan se apreciaba en algo más “bajada a tierra”, como Tiempos de Retorno (TR), por ejemplo.**
- Siendo datos simulados podemos calcular exactamente los TR y los modelos que ajustan dan TR casi idénticos pero esa comparación no tiene gracia.
- Lo que veremos aquí es la comparación “justa”, que es entre los modelos estimados, comparando los TR del que no ajusta con los TR del que sí ajusta.
- **Veremos diferencia apreciables, pero de tipo distinto en cada caso.**

Una vs Mezcla de 2.

x	TR extremal	TR MF	% ext/MF
1	1,5819767069	2,1796893163	72,578082347
2	2,5414940825	3,7468504516	67,830144687
3	3,5277264732	5,3314019747	66,168833074
4	4,5208116642	6,9204152266	65,325728532
5	5,5166555661	8,5112285463	64,816207626
6	6,5138824631	10,102945813	64,475080668
7	7,5119007146	11,695180982	64,230735086
8	8,510413955	13,287740413	64,047111778
9	9,5092573546	14,880516292	63,904082142
10	10,508331945	16,473443832	63,789527267
11	11,507574714	18,066481751	63,69571493
12	12,506943641	19,659602506	63,617479738
13	13,506409624	21,252787012	63,551239736
14	14,505951875	22,846021627	63,49443291
15	15,505555144	24,439296346	63,445178309
16	16,505207994	26,032603657	63,402063858
17	17,504901678	27,625937816	63,364008833
18	18,504629391	29,219294354	63,330172068
19	19,504385762	30,812669741	63,29988906
20	20,504166493	32,406061152	63,27262791
21	21,503968104	33,999466301	63,247957817

Una sola es alarmista, pues da TR mucho más cortos. De hecho, variando x de 1 a 200 los TR de Una sola son proomedialmente el 63% de los TR de la mezcla de 2 (consecuencias: descrédito, sobreinversión o recurso ocioso, etc.)

