

S10. Estadística. SALA M2

Coordinada por: **Javier Girón**, U. Málaga; **Graciela González Farias**, CIMAT-Monterrey.

PROGRAMA

jue19 15:00-15:40 → CARLES CUADRAS , <i>Sobre la dimensión geométrica en cópulas bivariantes.</i>	U. Barcelona
jue19 15:40-16:20 → RODRIGO MACÍAS , <i>Cluster selection criteria based on a block-shaped partition for one-mode dissimilarity matrix.</i>	CIMAT
jue19 16:20-17:00 → JOSÉ MIGUEL BERNARDO , <i>Contraste Objetivo de Hipótesis Científicas.</i>	R.A.C.
jue19 17:00-17:40 → RAÚL RUEDA , <i>Contraste de hipótesis en modelo lineales jerárquicos.</i>	IIMAS
jue19 17:40-18:20 → MANUEL MENDOZA , <i>Familias Exponenciales e Inferencia Bayesiana.</i>	ITAM
vie20 09:00-09:40 → ENRIQUE CASTILLO , <i>Dos Aplicaciones de la Estadística a Modelos de Tráfico.</i>	U. Cantabria y R.A.C.
vie20 09:40-10:20 → JORGE ARGÁEZ SOSA , <i>Usando mapas a priori para modelar una variable discreta sobre una región geográfica.</i>	U. Autónoma de Yucatán
vie20 10:20-11:00 → JAVIER GIRÓN , <i>Análisis bayesiano objetivo de coste-efectividad para tratamientos clínicos en presencia de covariables.</i>	U. Málaga y R.A.C.

RESÚMENES

Ponente: CARLES CUADRAS U. Barcelona

Título: *Sobre la dimensión geométrica en cópulas bivariantes*

Hora: (M2) jue19 15:00-15:40

Resumen: Es una práctica usual del análisis multivariante determinar el número de dimensiones que deben considerarse en un estudio estadístico. Las primeras dimensiones en análisis de componentes principales, análisis de correspondencias, análisis de correlación canónica, etc., explican una parte importante de la variabilidad.

Se puede llevar a cabo una descomposición similar en cópulas bivariantes, donde la dimensión geométrica viene determinada por el número de correlaciones canónicas. La dimensión geométrica reducida se calcula en el sentido de que proporciona coordenadas en relación a la distancia j_i -cuadrado, y la variabilidad viene representada por el coeficiente de contingencia de Pearson.

La cópula FGM (Farlie-Gumbel-Morgenstern) es un ejemplo de familia con dimensión uno. La cópula de Fréchet tiene dimensión numerable pero sin direcciones principales. Algunas cópulas con dependencia estocástica débil (AMH, Gumbel-Barnett) son prácticamente de dimensión uno. Otras cópulas (Clayton-Oakes, Joe), también de dimensión geométrica numerable, se describen con dos o tres dimensiones.

Un caso diferente lo constituye la cópula Cuadras-Augé. Para esta familia las correlaciones canónicas constituyen un conjunto no numerable, es decir, la dimensión es continua.

[1] C. M. Cuadras, *On the covariance between functions*, Journal of Multivariate Analysis **81** (2002), 19–27.

[2] ———, *Constructing copula functions with weighted geometric means*, Journal of Statistical Planning and Inference **139** (2009), 3766–3772.

[3] C. M. Cuadras and D. Cuadras, *Eigenanalysis on a bivariate covariance kernel*, Journal of Multivariate Analysis **99** (2008), 2497–2507.

[4] C. M. Cuadras, J. Fortiana, and M. Greenacre, *Continuous extensions of matrix formulations in correspondence analysis*. In: *Innovations in Multivariate Statistical Analysis*, 2000.

ccuadras@ub.edu

Ponente: RODRIGO MACÍAS CIMAT

Título: *Cluster selection criteria based on a block-shaped partition for one-mode dissimilarity matrix*

Hora: (M2) jue19 15:40-16:20

Resumen: Joint work with J. Fernando Vera.

One of the most important problems in clusters analysis and that in general is not solved is the determination of the number of clusters underlying in the structure of a data set. Several criteria have been proposed, among them an efficient criterion for rectangular matrix considered as a set of p -dimensional vectors-in K clusters, that presents a very effective performance. However, in many situations the dimensionality of the observations is often unknown, and/or only a proximity relation between the elements to be clustered is known. In this paper we propose an extension of some of the most employed criteria for their direct application to a dissimilarity matrix. A simulation study to compare the performance of this criteria in the dissimilarity space as well as in the original p -dimensional space is developed to show the efficiency of the proposed procedure.

rmaciasp@gmail.com

Ponente: JOSÉ MIGUEL BERNARDO

R.A.C.