Gráficamente



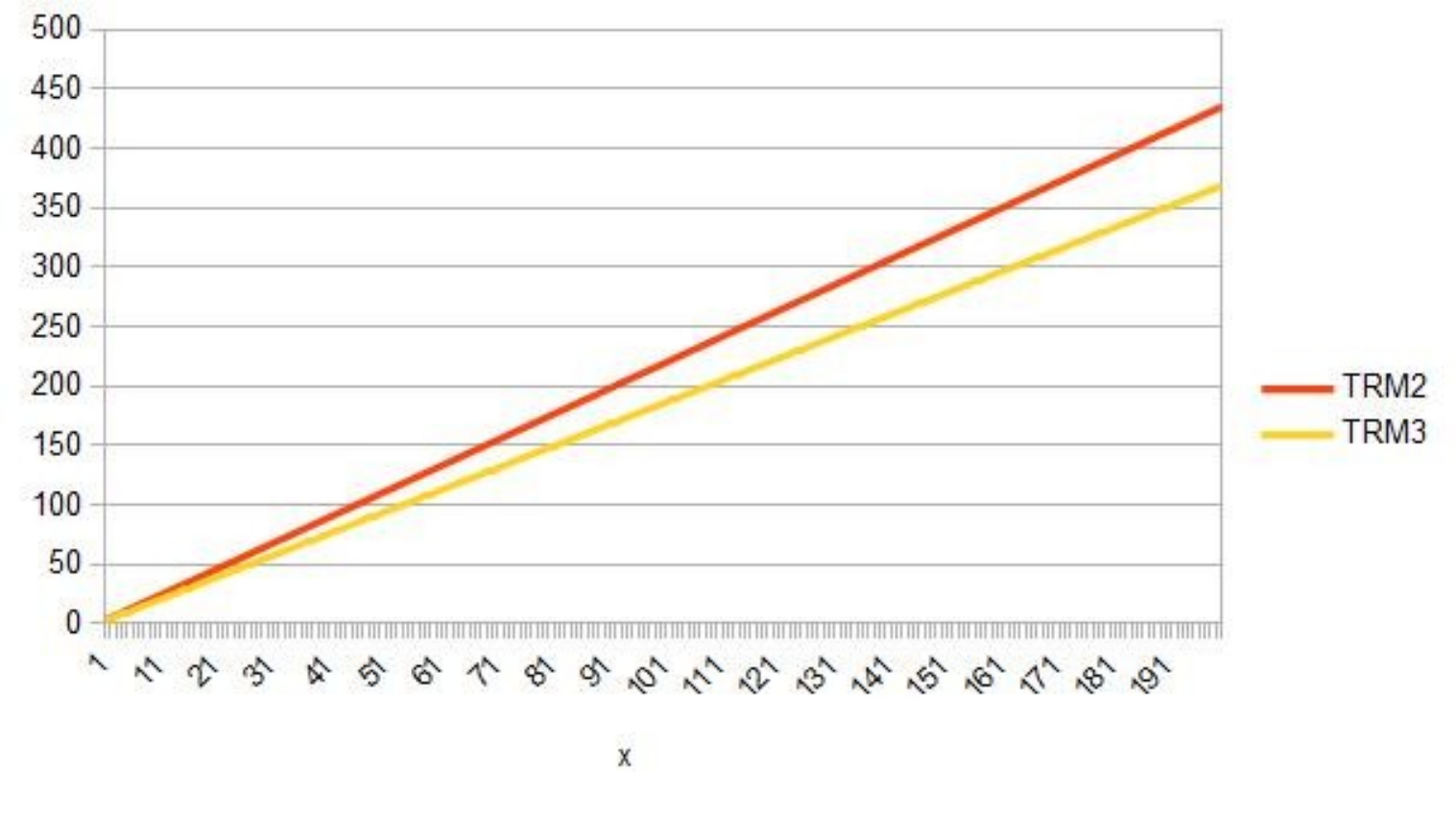
En este caso es realmente grosero el error del modelo que no ajustaba, y alarmista.

5.Mezcla de 2 vs. Mezcla de 3.

x	TRM2	TRM3	% M2/M3
1	2,8388102145	2,4848782406	114,24343326
2	4,9949282891	4,304680707	116,03481487
3	7,1625311176	6,1384745909	116,68258965
4	9,3332174607	7,9759772504	117,01660082
5	11,505171574	9,8149967376	117,22033008
6	13,677769125	11,654783801	117,35755342
7	15,850737864	13,495012821	117,45626384
8	18,023940128	15,335519519	117,53067841
9	20,197298833	17,176212051	117,58878368
10	22,370767454	19,017035052	117,63541158
11	24,544316251	20,857953162	117,67365695
12	26,717925323	22,698942739	117,70559374
13	28,891580854	24,539987378	117,73266387
14	31,065272949	26,381075337	117,75590097
15	33,238994336	28,22219799	117,77606531
16	35,412739553	30,063348861	117,79372856
17	37,586504416	31,90452299	117,80932888
18	39,760285666	33,745716514	117,82320773
19	41,934080728	35,586926382	117,83563514
20	44,107887537	37,42815015	117,84682748
21	46,281704423	39,269385841	117,85696016

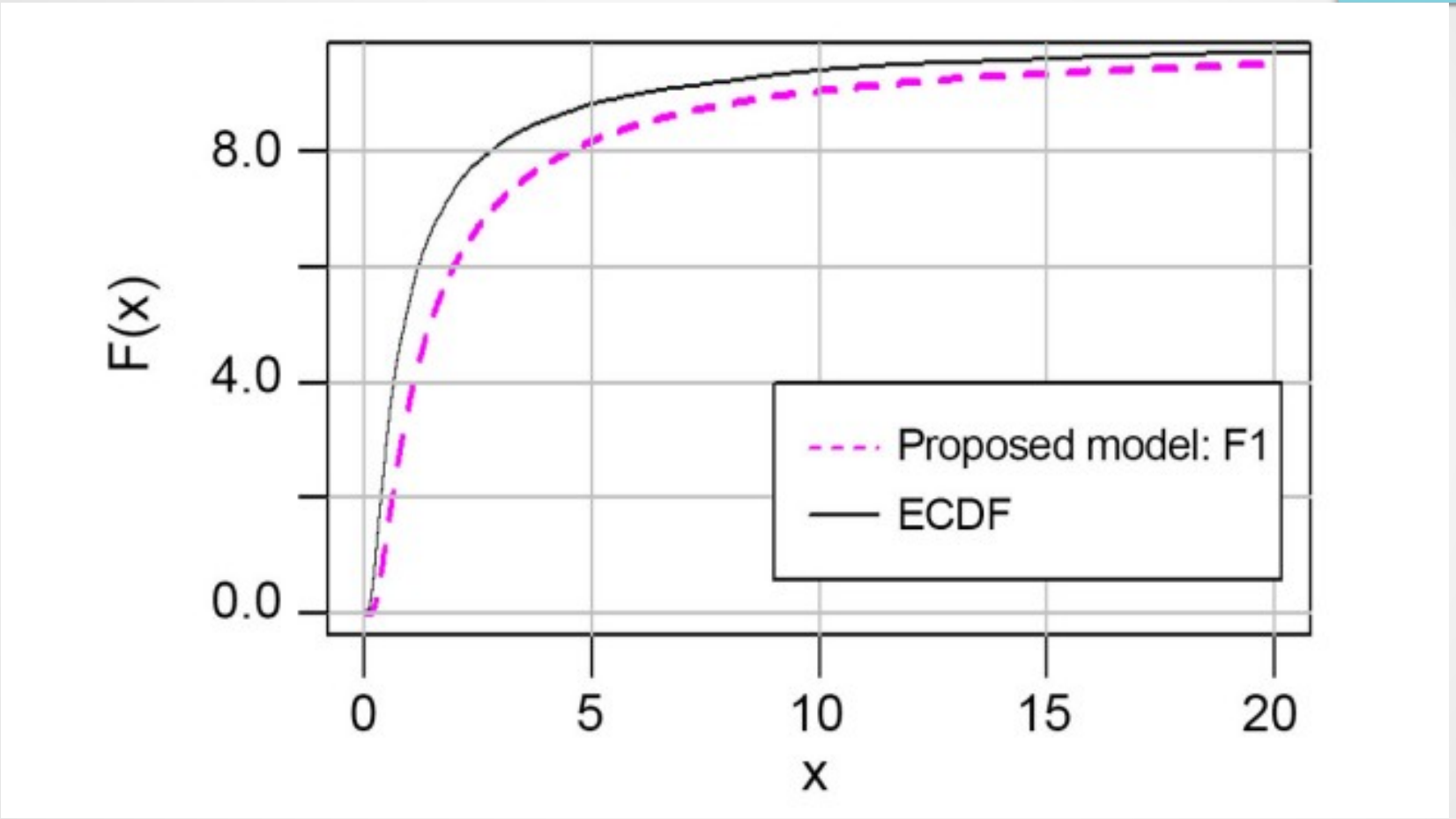
Los TR de la Mezcla de 2 son mayores a los de la Mezcla de 3, con x de 1 a 200, promedialmente 118% mayores, Erran menos que antes, pero arriesgan seguridad