Título: *Contraste Objetivo de Hipótesis Científicas*

Hora: (M2) jue19 16:20-17:00

Resumen: El resultado final de un análisis inferencial Bayesiano es la distribución final conjunta de todos los parámetros de interés, pero resulta frecuentemente imprescindible resumir su contenido para la asimilación pública de sus implicaciones. En particular, el análisis comparativo de los resultados obtenidos con dos estrategias diferentes aplicadas al mismo problema suele centrarse en el análisis de la diferencia o del cociente de alguno de sus parámetros básicos (*estimación*), con especial atención a la compatibilidad de los datos con la posibilidad de que esa diferencia sea nula o de que ese cociente sea la unidad (*contraste de hipótesis*).

El uso de la teoría de la decisión con funciones de pérdida continuas permite una solución integrada a los problemas de estimación y de contraste, en los que puede ser utilizada la *misma* distribución inicial que, además, puede ser impropia. En este trabajo se describe como el uso simultáneo de distribuciones iniciales de referencia y de funciones de pérdida invariantes, basadas en una distancia funcional entre modelos (en lugar de una distancia entre sus parámetros), permite dar una solución Bayesiana objetiva unificada tanto a los problemas de estimación, como a los problemas de contraste de hipótesis.

Los métodos descritos se ilustran con un ejemplo importante, la comparación de las proporciones de éxito asociadas a dos poblaciones binomiales, un problema relativamente elemental sobre cuya solución dista de existir un consenso, como demuestra la polémica recientemente generada en los medios científicos sobre la posible eficacia de la nueva vacuna RV144 contra el virus de la inmunodeficiencia humana.

jose.m.bernardo@uv.es

Ponente: RAÚL RUEDA

IIMAS

Título: *Contraste de hipótesis en modelo lineales jerárquicos*

Hora: (M2) jue19 17:00-17:40

Resumen: Trabajo conjunto con Patricia Romero.

El contraste bayesiano de hipótesis puntuales puede abordarse desde diferentes puntos de vista; desde la perspectiva de la teoría de decisión, debemos especificar una función de pérdida y elegir la decisión con pérdida esperada mínima. En esta plática se propone una solución basada en la discrepancia intrínseca como función de pérdida, en el caso de modelos lineales jerárquicos. Un estudio sobre el desempeño académico de alumnos de nivel preuniversitario, servirá para ilustrar estas ideas.

pinky@sigma.iimas.unam.mx

Ponente: MANUEL MENDOZA

ITAM

Título: *Familias Exponenciales e Inferencia Bayesiana*

Hora: (M2) jue19 17:40-18:20

Resumen: Trabajo conjunto con Eduardo Gutiérrez-Peña.

Las familias exponenciales de distribuciones han sido ampliamente estudiadas en la literatura estadística y siguen ofreciendo un campo fértil para la investigación. En este trabajo se examina, de nuevo, este tipo de modelos como solución a distintos problemas de optimización. En particular, esta idea puede utilizarse para justificar la elección de este tipo de distribuciones cuando se trata de describir fenómenos aleatorios con una cantidad mínima de supuestos. También se muestra que, en el contexto de un análisis estadístico Bayesiano, resulta natural el empleo de distribuciones conjugadas y, además, se muestra como a través de esta idea se pueden inducir diversas distribuciones iniciales no informativas.

mendoza@itam.mx

Ponente: ENRIQUE CASTILLO

U. Cantabria y R.A.C.

Título: *Dos Aplicaciones de la Estadística a Modelos de Tráfico*

Hora: (M2) vie20 09:00-09:40

Resumen: Trabajo conjunto con María Nogal y Aida Calviño.

Se presentan dos aplicaciones de la Estadística a modelos de tráfico. En la primera se supone que los flujos de las diferentes rutas son variables independientes de la familia Gamma desplazada $\mathcal{H}(\alpha, \theta, \lambda_0)$, donde el parámetro λ_0 es común a todas las rutas. Como consecuencia, los flujos de los arcos, nodos, pares OD (origen-destino), etc. pertenecen también a la familia $\mathcal{H}(\alpha, \theta, \lambda_0)$. Se presenta un método bayesiano que utiliza la familia de distribuciones conjugada para estimar los diferentes flujos anteriores. Se describe en detalle la asignación de la distribución *a priori*, el muestreo, la actualización de la distribución *a posteriori* y de su moda, que se propone para sustituir a la media. Como método de muestreo se utiliza principalmente el escaneo de matrículas, aunque éste puede ser complementado por aforadores estándar.