Gráficamente

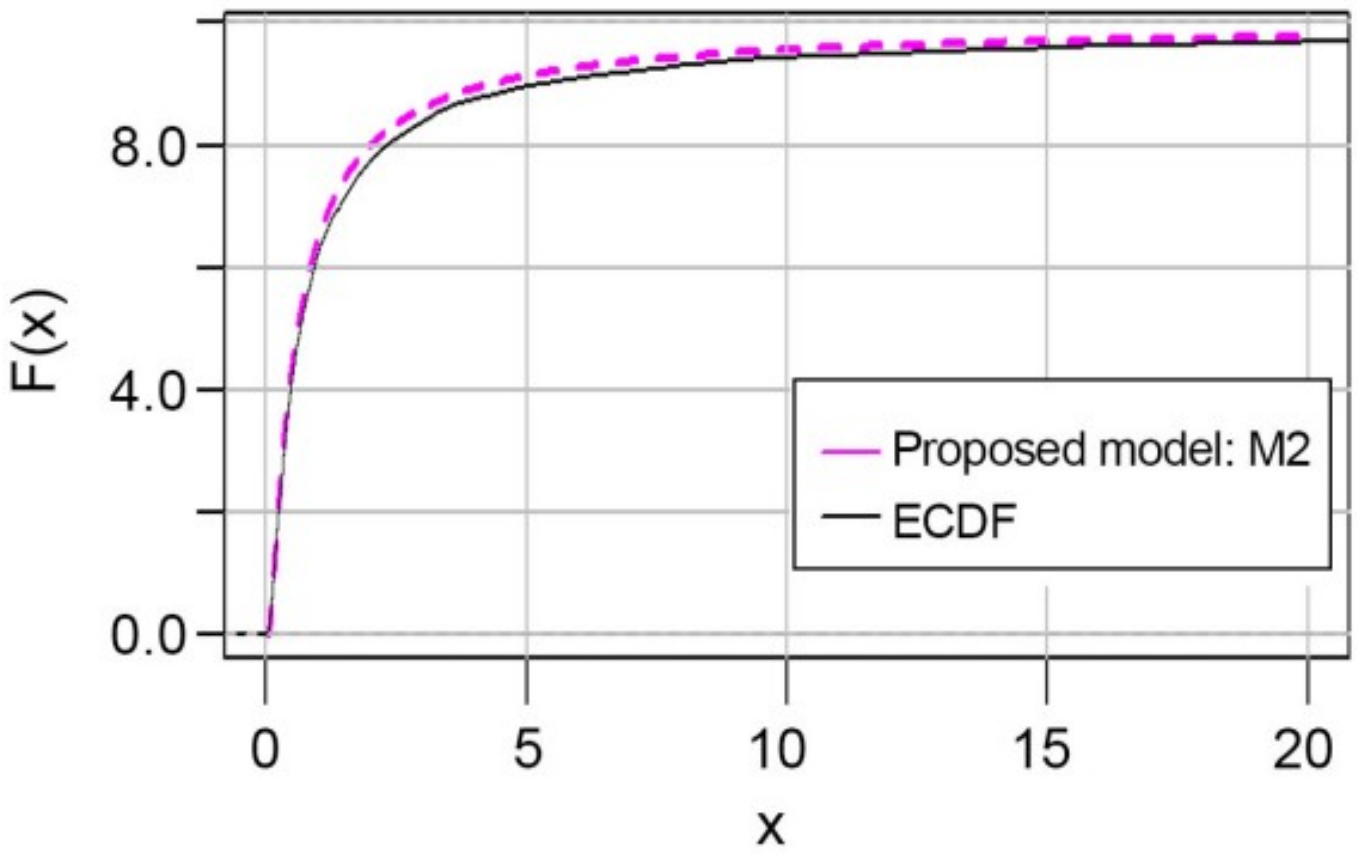


El modelo que no ajusta erra de forma clara, y con consecuencias graves.

El tipo de error (no su cuantía) se ve en las gráficas de ajuste de la clase anterior



Cola (1-distribución) de Una Extremal mucho más pesada que la de la empírica!!!



Colas de la Mezcla de 2 más livianas que las de la empírica.

3. La Teoría de Grandes Desviaciones y algunas aplicaciones.

Básicamente la Teoría de Grandes Desviaciones (Large Deviations) nación al impulso de los cálculos de riesgos en seguros y recibió un fuerte impulso en la Teoría de la Información, Física de Partículas y Mecánica Estadística y ha tenido relevantes aplicaciones en áreas como Bioinformática y Telecomunicaciones y que fue sistematizada como teoría por primera vez por Varadhan en 1966. En términos cualitativos, de qué se trata? Pues se trata de estudiar el orden magnitud de la probabilidad de un evento tan extremo que es sumamente raro, con probabilidad bajísima. Esto es probabilidades de 10^{-6}, 10^{-9} o 10^{-12} son todas bajísimas, pero no tienen nada que ver entre sí. Si partiéramos de la base que la probabilidad es del orden de 10^{-k}, es decir que la probabilidad es $C10^{-k}$, con C al menos 1 y C<10, la pregunta es cuánto es k.

Para que ésta pregunta tenga sentido, tengo que lidiar con un área con eventos extremos de bajísima probabilidad y con un volumen de datos realmente masivo (por este motivo no es una técnica que desarrollemos en el curso). .El análisis de una rar mutación en datos genómicos o del colapso o congestión aleatoria de una red de Telecomunicaciones monitoreada cada fracciones de segundo (distinguir de las congestiones sistemáticas, como las de medianoche de las fiestas de fin de año) son ejemplos. Describiremos pues someramente algunos resultados que van permitiendo avanzar en esta dirección y las aplicaciones que conocemos mejor (Telecomunicaciones, Redes IP)

- La referencia “bíblica”para este tema (expuesto con mucho tecnicismo matemático) es:
- ***Dembo, A.; Zeitouni, O (1993), Large Deviations Techniques and its Applications, Jones and Bartlett, New York.***

Lo describiremos de manera muy cualitativa. El primer gran resultado de esta Teoría es el ***Teorema de Crámer-Chernov***.

La descripción es la siguiente supongamos que tenemos n datos iid cuya distribución tienen colas livianas y por ende tienen media μ, varianza finita y si X tiene dicha distribución M(s)= E(exp(sX)) es finita para todo valor de s (función generatriz de momentos). Como por la Ley de los Grandes N;umeros el promedio empírico de los n datos tiende a la media cuando n tiende a infinito, si a > 0 es fijo (no varía con n), entonces que para n grande el promedio empírico se aparte de la media μ más de a constituye una **gran desviación**, evento al que llamaremos E(n).

El teorema de Crámer Chernov establece que:

El límite para n tendiendo a infinito, de (1/n) log P{E(n)}=C(a) <0, donde C(a) se calcula explícitamente a partir de la función generadora de momentos.

Esto puede reescribirse como

P{E(n)}=b(n)exp(C(a)n), donde b(n(sucesión determinística tal que (1/n) log{b(n)} tiende a 0 cuando n tiende a infinito.

Pero el ***Teorema de Bahadur-Rao*** aporta un conocimiento muy preciso de b(n), habilitando el desarrollo de aplicaciones como las que suscintamente veremos acá.

Ejemplo: Telecomunicaciones.

El tráfico IP es un claro ejemplo de recurso compartido y basado en el multiplexado: mandar muchos tráfico individuales por un mismo enlace para que cuando no usa el otro aproveche.

Un nodo de una red IP , sobre todo en su ‘backbone” (columna vertebral, línea principal de conexión) tiene mútiples enlaces entrantes, algunos de salida con ciertas capacidades (podría pensarse en la capacidad total de salida como primera aproximación, C) y un “buffer”, que es donde van esperando los paquetes de información que entraron en la cola y no pudieron ser despachados. Si el buffer, cuya capcidad es B, se llena, hay **congestión o saturación**, hay que tirar paquetes y mandar mensajes de falla o solicitud de reintento. Diseñar una red (determinar C y B) de forma que la probabilidad de saturación sea menor que un nivel dado, muy chiquito y considerado tolerable, es un problema central desde los inicios de las Telecomunicaciones con el gran Agner Erlang.