En la segunda aplicación, se introduce una red bayesiana para modelizar un modelo dinámico de tráfico con demanda estocástica, en el que las variables objetivo son los tiempos de viaje, los flujos o las intensidades de flujo en los arcos y su evolución temporal con objeto de predecir el tráfico futuro en determinados lugares e instantes. Se supone que las variables marginales son beta generalizadas, en el sentido de que se incorporan parámetros de localización y escala a la beta estándar, y que cuando se transforman marginalmente a normales, se convierten en normales multivariadas. El aprendizaje de la red bayesiana se hace mediante la observación de datos en diferentes lugares y tiempos. Para obtener una muestra homogénea y significativa, se seleccionan los datos con la idea de eliminar datos atípicos (normalmente resultantes de días festivos o con ocurrencias de accidentes o incidentes graves). El modelo permite suministrar estimadores puntuales, intervalos de confianza y densidades completas de las variables condicionadas a las observaciones. Dado que sólo es necesaria información local para las predicciones, el modelo se muestra útil para predecir el tráfico en grandes redes.

Finalmente, los modelos se aplican a dos redes, una sencilla para ilustrar, y otra real en la que se muestra su aplicabilidad a redes de tamaño medio. Los tiempos de cálculo resultantes, al ser muy reducidos, prueban la utilidad práctica de los modelos propuestos.

castie@unican.es

Ponente: JORGE ARGÁEZ SOSA

U. Autónoma de Yucatán

Título: *Usando mapas a priori para modelar una variable discreta sobre una región geográfica*

Hora: (M2) vie20 09:40-10:20

Resumen: La figura de un experto como fuente de información se encuentra presente en muchas áreas donde se utiliza la Estadística. En diversas aplicaciones un experto es capaz de proporcionar información relevante con respecto a un parámetro de interés, información que se puede utilizar para postular un modelo adecuado de acuerdo con el contexto del estudio. Bajo el enfoque de la estadística Bayesiana, un experto puede proporcionar información útil para determinar la denominada distribución *a priori* para el parámetro de interés, que conjuntamente con un modelo para los datos, permite obtener la distribución posterior (Teorema de Bayes) a partir de la cual se realiza el proceso de inferencia.

En este trabajo se considera una manera en la que un experto puede proporcionar información, la cual ha sido utilizada en la práctica, aunque de manera informal. Suponga que el objetivo de un estudio es estimar la probabilidad de que ocurra un valor particular de una variable nominal sobre una región geográfica de interés. Por ejemplo, suponga que el objetivo es estimar la probabilidad de presencia de una especie en un sitio particular de una región. En este caso un experto en la especie es capaz de delimitar sobre un mapa de la región de interés, subregiones en las que, según su conocimiento, es altamente probable o improbable que la especie se encuentre presente. A las regiones que proporcione un experto se les denomina mapas *a priori*. Usando los mapas *a priori* conjuntamente con variables relacionadas con la ocurrencia del fenómeno bajo estudio, en este trabajo se presenta una forma de cuantificar la información dada por el experto, así como una manera de postular los parámetros de la distribución Dirichlet, comúnmente utilizada como distribución inicial para estimar la probabilidad de ocurrencia de cada uno de los posibles valores (categorías) de una variables nominal (por ejemplo, la probabilidad de presencia o ausencia de una especie).

Con fines ilustrativos se presenta un ejemplo en el que se aplica la metodología propuesta a una especie endémica de la península de Yucatán, México, usando los mapas *a priori* proporcionados por un ecólogo experto en la especie, conjuntamente con información de variables identificadas como relevantes para la presencia de dicha especie.

jargs13@gmail.com

Ponente: JAVIER GIRÓN

U. Málaga y R.A.C.

Título: *Análisis bayesiano objetivo de coste-efectividad para tratamientos clínicos en presencia de co-variables*

Hora: (M2) vie20 10:20-11:00

Resumen: Trabajo conjunto con María Lina Martínez y Elías Moreno.

En un artículo reciente hemos extendido el factor de Bayes para distribuciones intrínsecas del modelo lineal normal al caso multivariante. Por otra parte, también hemos presentado una solución al problema de selección de variables en los problemas de coste-efectividad en presencia de covariables, utilizando la descomposición de la distribución conjunta del coste y la efectividad como una marginal y una condicionada. Esta solución, aunque relativamente satisfactoria, puede mejorarse considerando directamente la distribución bivalente del coste y la efectividad.

Este último procedimiento, no solamente es más natural para seleccionar variables, sino que, en general, produce una mejor selección de las variables influyentes.

Además, el procedimiento de selección previa de variables conjuntamente influyentes permite la identificación correcta de los posibles subgrupos de la población bajo estudio en términos de las covariables influyentes, como demuestran algunos estudios de simulación que hemos efectuado.

El nuevo análisis basado en la distribución predictiva conjunta del coste y la efectividad se aplica a dos ejemplos: uno simulado en el que hay cuatro tratamientos y otro real, con dos tratamientos, en los que se pone de manifiesto la importancia de la identificación de las variables influyentes para la selección del mejor tratamiento para los individuos de cada uno de los posibles subgrupos.

fj_giron@uma.es