Si yo tengo una conexión a internet en mi casa (por la vía que sea, asimétrica o simétrica respecto a velocidades de “subida”y “bajada” recordar la A de ADSL y evolucionm: clouding, etc.), tengo velocidades CONTRATADAS de conexión pero las realmente disponibles, las “efectivas” dependen del TRÁFICO CRUZADO en toda la red.

Esto llevó a Frank Kelly a definir el concepto de ***Ancho de Banda Efectivo*** (*Effective Bandwidth*), dándole una formulación matemática precisa que obviamente se apoya en la Teoría de Grandes Desviaciones, pues la saturación de un nodo troncal de una red es una Gran Desviación, claramente.

Kelly, F. (1996) “Notes on Effective Bandwidth”, in Stochastic Networks: Theory and Applications, edited by Kelly, Zachary and Ziedins, Oxford University Press.

A partir de este enfoque y resultados muy finos como los Courcoubetis-Weber, Ravi Mazumdar, etc., de Grandes Desviaciones en sistemas muy complejos como el arribo aleatorio de diversos tipos de tráfico al backbone de la red, han permitido hacer efectivamente la determinación de C, B, etc.

Un ejemplo de paper con tal tipo de resultados que conozco bien:
Laura Aspirot, Pablo Belzarena, Paola Bermolen, Andrés Ferragut, Gonzalo Perera, María Simon, (2005). Quality of service parameters and link operating point estimation based on effective bandwidths, Performance Evaluation, Volume 59, Issues 2–3, Pages 103-120, ISSN 0166-5316. https://doi.org/10.1016/j.peva.2004.07.006.

4. El caso multivariado y las Cópulas Extremales.

Como primer punto: googleen tranquilxs “Ëxtremal Copula”…….

En 1959 Sklar publicó una breve nota estableciendo una forma funcional de determinar la dependencia entre dos o más variables independientemente de sus distribuciones marginales, en un resultado que pasó casi inadvertido, como una curiosidad.

Sklar, A. (1959), "Fonctions de répartition à n dimensions et leurs marges", *Publ. Inst. Statist. Univ. Paris,* **8**: 229–231.

A partir de 1999, Paul Embrechts en Zürich para la Banca Suiza, luego Fermanian en Francia para el Crédit Lyonnais (resultados que se divulgaban con delay) se dieron cuenta que el Teorema de Sjk;ar proveái el concepto de Copula, una forma de trabajar con datos multivariados con énfasis en su interdependencia. La teoría sistematizada y entre los muchos tipos de Copula (arquimedianas, etc.) en Finanzas tenía particular interés las Copulas Extremales, que son las que se corresponden a variables que no dependen mucho en sus valores moderados pero que sincronizan o tienen ferte asociación en sus extremos. En Suiza, el quiebre de un banco no quiebra el sistema, el quiebre (evento extremo si los hay) de varias a la vez o muy cercanos en el tiempo, quiebra el sistema entero . En su tesis de maestría, Gabriel Illanes hizo un estudio de este tipo para el BCU.

Es particularmente importante el ö cerca en el tiempo”. Ejemplifiquemos con un problema ambiental. En Uruguay, para estudiar curvas IDF en las costas del Plata, había polémica sobre si lluvias y mareas intensas estaban de alguna manera asociadas o si por el contrario eran independientes. A su modo, ambos tenían razón. Mediante el uso de mezclas de cópulas (ptra vez mezclas!) se pudo probar que: en el mismo día, niveles mareales y lluvias no presentaban asociación significativa, pero que si desfasaban 24 hrs “lluvias ayer, mareas hoy”, había una muy alta asociación, y se constató que eso tenía una explicación física por los efectos de ciertas tormentas con origen en la Patagonia y que ascienden hacia la boca del estuario platense. Para curvas IDF, TR, etc., ver:

Silveira Luis, Usera Gabriel, Alonso Jimena, Scavone Martín, Chreties Christian, Perera Gonzalo, González Meliza (2014). Nuevas curvas intensidad-duración-frecuencia de precipitación para el departamento de Montevideo, Uruguay. Agrociencia Uruguay - Volumen 18 1:113-125.

Para cópulas extremales en general ver:

Gudendorf, Gordon & Segers, Johan. (2010). Extreme-Value Copulas. ISBN 978-3-642-12464-8, doi 10.1007/978-3-642-12465-5_6.

5. El caso de parámetro continuo: el método de Rice-Wschebor.

- Un proceso estocástico de parámetro continuo es una curva sorteada al azar. Sus trayectorias asumamos que son continuas, pero puede que sean diferenciables (Regresión de Nadaraya-Watson, por ejemplo) o que no lo sean en ningún punto (movimiento browniano y derivados). Hablaremos de procesos regulares e irregulares para referir a uno u otro tipo. En un intervalo de tiempo acotado, a un nivel dado u, las curvas regulares lo cortan un número finito de veces, las irregulares en general infinitas veces. El estudio del máximo de uno u otro tipo de curvas es un problema muy viejo, que puede pensarse como “el problema de barrera”, que se ha abordado y aborda desde diversas perspectivas, y donde el nombre de Enrique Cabaña en Uruguay es referencial, incidiendo en su trabajo conjunto muchísimo sobre el desarrollo académico de Mario Wschebor.
-

Entre 1944 y 1945 Rice desarrolló fórmulas para el cálculo de los momentos del número de cruce, durante un inetrvalo acotado dado, de un nivel u, de un proceso regular.

En su libro de 1985, Surfaces Aléatoires (y con el previo referido aporte de Cabaña), Wschebor, utilizando herramientas muy finas pocesos estocásticos y de la escuela italiana de geometría integral (medida de De Giorgi, etc.) lleva la fórmula de Rice a grandes niveles de generalidad, comienza a vincularla con el estudio de valores extremos y a través de un concepto que luego desarrollar;ia intuitivamente y que es intuición física pura.

Normalmente, una muestra de un proceso estocástico de parámetro continuo, para ser realista se pensaba en términos de discretizaciones”tomar observaciones en instantes equiespaciados, etc. Wschebor pensaba que si veía una curva en una pantalla, entonces estaba viendo el pasaje de la curva por un filtro suavizador, cuto grado de suavización tiene que ver con la escala de definición de la pantalla. De hecho un proceso irregular es imposible de observar, todo lo que podemos ver en todo caso es su suavización por un filtro. Para Wschebor ése era el muestreo natural de proceso irregular: su suavización por un flitro regularizador, que al tender el suavizado a cero, tiende a la irregularidad.

Para un proceso irregular, digamos que definido en el intervalo de tiempo [0,1], se define la medida de ocupación m de un conjunto B (sobre el eje de las y) cómo la cantidad de tiempo que la trayectoria del proceso permanece en B. Si, por ejemplo, $B=(u,+\infty)$, la medida de ocupación de B da cuánto tiempo la trayectoria está por arriba de u y eso obviamente tiene que ver con qué tan probable es que el proceso esté por arriba de ur, y por lo tanto, con su comportamiento extremal. A su vez, bajo condiciones bastante generales, en procesos irregulares, esta medida tiene una densidad que se llama tiempo local, L. Si $L(u)> L(v)$, la interpretación intuitiva es que el proceso “pasa más por la altura u que por la altura v.

.

Wschebor mostró que tomando un proceso irregular y suavizándolo con un filtro, el número de cruces del suavizado con el nivel u, en la medida que el suavizado tendiera a cero, con un reescalamiento que depende del suavizado, tiende al tiempo local. A partir de allí encontró estimaciones para la medida de ocuación y para integrales de la medida de ocupación, siempre basándose en el número de cruces del suavizado. UN ejemplo de estimación de integrales de la medida de ocupación es:

Perera, G. and Wschebor, M. (1998). Crossings and occupation measures for a class of semimartingales. *The Annals of Probability*, *26*(1), pp.253-266.

“

Expliquemos por qué las integrales respecto a la medida de la ocupación son tan relevantes. Supongamos que el proceso que describe lo que estamos midiendo en continuo es irregular y con medida de ocupación donde los resultados de Wschebor son aplicables. Imaginemos que si el proceso está por debajo de un nivel u no hay riegos, pero cuando pasamos por arriba de un, cada vez hay más riesgos y que lo podemos cuantificar. Esto es,si v>u, f(v) es el riesgo que produce que el proceso alcance el nivel v. Naturalmente , completemos la definición de f estableciendo que si $w \leq u$ entonces $f(w)=0$. Si queremos una medida del riesgo global de la trayectoria observada del proceso en cuestión, claramente es la integral de f respecto a la medida de ocupación en toda la recta (en todo el eje de las y, digamos). Los resultados de Wschebor que permiten estimar dichas integrales a partir de los cruces de regularizaciones del proceso dan pues una estimación del riesgo, donde se puede además establecer inetrvalos de confianza por conocerse la distribución asintótica.

:La obra medular de Wschebor es el libro escrito con su principal colaborador:
Azaïs, Jean-Marc, and Mario Wschebor (2009). *Level sets and extrema of random processes and fields*. John Wiley & Sons,

- Además de Azaïs, por supuesto Cabaña, y GP trabajaron con Mario Wschebor, o usaron sus técnicas de manera relevante una destacada lista de colegas como (lista para nada exhaustiva (en rojo quienes fueron dirigidos en sus doctorados)
- José Rafael León (Chichi), Joaquín Ortega, Corine Berzin, Marie Kratz, Ernesto Mordecki, Diego Armentano, Felipe Cucker,Juan Cuesta-Alberto,Jean-Marc Bardet, Serge Cohen...

6. Saliendo del horno: como lo sospechaba, HLE en ciertos contextos conduce a un PPNH

Bajo la convicción de que PPNH y derivados es una herramienta subexplotada para analizar eventos extremos espacio-temporales, hace tiempo buscamos vincular técnicas reconocidamente ligadas al Análisis de Extremos y en los días recientes surgió::

Bajo ciertas condiciones muy razonables, una versión generalizada de HLE en una superficie tiene como límite un PPNH (GP, recién)

Esta suerte de “trailer”esperemos sea desarrollado en un seminario o en una mucho mejor próxima edición de este curso.

Vayamos a las condiciones de evaluación del curso.

7. Propuesta de condiciones para trabajo final y defensa.

- 1. Quien se haya comunicado para mostrar o pedir datos y no haya recibido al menos un mail de nuestra parte, mil disculpas y nos contactamos.
- 2. El trabajo sobre eso datos debe escribirse como un informe o short note, con una Intro explicando qué son los datos y cuál es el problema, con descripción de métodos usados y resultados obtenidos, tablas, gráficas y breve bibliografía (incluyendo links si están en Internet). Máximo de extensión: 12 páginas en tipo de letra y espaciado razonable.
- 3. A eso debe agregarse los scripts de R usados para obtener los resultados antes mencionados.
- 4. Ambas cosas deben enviarse a los tres mails (GP, CC, AS)
- 5. La fecha límite para dicho envío es el Lunes 8 de abril del 2024 (inclusive). Tener presente que en 2024 semana de Turismo va del Lunes 25 de marzo al Viernes 29 de marzo.
- 6. En la última semana de abril, en fecha a precisar con quienes cumplan el punto anterior, se harán las defensas orales. Cada defensa consiste en un exposición de máximo 20 minutos basada en una presentación de máximo 6 slides. Pueden hacerse las presentaciones por Zoom y todos pueden asistir a las presentaciones. Terminadas la exposición los docentes haremos preguntas.
- 7. Las calificaciones se enviar;an individualmente por mail posteriormente a todas las defensas.
-
-

Por toda su paciencia…..MUCHAS GRACIAS!!!!